

ICRAT 2008

Third International Conference on
Research in Air Transportation

June 2008
Fairfax, Virginia, US



Third International Conference on Research in Air Transportation



1st - 4th June 2008
Fairfax, Virginia, US

Table of Contents

Message from Program and Conference Chairs	viii
Chairs and Committees	ix
 <i>Track 1: Advanced Modelling</i>	
Passenger Trip Delays in the U.S. Airline Transportation System in 2007	3
G. Calderón-Meza, L. Sherry and G. Donohue	
Constructing a Passenger Trip Delay Metric: An Aggregate Level Approach	13
S.-L. Tien, B. Subramanian and M. Ball	
Filtering and Aggregation Schemes for Delay Model Calibration	21
A. M. Churchill, K. Vlachou and D. Lovell	
The Impact of Ground Delay Program: Rationing Rules on Passengers and Airline Equity	29
B. Manley and L. Sherry	
Passenger Flow Simulation in a Complex Networked Transportation System	39
D. Wang, L. Sherry, G. Donohue and B. Hackney	
Modelling Stochastic Evolution of Runway Capacity Using Data Mining Concepts: Case Study of San Francisco International Airport (SFO)	51
Y. Yoon and M. Hansen	
Deconstructing Delay: A Case Study of Demand and Throughput at New York Airports	59
A. Kim and M. Hansen	
Propagation of Airspace Congestion. An Exploratory Correlation Analysis	69
C. Gwiggner, K. Kageyama and S. Nagaoka	
Smoothed Traffic Complexity Metrics for Airspace Configuration Schedules	77
D. Giannazza	
Network Restructuring Models for Improved ATS Forecasts	87
D. A. De Laurentis, T. Kotegawa and A. Sengstacken	
Resource Allocation in Flow-Constrained Areas with Stochastic Termination Times	95
M. Ganji, A. Nguyen, D. Lovell and M. Ball	
Trajectory Prediction: a Functional Regression Approach	103
S. Puechmorel, D. Delahaye and L. Boussouf	

The Robust Flight Level Assignment Problem	111
O. Klopfenstein and D. Nace	
Stochastic Airspace Demand for Strategic Traffic Flow Management	117
J. Henderson and A. Trani	
Modelling the Operational Impact of Air Traffic Control Automation Tools: A Case Study of Traffic Management Advisor	127
M. Smirti, M. Hansen and X. Chen	
<i>Track 2: CNS/ATM</i>	
Limitations of Subliminal Control in Air Traffic Management	137
G. Chaloulos and J. Lygeros	
Distribution of Longitudinal Speed Prediction Error of ADS-C System	143
M. Fujita	
Three-Degree Decelerating Approaches in Arrival Streams: Continuous Descent Approaches in High Traffic Density	151
A. de Leege, A. in't Veld, M. Mulder and R. van Paassen	
Separation Minima Standards: Research of Current Applicable Minima Laid Down and Foundations	159
M. Benitez and G. C. Angulo	
<i>Track 3: Safety & Security</i>	
A Review of the Research on Risk and Safety Modelling in Civil Aviation	169
F. Netjasov and M. Janic	
Accident Risk Analysis Benchmarking: Monte Carlo Simulation versus Event Sequences	177
H. A. P. Blom, S. H. Stroeve, J. J. Scholte and H. H. de Jong	
Analyzing Relationships Between Aircraft Accidents and Incidents	185
Z. Nazeri, G. Donohue and L. Sherry	
ATC Complexity as Workload and Safety Driver	191
J. Djokic, B. Lorenz and H. Fricke	
Assessment of Local Aircraft Crash Risk: Application of a Cluster Analysis as a Statistical Method for Detecting Similar Airports	199
C. Thiel and H. Fricke	

Hybrid System Framework for the Safety Modelling of the In Trail Procedure	207
M. Colageo and A. Di Francesco	
Proactive, Reactive, and Interactive Risk Assessment and Management of URET Implementation in Air Route Traffic Control Centers	215
G. Uang, J. Rakas and T. Bolic	
<i>Track 4: Decision Support Tools</i>	
Decision Support Tool for Predicting Aircraft Arrival Rates, Ground Delay Programs, and Airport Delays from Weather Forecasts	227
D. A. Smith and L. Sherry	
Pilot Support for Flying Curved Decelerating Approaches in Realistic Wind Conditions	235
A. in't Veld, R. Groenouwe, M. Mulder and R. van Paassen	
Developing a Decision-Support-Tool for an Air Taxi Service: A Research Proposal to Develop a Decision Support Tool to Analyze an Air Taxi Service on Strategic and Operational Level	245
P. A. Sengers and S. S. Ghijs	
Route Preliminary Demand Forecast Model for All-Business Airlines or Flights	253
N. H. Elferink and S. S. Ghijs	
<i>Track 5: Human Factors & Interfaces</i>	
An Infovis Approach to Compare ATC Comets: Comparing Visual Entities with a Theoretical Foundation	263
C. Hurter, V. Kapp and S. Conversy	
Functional Analysis of Human-Human Interactions during Collaborative Decision Making in Flight Operation	271
M. Groppe, M. Bui and R. Pagliari	
Responding to Uncertainty on Approach in Hazardous Situations	281
R. T. Joyekurun, W. Wong and P. Amaldi	
Automation for Task Analysis of Next Generation Air Traffic Management Systems	287
M. Medina and L. Sherry	
Human-Centred Innovation: Developing 3D-in-2D Displays for ATC	295
W. Wong, S. Rozzi, S. Gaukrodger, A. Boccalatte, P. Amaldi, B. Fields, M. Loomes and P. Martin	

Embedded Eye Tracker in a Real Aircraft: New Perspectives on Pilot/Aircraft Interaction Monitoring	303
F. Dehais, M. Causse and J. Pastor	
Visual Cognition Abilities in X-Ray Screening	311
D. Hardmeier and A. Schwaninger	
The Impact of Image Based Factors and Training on Threat Detection Performance in X-ray Screening	317
A. Schwaninger, A. Bolfig, T. Halbherr, S. Helman, A. Belyavin and L. Hay	
 <i>Track 6: Airport Operations</i>	
Comparison of Data Envelopment Analysis Methods Used in Airport Benchmarking	327
D. Schaar and L. Sherry	
Improving Aircraft Turnaround Reliability: Specific Aircraft Ground Parts Hamper Ground Handling and Airport Performance	335
H. Fricke and M. Schultz	
Runway Sequence and Ground Traffic Optimization	345
R. Deau, J.-B. Gotteland and N. Durand	
 <i>Track 7: Airline Operations</i>	
Accuracy of Reinforcement Learning Algorithms for Predicting Aircraft Taxi-out Times: A Case Study of Tampa Bay Departures	355
P. Balakrishna, R. Ganesan and L. Sherry	
Research of the Relation Between the Sustainability of Hourly Capacity at Schiphol Airport, KLM Arrival Punctuality and the Percentage of KLM Transfer Passengers at Risk of Losing their Connection	363
D. Mijatovic and M. Meert	
Efficiency of Aircraft Boarding Procedures	371
M. Shultz, C. Shulz and H. Fricke	
A Preliminary Evaluation of Potential Cargo Demand for Very Light Jets	379
Y. Xu and A. Trani	

Track 8: Airlines and ATC Concerns

- Untapped Potential of On-Board Advertising** 389
B. Badanik, H. Fakhri and M. Stefanik
- Pricing Schemes Based on Air Navigation Service Charges to Reduce En-Route ATFM Delays: Preliminary Results** 397
A. Ranieri and L. Castelli
- Estimation of Infrastructure Condition from a Biased Sample** 403
G. Gupta, J. Rakas and M. Hansen

Track 9: Environment, Weather & Economy

- On the Use of Near Field Computational Fluid Dynamics for Improving Airport Related Dispersion Models** 413
S. S. Aloysius and L. Wrobel
- Optimal Departure Aircraft Trajectories Minimising Population Annoyance** 421
X. Prats, V. Puig, J. Quevedo and F. Nejjar
- An Artificial Intelligence Approach to Operational Aviation Turbulence Forecasting** 429
J. Abernethy, R. Sharman and E. Bradley
- An Environmental Airport ATM Modelling Support Tool and its Use in Stacking and CDA Scenarios** 437
A. Goman and D. Atkins
- Analysis of Emissions Inventory for Single-Engine Taxi-Out Operations** 445
V. Kumar, L. Sherry and T. Thompson
- On the Use of Visualization Tools to Present Complex Simulated Environmental Data for Policy Making** 451
D. Pearce, L. Wrobel and I. Fuller
- Peak Oil, Fuel Costs and the Future of Aviation** 459
R. Klopheus
- Analysis of Air Transportation for the New York Metroplex: Summer 2007** 467
L. Wang, G. Donohue and R. Oseguera-Lohr

Doctoral Symposium

Demand Modelling for Low-Cost Airlines in Australia P. Srisaeng, C. Bil and M. Tein	477
Information Display for Collaboration and Distributed Team Work S. Rozzi, P. Amaldi and B. Fields	481
Optimizing the Predictability and Flexibility of Dynamic System: Case of 4D Aircraft Trajectory of Air Traffic Management T. T. Hoang	485
The Integrator Market: Actors and Their Strategies E. Onghena	489
A Construction Rationale to Tailor Crew Resource Management Training to Target Audiences V. Hagemann, S. Ritzmann and A. Kluge	497
Technology Assessment and Prioritization for Small and Medium Airports: A Methodological Approach O. Pinon, D. Mavris and E. Garcia	501
Evaluating Aeronautical Regulations Using Rigorous Specifications: Safeguarding Against Unintended Consequences L. Ruiz	505

Welcome from Programme and Conference Chairs

Welcome to the Third International Conference on Research in Air Transportation!

On the behalf of the ICRAT 2008 Organization Committee, we would like to express here our deep gratitude to the senior and young researchers in Air Transportation for having contributed to this young but challenging and exciting conference.

For this third edition of ICRAT, there were 77 qualified submissions by authors from 19 countries. The referee process resulted in 57 acceptances, for an acceptance rate of about 75%, among which 38 submissions were selected as standard papers, and 19 as short papers, representing respectively 50% and 25%. All selected papers, long and short, are of good quality, and we are very proud of the professionalism of all authors, reviewers, and of all Program Committee members. Thank you so much for your contributions and collaborations.

This is also the second year that Tutorials and a Doctoral Symposium are included in the conference program. Seven tutorials on the practice Air Transport are expected to bring up the understanding of how things work for the young scientists. The Doctoral Symposium is expected to create a forum for young researchers to discuss their research approaches with senior researchers to obtain guidelines and supports. The program is even more exciting with the six invited keynote speakers, all senior research scientists or strategists in Air Transportation. We are very grateful for their presence, contributions, and support.

The proceedings you are handling are the result of much hard work from many people. We would like to thank:

- The authors and co-authors of the paper submissions. They are, of course, what makes the conference program great.
- The invisible tertiary reviewers, who often supply the most expert and informed comments on their review, and the ICRAT'08 Scientific Program Committee. There were 40 members who had spent most of their free time during the referee process to review the submitted papers, and to return with careful comments. They are the guardians for the quality of the conference.
- The logistic team, also known as the conference secretariat team and the Webmaster team who worked hard to ensure the on-line processes with the authors, to collect, compile, and edit the final camera-ready proceedings.
- Telecom-Paris Tech with the support to host the website as well as for the time of Pr. Patrick Bellot and Loic Baud, who have worked pro-actively on the development and maintenance of the conference website.
- The Local Organising committee members and volunteers, for the local arrangements, the printing of the proceedings, and all the logistics at the conference place.
- The various institutions that provided the support for the paper process. The list includes the employers of all authors and co-authors and the employers of all reviewers and committee members.

Thank you all again, authors and reviewers, for your contribution to ICRAT'08 that surely be exciting. Thanks once more to the conference secretaries: Loic Baud, Simone Rozzi, Andrea Ranieri, Stephen Peterson, Ronish Joyekerun and the Publication Chair John Shortle to be the bridge between the Program Committee, the authors, and the Local Organisers. The success of this conference will be yours!

Andres Zellweger, General Chair,
George Donohue, Conference Chair,
Vu Duong, Program Chair.

Chairs & Committees

General Chair

Andres Zellweger ATC Quaterly, US

Conference Chairs

George Donohue GMU, US
Jan Van Doorn EUROCONTROL, FR

Past Chairs

Antonin Kazda Uni. Zilina, SK
Jean-Marc Garot Ministry of Trans., FR
Vojin Tosic Uni. Belgrade, SB

Program Chair

Vu N. Duong EUROCONTROL, FR

Registration Chair

Patrick Bellot ENST, FR

Tutorial Chair

John-Paul Clark Georgia Tech, US

Doctoral Symposium Chair

Marc Hansen UC Berkeley, US

Publication Chair

John Shortle GMU, US

Grants & Award Chair

Andres Zellweger ATC Quaterly, US
Christian Pusch EUROCONTROL, FR

Local Organizing Chair

Lance Sherry GMU, US

Conference Secretary

Loic Baud ENST, FR
Stephen Peterson Linkoping Uni., SE
Ronish Joyekerun Middlesex Uni., UK
Simone Rozzi Middlesex Uni., UK
Andrea Ranieri Uni. Trieste, IT

Program committee

Giovanni Andreatta Uni. Padova, IT
Alex d'Aspremont Princeton Uni., US
Michael Ball UMD, US
Philippe Baptiste Ecole Polytech., FR
Patrick Bellot ENST, FR
Marc Bourgois EUROCONTROL, FR
Steve Bradford FAA, US
Marc Brochard EUROCONTROL, FR
Marc Bui EPHE-Sorbonne, FR
Lorenzo Castelli Uni. Trieste, IT
Peter Choroba Helios, UK
John-Paul Clarke Georgia Tech, US
Dominique Collin de Verdier DGAC, FR
Kevin Corker SJSU, US
Nicolas Durand DSNA, FR
Jeffery Schroeder NASA, US
Heinz Erzberger NASA Ames, US
Hartmut Fricke TU Dresden, DE
Mark Hansen UC Berkeley, US
John Hansman MIT, US
Volker Heil DFS, DE
Jacco Hoekstra TU Delft, NL
Eric Hoffman EUROCONTROL, FR
Antonin Kazda Uni. Zilina, SK
Peter Lindsay Uni. Queensland, AU
Sandy Lozito NASA Ames, US
John Lygeros ETH Zurich, CH
Sakae Nagaoka ENRI, JP
Shinichi Nakasuka Uni. Tokyo, JP
Jose Miguel de Pablo Guerrero AENA, SP
Wim Post EUROCONTROL, BE
Amy Pritchett Georgia Tech, US
Jasenka Rakas UC Berkeley, US
Jozsef Rohacs Budapest UTE, HU
Colin Smith NATS, UK
Monica Tavanti Linkoping Uni., SE
William Wong Middlesex Uni., UK

Track 1

Advanced Modeling

PASSENGER TRIP DELAYS IN THE U.S. AIRLINE TRANSPORTATION SYSTEM IN 2007

Guillermo Calderón-Meza

PhD candidate
Center for Air Transportation System
and Research/GMU
Fairfax, VA, USA
gcaldero@gmu.edu

Lance Sherry, PhD

Center for Air Transportation System
and Research/GMU
Fairfax, VA, USA
lsherry@gmu.edu

George Donohue

Center for Air Transportation System
and Research
/GMU
Fairfax, VA, USA
gdonohue@gmu.edu

Abstract— The value of the air transportation system is the transportation of light-weight, high-value cargo, and passengers. Industry and government metrics for the performance of the air transportation focus on the performance of the flights. Previous research has identified the discrepancy between flight performance and passenger trip performance, and has developed algorithms for the estimation of passenger trip performance from publicly available data.

This paper describes an analysis of passenger trip delays for 5224 routes between 309 air ports in the U.S. air transportation system for 2007. The average trip delay experienced by passengers was 24.3 minutes for nationwide total of 247 Million hours. Flights delayed 15 minutes or more contributed 48% of the total delays, cancelled flights 43%, diverted flights 3%, and flights delayed less than 15 minutes contributed the remaining 6%. Passenger trip delays for oversold flights were negligible. Analysis of passenger trip delays for routes and airports, and the implications of these results are also discussed.

Keywords- *passenger trip delay; flight delay, airport delay.*

I. INTRODUCTION

The value proposition of the air transportation system is the rapid, safe, and cost effective transportation of high-value, lightweight cargo, and human passengers. This transportation is achieved by combining air transportation between airport terminals with ground transportation between origin (e.g. home)/destination (e.g. meeting) and the airport. The air component of the transportation is achieved through via single segment or multiple connecting segment scheduled airline operations.

To leverage economies of scale, airlines schedule and operate a daily itinerary that networks passengers, aircraft, flight, and cabin crews in connecting segments throughout the day. Individual flights on a segment may be delayed for several reasons such as: (e.g. mechanical) problems, weather, or traffic congestion. To maintain integrity of their networks in the presence of individually delayed flights, airlines may choose to delay, divert, or cancel flights.

When flights are delayed, the passenger trip for this segment is also delayed for the duration of the flight delay. When flights are cancelled or diverted, or passengers are bumped for overbooking, the passenger trip delay includes the duration of delay accrued waiting for the re-booked flight. All of these delays represent passenger trip delays.

Previous research by Bratu & Barnhart [2005] identified the discrepancy between flight performance and passenger trip performance. Wang [2007] showed that the 2% of passengers experiencing cancelled flights accrued delays of approximately 10 hours each, and that the total delays experienced by these passengers accounted for 40% of the total passenger trip delays.

This research provides the results of analysis of the U.S. air transportation system in 2007. The results are summarized as follows:

1. Passengers experienced a total of 247 Million hours of delays. The average delay was 24.3 minutes. Flights delayed 15 minutes or more accounted for 48% of the total delays, cancelled flight 43%, diverted flights 3%, and flights delayed less than 15 minutes accounted for almost all the remaining 6%. Passenger trip delays for overbooked passengers were less than 1%.
2. For flights on the 5224 routes between 309 airports, 50% of the routes experience an average passenger trip delay less than 15 minutes. 90% of the routes experience an average trip delay of less than 30 minutes.
3. For flights inbound and outbound of the 309 airports, 40% of the airports experience an average passenger trip delay of less than 15 minutes, 90% less than 30 minutes. Poorly performing airports included major hub airports as well as small commuter airports.
4. Passenger trip delay exhibited similar performance on routes of different stage-lengths¹.

The paper is organized as follows: Section 2 provides a summary of previous research. Section 3 describes the

¹ Stage-length is the *great-circle* distance of a flight.

algorithm and database structure used to compute estimates of passenger trip delay in 2007. Section 4 describes the results of the analysis. Section 5, Conclusions, discusses the implications of these results.

II. PREVIOUS RESEARCH

Researchers have shown that flight-based metrics, like the metrics reported in the Department of Transportation’s Airline Travel Consumer Reports (ATCR) [DOT, 2007] are a poor proxy for passenger experience [Wang, Schaefer, Wojik, 2003; Mukherjee, Ball, Subramanian, 2006; Ball et al., 2006; Bratu & Barnhart, 2005]. Bratu & Barnhart [2005] used proprietary airline data to study passenger trip times from a hub of a major U.S. airline.

This study showed that that flight-based metrics are poor surrogates for passenger delays for hub-and-spoke airlines as they do not capture the effect of missed connections, and flight cancellations. For example, for a 10 day period in August 2000, Bratu & Barnhart [2005] cite that 85.7% of passengers that are not disrupted by missed connections and cancelled flights arrive within one hour of their scheduled arrival time and experience an average delay of 16 minutes. This is roughly equivalent to the average flight delay of 15.4 minutes for this period. In contrast, the 14.3% of the passengers that are

disrupted by missed connections or cancelled flights experienced an average delay of 303 minutes.

Wang [2007], Sherry, Wang & Donohue [2006] developed an algorithm to estimate passenger trip delay for publicly available data from the Bureau of Transportation Statistics (<http://www.bts.gov>). One part of the algorithm joins separate databases with secondary data to derive the parameters to perform the passenger trip delay analysis. The next part of the algorithm computes an estimate of passenger trip delay for each scheduled flight. Key among those parameters used in the algorithm is the Passenger Load Factor for a flight. This algorithm uses the quarterly average Passenger Load Factor for flights on a given route. This results in undercounting for peak operations, and possible overcounting for non-peak operations. Further this analysis accounts for flight delays and cancelled flights only for routes between the OEP-35 airports.

The main results of this analysis are that passenger trip delays are disproportionately generated by cancelled flights. Passengers scheduled on cancelled flights represent 3 percent of total enplanements, but generated 45 percent of total passenger trip delay.

On average, passengers scheduled on cancelled flights experienced 607 minutes delay, and passengers who missed the

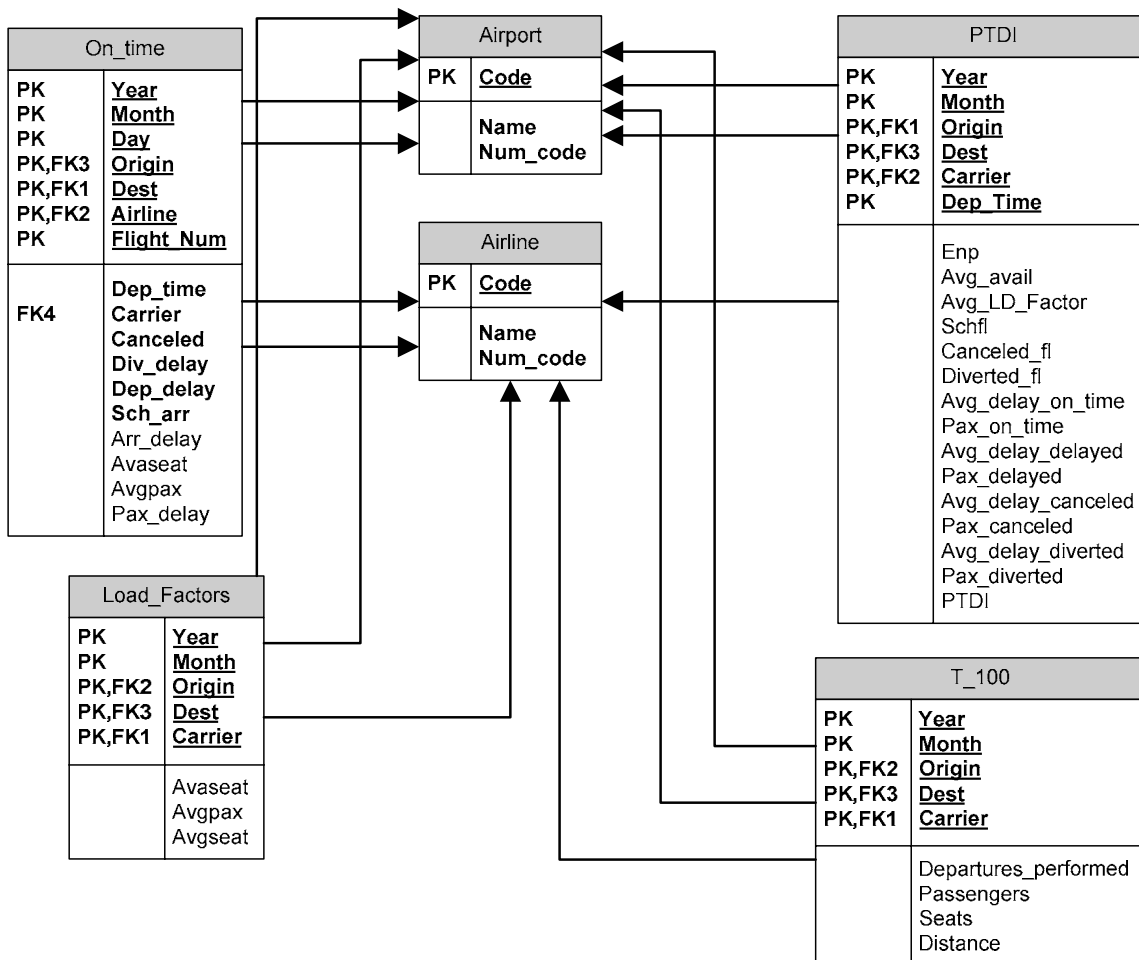


Figure 1. ER diagram of the local database

connections experienced 341 minutes delay in 2006.

The analysis described in this paper improved the algorithm by increasing the pre-processing of data to eliminate infeasible data and check for referential integrity. Further improvements were made to the algorithm to include diverted flights, improve processing throughput and automating manual steps in the processing.

III. DATABASE AND ALGORITHM

A. The local database

A local relational database stores data imported from public databases. The data consist of actual flight and performance values collected by competent institutions. Being as massive as they are, the raw data contain errors. Because of that, the database includes constraints to improve the quality of the input data. The design of the local database is illustrated by an ER diagram as shown in Fig. 1; it consists of six entities and thirteen integrity and referential constraints. Since the data are time dependent all several entities identify the tuples using year and month among other attributes. Other attributes that identify tuples in the entities are the carrier or airline, and the route (composed of one origin airport and one destination airport).

The Airport and Airline entities make sure that the other entities contain only known airports and airline codes: all of the other entities have foreign keys referring to Airport and Airline.

The *On_Time* entity contains the data about each individual flight. In particular, the attribute *canceled*, if its value is one, indicates that the flight was canceled (a value of one); otherwise, its value is zero. The attribute *div_delay* is either 0 for not diverted flights or 360 (min) for diverted flights. The attributes *avaseat* and *avgpax* are only used as temporal variables during the computation of *Estimated Passenger Trip Delay, EPTD* [Wang, 2007, Sherry, Wang & Donohue, 2006]. The attribute *pax_delay* (min) is the cumulated EPTD for all the passengers of the flight. Clearly, if *canceled* is 1, *div_delay* must be 0, and if *div_delay* is not 0, then *canceled* must be 0. The attributes *carrier* and *airline* are only different when the actual carrier is a subsidiary of an airline.

The *T_100* entity contains the input data concerning performance of pairs of route and carrier for domestic flights only. There are no data for individual flights. The entity includes information about the total number of departures done for a route and a carrier in the particular month (*departures_performed*), the total number of passengers transported (*passengers*), the total number of seats including all the flights (*seats*), and the distance of the particular route (in miles).

The entity *Load_factors* contains data derived from *T_100*. For a particular route and airline, the each record contains the average number of unoccupied (available) seats in the flights (*avaseat*), the average number of passengers per flight (*avgpax*), and the average number seats in the plane -the size of the plane- (*avgseat*). Clearly, the following conditions must be true at all times: $avgseat \geq avaseat$ and $avgseat \geq avgpax$.

The entity *PTDI* contains the result of the *Passenger Trip Delay Index (PTDI)* computation. In this case, flights are identified by their route, carrier, and departure time: no individual flights are recorded in this entity, but only averages of the flights that occur periodically at the given route, carrier, and departure time. The entity also includes data about the total number of *enplanements*² (*enp*), the average total number of seats available (*avg_avail*), the average load factor of this flight (*avg_LD_factor*), the number of scheduled flights (*schfl*), the number of canceled flights (*canceled_fl*), the number of diverted flights (*diverted_fl*), and the average delay time in minutes and number of passengers delayed for each category (canceled, diverted, delayed, and on-time) of flight. Finally, the entity also contains (though redundantly because it can be derived from the other attributes) the PTDI value in minutes. Notice that the delays can be zero, negative or positive real numbers. Negative numbers indicate that the passengers were not delayed but they arrived early. The number of enplanements must be greater than zero for the PTDI to make sense. The same happens with the number of scheduled flights. Clearly, the condition $canceled_fl + diverted_fl \leq Schfl$ must be true at all times.

B. Input data

The computation of the PTDI uses data from the Bureau of Transportation Statistics (BTS); particularly from two databases that are available on-line to download.

The first database is the T-100 for the domestic segment [BTS, 2006b]. This database allows the download of a whole year for all the carriers in the domestic (USA) segment. The fields selected to download are: year, month, origin, dest, carrier, seats, departures performed, passengers, carrier region, and distance. This experiment uses a single file containing data for the year 2007 from January to October³. The file contains 277870 records for 203 different carriers, 1142 airports⁴, and 23507 routes. The process to compute load factors for the flights and distance information for the routes uses these values. Every record of this file must comply with the conditions states in Table I to enter the local database.

TABLE I. CONDITIONS FOR EACH RECORD OF THE T_100 DATABASE

Field	Condition
Year	Equal to 2007
Month	In range [1, 10]
Origin	The value must be already in the Airport table
Dest	The value must be already in the Airport table
Carrier	The value must be already in the Airline table
Seats	An integer number that is greater than or equal to Passengers
Departures performed	A positive integer number
Passengers	A positive integer number
Carrier region	Only the value "D" (for domestic) is accepted
Distance	A positive real number

² An enplanement is a transported passenger.

³ November and December were not available at the time of the experiment.

⁴ These data include airports in Puerto Rico, and airports in project that are being used already.

A record that does not comply with all the conditions does not enter the local database, so that it is not used during the computation of the PTDIs. A total of 134111 records actually entered the local database including 932 airports, 115 carriers, and 17493 routes. Notice that some of the airports, carriers and routes are not actually referred in the On-Time database for the same period of time. These extra records in T_100 have no effect in the final results because the algorithm does not use them. The values for seats, passengers, and departures performed are monthly totals. There are no data for individual flights; therefore, average values are used in this experiment to approximate the actual values. The local database derives and stores the following values concerning load factors per year, month, route, and carrier:

- average number of seats, $avgseat = seats / departures\ performed$
- average number of passengers, $avgpax = passengers / departures\ performed$
- average number of available seats, $avaseat = (seats - passengers) / departures\ performed$

Therefore, the average load factor for a year, month, route, and carrier is: $lf = avgpax / avgseat$.

The second database is the so-called Airline On-Time Performance [BTS, 2006a]. This database allows the download of individual months of a particular year for all the airports and carriers in the USA. The fields selected to download are: flight_date, carrier, origin, dest, arr delay, crs arr time, dep delay, crs dep time, cancelled, diverted, fl_num, and tail_num. This experiment uses ten separate files for the year 2007, one for each month from January to October. Table II summarizes the figures for each one of the files.

TABLE II. STATISTICS FOR EACH OF THE ON-TIME INPUT FILES

Month	Records	Carriers	Airports	Routes
January	621555	20	289	4436
February	565602	20	288	4411
March	639209	20	288	4396
April	614648	20	289	4504
May	631609	20	294	4476
June	629280	20	298	4599
July	648542	20	300	4569
August	653276	20	298	4606
September	600186	20	298	4568
October	629990	20	292	4554
Total entered	6233873	17	309	5224

Notice that only 17 of the 20 carriers entered the local database. It is because the records with the three missing carriers did not comply with the conditions stated below. To enter the local database, each record must comply with the conditions stated in Table III.

TABLE III. CONDITIONS FOR EACH RECORD OF THE ON-TIME DATABASE

Field	Condition
Flight date	Any valid date for the year 2007
Origin	The value must be already in the Airport table
Dest	The value must be already in the Airport table
Carrier	The value must be already in the Airline table
Arrival delay	Any integer number (including 0 and negative ones).
Scheduled arrival time	A four digit positive integer number. The two left-most digits represent the hour in 24 hr format. The two right-most digits represent the minutes.
Departure delay	Any integer number (including 0 and negative ones).
Scheduled departure time	A four digit positive integer number. The two left-most digits represent the hour in 24 hr format. The two right-most digits represent the minutes.
Cancelled	Either 0 (not cancelled) or 1 (cancelled)
Diverted	Either 0 (not diverted) or 360 (6 hrs in minutes)
Flight number	Any value, but usually a three or four digit integer number.
Tail number	Any value. Used only to filter invalid records.

Each record must be unique with respect to flight date, origin, destination, carrier, and flight number. If there are repeated records, only one of them enters the local database. When the repeated records show differences in other fields, the user decides which one to keep. For instance, one of the records states that the flight was delayed and the other, that it was cancelled. The cancelled flight enters the local database in this case. Situations like this are not frequent: for the current input data only 53 records were repeated.

C. The algorithm

At a very high level of abstraction the algorithm to compute the PTDI is as follows:

- Import the T_100 data into the local database. This implies the computation of the load factor-related values.
- Import the on-time data into the local database. This implies the consideration of the carrier / subsidiaries relations. This means that subsidiaries are changed to their “parent” carrier every time they appear.
- Compute the EPTD based on the local load factor values and the local on-time data. This is done flight-by-flight, one month at a time. Fig. 2 illustrates the computation process of the EPTD.
- Compute the PTDI based on the EPTD, the delay, cancellation, and diversion data.

The following formulas compute the EPTD for each category of passengers:

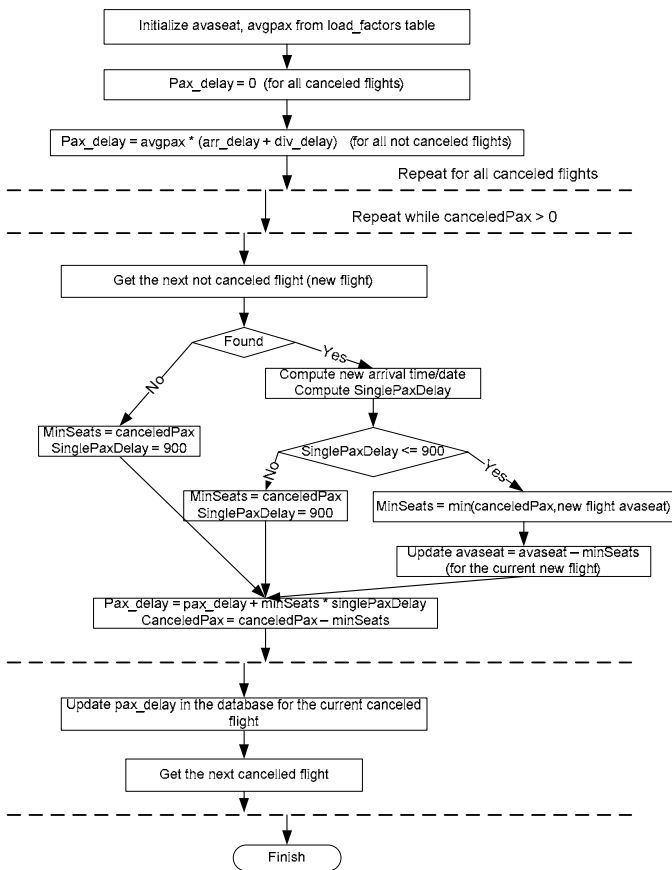


Figure 2. Algorithm to compute the EPTD

$$EPTD_{on-time}(f) = Pax(f) * ArrDelay_{<15}(f)$$

$$EPTD_{delayed}(f) = Pax(f) * ArrDelay_{\geq 15}(f)$$

$$EPTD_{cancelled}(f) = \sum_j Pax(f, j) *$$

$$\max(15 * 60, SchArr(j) - SchArr(f) + ArrDelay(j))$$

$$EPTD_{diverted}(f) = Pax(f) * 6 * 60$$

Where $Pax(f)$ is the number of passenger in the flight f . $Pax(f, j)$ is the number of passenger from flight f , that were reloaded on flight j . $ArrDelay_{<15}(f)$ is the arrival delay of flight f (in minutes) when it is less than 15 minutes (flight arrives on-time). $ArrDelay_{\geq 15}(f)$ is the arrival delay of flight f (in minutes) when it is delayed (15 minutes or more delay). $SchArr(f)$ is the scheduled arrival time of flight f . The constant $15*60$ represents the maximum wait time (assumed) the passengers will tolerate before changing to another airline or transportation means, it equals 15 hours (in minutes). The constant $6 * 60$ is the estimated delay time for a diverted flight; it equals 6 hours (in minutes).

At a high level of abstraction, the computation of the PTDI consists of eight steps:

- Compute the passenger delay for on-time flights: those arriving early or up to 15 minutes after the scheduled arrival time⁵.
- Compute the passenger delay for delayed flights: those arriving 15 or more minutes after the scheduled arrival time.
- Compute the passenger delay for canceled flights.
- Compute the passenger delay for diverted flights.
- Compute the number of enplanements.
- Compute the PTDI-related load factors.
- Eliminate null values (if any) and merge flights that depart less than 40 minutes after another flight of the same carrier on the same route.
- Compute the PTDI. Fig. 3 illustrates the computation process of the PTDI.

The following formula computes the PTDI:

$$PTDI_{r,a,t} = \frac{\sum_{r,a,t} Pax_{on-time}}{\sum_{r,a,t} Pax} * EPTD_{on-time}^{r,a,t} + \frac{\sum_{r,a,t} Pax_{delayed}}{\sum_{r,a,t} Pax} * EPTD_{delayed}^{r,a,t} + \frac{\sum_{r,a,t} Pax_{cancelled}}{\sum_{r,a,t} Pax} * EPTD_{cancelled}^{r,a,t} + \frac{\sum_{r,a,t} Pax_{diverted}}{\sum_{r,a,t} Pax} * EPTD_{diverted}^{r,a,t}$$

Where $Pax_{on-time}$ is the number of passenger on-time (less than 15 minutes delay), $Pax_{delayed}$ is the number of passengers delayed, $Pax_{cancelled}$ is the number of passengers in canceled flights, and $Pax_{diverted}$ is the number of passengers in diverted flights. Notice that the summations are performed after grouping the flights by route (r), airline (a), and departure time (t). Corresponding definitions are valid for the EPTD. Sub or superscripts r,a,t indicate that the associated values correspond to the average EPTD for the category (on-time, delayed, canceled, diverted) after grouping by route, airline, and departure time.

⁵ The convention is that flights arriving with less than 15 minutes of delay are on-time.

IV. RESULTS

The following analysis was conducted for 2007 for the months January through October using data derived from the BTS database for those months and year. The data included 512.8M passengers on 6.2 million flights on 5224 routes between 309 airports. The passenger trip delay includes an estimate of the total number of delay hours for on-time, delayed, cancelled, and diverted flights.

- Estimated total passenger trip was 247.08M hours. The average trip delay was 24.33 minutes.
- Estimated total passenger trip delay for passengers on flights delayed more than 15 minutes 119.44 M hours. The average trip delay for these passengers was 56.19 minutes.
- Estimated total passenger trip delay for passengers on cancelled flights was 107.39M hours. The average trip delay for these passengers was 667.93 minutes.
- Estimated total passenger trip delay for passengers on diverted flights was 7.77 M hours. The average trip delay for these passengers was 360 minutes.
- Estimated for passenger trip delay for over-booked passengers was negligible.

A. Comparison of flight delay and passenger delay

Fig. 4 shows a graphical comparison of flight delay and passenger trip delay (PTD). The y axis of the chart shows percentage of the total delay hours. The categories included are delayed, cancelled, and diverted flights. Flights that arrived early or with less than 15 minutes of delay are not included in the chart: they are considered on-time. Because of these on-time flights, the bars do not add to 100%. In other words, the on-time flights can also generate delays, but they are low enough to consider them as negligible.

The total delay measured using flight delay is 1.63 million hours as indicated in the chart. Notice that the flight delay metric does not consider canceled flights because those flights do not incur in delays.

On the other hand, the total delay measured using PTD is 240.08 million hours. This amount is very different from the 1.63 million of the other metric. In this case the total also considers the delays due to canceled flights, and not only diverted and delayed flights.

The PTD metric is more detailed and faithful to the real situation: passengers from a canceled flight experience considerable delays. In fact, the delays for passenger from canceled flights amount for about 43% of the total delay. About 48% of the total delay is due to delayed flights, and the rest of the delay is distributed among diverted and on-time flights.

B. Comparison of routes

Fig. 5 compares the histograms and cumulative distributions of the average PTDI and the maximum PTDI for all the routes with respect to the delay ranges (15 minutes each

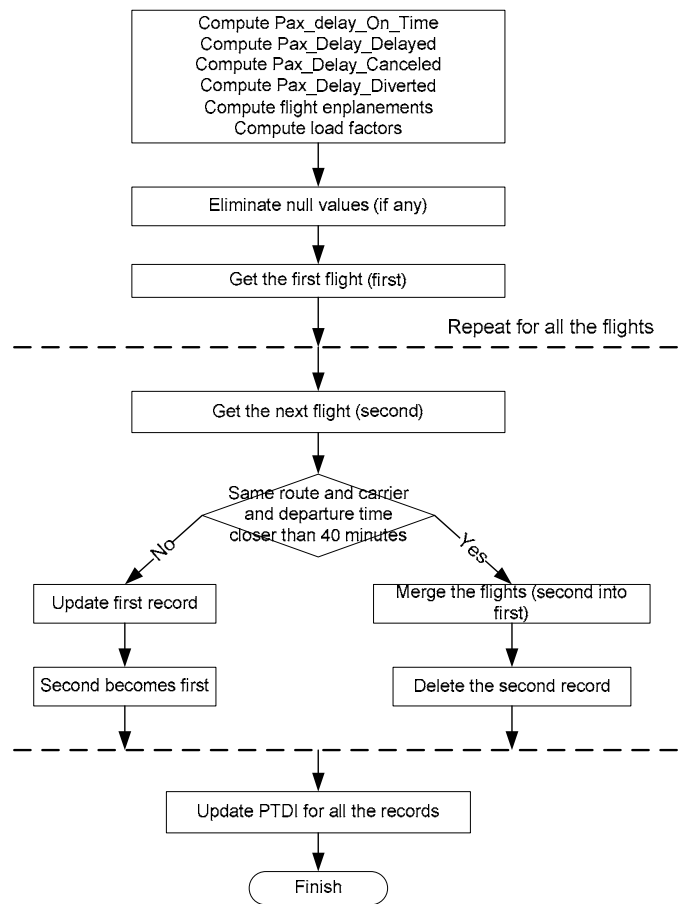


Figure 3. Algorithm to compute the PTDI

range). From the point of view of the average PTDI, 50% of the routes show on-time flights; and 90% of them show flights that are delayed less than 30 minutes. In extreme situations (maximum PTDI) about 20% of the routes show on-time flights and 50% show flights delayed 30 minutes or less. This distribution shows a peak not at the 0-15 minute range as the one for the average PTDI, but at the 30-45 minutes range. After the peak, the distribution descends monotonically slower than it the distribution of the average PTDI.

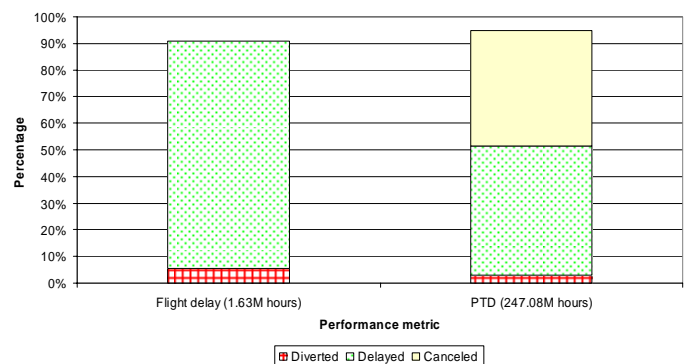


Figure 4. Comparison of flight delay and passenger delay as performance metrics

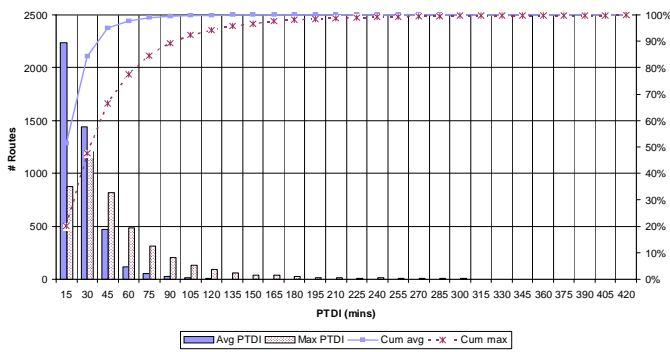


Figure 5. Distribution of routes with respect to the delay range

Comparison of route distance

The distribution of routes is similar for each distance range as shown in Fig. 6. Notice that the distance ranges are given in

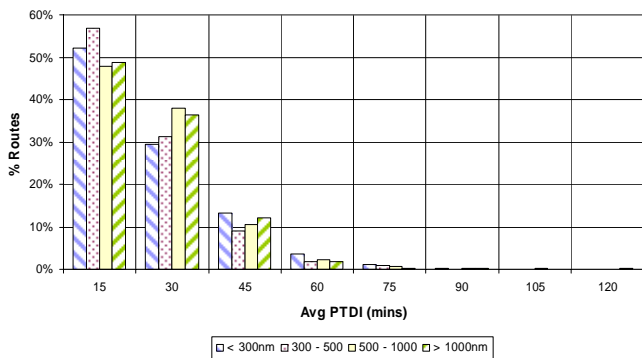


Figure 6. Percentage of routes grouped by distance per delay range

nautical miles (nm). All the ranges show between 55 and 47 percent of on-time routes. Between 29 and 38 percent of the routes show delays of 15 to 30 minutes. For the delay of 45 minutes the percentages are between 8 and 12. For the other distance ranges the behavior is also similar though with smaller percentage values. Though the differences are not big (8% at most), shorter routes tend to perform better: most of the routes of 500 nm and less are on-time (delay smaller than 15 minutes). Longer routes tend to delay more often. A significant part of the routes longer than 500 nm delay 30 minutes.

The informal comparison of the distribution of delays across distance ranges shows that the distribution has the same shape for all the distance ranges as shown in Fig. 7. In all the cases most of the flights are on-time and then the number of delayed flights decreases with each increase in the delay range. But, this chart also says that for shorter routes, is less probable to have long delay than it is for longer routes. For instance, the ratio of on-time to 30 minutes delay is about $17/10 = 1.7$ for routes of 300 nm or less, but it is $31/24.5 = 1.2$ for routes of 500 to 1000 nm.

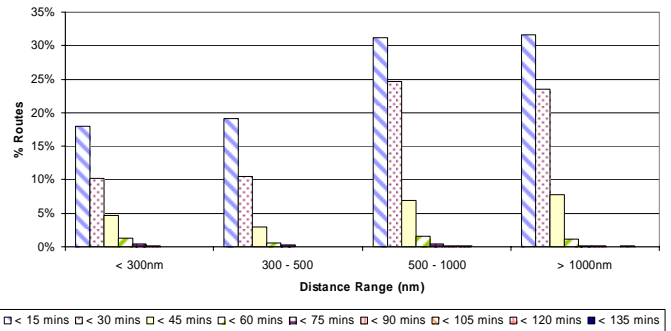


Figure 7. Percentage of routes grouped by distance range and delay range

C. Comparison of airports

The next step after comparing the routes is the comparison of the airports. In the case of inbound airports, Fig. 8 shows that most of them receive flights on-time or with 30 minutes

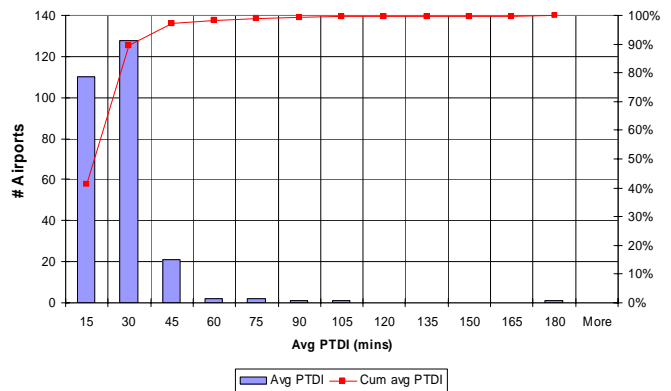


Figure 8. Inbound airport performance

delays: 40% of the airports show on-time flights, and 90% show delays of 30 minutes or less. Only few airports show average delays of 45 minutes or longer.

Table 4 summarizes a ranking of all the inbound airports in the database with respect to the average delay.

TABLE IV. BEST AND WORST INBOUND AIRPORTS RANKED ACCORDING TO PTDI

Best		Worst	
Rank	Airport (delay)	Rank	Airport (delay)
1	Greenville, MS	202	PHL (23)
2	Hilo, HN	226	IAD (26)
3	Pocatello, ID	239	DFW (31)
22	HNL	241	EWR (31)
31	SJC	245	LGA (33)
35	HOU	248	ORD (33)
39	OAK (10)	255	JFK (37)
40	MDW (10)	268	Meridian Regional (95)
59	LAS (11)	269	Rhineland-Oneida (171)
61	DAL (11)		
75	BWI (12)		

This ranking is based on the average PTDI for the airport. Rank ties are possible as shown in the table. Airports in bold belong to the OEP-35. Notice that some of the OEP-35

airports are ranked among the 75 best ones with respect the PTDI values.

The outbound airports behave as the inbound ones with respect to PTDI (see Fig. 9). About 90% the of the airports show delays of 30 minutes or less, and 40% show delays of 15 minutes or less. Again, only few airports show average delays of 45 minutes or more.

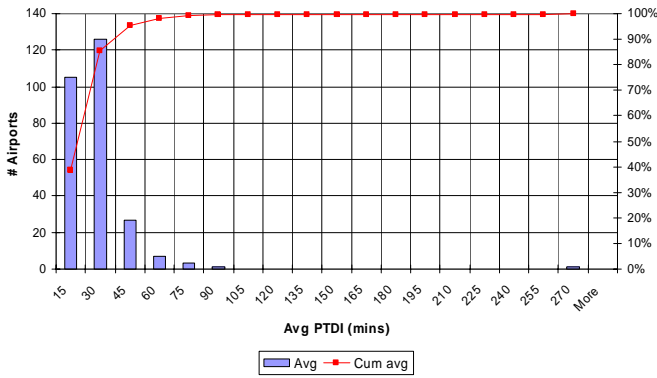


Figure 9. Outbound airport performance

Table 5 summarizes the ranking of all the outbound airports with respect to the average PTDI.

TABLE V. BEST AND WORST OUTBOUND AIRPORTS RANKED ACCORDING TO PTDI

Best		Worst	
Rank	Airport (delay)	Rank	Airport (delay)
1	Bristol/Johnson, TN	194	PHL (23)
2	Pocatello, ID	214	IAD (26)
6	Greenville, MS	229	EWR (29)
25	SJC	238	DFW (31)
28	HNL	239	ORD (32)
36	OAK (9)	248	LGA (34)
38	HOU (9)	249	JFK (35)
42	DAL (11)	265	Rhineland-er-Oneida (55)
53	MDW (11)	270	Middle GA Reg (260)
89	BWI (13)		
109	LAS (15)		

This ranking is based on the average PTDI for the airport. Rank ties are possible as shown in the table. Airports in bold belong to the OEP-35. Notice that some of the OEP-35 airports are ranked among the 89 best ones with respect the PTDI values.

D. Comparison of airlines

Finally, Table 6 summarizes the ranking of the airlines with respect to the average and maximum PTDI. Notice that in the case of average PTDI the difference is at most 27 minutes. In the case of the maximum PTDI, the difference is at most 700 minutes.

This ranking is based on either the average or the maximum PTDI for the airport as indicated in the column headings of the table. Rank ties are possible as shown in the table.

TABLE VI. AIRLINES RANKED BY PTDI

Average PTDI		Maximum PTDI	
Rank	Airline (delay)	Rank	Airline (delay)
1	Hawaiin (5)	1	Alaska (50)
2	Aloha	2	Aloha
3	Southwest	3	Hawaiin
4	Frontier	4	Frontier
5	Air Tran	5	Southwest (200)
6	Continental	6	USAirways
7	Alaska	7	Air Tran
8	ExpressJet (19)	8	Continental (250)
9	United (19)	9	JetBlue
10	SkyWest (19)	10	SkyWest
11	USAirways	11	United
12	Delta	12	ExpressJet
13	Northwest	13	Northwest/Airlink (291)
14	Northwest/Airlink	14	Mesa
15	Mesa	15	American
16	JetBlue	16	Delta
17	American (32)	17	Northwest (750)

V. CONCLUSIONS

Passenger trip delay is a critical performance metric for the airline transportation system. This metric assesses the performance of the true end-users of the system, and provides a measure of the true cost of delays.

Future research is planned to: (1) extend the algorithm to include lost luggage and refine the overbooked passenger algorithm, (2) add an algorithm to adjust the load factor for peak and non-peak periods, (3) continue to refine the automation of data retrieval and processing.

ACKNOWLEDGMENT

The authors would like to acknowledge the technical assistance and suggestions from Maria Consiglio, Brian Baxley, and Kurt Nietzke (NASA-LaRC), Todd Farley (NASA-ARC), Joe Post, Dan Murhy, Stephanie Chung, Dave Knorr, Anne Suissa (FAA, ATO-P), John Shortle, Rajesh Ganesan, Melanie Larson, Loni Nath, and Bengi Manley (GMU). This research was funded by NASA NRA NNN and Center for Air Transportation - George Mason University Research Foundation.

REFERENCES

Ball, M., D. Lovell, A. Mukherjee, and A. Subramanian. (2006) "Analysis of Passenger Delays: developing a passenger delay metric," NEXTOR NAS Performance Metrics Conference. ASilomar, CA. March

Bratu, and C. Barnhart, (2005) "An Analysis of Passenger Delays Using Flight Operations and Passenger Booking Data," Air Traffic Control, Volume 13, Number. 1, 1-27

Bureau of Transportation and Statistics, (2006a). Airline On-Time Performance. Data. Available: http://www.transtats.bts.gov/Tables.asp?DB_ID=120&DB_Name=Airlines%20OnTime%20Performance%20Data&DB_Short_Name=On-Time

Bureau of Transportation and Statistics, (2006b). Form 41 Traffic T-100 Domestic Segment Data. Available: http://www.transtats.bts.gov/Tables.asp?DB_ID=110&DB_Name=Air%20Carrier%20Statistics%20%28Form%2041%20Traffic%29&DB_Short_Name=Air%20Carries

Department of Transportation, (2006). Air Travel Consumer Report. Available: <http://airconsumer.ost.dot.gov/reports/index.htm>

- Mukherjee, M. Ball, B. Subramanian (2006) "Models for Estimating Monthly Delays and Cancellations in the NAS". NEXTOR NAS Performance Metrics Conference, ASilomar, CA. March 2006.
- Wang, D.(2007) "Methods for Analysis of Passenger Trip Performance In a Complex Networked Transportation System", Dissertation, George Mason University. Available <http://catsr.ite.gmu.edu>.
- Wang, D., Sherry, L. & Donohue, G. (2006). "Passenger Trip Time Metric for Air Transportation". The 2nd International Conference on Research in Air Transportation.
- Wang, P., Schaefer, L. & Wojcik, L. (2003). "Flight Connections and Their Impacts on Delay Propagation". In Proceedings of the 22nd Digital Avionics Systems Conference. Volume 1, 12-16.
- Wang, D. L. Sherry, Ning Xu, and M. Larson. (2008) "Statistical Comparison of Passenger Trip Delay and Flight Delay Metrics." In Proceedings Transportation Review Board 26th Annual Conference, Washington D.C.

Constructing a Passenger Trip Delay Metric

An Aggregate-level Approach

Shin-Lai Tien

Department of Civil & Environmental Engineering
University of Maryland
College Park, MD 20742, U.S.A.
alex tien (at) umd.edu

Bargava Subramanian

Institute for Systems Research
University of Maryland
College Park, MD 20742, U.S.A.
sbargava (at) umd.edu

Michael Ball

Robert H. Smith School of Business and Institute for
Systems Research
University of Maryland
College Park, MD 20742, U.S.A.
mball (at) rhsmith.umd.edu

Abstract—The on-time performance of passenger trips has received a great attention from government agencies in recent years but lacks a systematic metric to measure or trace the impact of flight delay to air travelers. The proposed model considers possible trip types of a passenger, utilizes system-wide flight-based performance metrics, and employs statistical approaches in order to develop an aggregate delay metric from passenger's perspective. Its results can be used to analyze historical passenger schedule reliability and can also be used to predict passenger experience for future aviation system.

Keywords—delay, passenger trip, performance metric, air travel

I. INTRODUCTION

The on-time performance of flights is a key concern of carriers and administrative agencies of aviation worldwide. It can be easily quantified for the U.S. National Airspace System (NAS) because all flight arrival and departure information is well recorded and disclosed by the Federal Aviation Administration (FAA). For example, the FAA's Aviation System Performance Metrics database (ASPM) provides individual flight information from all participating carriers at 75 major U.S. airports. Arrival delay of flights can thus be calculated by comparing scheduled and actual arrival time [10]. With suitable aggregation methods, delay metrics at airports or at the NAS-wide level can easily be constructed.

While flight delay statistics are well-recorded and well-publicized, they are not necessarily an accurate measure of a passenger's level of satisfaction. In particular, a passenger's average trip delay can vary substantially from average flight delay due to trip disruptions due to cancelled flights or missed connections. Bratu and Barnhart [1] analyzed proprietary airline data and indicated that the average time penalty on passenger trip time due to flight cancellations and missed connections is 303 minutes, while the average delay for non-disrupted passengers was only 16 minutes. However, acknowledging that passenger delay is also an important factor

of system performance, it is not easily measurable from any publicly accessible data. Since ticket information is not released by airlines nor collected by the government due to privacy concerns, the delay of multiple-leg passenger trips can be traced only with great difficulty. Even through proper sampling and survey techniques, passenger delay can only be observed during the selected survey period. Considering a long term objective of quality assurance of air travel, it would appear that there exists a need for defining a passenger oriented metric to be used as a quantitative measure of system-wide flight delay impact on passenger trips.

There is limited research that models passenger delay most likely because of the relative unavailability of individual passenger trip information. The essential challenge is to quantify the impact of flight delay on passenger trip disruption. Wang [3] treated passenger delay by its causes: delay due to delayed flights and due to cancelled flights. To estimate the passenger delay from cancelled flights, an algorithm was proposed that processed single-segment flight data. The underlying idea was to assign cancelled seats to the temporally closest available flights. Intuitively, this approach should work well in cases where only direct flights are being considered or under the assumption that on multi-leg passenger trips, the passenger always maintains the same intermediate stopping point. It should also be noted that this research does not model the possibility of missed connections on multi-leg flights.

The common characteristics of our paper and Wang [3] are: 1) both develop passenger-based performance metrics, and 2) both quantify the impact of flight cancellations on passenger delays. However, while the Wang model is a detailed "microscopic" model that estimates delays at a flight level, our model is macroscopic scope, attempting to directly estimate overall averages. The FAA's NAS Strategy Simulator (NSS) is a high-level policy analysis tool that predicts the impacts of future demand growth, policy changes, increasing fuel price, etc. [7]. Our research was specifically aimed at producing a performance module for the NSS. In the NSS context, all input

and output data are maintained at an aggregate level and so it is assumed that flight-level data are not available. Likewise, the required output should be NAS-wide average flight delay and cancellation rates rather than similar flight-specific metrics.

This paper is organized as follows. In Section 2, the concepts of the proposed model are discussed, and statistical methods are performed to estimate the probability of missing a connection flight. Numerical examples are constructed to illustrate the model, and the trend of passenger delay since 2000 is presented in Section 3. In Section 4, sensitivity analysis is conducted by analyzing the impact of key parameters on passenger delay. In Section 5, the potential usages and limitation of the proposed model are discussed.

II. MODEL CONSTRUCTION AND ESTIMATION

Passenger delays can be “inherited” directly from delayed flights but also can result from cancelled flights. Further, on multi-leg passenger trips, long flight delays on the initial leg can result in missed connections and induced delays not equal to, or even proportional to the original flight delay. In fact, cancellations and missed connections very often result in the most severe passenger delays. With these effects in mind, it can be seen that passenger delays depend on:

- Distribution of flight delays
- Flight cancellation rate
- Average load factor
- Percentage of passengers with 2 or more flight legs in their itinerary

In order to accurately address the actual delay experienced by passengers, models and statistical analysis are required that transform statistics related to these factors to passenger delay measures.

A. Scenario Tree Model of Passenger Delay

Our passenger delay model employs in a fundamental way, the concept of a *disrupted passenger*, which was introduced in Bratu and Barnhart [1]. A disrupted passenger is a customer who must use a flight other than the one on which the customer was originally scheduled due to a missed connection or flight cancellation. Disrupted passengers incur delays not related in a direct way to the delays on any of the flights in their original itinerary. Such passengers might be able to recover quickly, e.g. by taking the “next” flight scheduled to the missed destination or might incur a very long delay, e.g. requiring an unplanned overnight stay.

In order to model passenger delays, we create a scenario tree that represents all possible outcomes of a passenger’s trip. The database of Airline Origin and Destination Survey (DB1BMarket) contains directional market characteristics of each domestic itinerary of the quarterly Origin and Destination Survey [11]. The trip leg information of domestic markets from 2000 to 2007 is summarized in Figure 1, indicating that on average over 97% of the passengers chose direct or two-leg flights. Thus, because of the relative infrequency of three or

more leg trips in the U.S., we will represent itineraries as consisting of either one or two flight-legs.

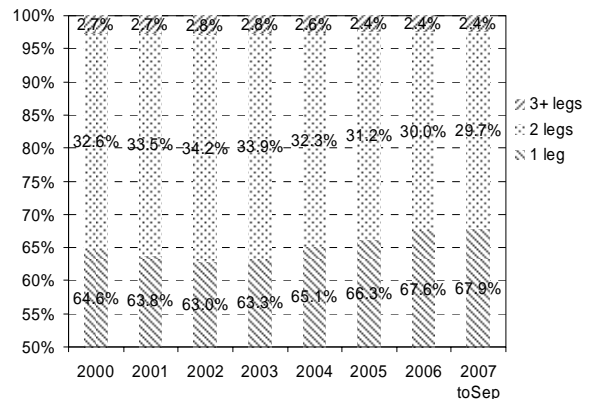


Figure 1. BTS Survey Results on Passenger Trip Leg Information

Our scenario tree is given in Figure 2. It represents the various events that can occur on a passenger itinerary, where for a 1-leg trip, the flight is denoted by f_1 and for a 2-leg trip the first flight is f_1 and the second is f_2 . Each leaf of the scenario tree represents a different outcome of a passenger trip and leads to a different “type” of passenger delay.

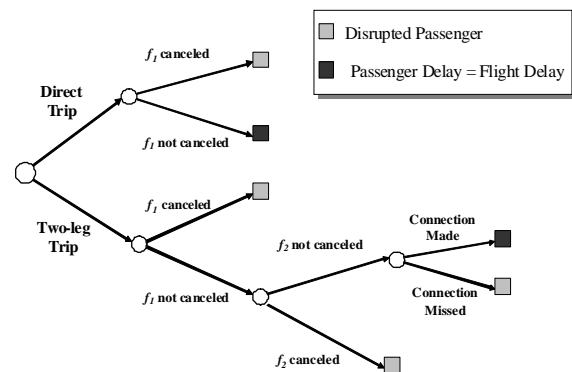


Figure 2. Scenarios Tree for a Passenger Trip

Expected passenger delay could be computed by computing the expected passenger delay at each leaf node in this tree and the probability of reaching each leaf node. The sum of the product of the leaf node probabilities times their expected delays would give the expected passenger delay. This would accurately compute expected passenger delay given the restriction to one and two leg trips. This is the approach we take; however, we must make several approximations in order to estimate the various probabilities and expectations. We hope that over time some of these approximations can be improved.

In computing our estimate of passenger delay, we use the following quantities:

- P_DIRECT: the fraction of passenger itineraries that are direct flights

- P_CANCEL: the fraction of scheduled flights that are canceled
- F_DELAY: Average flight delay
- DISRUPT: Average delay of disrupted passengers
- P_MISS: An estimate of the probability that a passenger misses connecting flight (the method for computing this estimate is discussed in the next section)

We now list all leaf nodes in the scenario tree, give our approximations of the expected passenger delay at that node and the probability of reaching that node, and discuss the accuracy of these approximations.

The various possibilities that can arise are:

1) Direct Trip, f_1 canceled:

Probability estimate: $P_DIRECT * P_CANCEL$

Delay estimate: DISRUPT

Discussion: The probability estimate is fairly accurate; however, P_CANCEL is actually a surrogate for the probability that a passenger is booked on a canceled flight. To the extent that there is a greater propensity for airlines to cancel flights with fewer passengers, a more accurate estimate could be obtained by doing a calculation that weights flights by the number of passengers (or seats). In fact there is not source of accurate statistics on the delay of disrupted passengers so the value we use for DISRUPT is a very rough estimate. Further, models could take into account whether a passenger is disrupted by a cancellation or a missed connection. DISRUPT also would be impacted by changes in airline policies and flight characteristics, such as load factor, so these could be used in improving estimates.

2) Direct Trip, f_1 not canceled:

Probability estimate: $P_DIRECT * (1 - P_CANCEL)$

Delay estimate: F_DELAY

Discussion: Subject to the caveats related to P_CANCEL mentioned above, both the probability estimate and the delay estimate should be highly accurate in this case.

3) Two-leg Trip, f_1 canceled:

Probability estimate: $(1 - P_DIRECT) * P_CANCEL$

Delay estimate: DISRUPT

Discussion: See discussion for previous two cases.

4) Two-leg Trip, f_1 not canceled, f_2 canceled:

Probability estimate: $(1 - P_DIRECT) * (1 - P_CANCEL) * P_CANCEL$

Delay estimate: DISRUPT

Discussion: See discussion for previous two cases.

5) Two-leg Trip, f_1 not canceled, f_2 not canceled, connection made:

Probability estimate: $(1 - P_DIRECT) * (1 - P_CANCEL) * (1 - P_CANCEL) * (1 - P_MISS)$

Delay estimate: F_DELAY

Discussion: As will be discussed later, estimating P_MISS can be very challenging. Our approach is to estimate the probability that flight delay exceeds a certain (constant) threshold. Clearly the required connection time varies substantially by flight so in reality the required threshold itself is a random variable. Further, it can be the case that both f_1 and f_2 are delayed so that even with a large delay on f_1 the connection can be made. Assuming the connection is made the passenger delay equals the delay on f_2 so that F_DELAY is a good estimate of passenger delay in this case.

6) Two-leg Trip, f_1 not canceled, f_2 not canceled, connection missed:

Probability estimate: $(1 - P_DIRECT) * (1 - P_CANCEL) * (1 - P_CANCEL) * P_MISS$

Delay estimate: DISRUPT

Discussion: See discussion in previous case regarding P_MISS. As discussed earlier it is certainly the case that the expected delay experienced by a disrupted passenger could vary depending on whether a canceled flight or missed connection was involved.

Based on this scenario tree and the preceding analysis, our estimate of average passenger delay, Pax_DELAY can be computed as:

Pax_DELAY =

$$\begin{aligned} & (P_DIRECT * (P_CANCEL) * DISRUPT + \\ & (P_DIRECT * (1 - P_CANCEL) * F_DELAY + \\ & (1 - P_DIRECT) * (P_CANCEL) * DISRUPT + \\ & (1 - P_DIRECT) * (1 - P_CANCEL) * (P_CANCEL) * DISRUPT + \\ & (1 - P_DIRECT) * (1 - P_CANCEL) * (1 - P_CANCEL) * (1 - \\ & P_MISS) * F_DELAY + \\ & (1 - P_DIRECT) * (1 - P_CANCEL) * (1 - P_CANCEL) \\ & * P_MISS * DISRUPT \end{aligned}$$

B. Probability of Passenger Missing Connection

Three of the inputs in the Pax_DELAY equation, i.e. F_DELAY, P_CANCEL and P_DIRECT, can be easily obtained from historical NAS performance statistics. For example, the monthly flight arrival delay and cancellation rate for the NAS can be calculated from ASPM individual flight data; the percentage of direct trips can be estimated from the quarterly market survey provided by the Bureau of Transportation Statistics, as shown in Figure 1. However, two inputs, i.e. DISRUPT and P_MISS, require reasonable approximation or further modeling efforts since they are not readily available in any data sources or previous research.

In order to provide a reliable estimate of P_MISS, we conduct a statistical analysis on the composition of P_MISS. If we denote by D_f , the random flight delay, then we define our estimate of the probability that a connection is missed because of a delayed flight by:

$$P_MISS = \text{Prob} \{ D_f > \text{Threshold} \}$$

where $\text{Threshold} = \text{LAY} - \text{CONNECT}$, LAY is a nominal flight layover time for connecting flights, and CONNECT is an estimated minimum time required to connect between two flights.

We assume that schedules are created so that if a flight arrives “on-time” then it makes its connection. Here on-time is defined relative to the U.S. Department of Transportation standard so that a flight is not classified as delayed if it is no more than 15 minute late. Thus, if D_f is less than or equal to 15 minutes, then we assume the passenger makes the connection successfully to the second flight leg. The probability of passenger missing connecting flight can thus be modeled as a conditional probability. Specifically, the probability that the connection is missed “given that” the flight is delayed (more than 15 minutes) is represented as:

$$\begin{aligned} & \text{Prob} \{ D_f > \text{Threshold} \mid \text{Flight being Delayed} \} \\ &= \frac{\text{Prob} \{ D_f > \text{Threshold} \cap D_f > 15 \}}{\text{Prob} (D_f > 15)} \\ &= \frac{P_MISS}{P_DELAY} \end{aligned}$$

where $P_DELAY = \text{the probability that a flight's delay} > 15 = \text{Prob} (D_f > 15)$. The probability of missing a connecting flight can thus be represented as:

$$P_MISS = P_DELAY \times \text{Prob} \{ D_f > \text{Threshold} \mid D_f > 15 \}$$

The first term is the probability that a flight is delayed more than 15 minutes. The second term is a conditional probability. P_DELAY can be estimated directly from flight delay data for the purposes of computing a metric. We also provide a way of estimating it using only an estimate of F_DELAY . This was done in order to derive estimates for future years in the context of the FAA Strategy Simulator. Our approach to estimating the second term for a time period, e.g. one month, will be to estimate the distribution: $\text{Prob} \{ D_f > D \mid D_f > 15 \}$ based on several years of historical data. The parameters of this distribution will be estimated as a function of F_DELAY and P_CANCEL . The value of Threshold and these flight performance statistics for the time period in question will be plugged into the distribution function to determine the estimate of the second term.

C. Probability of a Flight Being Delayed

As discussed above, P_DELAY can be computed directly from historical data. However we also provide a way of estimating it from flight delay statistic. From the ASPM database [10], for each month from January 2000 to December 2004, we computed the monthly values of F_DELAY and P_DELAY . Due to the obvious non-linearity in distribution functions, we postulated a quadratic relationship between F_DELAY and P_DELAY . A simple regression produced the following model with an R^2 of 0.9628.

$$P_DELAY = [(-0.0206) * (F_DELAY) * (F_DELAY) + 2.0431 * (F_DELAY)] / 100$$

D. Estimating Conditional Distribution of Flight Delays

In this section we describe our approach to estimating the conditional distribution function: $\text{Prob} \{ D_f > D \mid D_f > 15 \}$. Individual flight information stored in APSM database was used to compute the arrival delay of flights, which is defined as the difference between actual and scheduled arrival time. For each of month, an empirical distribution of flight delays > 15 was created. Specifically, each flight delayed over 15 minutes was placed into a 15 minute bin (15-30, 30-45, etc.) based on its delay value.

Empirical flight delay distributions were obtained in this way for each month from January 2000 to December 2004. These distributions were then fitted with the Bi-Weibull distribution. The Bi-Weibull, which is a combination of two Weibull distributions, is widely used in reliability applications. The Bi-Weibull distribution assumes a different form based on its shape parameters, which are:

- x_0 : the point at which the parameters change, and
- (α_1, β_1) and (α_2, β_2) : the parameters of the two Weibull distributions.

The parameter β_2 is a function of the other parameters, so there are four parameters in total to be estimated.

The fitted distributions gave 60 sets of observations of $(x_0, \alpha_1, \beta_1, \alpha_2)$. A regression was performed on each of these parameters, respectively, by using independent variables F_DELAY and P_CANCEL . The results from the regression are as follows:

- $x_0 = 11.1081 + 0.014 * F_DELAY * F_DELAY + 741.87 * P_CANCEL$ ($R^2 = .93$)
- $\alpha_1 = 0.37 + 0.00083 * F_DELAY * F_DELAY + 3.2 * P_CANCEL * P_CANCEL + 0.0032 * F_DELAY$ ($R^2 = .87$)
- $\beta_1 = 11 + 2.83 * F_DELAY + 112.12 * P_CANCEL * P_CANCEL$ ($R^2 = .901$)
- $\alpha_2 = 0.1143 + 0.0013 * F_DELAY * F_DELAY + 0.87 * P_CANCEL * P_CANCEL$ ($R^2 = .82$)

Thus, the distribution $\text{Prob} \{ D_f > D \mid D_f > 15 \}$ was estimated as a Bi-Weibull distribution whose parameters are given as functions of F_DELAY and P_CANCEL .

III. MODEL APPLICATION AND DATA ANALYSIS

The passenger delay model takes into account several major factors that impact passenger delay. Some model inputs are the results of aforementioned statistical models; some are available from reliable data source or analysis.

As the market survey results on trip leg information from 2000 to 2007 shown in Figure 1, it is observed that on average two-thirds of the passengers take direct flights. Hence, for model application purposes, we set

$$P_DIRECT = 66\%.$$

Disrupted passengers might be re-assigned to a later flight and often experience overnight stays. There are no publicly

available data about average delay of disrupted passengers. The research results of Bratu and Barnhart [1] based on a combination of proprietary data and simulation provide an estimate of 303 minutes as the average delay of disrupted passengers. Hence we set

$$\text{DISRUPT} = 303 \text{ minutes.}$$

The delay threshold of not missing a connection flight is the difference between the average flight layover time and minimum required connection time. Calculating average layover time experienced by a passenger requires detailed analysis on either passenger itinerary information or flight schedule along with seat information, which are not publicly accessible. The minimum required connection time can differ among individual airlines or even airports. Therefore, we take a conservative estimate on these two inputs based on empirical experience and assume that LAY = 45 minutes and CONNECT = 15 minutes. Thus, the delay threshold of not missing connecting flight is:

$$\text{Threshold} = \text{LAY} - \text{CONNECT} = 30 \text{ minutes.}$$

We now have provided models, estimation methods or approximation to obtain all required inputs for our metric. We use a simple example summarized in Table 1 to show how the passenger delay metric is computed. Given that Monthly NAS delay is 13.62 minutes and cancellation rate is 3.08%, the probability of a flight being delayed as well as the parameters of flight delay distribution is determined. The probability of flight delay more than connection threshold is computed by using the fitted distribution. As a result, the probability of missing a connection flight is 0.113, and the estimated monthly average passenger trip delay is 35.95 minutes. The relation among major model components is shown in Figure 3.

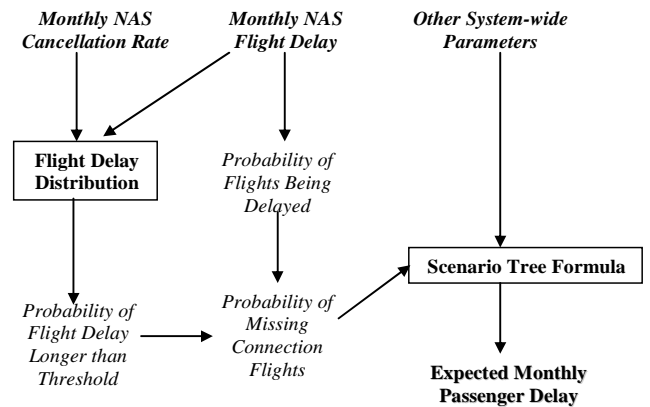


Figure 3. Application Procedures of Proposed Passenger Delay Model

Given the application procedures in Figure 3, monthly passenger delay metrics from January 2000 to May 2007 are computed by using ASPM flight delay and cancellation data. Figure 4 shows the time series of monthly passenger delay against flight delay and cancellation rate. Most of the spikes of passenger delay trend are due to high cancellation rates in those months as more passengers are disrupted. This suggests that there will be large penalty for passengers in terms of delay-minutes whenever a flight is cancelled, and also provides an explanation for why passenger experience varies from year to year as the overall cancellation rates change.

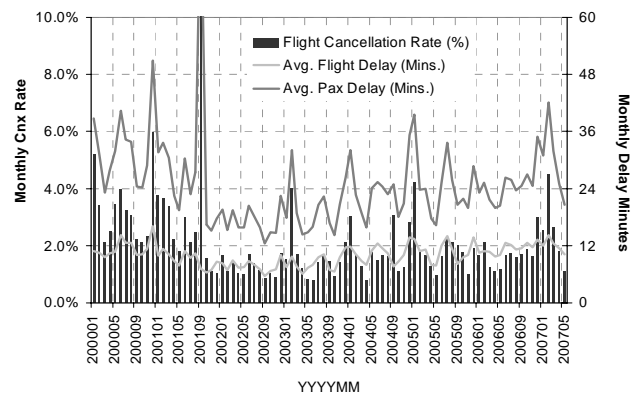


Figure 4. Time Series of Passenger Delays

The comparisons of modeled passenger delay against cancellation rate and average flight delay in the NAS are plotted in Figures 5 and 6, respectively. It can be also seen that as flight delay increases the passenger delay increases in more than a linear fashion. This validates our claim that as flight delays increase, more passengers are disrupted and the impact on passenger delays is much worse than actual flight delays.

TABLE I. A NUMERICAL EXAMPLE OF PASSENGER DELAY MODEL

Variable Name	Value	Source
Avg Monthly NAS Delay of Flights	13.62 mins.	Historical data or estimated from other models
Monthly NAS Cancellation Rate	3.08%	Historical data or estimated from other models
P_DIRECT	66%	BTS DB1B Database
DISRUPT	303 mins.	Result from Bratu's study
Threshold	30 mins.	Assumed
P_DELAY	24%	Estimated by this study
x_0	36.55	Estimated by this study
α_1	0.57	Estimated by this study
α_2	0.35	Estimated by this study
β_1	49.65	Estimated by this study
$\beta_2 = x_0 * \beta_1 / (\alpha_2 - \alpha_1)$	22.96	Estimated by this study
P_MISS	0.1134	Estimated by this study
Pax_DELAY	35.95 mins.	Calculated by using scenario tree formula

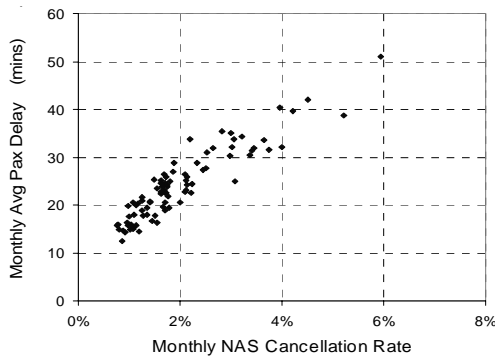


Figure 5. Monthly Passenger Delays vs. Cancellation Rates from Jan. 2000 to May 2007 (Sept. 2001 is excluded)

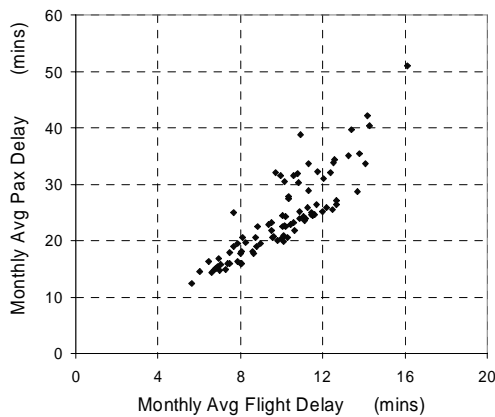


Figure 6. Monthly Passenger Delays vs. Flight Delays from Jan. 2000 to May 2007 (Sept. 2001 is excluded)

IV. SENSITIVITY ANALYSIS

As one of the modules in a high-level policy analysis tool, our model is designed to use flight performance statistics and to evaluate the passenger trip experience in response to changes in aviation system. The creditability of our model relies on proper inputs of parameters, either processed from historical data or calibrated from other modeling efforts. To better understand how passenger delays correspond to average flight delay, sensitivity analysis is conducted by varying the values of several key parameters. The parameters of our base scenario are summarized in Table 2.

TABLE II. PARAMTERS OF BASE SCENARIO

Variable Name	Value
P_CANCEL	2%
P_DIRECT	66%
DISRUPT	300 mins.
Threshold	35 mins.

Figure 7 illustrates the relation between flight delay and passenger delay with increasing values of DISRUPT, which is

the average delay of disrupted passengers. Certainly, DISRUPT is the most difficult to estimate input parameter. We see that P_DELAY increases with DISRUPT but that the sensitivity is fairly modest and the functional relationship between F_DELAY and P_DELAY generally retains its structure as DISRUPT changes. Figure 8 provides a similar sensitivity analysis for P_DIRECT. Note that the nonlinear structure of the curves in Figures 7 and 8 results from the fact that the probability of missing connections increases more than linearly with average flight delay.

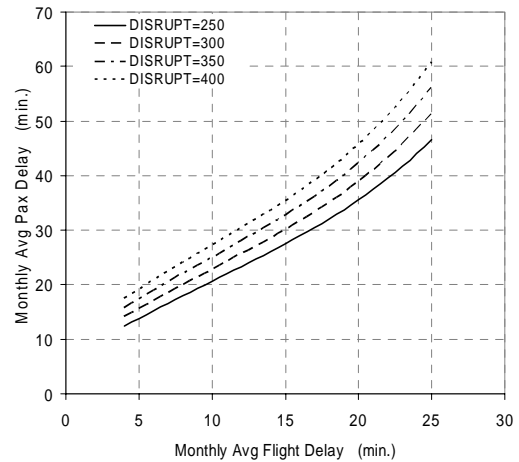


Figure 7. Sensitivity of Increasing Average Delay of Disrupted Passengers

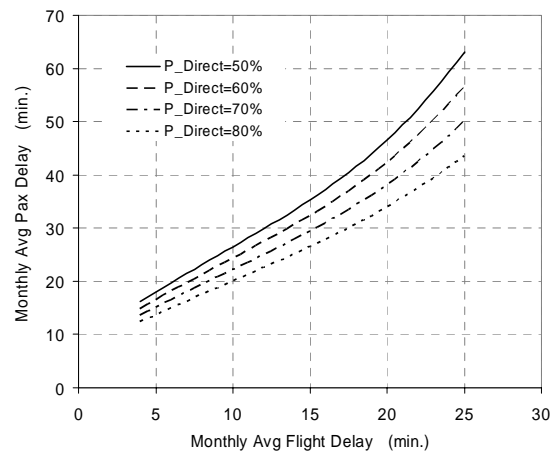


Figure 8. Sensitivity of Increasing Direct Trip Percentage

Figure 9 shows the sensitivity of varying the flight connection threshold. The choice of Threshold depends on the settings of minimum required connection time and average flight layover time, which are related to airlines' behaviors on schedule design and fleet management and require further exploration. This value is employed in determining the probability of missing a connection. The shorter the connection threshold, the greater the likelihood a flight is missed. The P_DELAY growth rate exhibited in Figure 9 is explained by the rather drastic growth rate in the probability of missing a connection for F_DELAY over 15 minutes as illustrated in Figure 10.

The multiplier effect of reducing Threshold on the probability of missing connection becomes more significant as flight delay increases. When system performance is getting worse, stringent connection times will increase the chances of missing connections and aggravate passenger trip delay. At 15 minutes of flight delay, the probability of missing connections with Threshold=40 is about 170% of that with Threshold=100. At 25 minutes of flight delay, the probability of missing connections with Threshold=40 is more than 220% of that with Threshold=100.

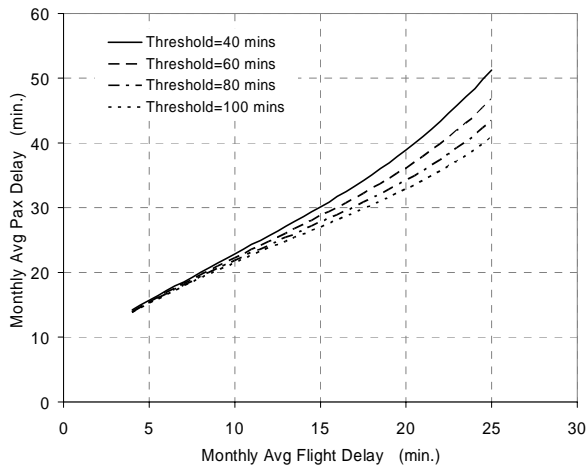


Figure 9. Sensitivity of Increasing Connection Threshold on Passenger Delays

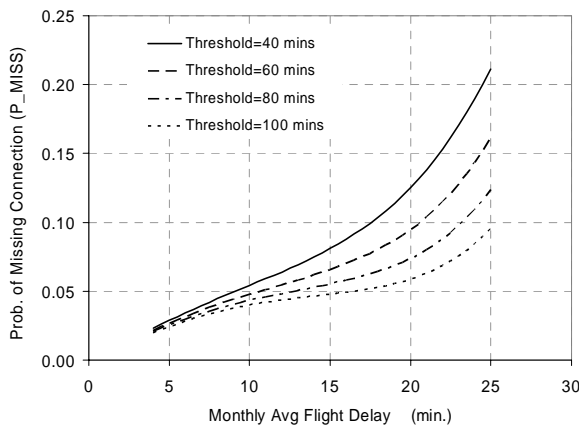


Figure 10. Sensitivity of Increasing Connection Threshold on Probability of Missing Connections

V. CONCLUDING REMARKS

It is generally agreed that flight-based delay metrics are not good surrogates of overall passenger experience of air transportation system. This study addresses the need for a quantitative measure of NAS passenger trip delays. The main contribution is that a passenger-based metric is modeled by considering a scenario tree for a passenger trip. This model allows the estimation of passenger delay based on existing flight-based performance metrics. A drawback of this study as well as other comparable research on passenger delays is that

the results can not be validated because of the unavailability of comprehensive passenger trip records.

The proposed model uses NAS-wide performance metrics, i.e. average flight delay and cancellations, in order to measure passenger delay from a strategic perspective. The inputs of the passenger delay metric are obtained from historical data analysis, statistical models, and reasonable approximation. Its intention is to provide an efficient but dependable estimate of passenger schedule reliability without much effort on analyzing detailed flight activities. Using models that forecast NAS-wide performance metrics, e.g. the flight delay models in Wieland [8] and Subramanian [9] and the cancellation rate model in Subramanian [9], the results of this research can also be used to predict passenger experience of future aviation system.

ACKNOWLEDGMENT

This work was supported by the National Center of Excellence for Aviation Operations Research (NEXTOR), under FAA cooperative agreement number 01CUMD1. Opinions expressed are solely those of the authors. They do not reflect any official opinions or policies the FAA or the U.S. Department of Transportation. The authors would like to thank Dave Knorr, Stephanie Chung, and Anne Suissa of the FAA as well as three anonymous reviewers for their valuable comments and suggestions.

REFERENCES

- [1] Bratu, S and C. Barnhart. An analysis of passenger delays using flight operations and passenger booking data. *Air Traffic Quarterly*, volume 13(1) 1-27, 2005.
- [2] Bratu, S. *Airline Passenger On-Time Schedule Reliability: Analysis, Algorithms and Optimization Decision Models*. PhD. Dissertation, Massachusetts Institute of Technology, Cambridge, MA, 2003.
- [3] Wang, D. *Methods for Analysis of Passenger Trip Performance in a Complex Networked Transportation System*. PhD. Dissertation, George Mason University, Fairfax, VA, 2007.
- [4] Wang, D and L. Sherry. Trend Analysis of Airline Passenger Trip Delays. In the Proceedings of the 86th Transportation Research Board Annual Meeting, Washington D.C., 2007.
- [5] Sherry, L., D. Wang, and G. Donohue. Air Travel Consumer Protection: Metric for Passenger On-Time Performance, *Transportation Research Record: Journal of the Transportation Research Board*, No. 2007, 2007, pp. 22-27.
- [6] Wang, D., L. Sherry and G. Donohue. Passenger Trip Time Metric for Air Transportation. The 2nd International Conference on Research in Air Transportation, Belgrade, Serbia and Montenegro, June 2006.
- [7] Suissa, A. FAA NAS Strategy Simulator (NSS): Overview. Presented at National Airspace System Performance Workshop, March 14-17, 2006, Pacific Grove, CA.
- [8] Wieland, F. Estimating the Capacity of the National Airspace System. The 24th Digital Avionics Systems Conference, Volume 1, November 2005.
- [9] Subramanian, B. *Aggregate Statistical Models for National Airspace System Performance*. Master Thesis, University of Maryland, College Park, 2007.
- [10] Federal Aviation Administration, *Aviation System Performance Metrics (ASPM) Online Database*. <http://aspm.faa.gov/>. Retrieved on July 31, 2007.
- [11] Bureau of Transportation Statistics. *Airline Origin and Destination Survey Database*. http://www.transtats.bts.gov/Tables.asp?DB_ID=125. Retrieved on January 24, 2007.

Filtering and Aggregation Schemes for Delay Model Calibration

Andrew M. Churchill (churchil@umd.edu)

Kleoniki Vlachou (kvlachou@umd.edu)

David J. Lovell (lovell@umd.edu)

Department of Civil and Environmental Engineering & Institute for Systems Research
University of Maryland College Park
College Park, MD

Abstract—This paper describes some methods for filtering and aggregating delay data from individual flights. The purpose for these transformations is to make the delay data more consistent with the outputs from queuing models. The transformed data can then be used to make much more relevant, and successful, comparisons against such models. This enables better calibration of the models, and helps to reveal what fraction of the total delay in a system might be generated solely from the consideration of congestion resulting from competition amongst aircraft for scarce airspace and airport resources. The paper describes the transformations in detail, and demonstrates their theoretical validity through examples. Real data are modified according to these transformations and are then compared against a stochastic queuing model to show the efficacy of the technique.

Keywords-queuing models, airport delay, delay filtering

I. INTRODUCTION

Queuing models, either deterministic or stochastic, are commonly used to predict delay statistics in the National Airspace System (NAS). The need for estimating delay is great, especially for busy airports. These models are particularly useful for studying future conditions that might include changes from the demand and capacity profiles expected under current operations. In some cases, such models can also be used to predict the effects of important infrastructure or policy changes, such as the addition of a runway or changing separation standards.

Queuing models are designed to estimate that component of delay that is incurred by aircraft as a result of competition, with other aircraft, for a capacitated resource, such as a portion of the airspace or a runway. Real delay data at a destination airport represent a broader collection of influences, and might include, among other things, upstream delays that accrued in previous flight legs, delays caused by late arrival of the crew, and delays caused by a mechanical problem with an aircraft.

Queuing models expect, as their inputs, nominal arrival times of flights; i.e., those arrival times that would prevail if other influences did not create delays. The differences between the nominal (or “desired”) arrival times and the actual arrival times are the statistics recorded as delays. Real demand data, in the form of flight schedules, do not represent this notion

exactly, primarily because air carriers include in their estimates of arrival time some expectation of delays. When executed properly, this is a perfectly reasonable practice, because it maximizes the likelihood that the actual performance of a flight will match the customers’ expectations for that flight. The problem, however, arises when an analyst tries to use the same data to populate queuing models, because their intent is to estimate those congestion-related delays themselves, rather than having them subsumed in the input data source. In this paper we offer a partial solution to this problem, although it is our belief that the general question of determining nominal (not padded for expected delay) arrival times for aircraft remains open.

All of the above constitute some of the reasons that the process necessary to facilitate proper comparisons of the outputs of queuing models with real data can be quite involved. Some form of comparison is essential, however, because the queuing models require calibration. It is also important to understand, once they are calibrated, that their delay predictions represent only a fraction of the total delay that might be expected when all of the other influences (which might be more difficult to model) are present. This proportion can be estimated as part of the overall calibration process and that is a valuable result in and of itself.

II. APPROACHES FOR MODIFYING DATA

The methods described in this paper for transforming individual flight data can be thought of as belonging to two classes of operations: filtering and aggregation. In the former case, we are attempting to subtract from real flight delay data the best estimates of delay components that are not directly attributable to congestion in the queuing sense. This step makes a direct comparison with delay data from queuing models much more valid.

Because the results from queuing models are most often shown in aggregate terms (e.g., the average delay incurred by all aircraft in the system during a particular time slice), the second step is to aggregate the real data, filtered accordingly, in a manner consistent with how queuing models tend to report their results. The need for this step is obvious, but its inclusion here is important because the paper illustrates how the

aggregate delay statistics available in the most common aviation databases are *not* averaged in a manner that allows for direct comparison with model results. Any attempt to make such a comparison, therefore, without following steps such as those outlined in this paper, is very likely to lead to a poor match between model results and real data, leading to the possible (and likely erroneous) conclusion that the model is doing a bad job or that queuing delays are not a significant component of the overall delays incurred by aircraft.

A. Filtering Schemes

The basic inputs to a queuing model are demands and capacities. A straightforward (although, we will argue, incorrect) method to use real data to feed such a model would be to use the collection of scheduled arrival times at an airport as the demand and a record of the declared airport arrival rates (AARs) as the capacity. The outputs from the queuing model might include average delay per time period, and it might be tempting to compare these directly against an aggregate average delay statistic in a database such as the FAA Aviation System Performance Metrics (ASPM) database, partly because the name of the metric is very similar. Again, this paper offers evidence that a more refined method is better for these purposes.

The filtering mechanism encapsulates two basic processes, one for the input data for the model, and one for real delay data to which output data will be compared. In the input data, rather than using scheduled arrival times directly, we develop a scheme for predicting the nominal or “best” arrival time for each flight being considered. Since the queuing model only represents congestion effects at the single airport in question, data of similar scope must be used for output comparisons. We take individual flight delay data from a real database and subtract an estimate of upstream propagated delays that would not be accountable for in the queuing models. These processes are described in detail in Section III.

B. Aggregation Schemes

When looking at one of the readily available aviation performance databases such as ASPM, one can find aggregate delay statistics recorded on an hourly (or sometimes quarter hourly) basis. For example, one could find a report of the average delay at Atlanta Hartsfield-Jackson International Airport (ATL) between 4 pm and 5 pm on some day. It is not clear simply from the title of the field, however, what the domain of aggregation is. In fact, what happens, using the above example, is that for all flights that landed at ATL between 4 pm and 5 pm, their delays (relative to schedule) were computed, and then these were averaged over these flights. A flight scheduled to land at 3 pm but landing at 5 pm would be assigned two hours of delay, but both of those hours of delay would be aggregated in the time window 4-5 pm, when actually only one of them actually occurred during that window. In fact, given the possibility of upstream propagated delays, the actual delays might have occurred considerably earlier in the day. This is not a flaw in the reporting mechanism, however; the way that ASPM (and other) delays are aggregated is simply the easiest and least ambiguous way to

record the ultimate differences between scheduled and actual arrival times.

The problem comes when trying to compare such data to the outputs of queuing models. In a deterministic queuing model, one can track the progress of individual aircraft, so it is possible to generate data that are consistent with this reporting mechanism. It is more common with such models, however, to use delay accounting practices taken directly from seminal sources on deterministic queuing (see for example [1]), where delays are accounted for *as they occur*, rather than after flights have landed. This difference can perhaps best be seen by graphical example; we call the mechanism used for reporting real data in places like ASPM “horizontal aggregation” and that typically used in deterministic queuing “vertical aggregation.”

Importantly, stochastic queuing models (not simulations) frequently do not allow for the tracking of individual aircraft. Instead, the state space consists of the range of possibilities of the length of the queue at any given time, and the differential equations of the state dynamics govern how this queue grows or shrinks over time. There is no accounting, however, for which particular aircraft are present at any given time. Thus, the horizontal aggregation mechanism is not possible. The vertical aggregation mechanism is possible, and in a stochastic model each possible queue length is assigned some probability of prevailing at any particular time, so the vertically aggregated delay statistic generated represents the expected value of the delay incurred by aircraft during that time slice.

Fig. 1 shows an example of a cumulative demand curve (the upper curve, representing the number of flights that wanted to land by a particular time) and a cumulative supply curve (the lower curve, representing the actual number of flights that were allowed to land by a particular time, as constrained by the arrival capacity). The abscissa represents time, while the ordinate represents flight count, and the flights can be considered to be sorted in order of their desired arrival times.

During the time slice t_1 to t_2 , flights labeled f_1 through f_2 landed. The total delay experienced by these flights over their lifetimes can be computed as the area of the horizontal band bounded by these two flight labels on the top and bottom, and by the two cumulative curves on the left and right. After dividing by the number of aircraft $f_2 - f_1$, the result is the average delay statistic that would have been reported in a database like ASPM. Because delays for individual flights are read from the figure as horizontal spacings between the two cumulative curves, we call this form of averaging delay “horizontal aggregation.”

If one looks vertically at the same time slice, however, the band between the curves represents the total quantity of delay incurred by flights whose desired landing times occurred prior to the time slice in question, but whose actual landing times occur (or will occur) during or after that time slice. In this case, the total number of flights represented is $f_3 - f_1$ and the average delay statistic can be computed as the area of the vertical band divided by this number of flights. Again, this is the statistic traditionally (but not necessarily) drawn from

deterministic queuing models, and necessarily drawn from stochastic queuing models.

It should be clear from the figure that the two quantities can be quite different. Perhaps only a few flights are figured into both calculations, and even then the entirety of a flight's delay experience would be horizontally aggregated while only a portion of the delay would be captured with vertical aggregation during that time slice. Another way of thinking of the two methods is temporally: the horizontal aggregation method looks to the past, recording statistics about delays that have already occurred, while the vertical method looks to the present, by recording delays as they occur, but also to the future, because delays components yet to occur for those flights will be reported in later time slices. It is extremely important to note that both methods represent the "truth"; neither is more or less accurate than the other. The difference is simply in deciding which domain, in terms of time and flight identification, will be considered for aggregation and reporting during any particular time slice.

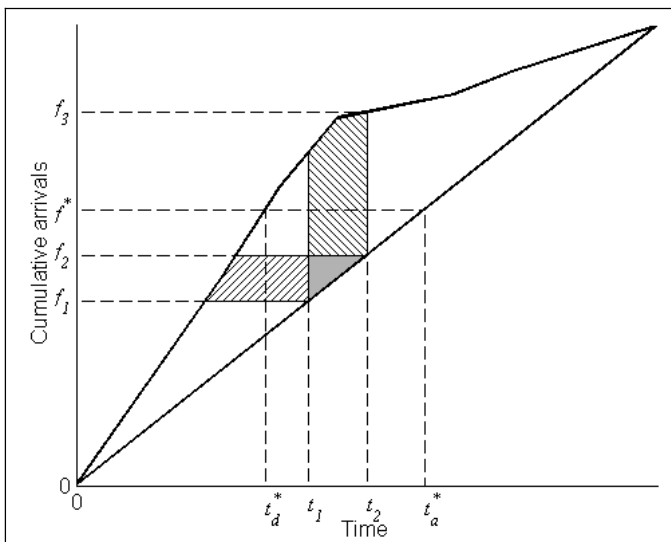


Figure 1. Vertical and horizontal integration of delay

When considering the specific flight f^* shown in Fig. 1, its desired landing time was t_d^* and its actual landing time was t_a^* . It would have contributed $t_2 - t_1$ units of delay to the vertical measure of aggregate average delay for that same time slice, and other amounts to other time slices. In the horizontal scheme, however, it would only have contributed to the measure recorded for the time slice containing time t_a^* , and the amount of delay contribution would have been $t_a^* - t_d^*$.

III. DELAY FILTERING

Because of the economic realities of the airline industry, individual aircraft are scheduled to operate several flights each day with little time between flights. Thus, if an aircraft suffers a delay early in the day, it becomes more likely that later flights operated by that same aircraft will also be delayed. When using real delay data to calibrate a queuing model, however, these propagated delays must be accounted for.

In this section, we describe an approach for identifying and removing these propagated delays from real delay data. The resulting statistics more clearly represent the queuing delay imposed on the aircraft. Additionally, we propose a technique to utilize this filtered data to produce a new "schedule" for each aircraft (and hence, for each airport). These schedules can be used as a better proxy for the true demand for resources as input to queuing models.

A. Procedure

The approach taken in this work and several others (see [2], [3]) has been to use individual flight records to trace aircraft by their tail numbers as they are routed from airport to airport over some period of time. These series of flights by a single aircraft are used to identify and remove propagated delay.

Thus, the first step in this process is to identify series of related flight data. Initially, data are grouped by tail number and sorted by departure day and time. However, because some flight records may be unavailable in the database, it may be infeasible to use all records for a single tail number as a single series. This phenomenon is evidenced by an arrival airport not matching the subsequent departure airport, indicating a missing flight record (e.g. caused by a "ferry" flight or data corruption). Data series are also considered broken if more than 24 hours elapse between an arrival and subsequent departures. From this procedure come several series of data for each tail number being examined.

Once these series of connected flights have been built, propagated delays must be distinguished from "new" delays. This process works by determining the best possible departure time for a flight, given the delay the previous flight experienced prior to its arrival. The best departure time is calculated as the maximum of two quantities: the scheduled departure time, or the previous (delayed) arrival time plus some minimum turn time, as shown in (2) and (3). The maximum of these two quantities is considered so as to prevent the best possible departure time from falling before the scheduled departure time. This would unfairly penalize flights relative to their schedule.

The minimum turn time is calculated in (1) as the minimum of the scheduled turn time and some parameter T_{turn} . In this analysis, only domestic flights were considered. Because these are generally operated by small or medium sized aircraft, the minimum turn time parameter T_{turn} was taken to be 40 minutes. An enhancement to be considered for future work using this delay filtering algorithm would be to consider variable minimum turn times, wherein the parameter might vary based upon the aircraft type, length of previous flight, airport in question, time of day, or some other factors.

Once the best departure time has been calculated, the best arrival time must be computed. In this work, the best arrival time was taken to be the sum of best departure time and the scheduled block time, as shown in (4). As mentioned previously, this block time does not represent the minimum time against which a queuing model might compare, as the scheduling carrier implicitly accounts for delay when scheduling the block time. However, estimating a true

minimum block time for a given flight may be a fairly complex endeavor, and as such, has been left for subsequent work.

Finally, the filtered arrival delay $D_{f,i}$ is computed as the maximum of zero and the difference between the best and actual arrival times. It is customary in aviation delay calculations to disregard negative delays, and we maintain that practice here, particularly because a queuing model would never predict negative delays.

This algorithm is stated below for a series of flights $i = 1, \dots, I$, given the input data listed previous described.

Compute minimum turn time:	
$T_{turn,i} = \min [T_{turn}, (t_{sd,i} - t_{sa,i-1})]$	$\forall i \in \{2, \dots, I\}$ (1)
Compute best departure time:	
$t_{bd,i} = t_{sd,i}$	$i = 1$ (2)
$t_{bd,i} = \max [t_{sd,i}, (t_{aa,i-1} + T_{turn,i})]$	$\forall i \in \{2, \dots, I\}$ (3)
Compute best arrival time:	
$t_{ba,i} = t_{bd,i} + T_{block,i}$	$\forall i \in \{2, \dots, I\}$ (4)
Compute filtered delay:	
$D_{f,i} = \max (0, t_{aa,i} - t_{ba,i})$	$\forall i \in \{2, \dots, I\}$ (5)
Input data:	
$t_{sd,i}$: Scheduled departure time for flight i	
$t_{sa,i}$: Scheduled arrival time for flight i	
$t_{aa,i}$: Actual arrival time for flight i	
$T_{block,i}$: Scheduled block time for flight i	

Once the filtered arrival delay $D_{f,i}$ has been computed for each flight, these records can be aggregated in either the horizontal or vertical methods previously mentioned. If they are aggregated horizontally by airport and time period, they will be comparable to those typically reported, but they will necessarily be lesser in magnitude. The case in which they are aggregated vertically will be discussed later.

If the aircraft counts are aggregated by best possible arrival time, it is possible to create a new "schedule" which better reflects the true demand for operations during that time period, from the perspective of the queuing model. For example, a queuing model being applied to ATL does not care if a flight had originally intended to arrive at ATL at 5 pm but due to delays two flight legs prior to that cannot even depart the airport immediately upstream of ATL until 5:30 pm. The real question for the queuing model is, given this penultimate status update, what would be the nominal arrival time for the aircraft at ATL. This data can be used as input for a delay prediction model to provide a better proxy for demand than the traditional schedule would.

B. Numerical Example

To illustrate the principles described above, a numerical example has been developed. The aircraft under consideration was routed as shown in Fig. 2. The detailed calculations for

one flight leg are shown in Table I, while the scheduled and actual performance for its entire itinerary are shown in Table II.



Figure 2. Example case routing

This aircraft suffers a fairly significant delay of 41 minutes on the second flight in this series. As a result, the best departure and arrival times computed for the subsequent flights are later than the scheduled ones. Further, the filtered flight delay must then be significantly less than the delay reported by traditional metrics.

To illustrate the steps of the algorithm described above, the calculations for the third flight leg (SFO – PHX) are presented here in Table I.

TABLE I. DELAY FILTERING COMPUTATIONS SFO – PHX FLIGHT

Compute minimum turn time:	
$T_{turn,i} = \min [40, (11:05 - 10:11)] = 40$	
Compute best departure time:	
$t_{bd,i} = \max [11:05, (10:52 + 40)] = 11:32$	
Compute best arrival time:	
$t_{ba,i} = 11:32 + 113 = 13:25$	
Compute filtered delay:	
$D_{f,i} = \max (0, 13:27 - 13:25) = 2$	

It is interesting to note that the schedule for each of these flights allowed for turns longer than the T_{turn} parameter of 40 minutes. As a result, each of the assumed turn times used to compute the best possible departure time was smaller than that which was scheduled. It should also be noted, however, that the average scheduled turn time was 48.5 minutes, while the average performed turn time was 44.3 minutes. Because of the time pressure of the delayed flight, the turns were performed faster than scheduled, and more closely matched the 40 minute parameter used in the algorithm.

TABLE II. DELAY FILTERING EXAMPLE DATA

From	To	Schedule			Actual			Best			Delay	
		Turn	Dep.	Arr.	Turn	Dep.	Arr.	Turn	Dep.	Arr.	Reported	Filtered
DEN	LAS	-	7:10	7:55	-	7:05	7:56	-	7:10	7:55	1	1
LAS	SFO	0:45	8:40	10:11	0:39	8:35	10:52	0:40	8:40	10:11	41	41
SFO	PHX	0:54	11:05	12:58	0:37	11:29	13:27	0:40	11:32	13:25	29	2
PHX	ORD	0:50	13:48	19:14	0:55	14:22	19:36	0:40	14:07	19:33	22	3
ORD	LAS	0:45	19:59	21:49	0:46	20:22	22:10	0:40	20:16	22:06	21	4

IV. VERTICAL INTEGRATION

There are many ways in which individual flight delays can be aggregated. The most familiar category of metrics involves summing delays across flights arriving at a given airport (or set of airports) during a particular time period. However, one must be precise when describing exactly which data are summed for the given time period.

In the traditional metrics reported in the ASPM and other systems, delays are grouped according to the time at which flights arrived. Regardless of when those delays were accrued, they are assigned to the period of arrival under consideration. The essence of the vertical integration technique, however, is to sum delays that are accrued during a given time period, regardless of when the affected flights arrive.

A. Procedure

The first part of this procedure is to establish at what time delay begins accruing on a flight. Establishing this baseline allows the delay to be assigned to bins beginning at that time. This assumption must be carefully examined, lest delay be assigned to the incorrect time bins. In this work, we assume that delay begins accruing when the nominal, or best possible, arrival time has passed, and the aircraft has not yet arrived at its destination. This best possible arrival could be calculated in many ways, depending upon the assumptions about departure and flight times that were applied. Based upon the delay filtering analysis presented previously, we will use the best possible arrival time calculated as part of that algorithm.

The first step in calculating these delays is to divide each day into a series of time bins, each bounded by some numbers t_p and t_{p+1} . Let L and U define the upper and lower bounds for the delay accrual period. In this case, these bounds are the best arrival time and the actual arrival time, respectively. Then, find the first bin l into which the flight i contributes delay, as shown in (6).

$$l = \max \{ p \mid L - t_p \geq 0 \} \quad (6)$$

Next, find the last bin u into which the flight i contributes delay, as shown in (7).

$$u = \max \{ p \mid U - t_p \geq 0 \} \quad (7)$$

Then, for each bin $p \in \{l, \dots, u\}$, apply the following four logical tests to determine the delay accrual $D_{p,i}$ from flight i into bin p .

- (a.) IF $L \geq t_p \ \& \ L < t_{p+1} \ \& \ U < t_{p+1} \ \& \ L < U$
THEN $D_{p,i} = U - L$
- (b.) ELSE IF $L \geq t_p \ \& \ L < t_{p+1} \ \& \ U \geq t_{p+1}$
THEN $D_{p,i} = t_{p+1} - L$
- (c.) ELSE IF $U \geq t_p \ \& \ U < t_{p+1} \ \& \ L < t_p \ \& \ L < U$
THEN $D_{p,i} = U - t_p$
- (d.) ELSE IF $L < t_p \ \& \ U \geq t_{p+1}$
THEN $D_{p,i} = (t_{p+1} - t_p)$

The $L < U$ condition is applied to exclude those cases in which the flight arrives before its best possible arrival time. In those cases, the new calculated delay would be negative. We treat these cases as having accrued zero delay.

Fig. 3 illustrates each of these logical tests, and the specific case of L and U that they approach. The hatched area in the figure shows the delay accrual period for the flight. Case (a.) applies when both L and U fall in the same time bin. Case (b.) applies when L is in the current bin, but U is in any later one. Case (c.) applies when L is in a previous time bin, but U is in the current one. Case (d.) applies when L is in an earlier time bin, and U is in a later one.

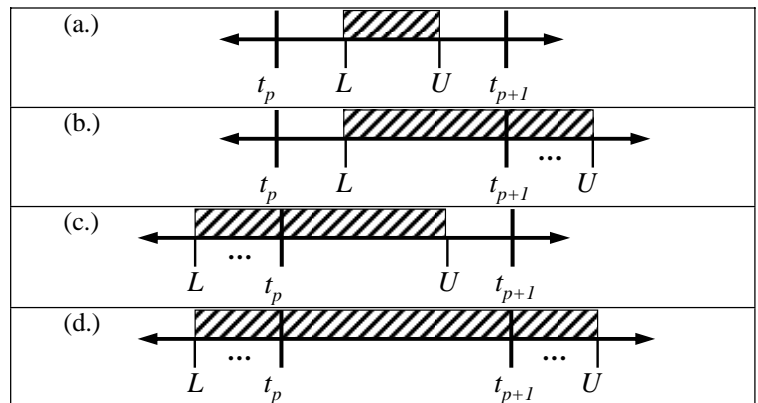


Figure 3. Logical cases for binning delays

B. Numerical Example

This algorithm is illustrated here by examining a fictitious set of flights shown in Table III, and represented graphically in Fig. 4. Assume that the delay filtering algorithm previously described has been applied to a larger dataset, and that these flights destined for ORD were extracted. The best departure and arrival times, as well as the actual arrival times, are shown.

The filtered delay is calculated as the difference between the best possible and actual arrival times.

TABLE III. VERTICAL INTEGRATION EXAMPLE DATA

From	To	Best		Actual Arrival	Filtered Delay
		Departure	Arrival		
ATL	ORD	12:19	1:23	1:38	15
BWI	ORD	12:11	1:26	1:32	6
LGA	ORD	12:05	1:29	1:41	12
CLT	ORD	12:03	1:00	1:21	21
OKC	ORD	12:00	1:15	1:30	15

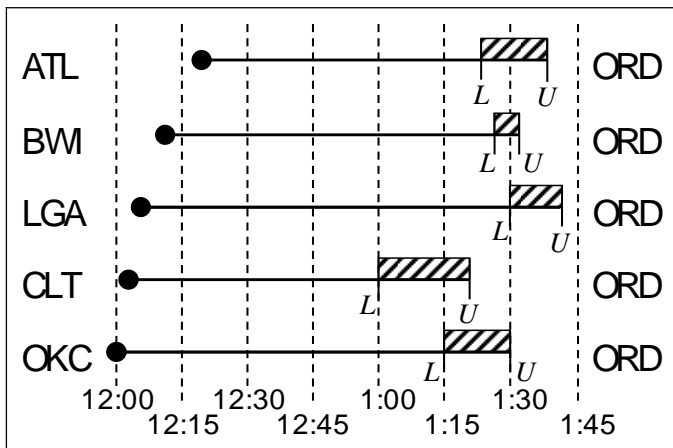


Figure 4. Vertical integration example data

As an example, apply the various tests on the first flight shown above, that from ATL to ORD. The first bin l to consider is the 1:15 bin, and the last bin u is the 1:30 bin. Each of the logical tests is evaluated for this flight and the results shown in Table IV.

TABLE IV. VERTICAL INTEGRATION COMPUTATIONS FOR ATL – ORD FLIGHT

Test	1:15 bin		1:30 bin	
	Test Result	$D_{p,i}$	Test Result	$D_{p,i}$
(a.)	FALSE	-	FALSE	-
(b.)	TRUE	7	FALSE	-
(c.)	FALSE	-	TRUE	8
(d.)	FALSE	-	FALSE	-

Upon evaluating the logical tests for each of these flights, the data are summed across time bins, and the results summarized in Table V. As expected, the reported delay differs significantly from the delay actually accrued by all flights in each period.

TABLE V. VERTICAL INTEGRATION EXAMPLE SUMMARY STATISTICS

Time period		Arrival Count		Total Arrival Delay	
Begin	End	Sch.	Actual	Reported	Accrued
1:00	1:14	1	0	0	15
1:15	1:29	3	1	21	31
1:30	1:44	1	4	48	22

V. RESULTS

The delay filtering and vertical integration algorithms were incorporated and applied to a large test dataset to provide comparison data for calibrating a queuing model. Several airports were considered.

A. Delay Models

For this work, we used a stochastic queuing model with non-homogeneous Poisson arrivals and Erlang services times $(M(t)/E_k(t)/1)$. This model is solved analytically using the DELAYS software developed at MIT (see [4], [5]). It is specifically designed to estimate the delay incurred by aircraft on landing at an individual airport given the capacity and demand profiles over specific time periods. Previous work has focused considerable attention on the accuracy of the approximation scheme used by this software. This work will demonstrate that DELAYS provides a suitable queuing model for airport arrival operations.

B. Input Data

The individual flight records were obtained from the Bureau of Transportation Statistics (BTS). The BTS database includes records for certified US air carriers that account for at least one percent of domestic scheduled passenger revenues. Other sources of individual flight records could be used as well, but the BTS data provides excellent coverage of operations at most of the largest airports in the US.

The Airport Arrival Rates used as capacities for the DELAYS model were drawn from the ASPM system. The demands used as input for the DELAYS model were not the scheduled demand, but rather were summed using the best possible arrival time as the scheduled arrival time.

C. Case Studies

Data from numerous airports were examined in this work. The results for several are shown here, but others are available from the authors.

The “Reported” category refers to the horizontally aggregated data typically reported. The “Filtered” category shows the results of filtering out propagated delays, but aggregating in the traditional horizontal manner. The “Filt/Vert” category shows the results of both filtering the data, and aggregating it by the period in which it was accrued.

1) Atlanta Hartsfield-Jackson International Airport (ATL)

ATL is a very large and busy airport serving as the hub for several carriers. Demand is frequently at or near capacity.

Fig. 5 shows results for a sample month during 2004. The first thing to note is that the reported delays are almost always higher than all other metrics. The filtered delays fall slightly below the reported delay, but follow the same series of peaks and valleys. Particularly at the end of the day, the gap between these two is large, as should be expected. The vertically integrated and filtered data suggest an amount of total delay similar to the filtered delay, but have peaks and valleys that more closely follow those of the DELAYS series, which are nearly almost lower.

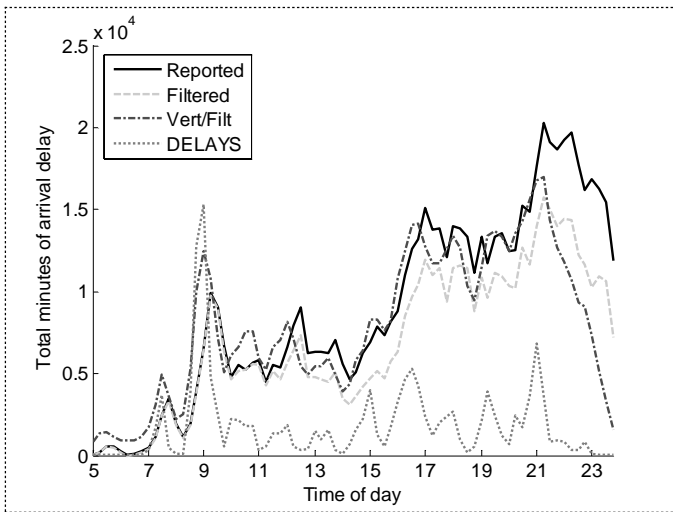


Figure 5. ATL: February, 2004

2) Detroit Metropolitan Wayne County Airport (DTW)

The results for DTW were fairly similar to those for ATL. An interesting feature of the DTW results, which was present to a much lesser degree for ATL, is the correspondence of peaks and valleys in the monthly data shown in Fig. 7. The peaks for the reported and filtered series correspond quite well, as should be expected, because they are both horizontally aggregated. In addition, the peaks for the vertically integrated filtered data and the DELAYS model correspond quite well. The interesting feature here, however, is that the peaks for the first pair of data lag those for the second pair. This exhibits the exact feature espoused earlier in the paper, which is that the delay model will show delay as it is accumulated, while the reported statistics will show it as the aircraft arrive.

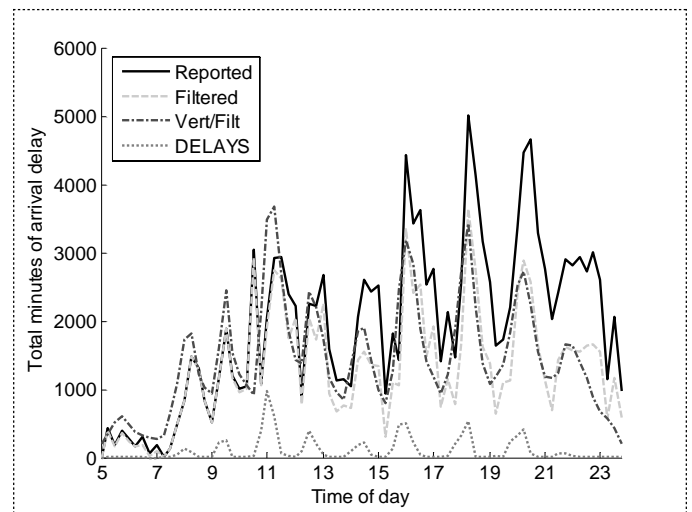


Figure 7. DTW: December, 2004

Fig. 6 shows three sets of pairwise correlations, between the three delay quantities mentioned above, and the delays predicted by the DELAYS queuing model over each month in 2004. The correlations between the vertically integrated and filtered data and the DELAYS output (the gray bars) were uniformly higher than those for all other metrics. The figure only shows results for ATL, but this conclusion held true for every airport examined.

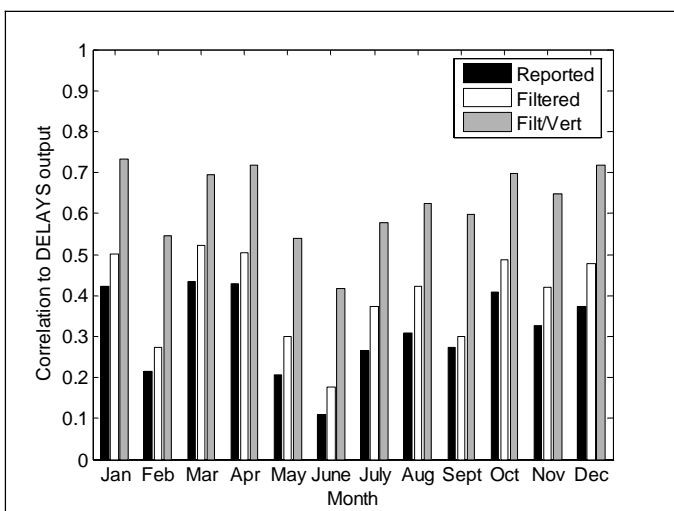


Figure 6. ATL 2004: Monthly Correlations

As was the case for ATL, the correlations between the vertically integrated filtered data and the DELAYS outputs are uniformly and significantly higher than those of any other metric. This suggests that the proposed methods provide data that corresponds better with the DELAYS model output, and we would expect this same conclusion to hold for other queuing models.

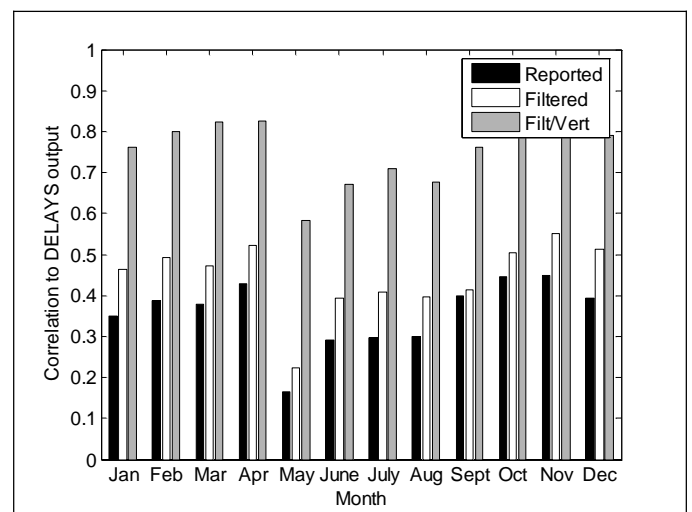


Figure 8. DTW 2004: Monthly correlations

VI. CONCLUSIONS

Several schemes were presented in this paper to help understand the relationship between operational and queuing model data in aviation systems. In their native formats, the data have slightly different contextual meanings, and this makes direct comparison troublesome. They can be rectified, however, by the methods discussed in the paper.

The first method discussed for bringing the data sources into agreement was the application of filtering techniques. These are useful in removing the effects of delay propagated between flights using the same aircraft. This technique removed some portion of this delay, and produced data that showed a stronger correlation with predicted results.

The second technique shown in this paper was a different scheme of aggregation than is typically used for aviation delay data. The methodology proposed allows for delays to be reported in the time bin in which they are accrued, rather than the time bin in which the flight arrives. This technique, combined with the first, produced results that show a very strong correlation to the predicted delays.

These two techniques have myriad applications in aviation system planning and modeling. Both are very useful in calibrating and understanding delay prediction models. In addition, they encourage the reader to consider the nature of the delay reporting mechanisms in use.

REFERENCES

- [1] Newell, G.F. 1982. Applications of queuing theory, 2nd Edition. Chapman and Hall, New York, NY.
- [2] Lovell, D.J., A.M. Churchill, A.R. Odoni, A. Mukherjee, Ball, M. O., 2007, "Calibrating Aggregate Models of Flight Delays and Cancellation Probabilities at Individual Airports," *Proceedings of 7th USA/Europe Air Traffic Management R&D Seminar*, Barcelona, Spain.
- [3] Mukherjee, A., D.J. Lovell, M.O. Ball, A.M. Churchill, A.R. Odoni. 2007. Estimating long-term average flight delays and cancellation rates for an airport, unpublished.
- [4] Kivestu, P. 1976, "Alternative Methods of Investigating the Time-dependent M/G/K Queue," S.M. Thesis, Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA.
- [5] Malone, K., 1995, "Dynamic Queuing Systems: Behavior and Approximations for Individual Queues and Networks," Ph.D. Thesis, Operations Research Center, Massachusetts Institute of Technology, Cambridge, MA.

The Impact of Ground Delay Program (GDP) Rationing Rules on Passenger and Airline Equity

Bengi Manley

Center for Air Transportation System Research
George Mason University
Fairfax, VA
bmanley@gmu.edu

Lance Sherry

Center for Air Transportation System Research
George Mason University
Fairfax, VA
lsherry@gmu.edu

Abstract— The discrepancy between the demand for arrival slots at an airport and the available arrival slots on a given day is resolved by the Ground Delay Program (GDP). The current GDP rations the available arrival slots at the affected airport by scheduled arrival time of the flights with some adjustments to balance the equity between airlines. Current rationing rules do not take into account passenger flow efficiency in the rationing assignment tradeoff.

This paper examines the tradeoff between flight delays and passenger delays as well as airline equity and passenger equity in GDP slot allocation. A GDP Rationing Rule Simulator (GDP-RRS) is developed to calculate efficiency and equity metrics for all stakeholders. A comparison of alternate GDP rationing rules identified that passenger delays can be significantly decreased with a slight increase in total flight delays. Compared to the traditional Ration-by-Schedule, Ration-by-Aircraft size (RBAC) decreased the total passenger delay by 10% with 0.4% increase in total flight delay, and Ration-by-Passengers (RBPax) decreased total passenger delay by 22% with only 1.1% increase in total flight delay. The disutility of implementing a GDP is minimized with Ration-by-Passengers (RBPax) when passengers as well as airlines are considered in the decision. The current scheme, Ration-by-Schedule (RBS), is preferred only when the system solely focus on airlines. The tradeoffs between airline and passenger equity, and the implications of these results are discussed.

I. INTRODUCTION

The purpose of the air transportation system is the cost-effective, rapid, safe transportation of passengers and cargo. In this way the air transportation system is a significant “engine” of the national economy and provides a service that cannot be achieved by other modes of transportation (Duke and Torres, 2005).

Passenger and cargo demand for air transportation has been growing steadily over the years and is forecast to grow at the same rate for several decades (FAA Forecast, 2007). The growth of air transportation capacity to meet this demand has been lagging (MITRE, 2007). Denver International (DEN), Dallas Fort Worth (DFW) and George Bush Intercontinental (IAH) airports are the only new airports opened in the last 40 years. The capacity of these airports is helpful, but does not solve the current congestion problems at the nation’s busiest airports, such as Newark (EWR) or Chicago O’Hare (ORD). The most congested airports cannot expand due to land and/or environmental problems (Howe et.al. 2003). Further, the full

capacity improvement benefits of Next Generation Air Transportation System are not expected to be operational before 2025.

This imbalance between demand for flights and available capacity is estimated to cost passengers \$3 billion to \$5 billion a year in trip delays (Robyn, 2007). Congestion related flight delays are estimated to cost the financially fragile U.S. airlines an estimated \$7.7 billion in direct operating costs in 2006 (MITRE, 2007). These delays also have environmental and climate change implications as well as regional economic repercussions (Miller and Clarke, 2003).

In the presence of over-scheduled arrivals at airports, Traffic Flow Management (TFM) initiatives are used to resolve the daily demand-capacity imbalance. In particular, the Ground Delay Program (GDP) collaborates with the airlines to manage the scheduled arrival flow into airports consistent with the airport’s arrival capacity. The current GDP rations the arrival slots according to the scheduled arrival time of the flights. This rationing scheme is adjusted to account for penalties suffered by long-distance (e.g. transcontinental flights) flights when arrival capacity increases (e.g. due to improving weather) and the GDP is cancelled. The rationing scheme is also adjusted to more equitably allocate arrival slots between airlines to ensure that one airline (e.g. with a hub operation) is not excessively penalized.

Previous research has examined alternative rationing schemes to: (i) maximize throughput while preserving equity amongst airlines (Hoffman, 2007), (ii) improve airline fairness (Vossen, 2002), and (iii) improve airline efficiency by trading departure and arrival slots (Hall, 1999, 2002).

This paper examines the impact of passenger flow efficiency during a GDP. Three alternate GDP rationing rules were applied to a GDP at Newark Airport. A comparison of the alternate GDP rationing rules identified that passenger delays can be significantly decreased with a slight increase in total flight delays. For example, compared to the traditional Ration-by-Schedule, Ration-by-Aircraft size (RBAC) decreased the total passenger delay by 10% and Ration-by-Passengers (RBPax) decreased total passenger delay by 22%. The tradeoffs in airline and passenger equity, and the implications of these results are discussed.

Section II provides an overview of the GDP’s and previous

research on GDP rationing rules. Section III describes the GDP Rationing Rule Simulator (GDP-RRS). Section IV describes the results of a case study of the alternate rationing rules for a GDP at Newark airport. Section V discusses the implications of these results and future work.

II. BACKGROUND

A. Ground Delay Program (GDP)

The Ground Delay Program (GDP) is a mechanism to decrease the rate of incoming flights to an airport when the arrival demand for that airport is projected to exceed the capacity for a certain period of time. The motivation behind GDP is to convert the foreseen airborne delays into cheaper and safer ground delays (Ball and Lulli, 2004).

FAA first implemented GDPs in times of major-weather-related-capacity reductions at airports after the air traffic controllers strike in 1981 (Donohue, Shaver and Edwards, 2008). Since 1998, GDPs have been implemented under Collaborative Decision Making (CDM). CDM is a joint government-industry effort, which tries to achieve a safer and more efficient Air Traffic Management through better information exchange, collaboration, and common situational awareness. Air Traffic Control (ATC) specialists and CDM participating airlines use Flight Scheduled Monitor (FSM), developed by Metron Aviation Inc., to monitor and model TFM initiatives and evaluate alternative approaches. Fig.1 shows a visualization of a demand-capacity imbalance that warrants a GDP similar to charts available in FSM. In the figure, the airport capacity drops from 100 flights to 75 flights per hour between hours of 17:00 and 22:00. Thus, demand is in excess of capacity during this time period. When GDP is implemented, it brings the scheduled demand to match the airport capacity by delaying flights on the ground. Blue bars in Fig.2 shows the delayed flights, which spill into the hours after the GDP program.

If the ATC specialist decides a GDP is needed, there are three parameters to be set before issuing the program. The first parameter is GDP Start Time and GDP End Time. These are the start and the end times of the program, and they are determined by the scheduled demand and forecasted weather profile at the time of the GDP planning. If a flight is scheduled to arrive at the constraint airport between these times, it will be controlled by the GDP. The second parameter is the "scope" of the program. It specifies the flights departing from which origin airports will be controlled by the GDP. There are two types of scope: 1) Tier-scope identifies the airports included in the program by ATC centers. 2) Distance scope specifies a radius around the GDP airport and exempts any flights departing from origins outside the specified radius. The third parameter is the GDP Program Airport Acceptance Rate (PAAR). It depicts the number of aircraft that can safely land in an hour during GDP.

The overall GDP process under CDM can be summarized as follows: ATC specialists continuously monitor the demand and capacity of airports. When an imbalance between demand and capacity exists for any reason, they model GDP using FSM. If time allows, they send an advisory to all airlines

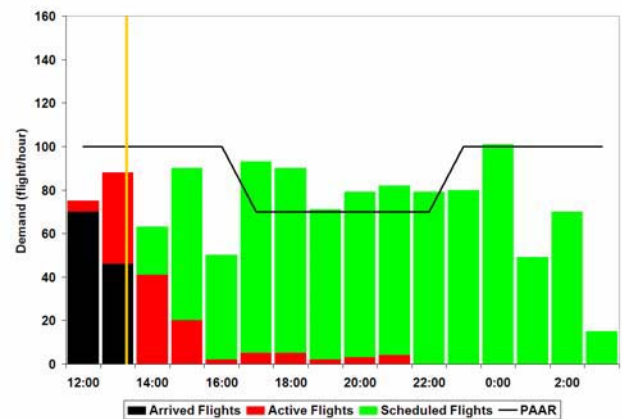


Figure 1: Demand Capacity before GDP (FSM View)

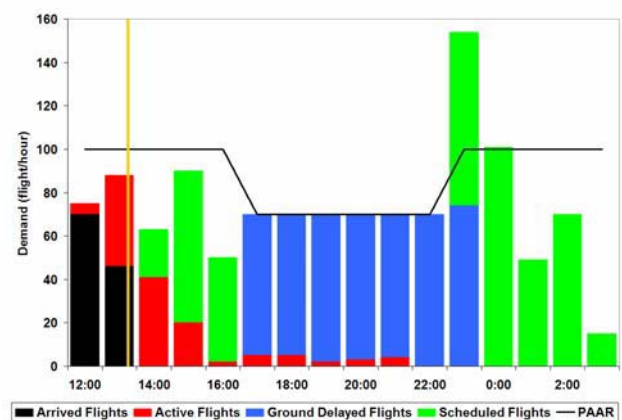


Figure 2: Demand Capacity after GDP (FSM View)

before implementing the program. Airlines check the impact of this proposed GDP on their operations and may opt to cancel some of their flights. Then, specialists reevaluate whether a GDP is still needed. If it is, they run a Ration-by-Schedule (RBS) algorithm and issue each flight its Controlled Time of Arrival (CTA) and Controlled Time of Departure (CTD). Once flight controlled times are received, airlines get a chance to respond by substitutions and cancellations. CTAs depict the arrival slots assigned to each airline, and these slots are now considered to be "owned" by that airline, and airlines can swap any two flights as it fits their business needs as long as both flights can depart by their new CTDs. Following the airline substitutions and cancellations, compression is run. Compression is an inter-airline slot swapping process that fills open slots that airlines are unable to fill through substitutions and cancellations. Compressions are now run automatically whenever an open slot is created. During the GDP, program parameters might need to be revised to account for changing conditions. GDP revisions may lead to further substitutions and cancellations, followed by compression. GDP ends when the GDP End Time is reached or the program is cancelled.

Arrival slots in a GDP are time intervals to achieve PAAR. If PAAR is set at 60 aircraft per hour, the airport can safely land 1 aircraft every minute; therefore, there will be 60 arrival slots to be allocated in an hour during GDP. These slots are uniformly spaced in an hour. The interpretation of an arrival

slot during GDPs is different than that of a “regular” arrival slot. International Air Transport Association (IATA) scheduling guidelines explicitly state that flight schedules planned at the biannual conferences for available airport slots has nothing to do with adjustments to these schedules on the day of operation for air traffic flow management, such as GDPs. The two types of slot allocation are quite different and unrelated (IATA, 2000). The slots owned by airlines under the High Density rule are often interpreted as “the right to schedule or advertise a flight at a specific time”, which entails no explicit connection to a right on the day of operation (Vossen, 2002). Thus, allocation of arrival slots during GDPs can be based on different rationing rules than every day operations.

In a GDP, the available arrival slots are allocated on a “first-scheduled, first-served” basis. This allocation scheme is called “Ration-by-Schedule” (RBS). In other words, arrival slots are allocated based on the flight’s original scheduled time of arrival as published in the Official Airline Guide (OAG) rather than reported departure time on the day of operation. When flights are cancelled or delayed, airlines retain their rights to these arrival slots and can assign flights to these slots based on their own business models. RBS algorithm creates three distinct queues; exempt flights are assigned to slots first, followed by previously GDP controlled flights, then non-exempt flights. A flight can be exempt because the flight is active when GDP is issued or the flight is departing from an origin outside the scope.

B. Trends in GDP Use

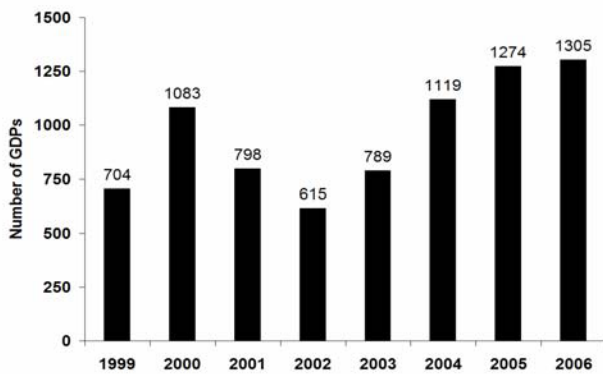


Figure 3: Total Number of GDPs by Year (1/1/1999-12/31/2006)

The use of GDPs has been growing over time as has the number of airports affected by GDPs. Fig.3 shows the growth in the number of GDPs per year as the growth in flight demand increased after 2001.

Fig.4 shows the number of GDPs implemented on a given day between 2000 and 2006. On any given day, there is an 86% probability that flights into at least one airport will experience a GDP. Note: the high number of GDPs per day (10 and above) were GDPs implemented to address airspace congestion due to rare national severe weather days. This use of the GDP is now obsolete and has been replaced by Airspace Flow Programs (AFP).

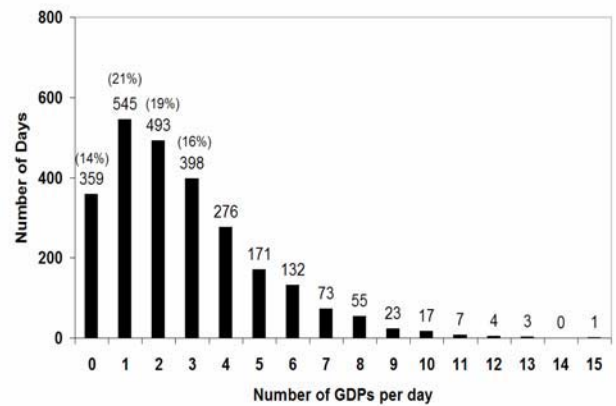


Figure 4: Histogram for Number of GDPs per Day (1/1/2000-12/31/2006)

C. Previous Research

Vossen (2002) examined different GDP rationing rules to achieve fairness among airlines. Fairness was interpreted as allocating delays equally among airlines. Several methods were used to decide how to distribute delays. The “Proportional Random Assignment (PRA)” scheme assigns an available slot to an airline with a probability that is proportional to the number of flights with earlier scheduled arrival times than the slot, following preset axioms. Results show that both RBS and PRA result in similar average airlines delays, even though their underlying philosophies are fundamentally different. PRA may introduce a substantial amount of variance in the assigned delays, which may not be acceptable by airlines. Vossen (2002) also examined methods to deal with achieving slot allocation fairness in the presence of flight cancellations, substitutions and GDP exemptions. These methods are alternatives to the compression where available slots are re-rationed whenever there is an open slot. The results indicate that Greedy Procedure (favors the airline with the earliest flight that can use the slot) and Compression result in very similar flight-slot assignments.

Hoffman (2007) developed a rationing scheme, known as “Ration-by-Distance (RBD)” to maximize airport arrival flight throughput while preserving equity among airlines under changing arrival capacity (due to improving weather). RBD puts flights in order of their distance from the GDP airport and gives preference to long-haul flights. Equity among airlines is total amount of delay assigned to each airline. Results show that if RBS assignment is assumed to have the “perfect” equity, then RBS with distance scope has perfect equity when the GDP is not cancelled, since RBS calculates the slots based on a GDP End Time. When a GDP is cancelled early, RBD significantly reduces delays. Both RBD delay and equity savings gets better when GDP is cancelled 3 or 4 hours early.

Hall (1999, 2002) examined “Arrival-Departure Capacity Allocation Method (ADCAM)”. This rationing method allocates both arrival and departure capacity to airlines according to the published schedule. Airlines can then trade arrivals for departures. The results show that airlines achieved a greater objective value with ADCAM compared to RBS, because it allows airlines to have better connectivity without using more airport capacity. However, some airlines with a

small number of operations can get penalized to a greater extent. Hall (1999, 2002) also examined “Objective-based Allocation Method (OBAM)”. This method assigns arrival slots to GDP flights by maximizing the collective value produced by the airlines. It uses airline objective functions to assign slots, but airlines cannot represent combinatorial or stochastic objectives directly. The motivation behind OBAM is to prevent airlines from scheduling flights they don’t intend to fly. In practice, OBAM requires airlines to pay fees for the slots they receive and these fees may be viewed by airlines as means to introduce new taxes.

Previous research has examined the impact of GDP rationing rules on only airline efficiency and equity. This research is directed toward examining the impact of GDP rules on passenger flow efficiency.

III. GDP RATIONING RULE SIMULATOR (GDP-RRS)

GDP Rationing Rule Simulator (GDP-RRS) developed by Center for Air Transportation Systems Research at George Mason University, investigates the impact of different GDP rationing rules on airlines, passengers, and airports. GDP-RRS calculates GDP efficiency and equity metrics that result from GDP planning for airlines, passengers and the GDP airport. Fig.5 shows three main components of the model.

First module inputs a flight schedule and airport capacity profile, and then determines whether a GDP is needed. This module captures the decision making process of an ATC specialist. If a GDP is needed, then the second module is activated.

“GDP Slot Assignment Module” assigns slots to flights that are scheduled to arrive at the GDP airport during the program. Fig.6 shows the pseudo algorithm with nine main steps. Steps 1-5 result in Planned CTDs and CTAs, which are sent to airlines for substitution and cancellations. Steps 6-7 show the simulated decision making for airlines in Airline Substitutions and Cancellations module. Steps 8-9 input airline updated CTDs and CTAs, and result in the main CTDs and CTAs that airlines are expected to comply with after the compression algorithm.

1. Calculate Required Variables for Each Flight: Scheduled Gate Time of Arrival and Scheduled Gate Time of Departure for each flight are inputs to the model. Scheduled runway times, which are used in the GDP slot assignment, are calculated from these inputs assuming 10 minute taxi times. Estimated Time Enroute (ETE) for each flight is the difference between Scheduled Runway Time of Arrival (SRTA) and Scheduled Runway Time of Departure (SRTD). “Available Seats” is the average yearly number of seats for a given aircraft type assigned to each flight (ETMS database). “PAX” is the number of passengers on-board and is calculated as Available Seats on a flight multiplied by its load factor. Load factor is the average yearly monthly load factor for a given airline from a given origin (BTS database). For international origins and airlines coming from unknown origins, the default load factor is 100%.

2. Find Flights in GDP: All flights going to the GDP airport are assigned control times. However, the delay as a

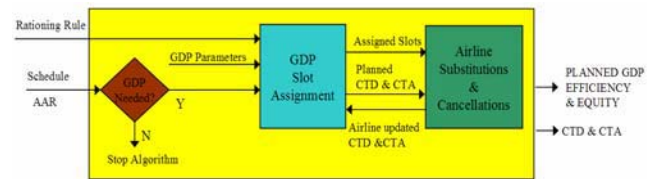


Figure 5: GDP Rationing Rule Simulator

result of the capacity reduction is only distributed among the flights that are controlled by the GDP. For a flight to be controlled, it needs to fulfill the below requirements:

- Flight’s SRTA is between GDP Start and End Time.
- Flight is not originated from an international airport.
- Flight’s departure airport is in GDP scope.

3. Create Priority Queues: Two priority queues are created for all flights scheduled to arrive at the airport between GDP Start and End Times. Exempt Flights queue has precedence over the remaining flights. Exempt Flights queue contains international flights and flights departing from airport outside the GDP scope.

4. Create Slots: The number of slots available for distribution depends on the PAAR. Airport capacity profile is an input to the model. Slot size is the time in minutes between two available slots. The number of slots created depends on the number of scheduled flights. Slot times are uniformly distanced based on Slot Size starting from GDP Start Time.

5. Assign Slots to Flights: The assignment of slots to flights is done by queue type. Exempt Flights are assigned their slots first based on an ordering of increasing SRTA. Then, non-exempt flights are assigned their slots based on an ordering depicted by the GDP rationing rule. For each flight, algorithm searches for the earliest slot which has the slot time equal to or later than the flight’s SRTA. When such a slot is found, if the flight’s SRTA is later than the slot time, the flight’s CTA is the same as the slot time. If the flight’s SRTA is between the chosen slot time and the next slot, then the flight’s CTA is the same as its SRTA. CTD is back-calculated using CTA and ETE for the flight. These CTAs and CTDs are sent to Airline Substitutions and Cancellations Module.

6. Cancel Flights: Each flight is cancelled randomly based on a probability distribution for a given airline from a given origin airport in the year that GDP is implemented.

7. Substitute Flights: Substitution for an airline is only possible if that airline has cancelled a flight. If there is a cancellation, the slot opened can be used by a flight from the same airline if the new flights CTA is later than the open slot time or the flight can arrive at the new assigned slot. If such a substitution is made, the flight’s CTA and CTD are recalculated and its previous slot is open for another possible substitution. Substitution algorithm stops when there is no further substitutions can be made. Substitution algorithm uses two different strategies to simulate airline behavior. Strategy 1 orders an airline’s all flights by increasing SRTA and gives earlier scheduled flights precedence for substitution. This strategy minimizes an airlines overall GDP flight delay. Strategy 2 orders an airline’s flights by decreasing PAX and

gives precedence to flights carrying more passengers. This strategy results in less overall GDP passenger delays. At the end of this step, passenger delays are calculated as well as flight delays. It is assumed that a cancelled flight's passengers will be transferred to another flight from the same origin. However, due to high load factors, some passengers may not be accommodated. It is assumed that these passengers will leave the airport the next day at 6am.

8. **Run Compression:** Compression tries to fill in the unused slots after airline substitutions and cancellations. All slots are sorted in order of their slot times. If an unassigned slot is found, algorithm checks if the delay of any non-cancelled flight can be reduced by assigning the flight to this slot instead. First, flights from CDM member airlines are considered in the order of their ranking due to the chosen GDP rationing rule, followed by the remaining flights. Assignment is done only if the flight can make it to its new assigned slot. If such a flight is found, flight's CTA and CTD are recalculated. If no such flight is found, then slot remains unassigned. Algorithm stops when all unassigned slots are checked.

9. **Issue CTA and CTD:** The last step in the algorithm is to validate the slot assignments before CTDs and CTAs are issued. Algorithm checks if each flight is assigned to only one slot, if each slot is assigned to only one flight, and if each flight's SRTA is equal to greater than assigned slot time. If there is a problem, algorithm goes back to Step-5. If not, Planned GDP efficiency and equity metrics are calculated.

Steps 1-5 and Steps 8-9 are simplified versions of the current GDP algorithm. Differences between GDP-RRS and the current GDP algorithm are shown in italics in Fig.6. These additions are required to simulate new GDP rationing rules and calculate passenger-based metrics. Current GDP algorithm only runs Ration-by-Schedule (RBS) scheme, and only calculates flight-based metrics.

IV. RESULTS

To examine the impact of passenger flow efficiency and airline equity in a GDP, three alternate rationing rules are examined.

1. **Ration-by Schedule (RBS)** allocates available slots among GDP flights in the order of their scheduled arrival times. The earlier flights are given precedence over later flights. If there are two flights scheduled to arrive at the same time, one of them is randomly selected to be the first for slot assignment.

2. **Ration-by-Aircraft Size (RBAC)** rations available slots by aircraft size. RBAC creates three priority queues for three categories of aircraft size considered: Heavy, Large and Small. Flights under Heavy category are assigned their slots first, followed by Large and Small categories. 23% of the flights in the study fall under Heavy, 77% in Large, 1% in Small category. Heavy, Large, and Small category flights are re-ordered by their scheduled arrival times in a given category. Thus, if there are two flights in the same category (Heavy-Heavy), RBAC chooses the flight with the earlier scheduled arrival time for slot assignment first. If two flights are in the

same category and are scheduled to arrive at the same time, one of them is picked randomly to be the first for the slot assignment.

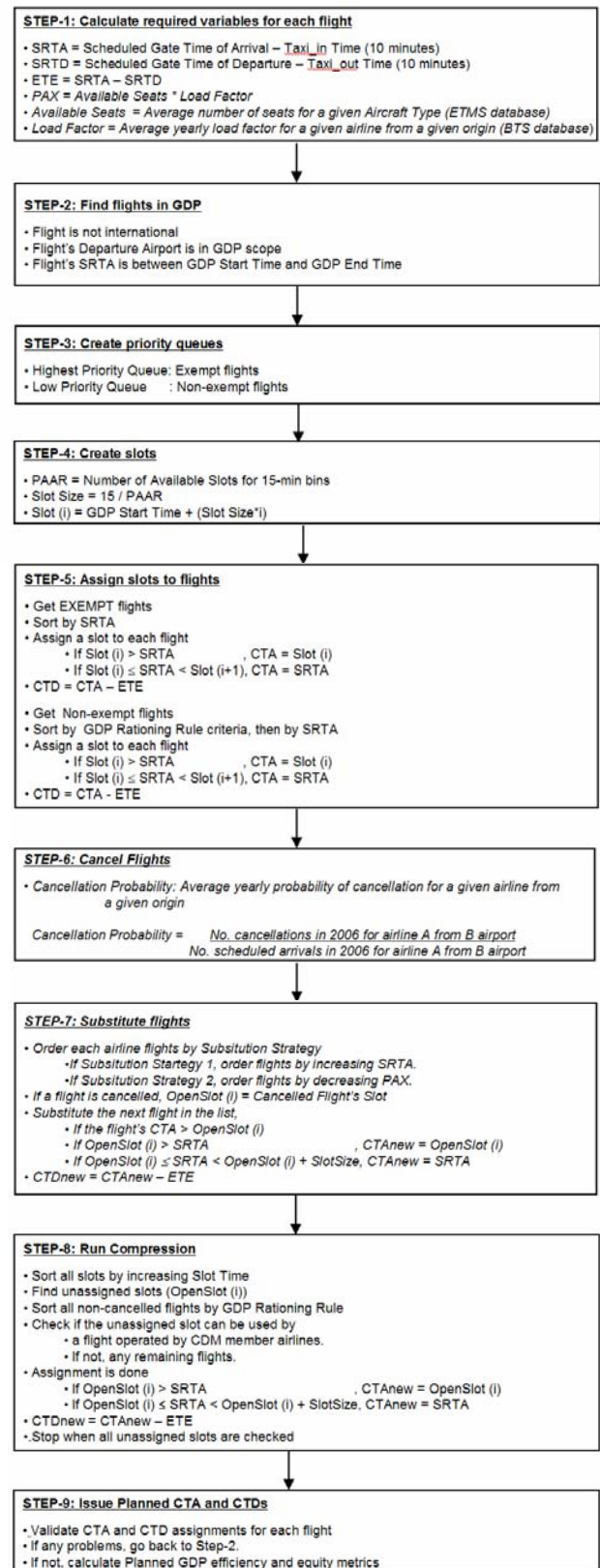


Figure 6: GDP Slot Assignment Module Pseudo Algorithm

3. Ration-by-Passengers (RBPax) rations available slots by the number of passengers carried on each flight. RBPax algorithm puts flights in the order of passengers on board. Flights carrying more passengers are given precedence over flights carrying fewer passengers. If there are two flights scheduled to arrive at the same time carrying the same number of passengers, RBPax chooses the flight with the earlier scheduled arrival time for slot assignment first. If two flights are in the same category and are scheduled to arrive at the same time, then one of them is chosen randomly to be the first for slot assignment.

Substitution strategy 1 is used in this case study.

Case Study GDP at Newark Liberty Airport

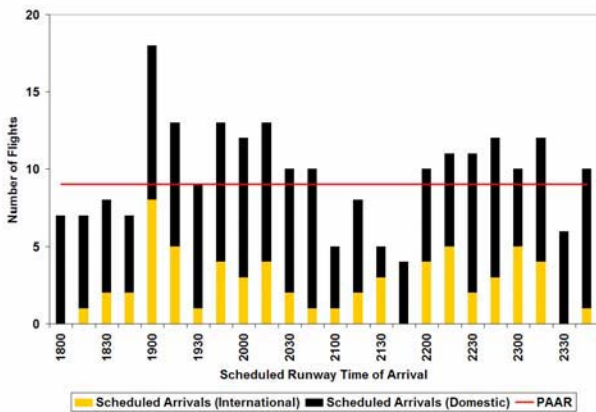


Figure 7: Flight List before GDP is Implemented

A GDP is implemented at EWR on June 10th, 2006 starting at 18:00 and ending at 23:59 GMT time. Fig.7 shows the scheduled flights arriving at EWR during this time before any cancellations or substitutions. Projected Airport Arrival Rate (PAAR), the red line, is set at 9 flights per 15 minutes. This PAAR value falls in the historic range for EWR GDPs. Scope is chosen as all domestic airports (Tier scope: All) with international flights being exempt. Yellow bars in Fig.7 represent the exempt flights.

There are 231 flights between GDP Start and End Time, 63 of which are international. GDP delay is split among 168 domestic flights. There are 26 major airlines coming from 109 different origins carrying 25,501 passengers (11,516 international and 13,985 domestic passengers). Among these origin airports, the most number of scheduled flights are from Chicago (ORD: 7), Atlanta (ATL: 6), Los Angeles (LAX: 6), and Boston (BOS: 6). It is interesting to see departure airports in 100 nautical miles radius from EWR (LGA, FRG, and BDL). Even though the number of passengers on these flights is not very large, the flight categories can be different. For example, the flight from La Guardia (LGA) is a “Small” carrying 8 passengers, whereas the flight from Windsor Locks, CT (BDL) carrying 32 passengers is a “Large”. There are 15 cancellations, and the cancelled flights are kept the same in all three rationing rule simulations.

The results of the case study are summarized in Fig.8. All three rationing rules result in different trade-offs for the system. Both RBPax (blue) and RBAc (red) trades off more

flight delays with less passenger delays compared to the current rationing rule (RBS). Compared to RBS, RBAc (red) decreases total passenger delay by **10%** (67,288 minutes less delay) with a 0.4% increase in total flight delay (12 minutes). The biggest improvement in efficiency is achieved by using RBPax. Moving to RBPax from RBS decreases total passenger delay by **22%** (144,407 minutes less delay) with only 1.1% increase in total flight delay (31 minutes).

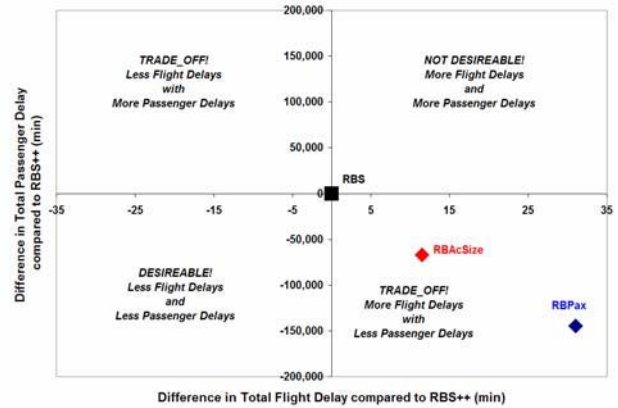


Figure 8: GDP Efficiency Comparison between RBS, RBPax and RBAc

Since all GDP rationing rules result in a trade-off, a decision can be reached using utility theory. Disutility of implementing a GDP can be calculated using different weights for two efficiency metrics calculated; Total flight delay and total passenger delay as a result of a chosen rationing rule. Below is the disutility calculation for RBS as an example:

$$\text{Disutility}_{\text{RBS}} = w_{\text{TFD}} * \left(\begin{matrix} \text{TFD}_{\text{RBS}} \\ \text{TFD}_{\text{RBS}} + \text{TFD}_{\text{RBAC}} + \text{TFD}_{\text{RBPax}} \end{matrix} \right) + (1 - w_{\text{TFD}}) * \left(\begin{matrix} \text{TPD}_{\text{RBS}} \\ \text{TPD}_{\text{RBS}} + \text{TPD}_{\text{RBAC}} + \text{TPD}_{\text{RBPax}} \end{matrix} \right)$$

- w_{TFD} : Weight of Total Flight Delay
- (1-w_{TFD}) : Weight of Total Passenger Delay
- TFD_{RBS} : Total Flight Delay as a result of RBS
- TFD_{RBS} : Total Passenger Delay as a result of RBS

Fig.9 shows the disutility calculated for EWR case study using different weights. As the weight of the total flight delay gets larger, the system focus moves further away from passengers to airlines and flights. Fig.9 shows that current rationing rule (RBS) is acceptable only when the system solely focus on flights. However, when the passengers are considered, RBPax gives the minimum system disutility, followed by RBAc.

Total flight and passenger delay values are important metrics. However, they don't imply any information about the fairness of the delay distribution. Equity becomes an issue whenever goods, in this case available arrival slots, which are held in common by a group of users, airlines, must be allotted to them individually (Young, 1994). In the case of GDPs, equity means distributing fairly among all involved stockholders. Airline Equity by Flights (Fig.10) and Passenger Equity by Distance (Fig.11) captures this from the view point of airlines and passengers.

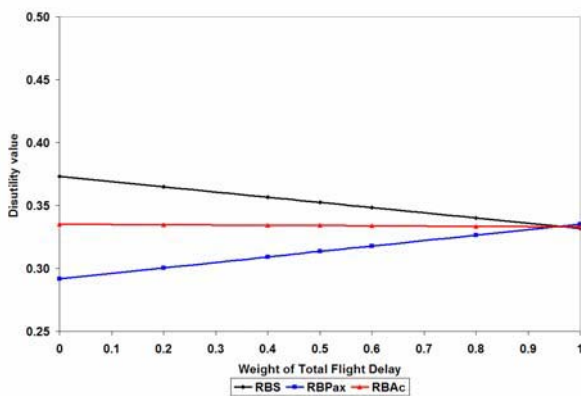


Figure 9: Disutility of GDP for RBS, RBPax and RBAC

From airlines' perspective, the more flights an airline has the more delay it should be assigned. Airline Equity by Flights is calculated as the ratio of an airline's flight delays over the total GDP flight delay divided by the ratio of that airline's flights in the GDP over all GDP flights. "Perfect equity" is represented as 1. If an airline's equity is smaller than 1, the airline is given less delays than is fair. Conversely, if an airline's equity is greater than 1, than the airline is given more delays than its fair share. Fig.10 shows the equity for airlines in the GDP. Airlines which have only international flights are omitted in this figure, since all their flights will be exempt. For each airline, the number of flights in GDP is also given in parenthesis.

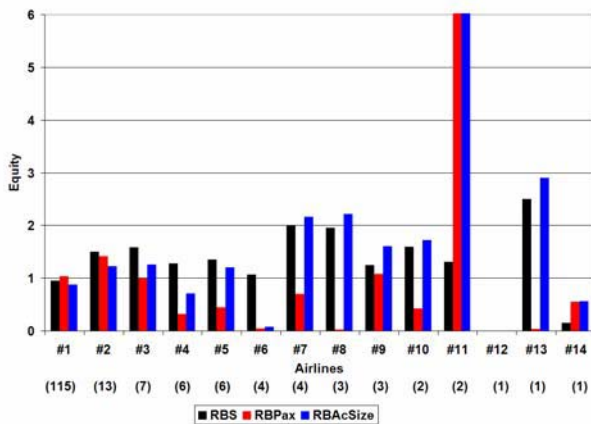


Figure 10: Airline Equity by Flights under RBS, RBPax and RBAC

As expected, the results are different for different airlines. For Airline 1, the dominant carrier, the three rationing rules do not make much difference in its overall delays. However, for Airline 2, the main competitor, the equity does not reach 1 but it moves in the right direction with new rationing rules. All airlines have less delay with RBPax, except for airlines 2 and 11. Airline 11 has two flights, one from Cleveland (CLE) and one from Long Island (FRG). FRG flight is assigned the same slot in all rationing rules and is cancelled later on. CLE flight is a general aviation flight carrying only 6 passengers, scheduled at 19:05. As seen from Fig.7, this is a very busy period for EWR, and RBPax assigned this flight a delay of 306 minutes. Since there are only 2 "small" category flights, RBAC also assigned a very high delay. On the other hand,

Airline 7 enjoys RBPax whereas RBS and RBAC results in the same unfair equity level. It has 4 flights, 3 of which are from DTW. All four flights are "large", carrying 71-95 passengers. RBS gives these flights higher delays than RBPax because they are scheduled at the busy times of EWR. RBAC also gave high delays because all flights are "large". Since 77% of all flights are large, the RBAC delay assignment is very close to RBS. However, RBPax further distinguishes flights with the number of passengers on board and assigned less delay to these flights. (Note: Equity metrics are calculated over a long-term. For the purposes of this case study, only one day of data is used.)

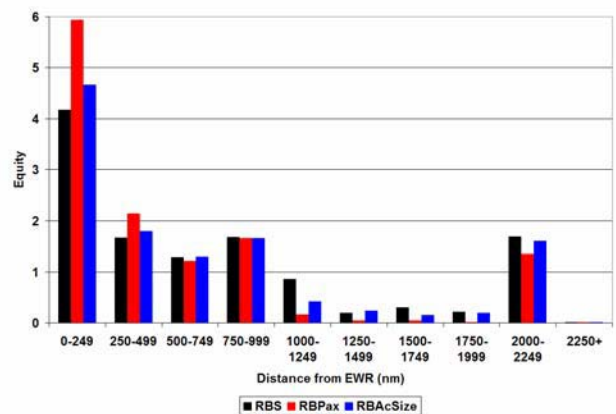


Figure 11: Passenger Equity by Distance under RBS, RBPax and RBAC

From passengers' perspective, the passenger delay they encounter is important rather than the flight delay itself. Flight-based metrics cannot accurately reflect passenger travel experience (Wang, 2007). Flight cancellations reduce total flight delay while increasing total passenger delays, especially when the load factors are high. Passenger Equity by Distance (Fig.11) compares how much passenger delay is assigned to passengers flying from a distance group compared to the total number of passengers in the GDP. In other words, the more passengers a distance group has, the more passenger delay it will be assigned. Passenger Equity by Distance is calculated as the ratio of passenger delays for a given distance group over the total GDP passenger delay divided by the ratio of the number of passengers from that distance group over all passengers encountering the GDP. "Perfect equity" is again represented as 1. Fig.11 shows that the long-haul passengers are encountering much less delays than short-haul passengers in all three rationing rules. This is due to the fact that longer-haul flights are scheduled less frequently with larger aircraft having more seats, whereas short-haul flights are scheduled more frequently with smaller aircraft. The higher load factors (100% for international flights) also result in favorable passenger delays for longer-haul flights. Difference in equity is more pronounced for RBPax than the other two rationing rules. (Note: Equity metrics are calculated over a long-term. For the purposes of this case study, only one day of data is used.)

Fig. 12 shows the total inequity as a result of all rationing rules. Total inequity for a given rationing rule is calculated as the sum of absolute differences between a category's equity and the "perfect" equity (1). Figure shows RBS clearly results

in the smallest airline inequity compared to the other rules. Passenger inequity is also smaller with RBS. However, the favorable equity by RBS is achieved at the expense of 144,407 minutes more passenger delays. Passenger equity values are very dependant on the airline scheduling choices and can be improved by upgauging.

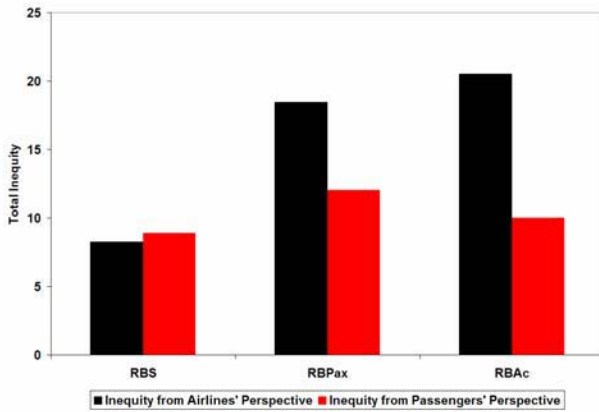


Figure 12: Inequity Comparison for RBS, RBPax and RBAC

V. CONCLUSION

The case study of GDP with alternate rationing rules at EWR demonstrates the impact of GDP rationing rules on passenger flow efficiency and on airline equity. Adjusting the rationing rules to maximize the flow of passengers (and cargo) results in significant reductions in overall passenger trip delays. These delays are achieved with small changes in overall flight delay. Airline equity is adjusted in favor of larger airlines. Addressing this issue is an area of future work.

The results of the case study at Newark Liberty International Airport (EWR) are as follows:

- All three GDP rationing rules resulted in the different trade-offs between airlines and passengers.
- Ration-by-Aircraft size (RBAC) decreased the total passenger delay by 10% compared to RBS with a 0.4% increase in total flight delay.
- Ration-by-Passengers (RBPax) decreased total passenger delay by 22% compared to RBS with 1.1% increase in total flight delay.
- Ration-by-Passengers (RBPax) results in the minimum disutility for the air transportation system when both airlines and passengers are concerned. RBS is preferred choice only if airlines are the main focus of the system.
- Ration-by-Schedule (RBS) results in the minimum total inequity for both airlines and passengers. However, this is achieved at the expense of a large efficiency loss due to high passenger delays.

The application of alternate GDP rationing rules has broader implications. In principle, GDP rationing rules create priority queues which give preference to the compliant flights. As a consequence the rationing rules incentivize airline behavior. For example, the Ration-by-Passengers rule could,

in the long-run, result in the migration of airline fleets to larger sized aircraft that would increase the passenger flow capacity. This would improve the efficiency of the air transportation system. This incentive does not directly result in reduced frequency, but reduced frequency may be a by-product of upgauging.

Results presented here are the outputs of the GDP Planner with substitution strategy 1. For future work, these results will be compared against the results of GDP Flight Simulator to see the differences between planned and actual metrics. Results can be further improved by comparing airport metrics to airline and passenger metrics.

ACKNOWLEDGMENT

Thanks for technical comments and suggestions from Mark Klopfenstein, Michael Brennan, Terry Thompson (Metron Aviation), Maria Consiglio, Brian Baxley, and Kurt Nietzsche (NASA-LaRC), George Donohue, John Shortle, Michael Bronzini, Rajesh Ganesan, David Smith, Vivek Kumar (GMU/CATSR), Joe Post, Dan Murphy, Stephanie Chung, Kimberly Noonan (FAA ATO-P). Thanks for program support to Harry Swenson, Michael Landis (NASA). This research was funded by NASA NRA NNA07CN23A, and internal GMU/CATSR Foundation Funds.

REFERENCES

- (Ball and Lulli, 2004) M.O. Ball, and G. Lulli, "Ground Delay Programs: Optimizing over the included flight set based on distance," *Air Traffic Control Quarterly*, vol.12, pp. 1-25, 2004.
- (Donohue, Shaver and Edwards, 2008) G.L. Donohue, R.D. Shaver, and E. Edwards, "Terminal chaos: Why U.S. air travel is broken and how can we fix it," AIAA press, forthcoming Spring 2008.
- (Duke and Torres, 2005) J. Duke, and V. Torres, "Multifactor productivity change in the air transportation industry," *Monthly Labor Review*, March 2005.
- (FAA Forecast, 2007) U.S. Department of Transportation Federal Aviation Administration Policy and Plans, "FAA aerospace forecast fiscal years 2007-2020," 2007.
- (Hall, 1999) W.D. Hall, "Efficient capacity allocation in a collaborative air transportation system," PhD Thesis in Operations Research, Supervisor: A. Odoni, Massachusetts Institute of Technology, Cambridge, MA, 1999.
- (Hall and Peterson, 2002) W.D. Hall, and E. Peterson, "ASCENT: A network-wide simulation of air traffic flow management incorporating airline decisions," *Proceedings of the American Control Conference*, Anchorage, AK May 8-10, 2002.
- (Hoffman, Ball, and Mukherjee, 2007) R. Hoffman, M. Ball, and A. Mukherjee, "Ration-by-Distance with equity guarantees: A new approach to Ground Delay Program planning and control," the 7th ATM R&D Seminar, Barcelona, Spain, 2007.
- (Howe et.al, 2003) B.Howe, C. Richardson, M. Gordon, S. Russell, S. Sprague, G.L. Donohue, and F. Wieland, "Potential air traffic congestion solution: Slot allocation by auction method," *Proceedings of the 2003 Systems and Information Engineering Design Symposium*, 2003.
- (IATA, 2000) International Air Transport Association, *Worldwide Scheduling Guidelines*, 3rd Edition, December 2000.
- (MITRE, 2007) The MITRE Corporation, Center for Advanced Aviation System Development, "Capacity needs in the national airspace system 2007-2025: An analysis of airports and metropolitan area demand and operational capacity in the future," May 2007.

(Miller and Clarke, 2003) B. Miller, and J.P. Clarke, "The hidden value of air transportation infrastructure," The 7th International Conference on Technology Policy and Innovation, Monterrey, Mexico, June 10-13, 2003.

(Robyn, 2007) D. Robyn, "Reforming the air traffic control system to promote efficiency and reduce delays," The Brattle Group, October 29, 2007.

(Vossen, 2002) T. W.M. Vossen, "Fair allocation concepts in air traffic management," PhD thesis, Supervisor: M.O. Ball, University of Maryland, College Park, Md, 2002.

(Wang, 2007) D. Wang, "Methods for analysis of passenger trip performance in a complex networked transportation system," PhD thesis in Information Technology, Supervisor: L.Sherry, George Mason University, Fairfax, VA, 2007.

(Young, 1994) H. P. Young, "Equity: In theory and practice," Princeton University Press, Princeton, New Jersey, 1994.

Passenger Flow Simulation In A Complex Networked Transportation System

Danyi Wang
Metron Aviation Inc.
Herndon, VA, U.S.A.
wang@metronaviation.com

Lance Sherry
CATSR
George Mason Univ.
lsherry@gmu.edu

George Donohue
CATSR
George Mason Univ.
gdonohue@gmu.edu

Bert Hackney
Metron Aviation Inc.
Herndon, VA, U.S.A.
hackney@metronaviation.com

Abstract—Passenger trip time performance is positively correlated with passenger satisfaction, airfare elasticity, and airline profits. Researchers have demonstrated that flight metrics are a poor proxy for passenger trip experience. Trip delays experienced by passengers due to missed connections and cancelled flights are not negligible.

This paper describes a passenger flow simulation which captures the asymmetric and unique passenger trip on-time performance and reflects the complexity and significance of the impact of a small set of cancelled flights and missed connections on passenger trip delays. It measures system performance from the flying public's view. Furthermore, it enables researchers to conduct experiments outside the range of historical data.

The results of this research provide decision makers with improved metrics for future investment decisions and better tools to manage the system. The passenger flow simulation model also provides the means to perform analysis for proposed changes to the system.

Keywords—on-time performance; passenger flow; performance metrics; passenger trip time

I. INTRODUCTION

The purpose of the Air Transportation System (ATS) is to provide safe and efficient transportation service of passengers and cargo. The on-time performance of a passenger's trip is a critical performance measurement of the Quality of Service (QoS) provided by any Air Transportation System. QoS has been correlated with airline profitability, productivity, customer loyalty, and customer satisfaction [1].

Bratu et al. have shown that official government and airline on-time performance metrics (i.e. flight-centric measures of air transportation) fail to accurately reflect the passenger experience and underestimate the disruption on passenger trip time caused by cancelled flights and missed connections [2] [3] [4]. Flight-based metrics do not include the trip delays accrued by passengers who were re-booked due to cancelled flights or missed connections. Also, flight-based metrics do not quantify the magnitude of the delay (only the likelihood) and thus fail to provide the consumer with a useful assessment of the impact of a delay [5].

Research on passenger trip delay is limited because of the unavailability of proprietary airline data, which is also protected by anti-trust collusion concerns and civil liberty privacy restrictions. Wang et al. developed a set of algorithms

designed to compute estimated passenger trip delay (EPTD) based on publicly available databases [6] [7] [8]. Results show disproportionately high passenger trip delays generated by cancelled flights. Cancelled flights accounted for only 1.4% of total scheduled flights in 2006, but they generated 39% of total EPTD. On average, passengers scheduled on cancelled flights in 2006 experienced 607 minutes of delay, while passengers scheduled on delayed flights experienced a much lower delay of 56 minutes. Except for the disproportionately high EPTD due to cancelled flights, Wang et al. proved passenger trip delay is a stochastic phenomenon that has asymmetric performance in terms of routes, airports, and time of year. Half of the total EPTD is generated by a smaller portion of routes (17%), airports (26%), and months (42%). Altogether, passenger behavior in the passenger tier of the air transportation system differs from flight behavior in the vehicle tier of the system.

Wang et al. designed the algorithm based on segment data, which doesn't contain flight connection information. As a consequence, the analysis does not include the passenger trip delay caused by missed connections. Moreover, expansion of the air transportation system is trending out of the historical operation range with record high load factors, operations, and enplanements. This trend prohibits using historical data for analysis, since historical data cannot predict the impact of future policy changes on passenger trip time. In this paper, a passenger flow simulation (PFS) is developed to perform "future option design evaluation." The PFS enables researchers to conduct experiments outside the range of historical data and estimates passenger trip delay not only due to delayed and cancelled flights, but also due to missed connections.

Section II of the paper describes the underlying concepts of the passenger flow simulation, PFS hierarchy, structure, algorithm, and results. Section III describes the experimental design for the PFS to identify significant factors for passenger trip performance and to perform sensitivity analysis.

II. PASSENGER FLOW SIMULATION

The operational evolution plan (OEP) 35 airports are the nation's busiest airports defined by the FAA [9]. They have the greatest number of operations and account for 73% of total enplanements and 79% of total operations in the air transportation system [10]. The passenger flow simulation is a closed network formed by 34 of the OEP-35 airports. Honolulu International Airport (HNL) is excluded due to its geographic location and negligible impact on the network.

A. The Underlying Concept

Air transportation simulations of flight movement do not capture the passenger flow and connecting process. In the air transportation network, passengers cluster together into groups to fly from one airport to another. After arrival at the destination airport, this group of passengers breaks up: nonstop passengers make connections to ground transportation, and connecting passengers continue their trips by re-clustering with other passengers. Compared with flight movement, passenger movement

- Simulates passenger behavior instead of flight behavior;
- Converts flight information, such as arrival time and origin and destination airports, into attributes of passengers or groups of passengers;
- Converts flight schedules into clustering and scattering rules followed by passengers.

The passenger flow simulation is built to simulate this dynamic clustering and scattering process of passenger flow in the system. Air transportation simulations of flight movement do not capture the passenger flow and connecting process.

B. Colored Petri Net Modeling Tool

A Petri Net is a graphical and mathematical modeling tool. It is well-suited to modeling public transportation networks [11] and has been used to model the passenger connecting process in a public bus transportation system [12].

For accurate modeling of a complex transportation system like the air transportation system, a more complicated extension of Petri Net is required. In this paper, a hierarchical, timed, Colored Petri Net (CPN) is built using CPN Tools to

simulate passenger flow and connecting processes in the system. CPN Tools is a graphical user interface for editing, simulating, and analyzing Colored Petri Nets [13]. CPN Tools can model the complex level of interactions in the air transportation system visually by creating nodes, transitions, and arcs in the model environment. This visual modeling environment allows users to track and understand the behavior of each passenger easily.

The concept of “color” distinguishes tokens (or resources) in the net. “PaxGroup” is defined as a color in PFS:

$$\text{Color: PaxGroup} = (\text{Origin}) * (\text{Dest}) * (\# \text{ of Pax Loaded}) * (\text{Aircraft Size}) * (\text{SchDepTime}) * (\text{SchArrTime}) * (\text{Carrier}) * (\text{FlightIndex}) * (\# \text{ of Local Pax}) \text{ timed};$$

For example, the PaxGroup (DCA, ORD, 165, 200, 730, 850, 13, 45, 165)@+750 in Figure 1 represents a group of 165 passengers, loaded on United Airlines flight 45 with 200 seats, scheduled to depart from DCA at system time 730, and arrived at ORD at system time 850. However, this flight actually departed at system time 750, which is 20 minutes later than scheduled.

Tokens in places (circles) represent available resources to enable a transition (rectangle). The left part of Figure 1 shows one group of passengers in place “local pax”¹ and two groups of connecting passengers in place “conn pax”. The first group of 20 connecting passengers arrived at the gate at time 715, and the second group of 15 connecting passengers arrived at the gate at 770. When the flight departed at time 750, the first group of 20 connecting passengers were loaded on time, whereas the second group of 15 connecting passengers missed their connections, since they arrived at the gate after the flight departed. The departing flight was scheduled to load 35 connecting passengers and depart with 200 total passengers at system time 730, but it actually loaded 20 connecting

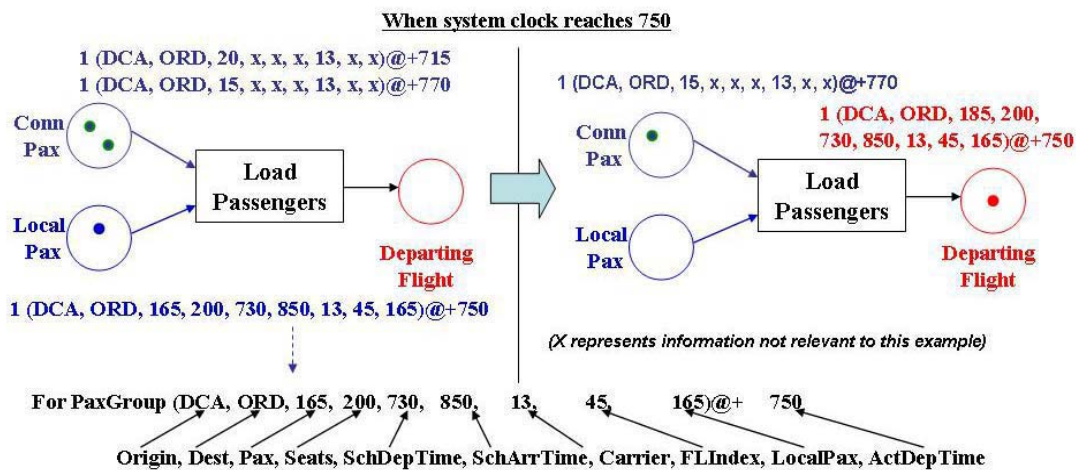


Figure 1 Example: passenger loading process in CPN

¹ Local passengers are passengers who have just appeared in the air transportation system. They could either be nonstop passengers from DCA to ORD or connecting passengers whose first leg flight is from DCA to ORD.

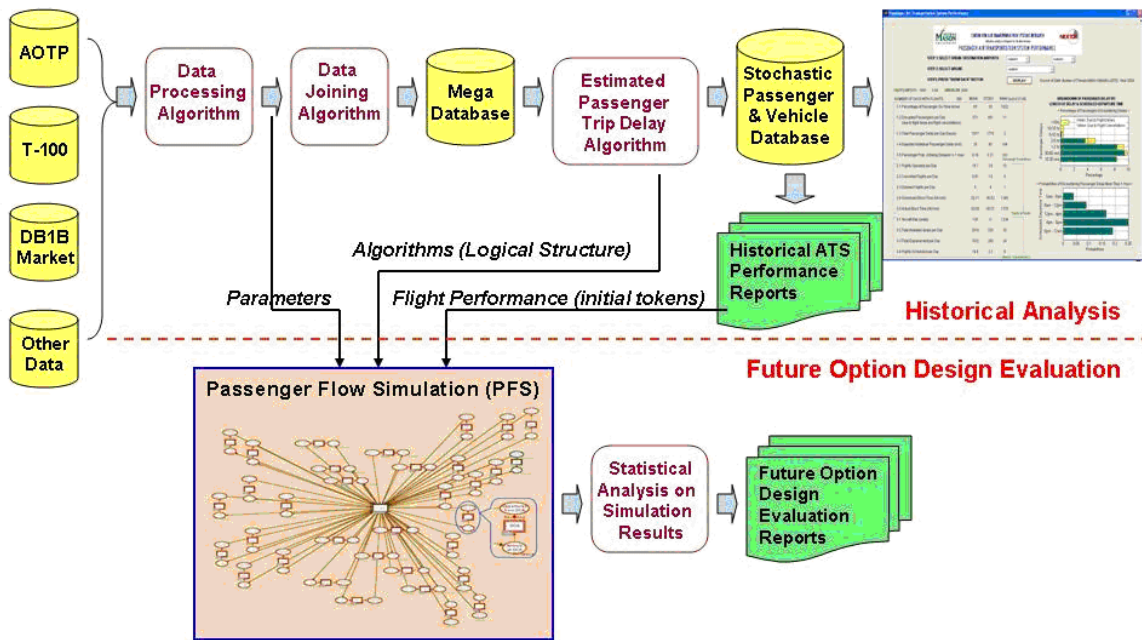


Figure 2. Big Picture: Correlation between Algorithms and PFS

passengers and departed at system time 750 with 185 total passengers. The right part of Figure 10 shows the CPN after the transition: one group of 15 connecting passengers missed their connecting flight, and a group of 185 passengers (165 local + 20 connecting) were ready to gate out.

C. PFS Overview

Figure 2 depicts the correlations between algorithms and PFS. The algorithm section above the dotted line targets the “historical analysis.” In this section, different algorithms are designed to manipulate the data in different data processing phases. The “historical analysis” section sets the stage for “future option design evaluation.” As shown in Figure 2, processed data, algorithms, and the analysis report are embedded into the passenger flow simulation model as parameters, logical structure, and initial tokens. In other words, the parameter setting and passenger flow control in the PFS are based on historical statistics calculated by algorithms.

TABLE 1 OVERVIEW OF PASSENGER FLOW SIMULATION (PFS) STRUCTURE

Aspects	Description
Airports	OEP-34 airports (excluding HNL)
Routes	1,030 routes formed by OEP-34 airports
Carriers	17 major carriers
Daily Flights	8,500
Daily Enplanements	900,000
PFS Modes	Deterministic and Stochastic
Hierarchy	3-level
Places	580
Transitions	343
Initial Tokens	20,000
Functions	42

The network structure of PFS is formed by 34 of the OEP-35 airports and the 1030 routes between pairs of these airports². Passengers flow from one airport to another through the existing routes. Table 1 gives an overview of the PFS structure.

The PFS has two modes: deterministic and stochastic. They share the same PFS structure but use different functions and parameter values. In the deterministic PFS, passengers scheduled on a specific flight (e.g. UAL123 on route ORD and PFS DCA) will arrive at the actual arrival time (e.g. 1145). But in the stochastic PFS, the flight time is determined by a set of -

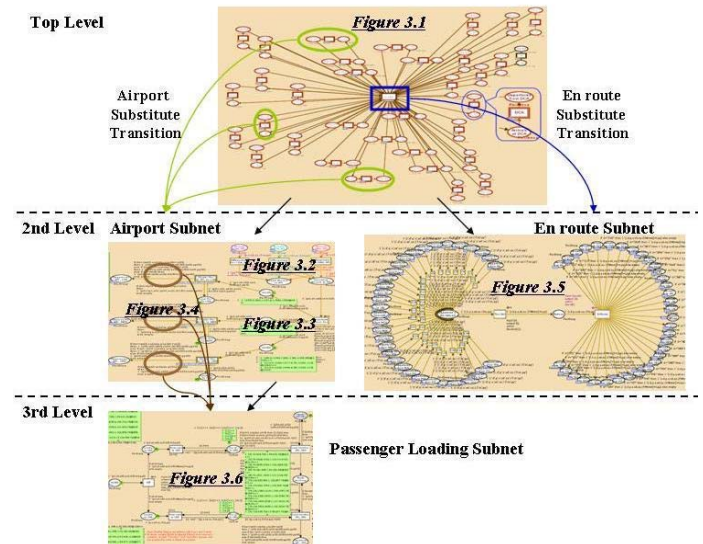


Figure 3. Big Picture: Correlation between Algorithms

² Not all of the 1122 possible city pairs are served by direct flights.

random number generators with specific means and standard deviations (e.g. flight time for UAL flights on route ORD-DCA is a normal random variable with $\mu=100$ minutes and $\sigma=15$ minutes). Thus the arrival time for flight UAL123 is a stochastic value generated as “DepTime + NormRNG ($\mu=100$, $\sigma=15$)”. In summary, the deterministic PFS is a pure conversion between flight performance and passenger performance, whereas the stochastic PFS allows flexibility and is more suitable for future option design evaluation.

D. PFS Hierarchy

The PFS has three levels, as shown in Figure 3. Airport and En Route subnets are represented as substitute transitions in the top-level net of PFS. These figures will be decomposed and described in the following paragraphs. Locations of the zoomed-in subfigures are labeled in Figure 3 (Figure 3.1 ~ 3.6). In addition, large figures showing the top-level, second-level and the third-level nets are available in Appendix A.

As shown in Figure 3.1, the top-level page depicts 34 airport substitute transitions, a single en route substitute transition, and the 68 ports connecting them. The zoomed-in inset shows an aggregated departure gate and an aggregated arrival gate for Washington-National Airport (DCA); there is one of these for each airport. An airport substitution transition is directionally connected to and from en route substitution through two ports, one representing an arrival gate and the other representing a departure gate (Zoomed-in figure is available in Appendix A).

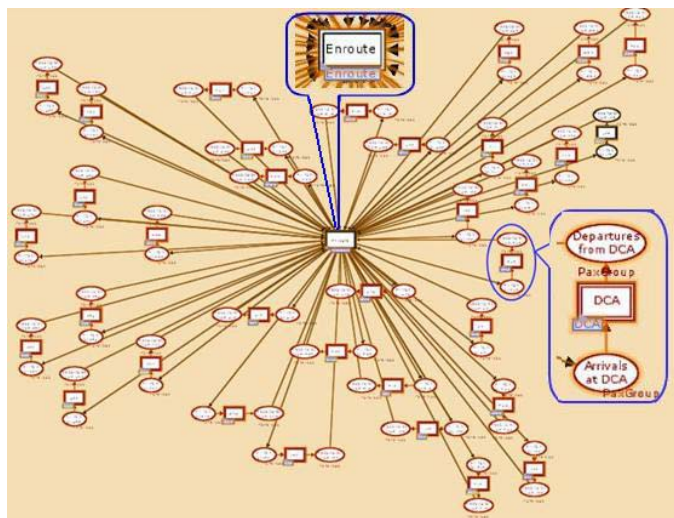


Figure 3.1 PFS Hierarchy, Top Level

In the second-level airport subnet, the flow process of passengers inside the airport boundary is divided into three steps: (1) splitting PaxGroup, (2) re-clustering of PaxGroup and (3) loading PaxGroup. In step 1, a PaxGroup arriving at the airport is split into two subgroups. The group of connecting passengers use the airport as a connecting hub, whereas the group of non-connecting passengers terminate their itineraries and leave the system at the airport. Different routes have different splitting rates for connecting and non-connecting passengers. In the example shown in Figure 3.2, 22% of passengers coming from ATL to ORD connect to another flight at ORD, while 78% terminate their trips at ORD.

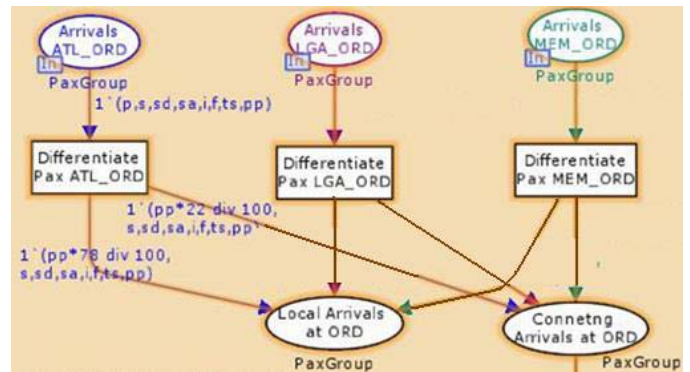


Figure 3.2 PFS Hierarchy, 2nd Level, Airport Subnet, Step 1

In step 2 (Figure 3.3), connecting passengers are re-clustered to form a new group for the second-leg flights. There are two functions involved in the simulation code. One of the functions returns the minimal connecting time (MCT) required for passengers between gates. The other function divides a group of connecting passengers into several subgroups and sends them to different gates [6].

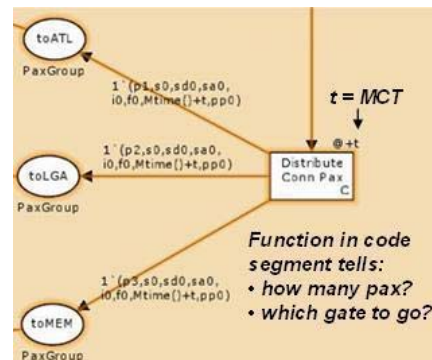


Figure 3.3 PFS Hierarchy, 2nd Level, Airport Subnet, Step 2

Finally, the newly formed PaxGroup is loaded onto their flights and ready to gate-out (Figure 3.4). The detailed loading process, which is the 3rd level subnet, will be explained in Figure 3.6.

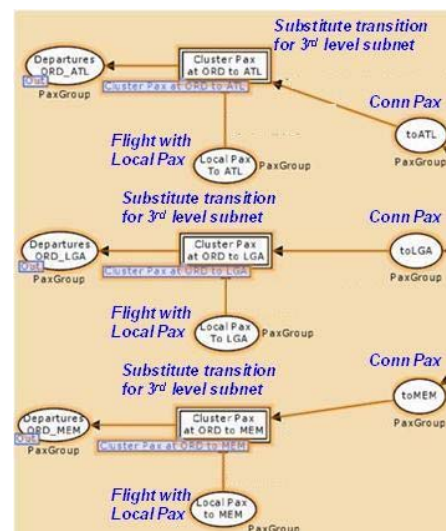


Figure 3.4 PFS Hierarchy, 2nd Level, Airport Subnet, Step 3

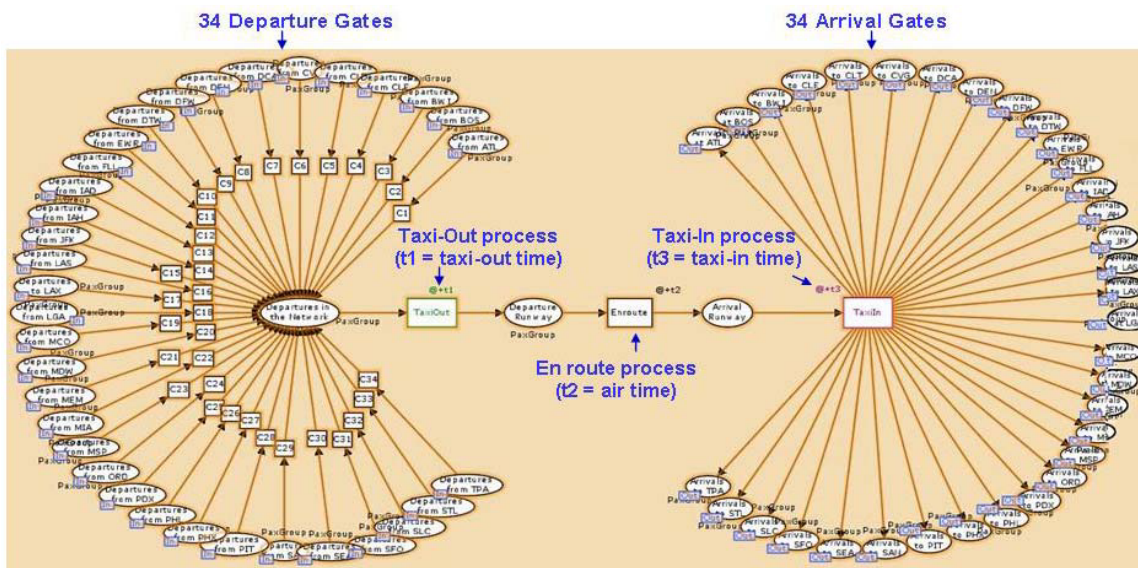


Figure 3.5 PFS Hierarchy, 2nd Level, Enroute Subnet

All the functions, statistics, and ratios written in the transition code segment and on the arcs are responsible for guiding the passenger flow according to historical statistics. In this specific PFS, all the statistics are obtained from 2006 historical data provided by BTS. There are three BTS databases involved: AOTP, T-100 and DB1B. A detailed explanation of how to calculate the statistics is available in reference papers [6] and [8].

In the second-level en route subnet, each PaxGroup goes through taxi-out, air time, taxi-in, and finally reaches the arriving gate at destination airports (Figure 3.5). Functions in this subnet are responsible for reading attributes of PaxGroup, transporting passengers on the correct route (gate-to-gate), and assigning correct taxi-out, en route, and taxi-in times to the

PaxGroup. The taxi-out time, air time, and taxi-in time are generated by random number generators, following some distributions with specific means and standard deviations calculated using 2006 data.

In the third-level passenger loading subnet (Figure 3.6), general connecting passengers flow to the upper branch and then are loaded onto the scheduled flights, while disrupted passengers (due to missed connections and cancels), flow to the lower branch and wait to be re-booked. Flights finished loading general connecting passengers will check for disrupted passengers before they depart. If disrupted passengers are detected, flights with available empty seats will load them until either no more seats are available or there are no more disrupted passengers. The general connecting passengers (upper branch) have higher priority than cancelled or missed

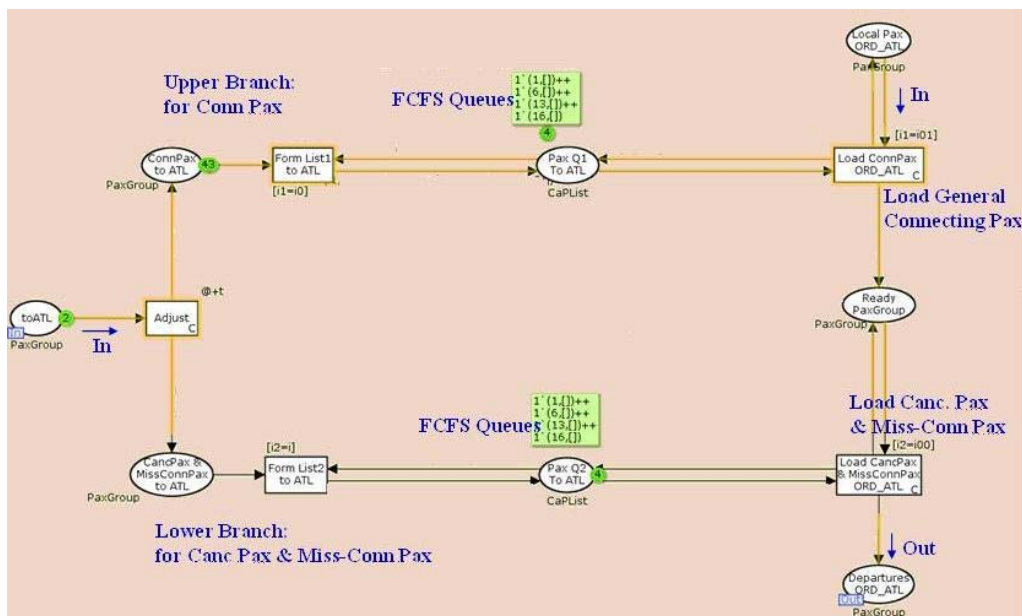


Figure 3.6 PFS Hierarchy, 3rd Level, Passenger Loading Subnet

connection passengers (lower branch). Passengers on both branches are sorted into first-come-first-serve airline queues to ensure passengers will be loaded on the correct flights (purchased flights) or will be re-booked by the same airline if disrupted (Appendix A Figure A.6).

E. Passenger Missed Connection Algorithm in PFS

Each experiment of PFS has two scenarios, the base scenario and the experimental scenario. The base scenario simulates passenger flow in an ideal environment without disruptions such as flight delays or cancellations. The goal of running the base scenario is to obtain passenger connecting information given a flight schedule. Passenger connecting information is then fed to the experimental scenario, which simulates delays, cancellations, and missed connections. As shown in Figure 4, passenger connecting information provided by the base scenario enables us to conduct research on missed connections in the experimental scenario. The simulation results of the experimental scenario estimate EPTD not only due to delayed and cancelled flights but also due to missed connections.

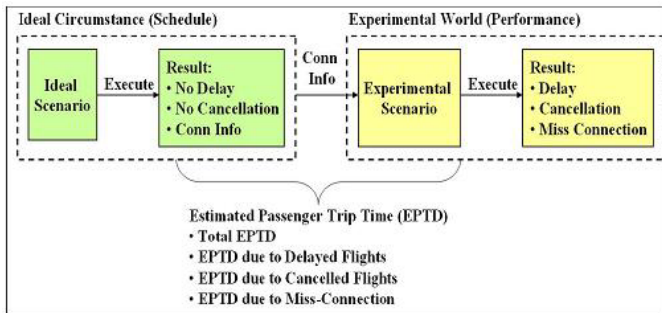


Figure 4. Passenger Missed Connection Algorithm in PFS

F. PFS Sample Results

July 6, 2005 is a randomly chosen weekday in summer 2005. Flight performance on July 6, 2005 was as follows:

- Scheduled Flights = 8,540;
- Delayed Flights = 1,764 = 21% of Scheduled Flights;
- Cancelled Flights = 176 = 2% of Scheduled Flights.

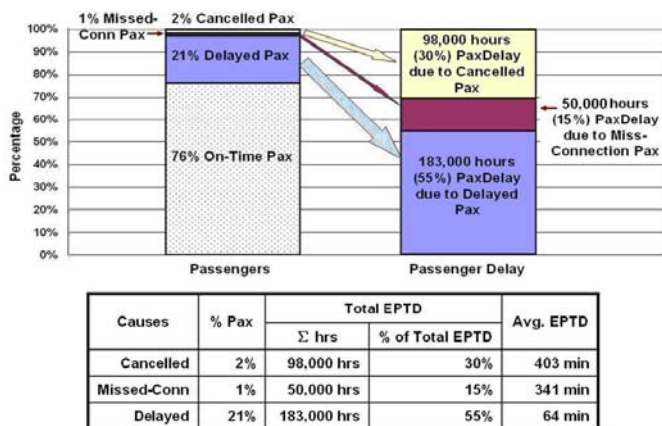


Figure 5. PFS Simulation Result for July 6, 2005

We used PFS to simulate passenger flow and calculate passenger trip delay on July 6, 2005. As shown in Figure 5, 2% of cancelled passengers generated 30% of total EPTD, 1% of missed connection passengers generated 15% of total EPTD, and 21% of delayed passengers generated 55% of total EPTD. On average, passengers scheduled on cancelled flights experienced 403 minutes of delay, missed connection passengers experienced 341 minutes of delay, and passengers scheduled on delayed flights experienced 64 minutes of delay.

III. EXPERIMENTAL DESIGN FOR PFS

The purpose of the experiment design is to identify and rank the significant factors that have strong impacts on passenger trip time and to analyze the sensitivity of EPTD given changes in these factors.

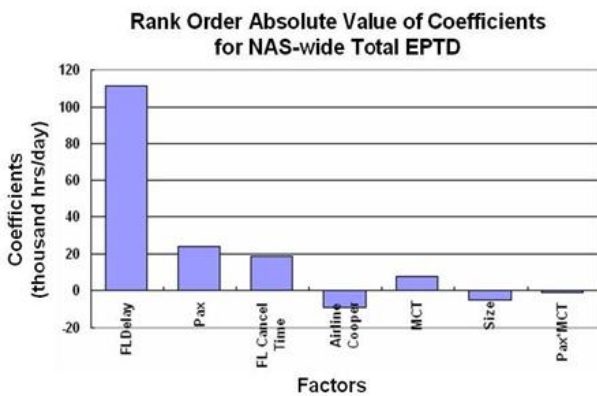
Based on experience and literature, six items are chosen as initial significant factors. These factors are shown in Table 2. Results of the experiments will prove how good the initial “guess” of significant factors is, and at what level they affect the passenger trip delay. A full factorial design for six factors, assuming a linear response function, needs 26 = 64 total runs, and each run requires two PFS scenarios (base and experimental scenarios). In total, 128 PFS models need to be built and executed for a full factorial design. Concerned about time, we performed a fractional factorial design with six factors, two levels (high and low) and 1/8 fraction. Table 2 lists the six factors and their high and low levels.

TABLE 2 HIGH AND LOW LEVEL SETTINGS FOR FACTORS

Factors	High	Low
# Passengers Loaded	Increased by 5%	Decreased by 15%
Aircraft Size (# of seats)	Increased by 15%	Decreased by 5%
Airline Cooperation Policy	Y	N
Flight Delay	+ 15 minutes	- 15 minutes
Cancellation Time	Cancelled four hours earlier	Remain the same cancellation time
Minimal Connecting Time	+ 15 minutes	- 15 minutes

The high and low levels of “# of passengers” and “aircraft size” are designed to keep load factor in the range of [61%, 92%]. The highest value of load factor (92%) occurs in experiments with “# pax”=H and “aircraft size”=L, while the lowest value of load factor (61%) occurs in experiments with “# pax”=L and “aircraft size”=H. Airline cooperation policy indicates whether airlines on the same route cooperate with each other on re-booking disrupted passengers. If not, disrupted passengers must stick with the same airline for re-booking.

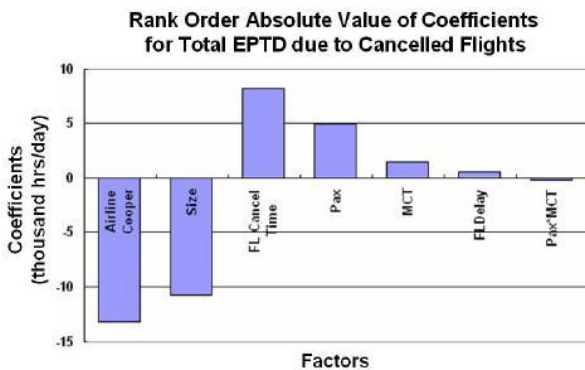
The rank order of significant factors in terms of the absolute value of coefficients for NAS-wide total EPTD is depicted in Figure 6. The most significant factor to total EPTD is flight delay, which is obvious, since more than half of the total EPTD is due to delayed flights. Along with flight delay, number of passengers, flight cancellation time, and airline cooperation policy also have significant impacts on total EPTD.



$$\text{Total EPTD} = 318236 + 111591 * \text{FLDelay} + 24049 * \text{Pax} + 18938 * \text{FLCancelTime} - 9110 * \text{AirlineCooper} + 7645 * \text{MCT} - 4778 * \text{Size} - 632 * \text{Pax} * \text{MCT}$$

Figure 6. Rank Order of Factors for Total EPTD (Delay+Cancel+MissConn)

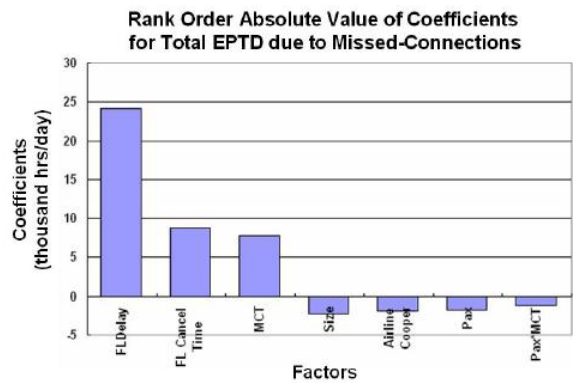
The rank order changes from case to case. For example, the rank order of factors in terms of EPTD due to cancelled flights is: airline cooperation policy, aircraft size and number of passengers (or load factor), flight cancellation time, as shown in Figure 7. These three factors have stronger impacts on EPTD due to cancelled flights than any other factors, since they are directly related to re-booking flexibility and resource availability.



$$\text{Total EPTD} = 62651 - 13099 * \text{AirlineCooper} - 10755 * \text{Size} + 8176 * \text{FLCancelTime} + 4929 * \text{Pax} + 1511 * \text{MCT} + 497 * \text{FLDelay} - 247 * \text{Pax} * \text{MCT}$$

Figure 7. Rank Order of Factors for EPTD due to Cancelled Flights

As shown in Figure 8, the rank order changes for total EPTD due to missed connections. The most significant factors affecting total EPTD due to missed connections are: flight delay, cancellation time, and minimal connecting time. The risk of missing a connecting flight increases if a previous flight leg is delayed. Affected passengers, whether due to flight cancellation or missed connections, compete for limited resources. As a consequence, cancellation time has a strong impact on EPTD due to missed connections. If the airport is poorly designed, connecting passengers may need longer minimal connecting time to travel from one gate to another, and this may result in missing connecting flights.



$$\text{Total EPTD} = 46658 + 24161 * \text{FLDelay} + 8780 * \text{FLCancelTime} + 7756 * \text{MCT} - 2231 * \text{Size} - 1879 * \text{AirlineCooper} - 1791 * \text{Pax} - 1102 * \text{Pax} * \text{MCT}$$

Figure 8. Rank Order of Factors for EPTD due to Missed Connections

In summary, the significant factors for different cases are as follows:

- To reduce total EPTD: decrease flight delay, encourage airline cooperation, earlier cancellation time, and lower load factor
- To reduce EPTD due to cancelled flights: encourage airline cooperation, lower load factor, and earlier cancellation time
- To reduce EPTD due to missed connections: decrease flight delay, decrease minimal connecting time required and encourage earlier cancellation time
- To reduce EPTD due to delayed flights: less flight delay and fewer passengers loaded.

A simple sensitivity analysis is done for a better understanding of the impact of factors on EPTD. As shown in Table 3, change in a single factor can result in 8% to 24% less total EPTD, thereby saving millions of dollars per day.

TABLE 3 SENSITIVITY OF THE TOTAL EPTD (DELAY+CANCEL+MISSCONN) TO CHANGES IN FACTORS

Changes in a single factor	Compared with Total EPTD on July 6, 2005	
	Decrease in total EPTD (hours per day)	Passenger Value of Time Saved (million \$ per day)
Reduce flight delay by 15 minutes	Decreased by 24%	Save \$2.3 million
Encourage airline cooperation	Decreased by 12%	Save \$1.1 million
Cancel flights 4 hrs earlier	Decreased by 10%	Save \$0.9 million
Reduce load factor from 83% to 70%	Decreased by 8%	Save \$0.7 million

Officials, operators, and service providers should consider the combined effect of factors on EPTD, which helps to achieve the strategic goals with minimal changes or costs.

IV. CONCLUSIONS

The goal of air transportation service is to provide safe, affordable, and convenient transport for passengers and cargo. As a consequence, the top level performance measures of the ATS should include the trip delays experienced by airline passengers. Passenger-based metrics, together with flight-based metrics, can give a more accurate and complete description of the ATS performance.

The passenger flow simulation captures the asymmetric and unique passenger trip on-time performance and reflects the complexity and significance of the impact of a small set of cancelled flights and missed connections on passenger trip delays. Major findings of this research are listed as follows:

1) High passenger trip delays are disproportionately generated by cancelled flights and missed connections.

2) Passenger-based metrics are needed to capture the passenger travel experience, since flight-based metrics can unintentionally distort the actual performance of the system and effectively “hide” explanatory and diagnostic system behavior.

3) Congestion flight delay, load factor, flight cancellation time, and airline cooperation policy are the most significant factors affecting total EPTD in the system. The combined effect of multiple factors should be investigated and used to support the decisions made by officials, policymakers, and researchers.

4) Passengers should treat trip time as a stochastic phenomenon that can be assigned a probability of occurrence but cannot be avoided entirely in any systematic manner. Simple strategies can be used by passengers to reduce the probability of occurrence, such as choice of departure airport and route. For example, for a trip from Washington, D.C. to Chicago, flights from DCA to MDW had a 5% probability of more than one hour delay, whereas flights from DCA to ORD had 12% probability of more than one hour delay.

REFERENCES

- [1] Heskett, J., Jones, T., Loveman, G., Sasser, W., and Schlesinger, L. (1994), Putting the Service-Profit Chain to Work, *Harvard Business Review*, March-April 1994, 164-174.
- [2] Bratu, S. (2003), *Airline Passenger On-Time Schedule Reliability: Analysis, Algorithms and Optimization Decision Models*, Ph.D Thesis, Supervisor: Cythia Barnhart, Massachusetts Institute of Technology, Cambridge, MA.
- [3] Bratu, S. and Barnhart, C. (2005), An Analysis of Passenger Delays Using Flight Operations and Passenger Booking Data, *Air Traffic Control Quarterly*, No. 1, Volume 13.
- [4] Bratu, S. and Barnhart, C. (2006), Flight Operations Recovery: New Approaches Considering Passenger Recovery, *Journal of Scheduling*, Volume 9, Issue 3, pp. 273-298, June 2006.
- [5] Wang, D., Sherry, L., Xu, N., and Donohue, G. (2007), System Analysis of Flight and Passenger Trip Delays in the National Airspace System, 26th Digital Avionics Systems Conference, October 2007.
- [6] Wang, D. (2007), *Methods For Analysis Of Passenger Trip Performance In A Complex Networked Transportation System*, PhD Thesis, Supervisor: Lance Sherry, Systems Engineering and Operations Research Department, George Mason University, Fairfax VA.
- [7] Sherry, L., Wang, D., and Donohue, G. (2007), Air Travel Consumer Protection: A Metric for Passenger On-Time Performance, *Journal of the Transportation Research Board*, ISSN: 0361-1981, pp. 22-27.
- [8] Wang, D., Sherry, L., and Donohue, G. (2006), Passenger Trip Time Metrics, *Proceedings of the 2nd International Conference on Research in Air Transportation (ICRAT)*, Belgrade.
- [9] Federal Aviation Administration (2004), *Airport Capacity Benchmark Report 2004*.
- [10] Bhadra, D., and Texter, P. (2005), Airline networks: an econometric framework to analyze domestic US air travel, *Journal of Transportation and Statistics*, Vol. 7, No. 1.
- [11] Castelain, E., and Mesghouni, K. (2002), Regulation of a public transport network with consideration of the passenger flow modeling of the system with high-level Petri nets, *Systems, Man and Cybernetics, IEEE International Conference*, Volume 6, 6-9.
- [12] Turki, A., Grunder, O. and Moudni, A. (2002), Public transportation systems modeling and analysis based on a new Petri net approach, in *Proceedings of the 2nd IEEE International Conference on Systems, Man and Cybernetics (SMC'02)*, October 6-9, 2002, Hammamet, Tunisia, Volume 5.
- [13] Colored Petri Net (CPN) Tools : <http://wiki.daimi.au.dk/cpntools/cpntools.wiki>

Appendix A

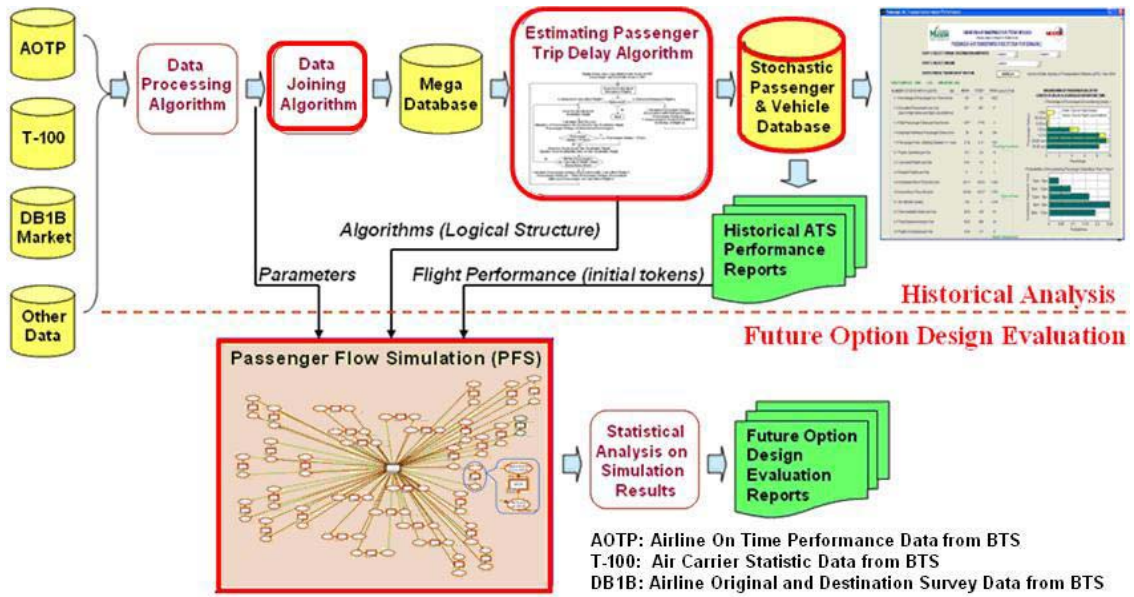


Figure A.1

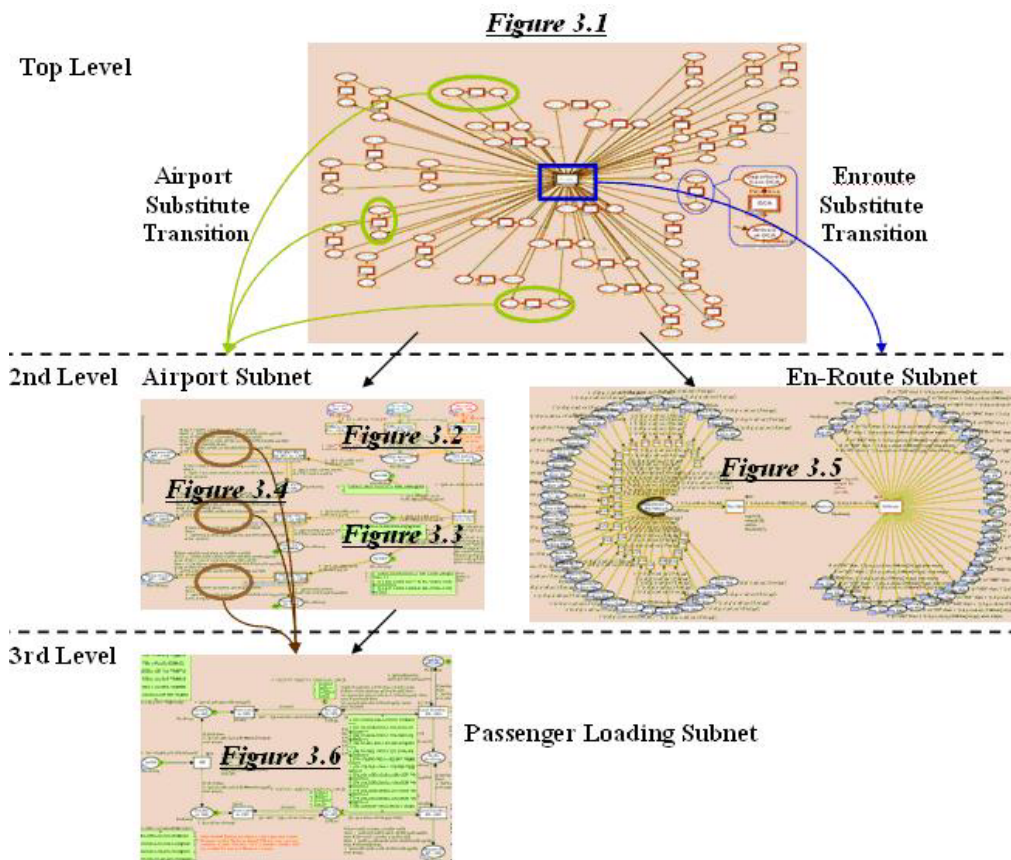


Figure A.2

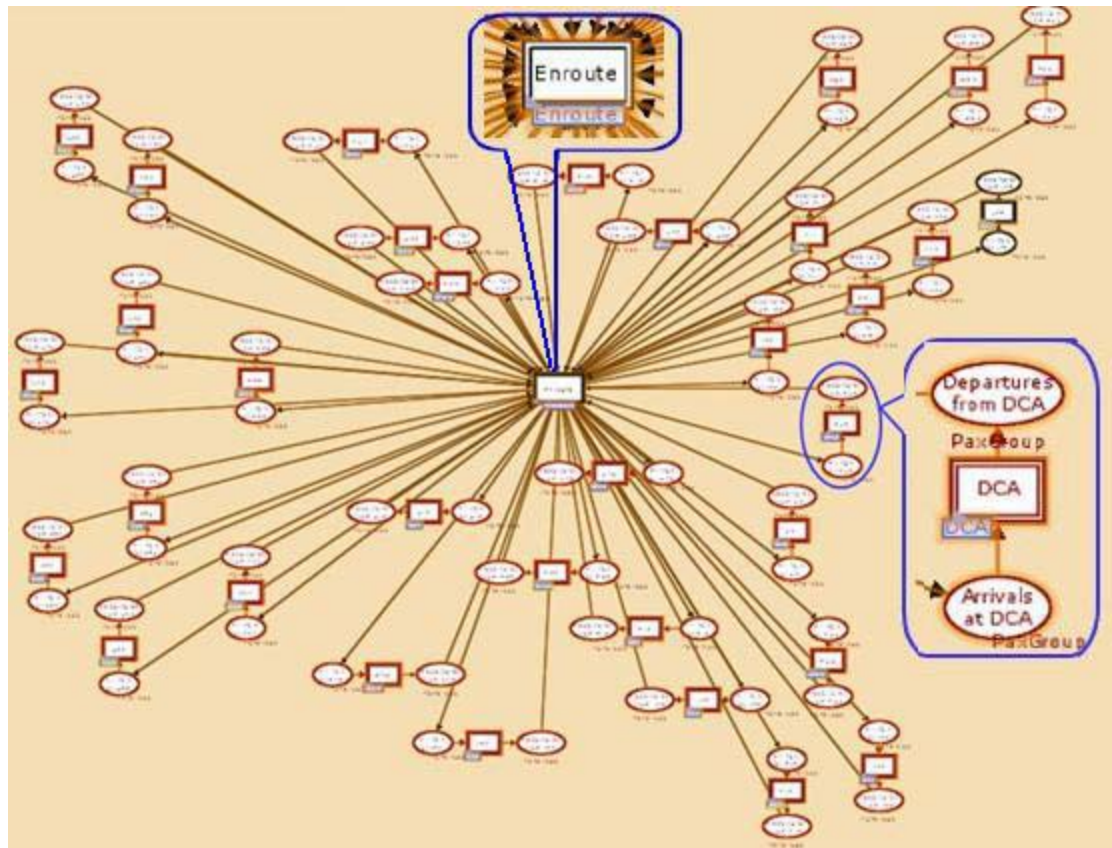


Figure A.3

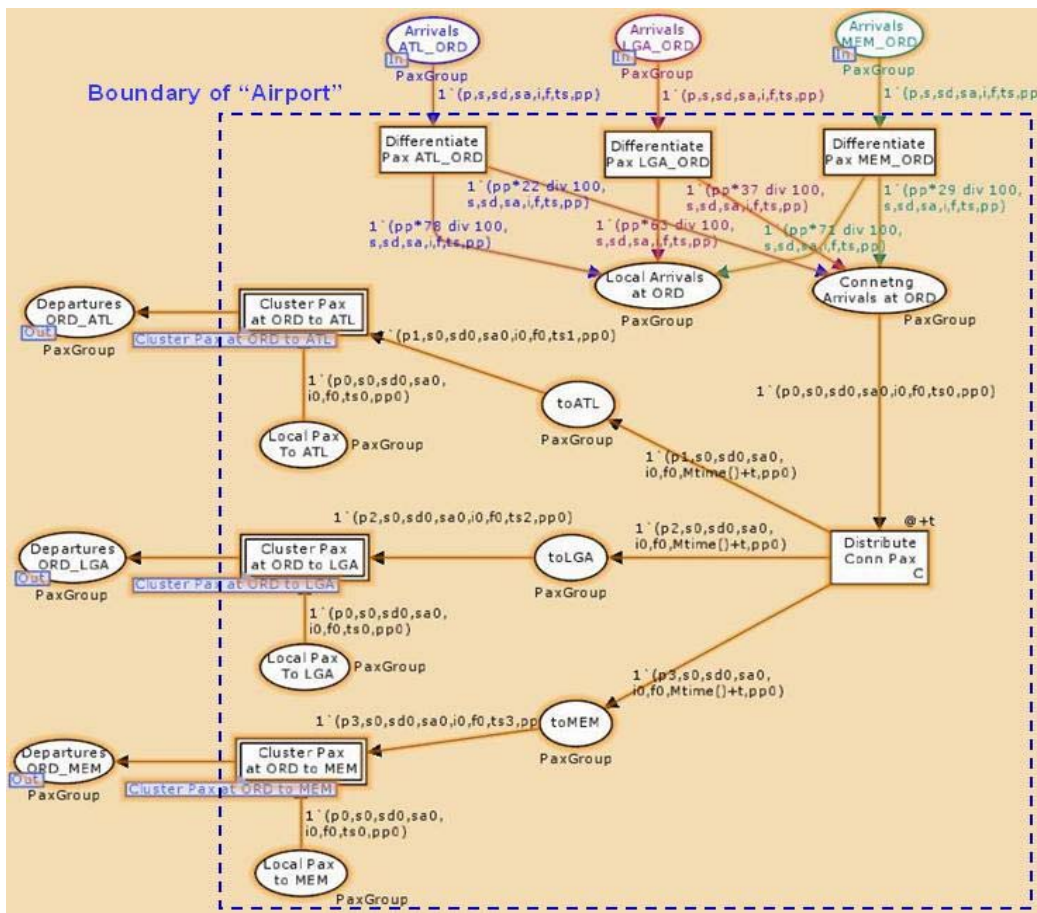


Figure A.4

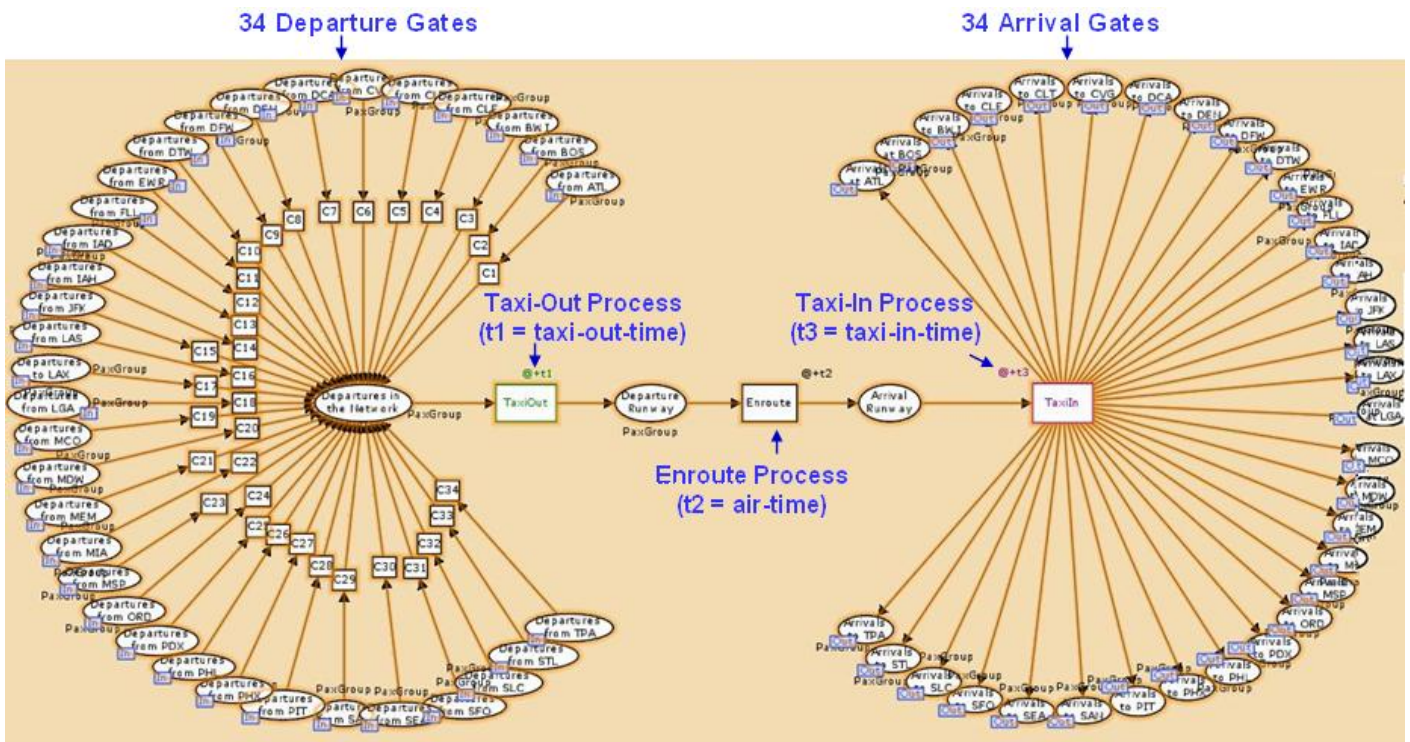


Figure A.5

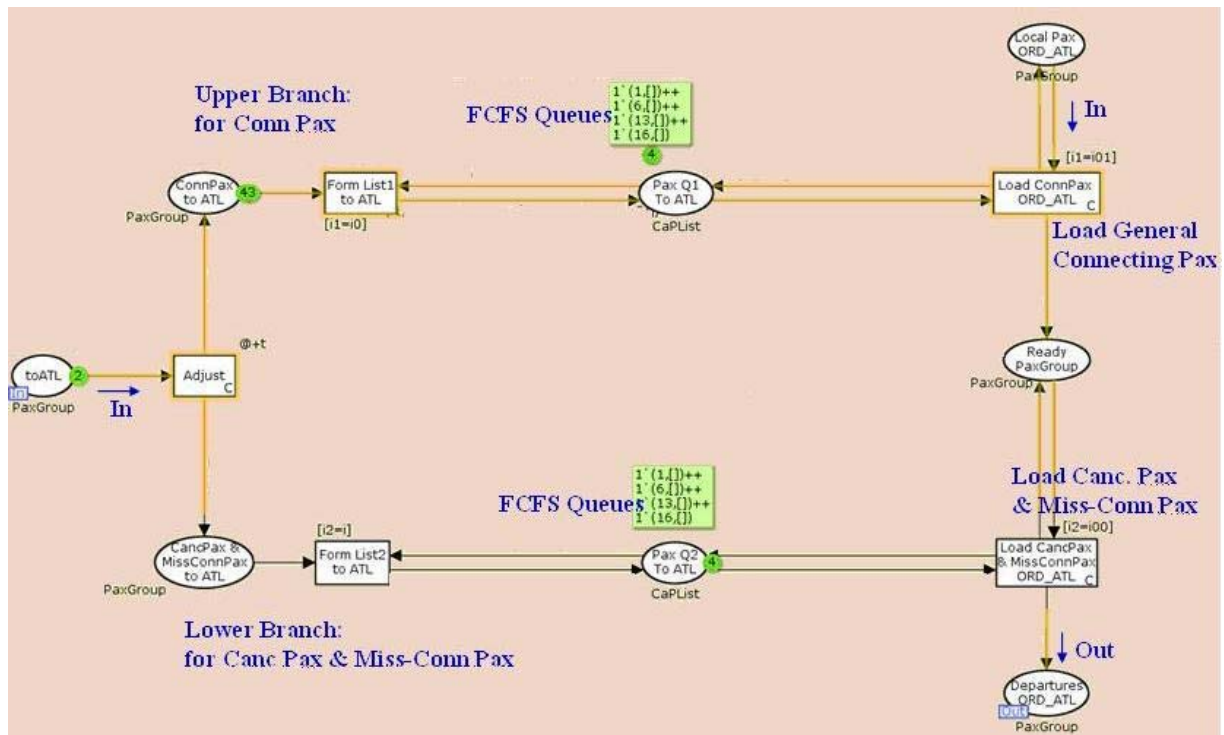


Figure A.6

Modeling Stochastic Evolution of Runway Capacity using Data Mining Concepts

Case Study of San Francisco International Airport (SFO)

Yoonjin Yoon

Mark Hansen

Graduate Student Researcher

Department of Civil and Environmental Engineering
University of California, Berkeley
Berkeley, U.S.A.
yoonjin@berkeley.edu

Professor

Department of Civil and Environmental Engineering
University of California, Berkeley
Berkeley, U.S.A.
mhansen@ce.berkeley.edu

Abstract— Variation in airport runway capacity, including arrival and departure is one of the main causes of operational disruptions such as flight delays and cancellation. In the ideal situation, we will know the exact timing and magnitude of such variations and plan accordingly to minimize such impacts. In reality however, capacity evolution process is probabilistic and determined by numerous factors. Capacity scenarios are the probabilistic representation of capacity variation at daily level. Scenarios provide probabilistic representation of capacity profiles to reduce modeling complexity of capacity prediction model. There are two data domains one can use to generate capacity scenarios; historical data, and day-of-operation information. While historical data provide long-term trend of capacity variation at an airport, day-of-operation information can increase the accuracy of the likelihood of each scenario on a given day. In this paper, we explore various Data Mining (DM) approaches to understand the historical trend of Airport Acceptance Rate (AAR) at San Francisco International Airport (SFO). We revisit earlier research based on k-means clustering. Among other shortcomings of k-means application, it lacks the sequential and time-dependent nature of AAR evolution. We first construct the Directed Acyclic Graph of AAR evolution to understand the conditional dependency among different time periods. Based on our observation that AAR change is mostly Markovian, we apply Sequence Clustering to properly address sequential nature of AAR evolution. In the later section, we include the preliminary result of Bayesian approach that utilizes weather information. In the last section we discuss the applicability of Data Mining concepts in aviation research, and future directions of our runway capacity modeling research.

Keywords-component; terminal capacity; data mining; bayesian learning; capacity prediction; AAR; scenario generation

I. INTRODUCTION

Variation in airport arrival and departure capacity, including arrival and departure, is one of the main causes of numerous operation disruptions, such as flight delays and

cancellations, as well as crew and aircraft rescheduling. In the perfect world, we will exactly know *when* and *how* such variation will occur, and be able to plan accordingly to reduce such disruptions. In reality however, capacity evolution process is probabilistic, which makes it harder to predict.

Capacity scenarios represent the probabilistic variation of airport terminal capacity at daily level. A good set of scenarios significantly reduces the modeling complexity that a capacity prediction model needs to handle, without compromising the integrity of original data. There are two data domains one can use in scenario generation: (1) historical data of capacity variation, and (2) day-of-operation information such as weather forecast of the day. While historical data provide information about the long-term trend at an airport, day-of-operation data can increase the accuracy of the likelihood of each scenario. In earlier models, scenario probabilities were based upon historical frequencies alone. While day-of-operation information, such as weather conditions and forecasts, is clearly relevant to predicting how capacity will evolve, our ability to harness this information is lacking.

Our main goal in this research is to understand the daily AAR evolution process, and to establish a Bayesian learning model. The main advantage of Bayesian approach is that capacity prediction is made not only based on the historical data, but also on day-of-operation information as the day unfolds. For example, if Air Traffic Control personnel are to make a decision on AAR changes at noon, Bayesian model utilizes realized AARs until noon, to make AAR prediction for the rest of the day. We consider several types of day-of-operation information; (1) realized AAR of the day, (2) weather forecasts, and (3) physical and other operational constraints at the airport.

We also identified two major factors that have to be captured in scenario generation; (1) the sequential and time dependent nature of capacity and forecast data, and (2) the

input-output relationship between the weather factors and capacity. Although intuitively natural, establishing a systematic way to capture such relationships is not an easy task. Data are, if available, are scattered in many sources, and the size and complexity of data available today are certainly beyond simple statistical interpretation. In addition, capacity variation is determined not just by weather factors, but also by numerous other factors, including human experience.

In this paper, we first review several Data Mining concepts in chapter II. The data collection and representation is explained in chapter III. In chapter IV, we revisit the evaluation result of earlier research based on k-means clustering, and discuss the challenges of this distance-based partitioning algorithm. We then present Graphical Model of historical AAR, and Sequence Clustering based on the homogeneous first-order Markov Chain. Graphical Model shows the complete hierarchy of conditional dependency of AARs. Sequence Clustering effectively captures sequential and time-dependent nature of AAR evolution. Finally Bayesian Network model is presented, in order to study the potential role of weather information in airport runway capacity forecasting, and how we can extend our model to incorporate such information.

II. METHODOLOGY

In this chapter, we will briefly review several Data Mining (DM) concepts used in our study, including; (1) Graphical Model, (2) k-means clustering, (3) Sequence clustering, and (4) Bayesian Network.

A. Graphical Model

Graphical Model, or graph theory, is a mathematical representation of conditional dependency of data objects. A graphical model consists of nodes and edges. Each node in a graphical model corresponds to a random variable, and contains a family of probability distributions associated with the node. Each edge, whether directed or undirected, represents conditional relationships between nodes it connects.

When applied to the historical airport capacity data, Graphical Model gives us a complete hierarchy of condition dependency of capacity evolution over time. Given only historical AAR values, we can still make a relatively sound prediction based on conditional dependency and probability distributions found in our Graphical Model. Although a very powerful way to understand data and make predictions, Graphical Model in general has a complexity that is exponential to the number of nodes. If we want to include additional factors such as weather to the model therefore, it will further increase computational complexity.

B. k-means Clustering

k-means clustering is one of the most widely available and used data mining concepts. k-means clustering is a partitioning method, which construct partitions of given dataset. It is an unsupervised data mining task, as each data point is not used in training process but treated equally. K-means clustering partitions objects into k nonempty subsets. Iterative assigning process puts each object to the cluster with nearest centeroid, until no more new assignment is possible.

The main draw of K-means is that it is relatively efficient, and the solution is readily available in conventional statistics packages, such as SAS, SPSS. On the other hand, k-means requires to preset the number of cluster, k, and unable to handle non-numerical data, and outliers. It is also not suitable to analyze high-dimension data and often terminates at a local optimum. There are several variations of k-means to address some of shortcomings mentioned above.

C. Sequence Clustering

Sequence clustering constructs clusters based on the transitional behavior of sequential data. It analyzes the state transitions in a sequence, and partitions data based on the similar transitional behaviors. Sequence analysis, including sequence clustering and sequence pattern recognition, is a relatively new data mining concept, which is becoming more and more important in areas such as web-log analysis and DNA analysis. We found this concept applicable to capacity evolution data, to address the sequential and time-dependent nature of capacity and weather data.

Among several algorithm choices, we adopted the model utilizing first-order Markov chains to capture transition behaviors among states. The algorithm works in a similar way to k-means clustering. The model starts with a specified of clusters, which can be preset or optimized, and then assigns each observation to one of the clusters. Instead of evaluating the centeroids and distance between the centeroids and data objects, Sequence Analysis model learns and updates the transition probability of Markov chains in each cluster. This is one of the soft clustering algorithms, yielding more flexibility in making predictions.

D. Bayesian Network (Bayesian Inference)

Bayesian Inference utilizes a combination of conditional and unconditional probabilities of evidence, along with the hypothesis one is interested. It is a straightforward and powerful classification data mining method, applicable to risk management, decision analysis, and many other areas. Classification data mining method such as Bayesian Inference and Neural Network requires specifying input and output to the model, which suits our need to establish the relationship between weather factors and airport capacity in a systematic manner.

III. DATA

Data collection covers three domains; (1) airport operational data, (2) airport weather observations, and (3) airport weather forecasts. In addition, considering the fact that decisions on capacity changes are a human-driven process, we conducted a series of interviews and meetings with a Traffic Management Coordinator at SFO Air Traffic Control Tower, Air Traffic Management Officers at Oakland ARTCC, and Northern California TRACON, to understand the human factors in the decision process, and how these data affect final decision.

A. Airport Terminal Operational Data

Our main source of airport operational data is the *Aviation System Performance Metrics (ASPM)*, published by FAA. ASPM contains a wide range of data including, but not limited to, Airport Acceptance Rate (AAR), Airport Departure Rate (ADR), ceiling, visibility, wind speed, wind angle, and runway configuration. Each data field can be retrieved in quarter-hourly or hourly level. Although arguably the most extensive and complete source of airport operational data, ASPM has its limitations. Since it reports data values in fixed time intervals, the actual times of operational changes are missing, and some numeric values are divided or summed over user-specific reporting intervals. For example, if AAR was 30 at noon, and then changed to 60 at 12:39, the 15 minutes report of AARs from ASPM between 12:00 and 1:00 looks as follows; 7@ 12:00, 8@ 12:15, 15@ 12:30, and 15@ 12:45. Rates such as 7, 8, or 15 are not real, as Air Traffic Controllers will never call rates on quarter-hourly basis, and change rates whenever necessary, not quarter-hourly. There are also instances of missing data in the data fields, and some hours are not reported at all. In our research, we retrieved operational data every 15 minutes, and post-processed them to be suitable for our needs. Format and unit conversion, and aggregation over multiple periods were necessary in most cases.

B. Airport Terminal Weather Observation Data

Our main source of airport terminal weather observations is the Hourly Surface Observations Summary, published by National Oceanic and Atmospheric Administration (NOAA). This observation data includes a wide range of aviation-related weather factors, such as sky condition, visibility, wind direction, wind speed, as well as more common factors such as temperature and precipitation. Surface observations are mostly automatic, and recorded every hour, unless there are significant changes that potentially affect aviation, such as ceiling reduction.

C. Airport Terminal Weather Forecast Data

Our main source of terminal weather forecast data is *Terminal Aerodrome Forecast (TAF)*, published by National Oceanic and Atmospheric Administration (NOAA). According to National Weather Service Aviation Weather Center, "Terminal Aerodrome Forecast (TAF) is a concise statement of expected meteorological conditions at an airport during a specified period, usually 24 hours." TAF is generated by human forecaster, and considered to be more accurate than model-generated weather forecasts. TAFs are produced four times a day starting at approximately 30 minutes before each main synoptic hour (00Z, 06Z, 12Z, and 18Z), and is valid as designated in each forecast. There are also cases when amendment is necessary to report temporary weather changes that affect airport operational condition.

TAF is a detailed forecast, covering various factors affecting airport operational condition. Meteorological condition includes wind – visibility – weather - sky condition – and other optional data. Wind, visibility, and sky condition are mandatory field in any forecast, while other conditions are included only when significant. TAF is only available in text format, and contains specific keywords for different weather

factors. We developed our custom parsing tool *TAFparser* to import text TAFs into our database.

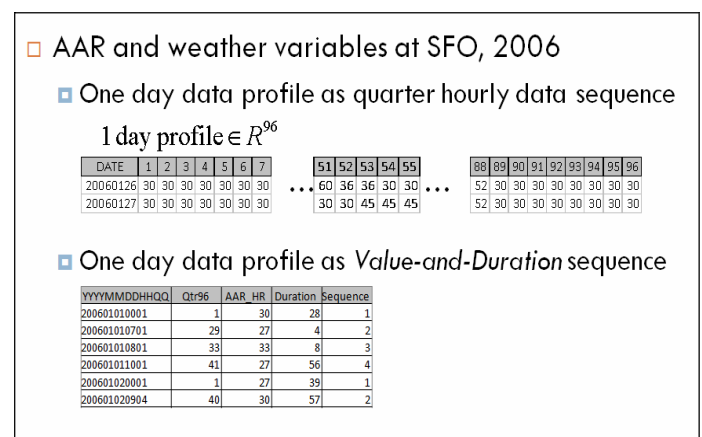
D. Data Manipulation and Representation

Finding the right representation of collected data is the first step in any data analysis, and in many cases, it is directly related to modeling choice. In the data preparation step, we converted quarter hourly AARs of ASPM, to hourly AARs, while maintaining the quarter hourly intervals. This gives us a better idea of how AAR evolves during the day. Also, we prepared two different representations of daily variation of AARs: one at a fixed time interval, and another as a sequence of changes, as shown in Figure 1.

Fixed time interval data representations are preferred in earlier researches, as they are readily available from ASPM, and easy to read and apply readily available statistical packages. However, it has certain limitations. First of all, fixed time interval representation emphasizes the continuation of one rate, rather than the changes in rates. Quarter hourly data gives a point in 96 dimensional space, and each point has the same degree of importance. Due to this fact, some of our early data analysis results suggested that the AARs tend not to change over time, and changes are pretty rare. Although this is an important fact, it is the main one we want to capture in our modeling.

Our goal of capturing the cause and trend of rate changes are better represented in a different format, *value-and-duration* representation. Value-and-duration representation shows daily AAR profile in a vector form. With this representation, we can more effectively detect the patterns of rate changes, as shown in the second table of Figure 1. This table shows for each day, what the total number of different rates called was, and what their sequence and duration were.

Figure 1. Two representations of one day AAR changes



IV. DATA MINING APPLICATIONS

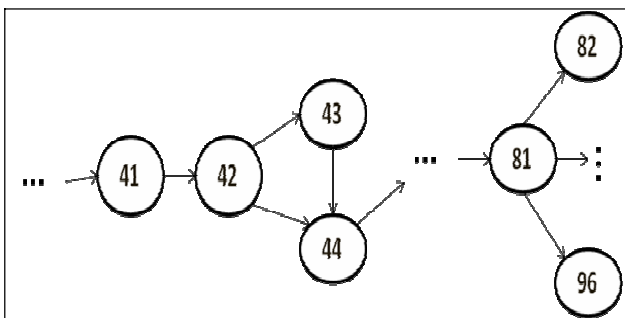
The focus of this chapter is studying AAR evolution process itself using several Data Mining (DM) methods. Studying historical AARs has two advantages. First, we can compare the several DM models to understand the advantage and disadvantages of different DM algorithms. Secondly, better understanding of capacity evolution process itself would give

us clues how and which additional information is crucial in generating representative scenarios, and making accurate prediction.

A. Graphical Model

The daily evolution of AAR can be represented as a Directed Acyclic Graph (DAG). Each node represents AAR at certain time period, and contains the family of probability distributions AARs at a given time. This graph provides complete information of conditional dependency of AAR at each time period. Figure 2 shows selected node-edge representation from the model output. One of the major finding is that most time periods exhibits Markovian property. For example, AAR of time period 42, or between 10:15am and 10:30 am, is only dependent on AAR of the previous time period 41. There were a few cases like time period 44, which exhibits second-order Markovian property, as it depends both on 42 and 43. This confirms our postulation that AAR evolution process is mostly Markovian. There are however, special cases such as time period 96, or last four hours of the day, which only depend on time period 81, or 8:00pm-8:15pm. This reflects an operational constraint specific to SFO, where AAR is lowered to the minimum 30 after 8:00 pm (9:00 pm during daylight saving), on most days.

Figure 2. Bayesian Network of AAR evolution at SFO

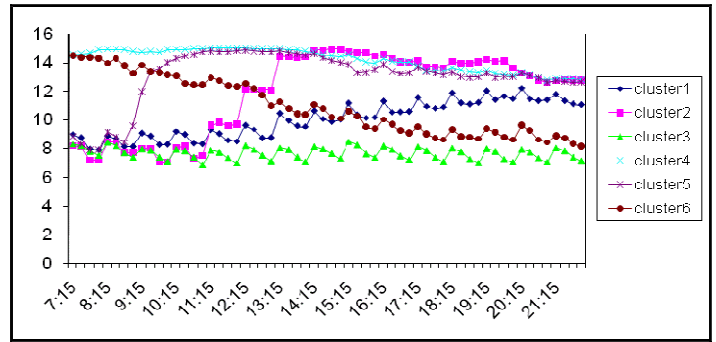


B. k-means Clustering

k-means Cluster analysis is a powerful and widely used data mining technique. Liu and Hansen (2006) applied this method to capacity scenario generation, by representing one day capacity profile as a point in 96-dimension space, and applied the K-means algorithm based on Euclidean distance. The result is shown in Figure 3.

As desired, the result appears to represent typical days for SFO. For example, cluster3 may represent days where the fog never burns off all day, and AAR remained at the minimum of 30 per hour.

Figure 3. k-means clustering of 15 minute AAR evolution



Although this is a reasonable approach, it may suffer from the general shortcomings of k-means clustering. First of all, given AARs every 15 minute, and thus 96 data points per day, K-means clustering treats one day capacity profile as a single point in the 96-dimension space. This has two potential problems: (1) it ignores the sequential and time-dependent nature of AAR; and (2) k-means clustering is subject to the *Curse of Dimensionality*. AAR at one time period is likely to be highly correlated with the previous and following ones. In addition, as we expand the dimensions of data, they become less concentrated and sparser, and the distance measure, which is the foundation of k-means clustering, becomes less meaningful. SFO airport capacity case, it is reasonable to assume the data points are well concentrated in certain regions, and dimensionality issue is not evident for certain clusters. However, it is desirable to develop a more robust solution that can be applied in the general case.

C. Sequence Clustering

Sequence clustering is a relatively new area in data mining, which captures the strength of partition-based clustering such as k-means, and applies it to sequential data. *Sequence* is a series of discrete events, or states, which are usually finite. Sequence data is ubiquitous in our everyday life. A series of book purchase you made at Amazon.com, the sequence of web sites you visited yesterday, the hourly temperature changes in San Francisco, and DNA sequences in gene expression. If the sequence is stochastic and has a Markovian property, then such characteristics are well modeled in Sequence clustering using Markov Chains.

Sequence clustering combines the strength of two techniques, by assigning each data object to specific cluster(s) with certain probabilities, while each cluster is characterized by a unique Markov chain. To determine which data object belongs to which cluster(s), Sequence clustering uses a probability measure, unlike distance measure of k-means clustering. Probability and likelihood of data objects are then calculated based on the transition matrix of each cluster. As data objects are added to clusters, transition probabilities are adjusted reflecting the new data members.

In our research, we can apply Sequence clustering to capture the time-dependant and sequential nature of daily airport capacity variation. Figure 4.1 is the cluster diagram we obtained with SFO 2006 data. Each cluster contains its own Markov chain that characterizes the cluster. Figure 4.2 shows Markov chain of Cluster 1. It is also extendable to higher-order

Markov Chains or Hidden Markov Chains, depending on the data characteristics.

Figure 4. Sequence Clustering Result

Figure 4.1. Cluster Diagram

Each node represents one cluster. Node with darker shades contains more data objects than lighter ones. Similar clusters (clusters with similar transition probabilities) are closer to each other, and the degrees of similarities are represented as edges connecting cluster. In this figure, cluster 5 has a rather unique transition behavior than other cluster, while cluster 2, 3, 4 share strong similarities.

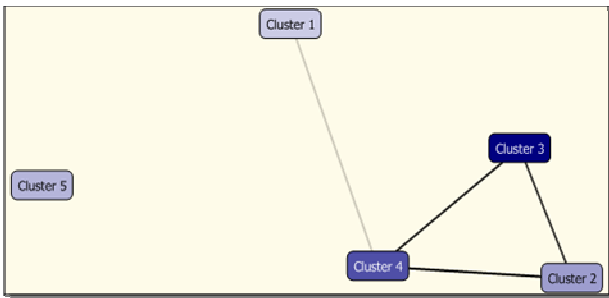


Figure 4.2. Markov Chain Associated with Cluster 2

Transition matrix associated with Cluster 2 is shown below. Cluster 2 contains AAR 20, 40, 45, 52, and 60. Days in Cluster 2 starts operation with AAR 30 with probability of 1. Once the rate is set, the rate tends to persist as high transition probability from one rate to the same rate suggests. Some of the characteristics of Cluster 2 include; (1) transition to rate 52 only occurs when the previous rate is 52, or 60; (2) rate 40 tends to make a transition to other rates, more than any other rate, suggesting that AAR 40 is used for short period of time; and (3) rate 60 tends to persist once the rate is set, with possibility to get reduced to 52.

Transition Matrix of Cluster 2						
	30	40	45	52	60	EOD
BOD	1					
30	0.95				0.02	0.03
40	0.06	0.70	0.06		0.18	
45			0.87		0.13	
52	0.08			0.92		
60				0.03	0.97	
<i>BOD: Beginning Of Day</i>						
<i>EOD: End Of Day</i>						

D. Distribution of Rate Changes

During the course of our DM applications, we observed that at SFO, there are handful of time windows and rates that are crucial in answering the question of when and what is the rate change going to be. For example, first rate change of the day is very likely to occur around 8:00 am, to decide whether to increase AAR from the early morning minimum rate of 30.

Also around 5:00pm, AAR is to be lowered to 52, even if weather permits the full capacity of 60, due to operational restriction such as noise abatement. Also, timing of recovery the full capacity of 60 depends on fog burn-off time, which is likely to burn off around 9:00-10:00, or 13:00-14:00. To investigate our observations further, we plotted probability distribution of AAR changes by absolute and relative frequency of time of such changes, as shown in Figure 5. It is also observed that first rate change mostly occurs at 8:00am, when the air traffic control personnel decide which rate they will start the day with.

From the probability distribution of time of change given AAR value (Figure 5.1), we can observe that the full capacity of 60 AAR per hour is most likely to kick in at the start of the operation at 8:00 am, followed by between 9:00 and 10:00 when early morning fog burns off, followed by between 13:00 and 14:00 when morning fog persists until afternoon and burns off late in the afternoon. Also, AAR change to 52 mostly happens between 17:00 and 21:00, as this rate is mandatory change during evening time, partly due to noise abatement issues.

An insight this observation provides us is that not all time periods are equal, and there are set of critical time periods that we want to model more accurately. For example, we might want to use most recent and accurate weather forecast available to predict capacity at 8:00 am. An interesting flip side of this observation is that AAR change might not be as dynamic as the weather change, and managed in a rather conservative manner.

Figure 5. Distribution of Rate Change

Figure 5.1. Probability Distribution of Time of AAR Changes, given AAR

This chart shows the relative frequency of time of change to a specific rate, or Probability(Time of Change|AAR). For example, change to AAR 60 is most likely to occur at time 8:00 am, followed by 10:00 am, and 2:00 pm.

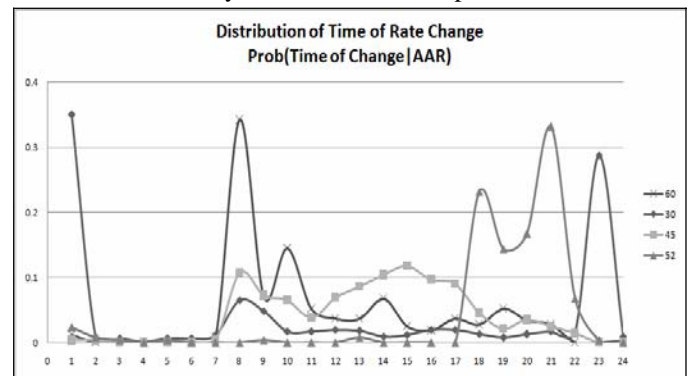
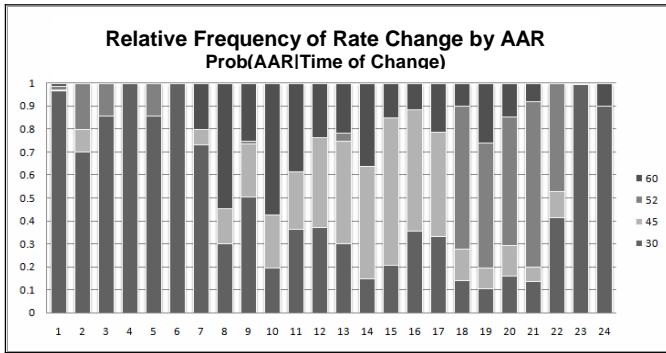


Figure 5.2. Relative Frequency of Rate Changes, given time of the day

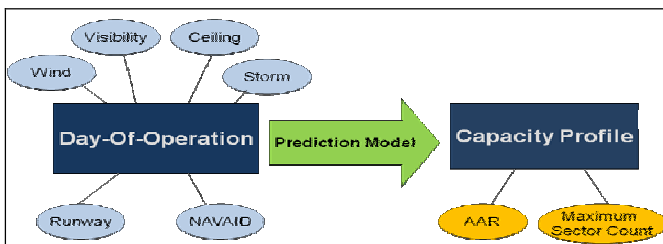
This chart shows the relative frequency of different AARs at a specific time of the day, or Probability(AAR|Time of Change). For example, rate change at 8:00am is most likely to be 60, followed by 30.



E. Bayesian Network

There has been recognition that it is necessary and possible to calibrate and extend the existing model, by incorporating weather information. By adding weather forecast information as additional model parameter, we might be able to obtain more accurate understanding, hence prediction, of airport capacity. Figure 6 shows the ideal modeling structure, incorporating weather information as well as historical AAR evolution process. In the ‘Prediction Model’ arrow, we model historical AAR, weather observations, and weather forecast data, as well as day-of-operation weather forecast, to generate day-of-operation capacity profiles.

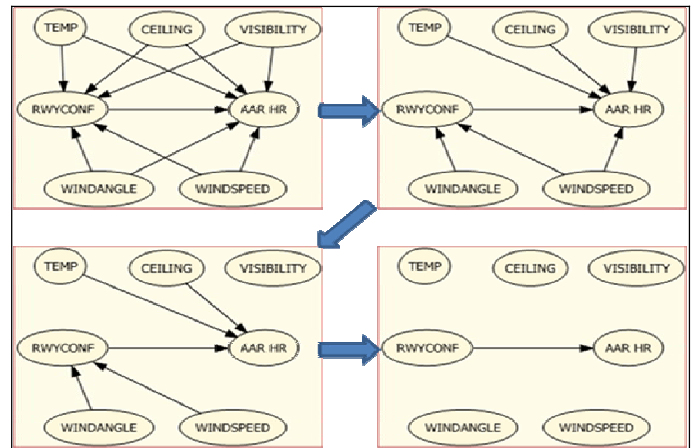
Figure 6. Capacity Prediction Modeling Structure



Bayesian Network is a simple yet powerful way to explore and understand input-output relationship of data with prediction ability. It is also powerful tool to analyze relationships among attributes when making a prediction, using conditional probability of observed events. Bayesian Network can be used as a preprocessing step to identify critical variables in predicting the variable of interest.

The implementation of Bayesian Inference is pretty much standard across different data mining platforms. Dependency Network from model output is shown in Figure 7. This figure illustrates the degree of dependency and predictability. From this diagram, we can see that AAR at SFO is most dependent on Runway Configuration, Ceiling, Wind Angle, Wind Speed, and Visibility in that order. It corresponds to widely recognized belief that Ceiling and Wind Angle have the biggest impact on airport capacity. Runway Configuration for SFO, obviously affects maximum AAR.

Figure 7. Dependency Network of Weather Factors and AAR



V. CONCLUSIONS AND FUTURE STUDIES

In this paper, several Data Mining concepts are introduced and applied to runway capacity scenarios generation. Data Mining concepts, while still at early stage of research and development, showed strong possibility to significantly contribute to aviation research, where complexity and size of available data are beyond the application of simple statistical methods.

Earlier research based on k-means clustering provided a reasonable mean to classify days with similar characteristics. We revisited k-means clustering and came to a conclusion that it lacks the sequential nature of AAR evolution, and is prone to dimensionality problem of k-means clustering. This review process also led us to explore different representation of daily AAR variation, such as in quarter hourly rate, hourly rate, and value-and-duration representation. Outcome of statistical analysis is heavily depends on how data is organized, and it is desirable to choose most suitable representation for each method. It is also observed that there is prevailing trend in the time and magnitude of the rate change at SFO, which may suggest that the rate change is not as dynamic as the weather change.

We first construct full Directed Acyclic Graph (DAG) of AAR change, to understand conditional dependency of AAR at each time period. At SFO, we found that AARs are mostly Markovian, which supports applications of algorithms such as Sequence clustering. We also observed that there are certain operational restrictions, not related to weather or previous AAR, such as mandatory rate reduction in the evening. Although providing complete hierarchy of conditional dependency among time periods, DAG has complexity that is exponential in the number of nodes, which makes it less attractive in making predictions.

To address the time dependent, sequential nature of AARs, Sequence Clustering with first-order Markov Chain is applied. This clustering method soft-partitions days based on transitional behaviors, which are captured in the transition matrix. The algorithm effectively captures the nearly Markovian property of AAR changes, as well as time of the day effect. Another advantage is that the analysis result represented in Markov Chain fits well as an input to stochastic optimization model.

Bayesian Network is applied to understand the relationship of different weather factors and AAR. Dependency Network confirms our prior belief at SFO, that runway configuration, ceiling, and wind conditions are most influential factors in AAR determination. Knowledge from Bayesian Network, combined with day-of-operation weather forecast can increase the accuracy and reliability of capacity prediction result.

As an extension of this study, authors continue focusing on relationship between AAR and weather factors, and on how to model such relationship.

REFERENCES

- [1] B. Liu, M. Hansen, "Scenario-Based Air Traffic Flow Management: From Theory To Practice"
- [2] Macqueen, 1967, Some Methods For Classification And Analysis Of Multivariate Observations
- [3] Department Of Commerce \$ National Oceanic & Atmospheric Administration \$ National Weather Service, National Weather Service Instruction 10-813, 2005
- [4] Keogh, "On The Need For Time Series Data Mining Benchmarks: A Survey And Empirical Demonstration"
- [5] J. Han, "Data Mining: Concepts And Techniques"
- [6] Z. Tang, "Data Mining With Sql Server 2005"
- [7] C. Bishop, "Pattern Recognition And Machine Learning"
- [8] T. Dietterich, "Machine Learning For Sequential Data: A Review"
- [9] O. Bousquet, "Introduction To Statistical Learning Theory"
- [10] T. Griffiths, A. Yuille, "Technical Introduction: A Primer On Probabilistic Inference"
- [11] P. Mafera, K. Smith, "Traffic Flow Management: Atc Coordinator's Information Requirements For The Nas"
- [12] A. Moore, "Reinforcement Learning: Survey"
- [13] Metropolitan Transportation Committee, " Bay Area Airport Operational Procedure Overview"
- [14] Faa Naimes Atcsc Domestic Web System
- [15] National Oceanic And Atmospheric Administration, Climatic Data Online
- [16] Federal Aviation Administration, Aviation System Performance Metrics
- [17] Federal Aviation Administration, Collaborative Decision Making

Deconstructing Delay:

A Case Study of Demand and Throughput at the New York Airports

Amy Kim

Department of Civil Engineering
University of California, Berkeley
Berkeley, CA
amy_kim@berkeley.edu

Mark Hansen

Department of Civil Engineering
University of California, Berkeley
Berkeley, CA
mhansen@ce.berkeley.edu

Abstract - This paper introduces an empirically driven, non-parametric method to isolate and estimate the effects of demand and throughput changes to observed changes in flight delay. Classical queuing model concepts were used to develop a method by which an intermediate queuing scenario could be constructed, in order to isolate the delay effects due to shifts in demand and throughput. This method includes the development of a stochastic throughput function that is based entirely on data and as a result has two advantages: it uses non-parametric, empirically-based probability distributions, and capacity need not be estimated explicitly. The method was applied to a case study of the three major New York airports of LaGuardia (LGA), John F. Kennedy (JFK), and Newark Liberty (EWR), for the peak summer travel seasons of 2006 and 2007, using data extracted from ASPM. This case study was of particular interest given that these airports experienced record levels of delay in 2007. The simulation results were consistent with both OPSNET and ASPM data, and were successful in quantifying the delay effects of demand and throughput changes from 2006 to 2007.

Keywords - delay; demand; throughput; capacity; runway operations; New York airports; simulation; probability; ASPM; OPSNET.

I. INTRODUCTION

This paper introduces a method for estimating the effects of demand and throughput changes to observed changes in flight delay. As the delay observed over days, weeks or years changes from one time period to the next, we would like to know how much its evolution can be attributed to demand and throughput changes. As a result, the motivation for this work is to address the following question: how can we isolate and measure shifts in delay caused by changes in demand and throughput when both are changing simultaneously?

There is an extensive body of literature and knowledge on methods to predict airport capacity and delay, both analytically [1] and through simulation. The purpose of this work is not to estimate the expected capacity outright [2], but to use empirical data that implicitly contains information about capacity to quantify how simultaneous changes in demand and throughput affect delay.

A new, empirically driven simulation procedure was developed from classical queuing concepts to address the question posed above. The main engine of this new procedure

is a stochastic throughput function that was developed to have two key advantages. Firstly, this throughput function is driven by non-parametric probability distributions of throughput constructed from available data. Secondly, capacity need not be explicitly estimated, as the capacity of the operation under analysis is implicitly included in the probability distributions. This is advantageous because operational capacity is subject to a wide variety of factors and can be quite difficult to estimate well.

The simulation method is then applied to a case study of flight delay at the three major New York area airports: LaGuardia (LGA), John F. Kennedy (JFK), and Newark Liberty (EWR). Specifically, the arrival and departure operations at these airports were analyzed in order to determine how demand and throughput affected the delay changes observed between the 2006 and 2007 summer travel seasons.

The main goal in applying this new procedure is to provide information about the causes of delay shifts at one greater level of detail. The ability to isolate individual contributions of demand and throughput mechanisms to delay could be helpful in creating more focused, effective strategies and policies to address the delay problem.

II. BACKGROUND

During the summer of 2007, flight delays reached record high levels throughout the National Airspace System (NAS) and beyond. National and international headlines reported story after story describing the extreme wait times and missed connections that air travelers were subject to during this peak travel season. The three airports of the New York area experienced some of the highest delays within the NAS, with travelers spending 3.9 million more hours waiting for their aircraft to take off after leaving their gates in 2007 as compared to a decade earlier [3]. The increase in total operations from 2006 to 2007 at these airports was approximately 3-4%, but the increase in delay was in the order of about 28% [4]. In addition, in 2007 the New York airports accounted for about 40% of all delay in the NAS; in 2004 they accounted for only 15% [5].

Delay metrics can be found and/or calculated with relative ease from several data sources. One such source is OPSNET, which is the official source of historical NAS air traffic delays and operations. In OPSNET, an airport picks up a delay each

time a flight is held up 15 or more minutes due to runway congestion, weather, air holding, traffic flow restrictions, or other event that would cause a flight's realized schedule to deviate from its flight plan. Table 1 contains the results of OPSNET airport delay data extracted for LGA, EWR, and JFK for May through September of 2006 and 2007. The first half of the table indicates that the total number of operations have decreased at LGA and EWR, but have increased significantly at JFK. The second half of the table shows the number of flights that were delayed more than 15 minutes from their flight plans; it can be observed that the number of delayed flights has almost doubled at JFK from 2006 to 2007.

TABLE I. OPSNET DATA

Total Number of Arrival and Departure Operations			
	May-Sept 2006	May-Sept 2007	% Change
LGA	172,142	168,616	-2.0%
EWR	191,531	188,211	-1.7%
JFK	169,957	197,626	+16.3%
Total Number of Flights Delayed >15 Minutes			
	May-Sept 2006	May-Sept 2007	% Change
LGA	14,119	15,810	+12.0%
EWR	21,707	19,809	-8.7%
JFK	8,276	15,065	+82.0%

III. METHODOLOGY

Delay can be estimated using the traditional queuing model, where a queuing scenario is constructed from a cumulative demand curve and cumulative throughput curve [6]. An example of a simplified fictional queuing scenario is shown in Figure 1. The demand and throughput curves are actually step functions because customers (or vehicles, aircraft, etc.) are discrete entities. However, demand and throughput can be approximated as continuous functions (smoothed curves) over sufficiently long periods of time, which simplifies calculations. Under a classical deterministic approach, the throughput function at some time t can be determined as follows:

$$Q(t) = d(t) \quad \text{if } d(t) < c \tag{1}$$

$$= c \quad \text{if } d(t) \geq c$$

Where $Q(t)$ is the throughput at time t
 $d(t)$ is the demand at time t
 c is the fixed service capacity, constant over all t

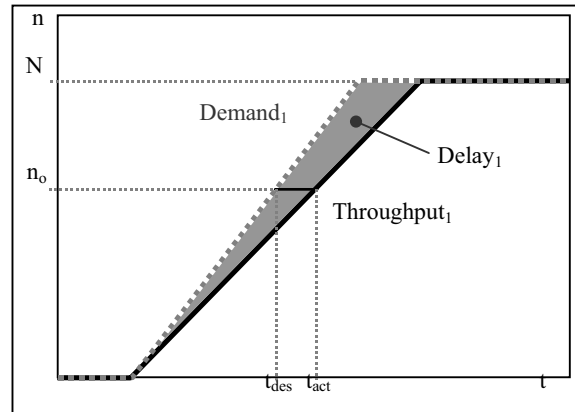


Figure 1. Queuing Scenario under Year 1 Demand and Year 1 Throughput

Note that in Figure 1 cumulative $Q(t)$ and $d(t)$ are shown to vary linearly with time. However, this is a simplification in that these quantities are most often non-linear, time-dependent functions.

Assuming first-in first-out (FIFO) conditions, the delay experienced by an arbitrary customer n is the difference between n 's desired service time (t_{des}) and actual service time (t_{act}). This is also the horizontal distance between the two curves. The number of customers queued for service at time t is the vertical distance between the curves at t . Where the demand and throughput curves meet, customers are being served without any delay and as a result there are no standing queues for service; when the curves are apart, customers must queue for service. The throughput curve cannot cross the demand curve as per Equation (1) because customers cannot be served until they demand service. The area between the demand and throughput curves is the total delay experienced by customers over the total observation time T (we assume that our observations begin at time 0):

$$\tau = \int_0^T [Q(t) - d(t)]dt \approx dt * \sum_{j=1}^{j=J} [d(j) - Q(j)] \tag{2}$$

Where τ is total delay over time period (0,T)
 $Q(t)$ is the throughput function at time t
 $d(t)$ is demand at time t
 T is total observation time
 dt is the duration of a small time slice
 j is the number of time slices over time T , from $j=1$ to $j=J$

An average delay per customer can then be determined by dividing this total delay by the total number of customers N that requested service over the observation time T . In these queuing diagrams, T could represent one day.

Figures 1 & 3 depict fictional queuing scenarios for an average day in an arbitrary year (Year 1) and the following year (Year 2), respectively. The areas between the demand and throughput curves represent the total delays in Year 1 and in Year 2. The change in total delay from Year 1 to Year 2 is the difference of the two areas; however, this difference could be

caused by changes in demand, changes in throughput, or both. In order to isolate the change in delay caused solely by a change in demand, we can construct a “counterfactual” scenario where the Year 2 demands are served using the Year 1 throughput function. The counterfactual scenario is represented in Figure 2, and the resulting delay is represented by the total area between the demand and throughput curves. The difference between the resulting counterfactual delay and the Year 1 delay (solid area in Figures 1 and 2) is the change in total delay due to the demand shift from Year 1 to Year 2 (depicted in cross-hatch). The Year 2 delay (total area between the curves in Figure 3) minus the counterfactual delay and Year 1 delay is the change in total delay due to the throughput shift from Year 1 to 2 (depicted by the unfilled area in Figure 3).

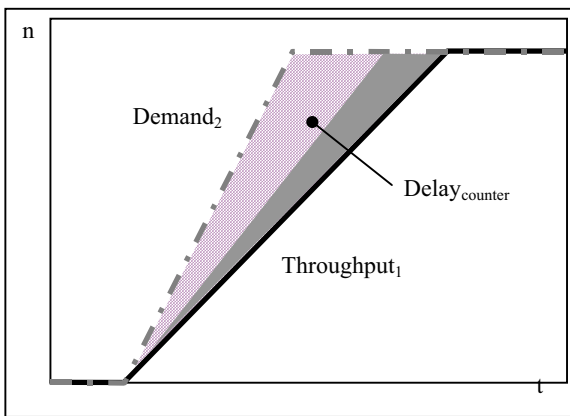


Figure 2. Queuing Scenario under Year 2 Demand and Year 1 Throughput (Counterfactual)

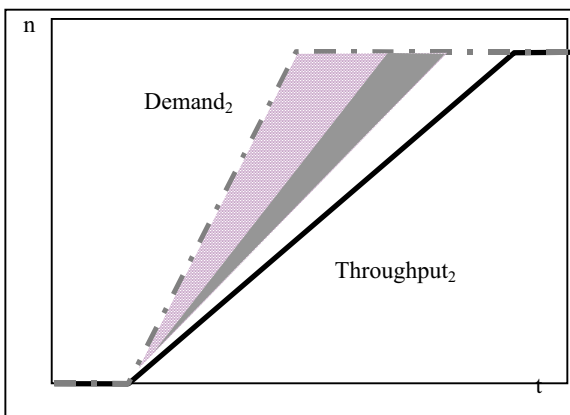


Figure 3. Queuing Scenario under Year 2 Demand and Year 2 Throughput

The figures show an increase in demand and a decrease in throughput from Year 1 to 2, but this trend was chosen for illustrative purposes only. The entire process is summarized in Table II.

TABLE II. DEMAND AND THROUGHPUT SCENARIOS

Demand	Throughput	Total Delay	Δ in Total Delay
<i>for an average day in...</i>			
Year 1	Year 1	(1) Year 1	n/a
Year 2	Year 1	(2) Counterfactual	(2)-(1); due to demand shift
Year 2	Year 2	(3) Year 2	(3)-(2); due to throughput shift

The Year 1 and Year 2 queuing scenarios can easily be constructed from available data (which will be discussed in detail later on), but the counterfactual scenario, because it does not actually exist, must be generated through simulation. The simulation is an iterative process that takes the demand in each time interval and, using a throughput function, assigns a throughput value. All aircraft not served in a time interval comprise the queue in that time interval, and from this a delay calculation can be made.

The classic definition of a deterministic throughput function was introduced in Equation (1). Based on available data sets that include arrival and departure demand, arrival and departure throughput counts, and weather information, we can construct a deterministic throughput function as follows:

$$q_o(t) = \min[d_o(t), c_o(w(t))] \tag{3}$$

- Where $q_o(t)$ is the actual recorded throughput for operation type o in time interval t
- $d_o(t)$ is the actual demand for operation o in time interval t
- c_o is the fixed capacity for operation o
- $w(t)$ is the weather condition at time t

Weather enters into the model as either visual or instrument flight rules (VFR or IFR), and is included as a factor in the model because of the significant impact it has on operational capacity. The operation types are either arrivals or departures.

The deterministic throughput function is an idealized situation and as such does not represent actual operations very well. c_o is a critical input to the function and several major assumptions are needed to determine its value(s). The alternative to the deterministic throughput function is a stochastic model that incorporates some levels of uncertainty. Based on the available data, a stochastic model that preserves the dependence of throughput on demand and weather can be constructed as follows:

$$P(Q_o(t) = q_o(t) | d_o(t), w(t)) = f_Q(q_o | d_o, w) \tag{4}$$

Where Q_o is a random variable representing throughput

f_Q is the conditional probability distribution function for throughput

The probability that Q_o takes some throughput value q_o , conditional on the demand and weather in time interval t , is taken from f_Q . It was necessary to include capacity as an explicit input to the deterministic model; however, in the stochastic throughput function it is implicitly captured in f_Q . f_Q can be constructed entirely from an appropriate data set without having to make assumptions about its shape and parameters. In fact, the non-parametric nature of f_Q is one of the main advantages of this model.

The counterfactual scenario discussed earlier was modeled using approximations of the stochastic throughput function, as described in Figure 4. The deterministic approximation uses mean throughput values conditional on demand, weather, and other known factors to simulate the counterfactual scenario. The stochastic approximation uses random number generation to simulate throughput values. For the purposes of the New York airports delay analysis, Year 1 will correspond to the time period of May through September 2006, while Year 2 represents that of the same months in 2007. Modeling the counterfactual scenario involves assigning simulated 2007 (Year 2) demand and 2006 (Year 1) throughput, calculating queue lengths in each time interval, and then calculating the average delay per flight over all time intervals from May through September. The following is the iterative procedure that was followed.

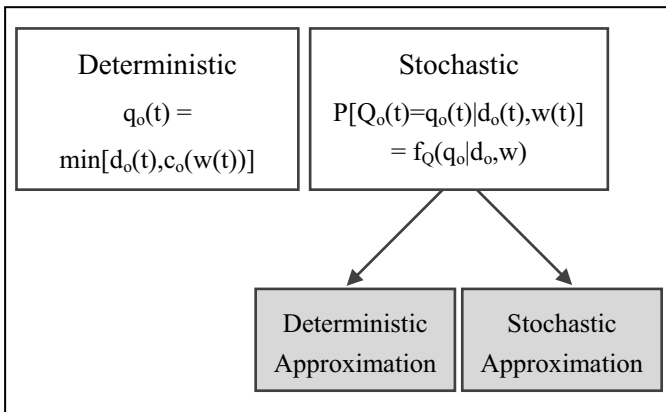


Figure 4. Specifications for the Throughput Function

1) At time interval $t=1$, initialize

$$\hat{D}_{o,07}(1) = D'_{o,07}(1) \tag{5}$$

Where $\hat{D}_{o,07}(t)$ is the simulated total (new & queued) 2007 demand for operation o in time interval t
 $D'_{o,07}(t)$ is the “new” 2007 demand for operation type o in time interval t .

2) Find $\hat{Q}_{o,06}(t)$ conditional on $\hat{D}_{o,07}(t)$, $Q_{o,07}(t)$, & w , where $\hat{Q}_{o,06}(t)$ is the simulated 2006 throughput for operation

type o in time interval t . $\hat{Q}_{o,06}(t)$ is determined using a stochastic throughput function.

3) If $t=T$, go to Step 4. Otherwise,

a) Set

$$\hat{D}_{o,07}(t) = D'_{o,07}(t) + [\hat{D}_{o,07}(t-1) - \hat{Q}_{o,06}(t-1)] \tag{6}$$

Where $\hat{D}_{o,07}(t)$ is comprised of the “new” demand of the current interval t in addition to the queued aircraft (those that are still waiting for service) from the previous time interval $(t-1)$.

b) Update $t=t+1$.

c) Repeat Step 2.

4) Calculate the average delay per flight for operation o for the simulated counterfactual scenario.

$$\hat{\tau}_o = \frac{\Delta t * \sum_{t=1}^T [\hat{D}_{o,07}(t) - \hat{Q}_{o,06}(t)]}{\sum_{t=1}^T \hat{Q}_{o,06}(t)} \tag{7}$$

Where $\hat{\tau}_o$ is the simulated average delay per flight for operation type o , from $t=1$ to $t=T$, in minutes
 Δt is the length of one time interval

The above procedure must be able to reproduce 2006 and 2007 operations as shown in the data such that when the counterfactual scenario is simulated using the same procedure, we can be confident of the results. In other words, the simulation method must produce good agreement between the actual and simulated baselines, which entirely depends on the specifications of the throughput function applied in Step 2. Deterministic approximations to the stochastic throughput function were first tested. These consisted of mean counts conditional on demand and weather, in addition to time of day effects and queue presence indicators, were first tested. Stochastic approximations of the throughput function, which involved randomly drawing from probability distributions of throughput conditional on demand and weather, were also tested. The methods above did not satisfactorily reproduce 2006 and 2007 operations, most likely due to underlying mechanisms not controlled for in the simulation. These phenomena might include serial correlation of demand and throughput between the quarter-hour intervals, arrival & departure interaction effects, and more. Finally, a stochastic approximation method that compares probability distributions of 2006 and 2007 counts, conditional on demand and weather, was tested. This approach, herein referred to as the “compared distribution” method, is able to, by design, identically replicates the baseline scenarios. As such the compared distribution method was chosen for use here.

In the compared distribution method, $\hat{Q}_{o,06}(t)$ is simulated in the following manner by starting with the 2007 (Year 2) data. All steps below are “substeps” of Step (2) from above.

The compared distribution method was used to generate the counterfactual scenarios for all New York airports under analysis. Note that the compared distribution method (as well as the other stochastic approximation method) preserves time of day, day of week, and monthly effects from one year to the next, because the simulation is run time sequentially from $t=1$ to T in both 2006 and 2007.

1) Construct cumulative probability distributions (cdf) of counts conditional on demand and weather, $F(Q_{o,07}|D_{o,07},w_{07})$, for 2006 and 2007.

2) Find the cumulative probability of the empirical 2007 count for operation type o , conditional on 2007 demand and 2007 weather condition for some time interval t , $F(Q_{o,07}|D_{o,07},w_{07})$.

3) Based on the simulated 2007 demand, $\hat{D}_{o,07}$, find the interval in the 2006 count cdf that the 2007 probability found in the previous step falls into. From this, lower and upper bounds ($F_L(Q_{o,06}|\hat{D}_{o,07},w_{06})$ and $F_U(Q_{o,06}|\hat{D}_{o,07},w_{06})$, respectively) of the 2006 cdf and corresponding 2006 simulated count values ($Q_{o,06,L}$ and $Q_{o,06,U}$, respectively) are obtained.

4) Construct a probability value, $f(x)$, for the simulated 2006 count based on the 2007 count cdf's position between the lower and upper bounds of the 2006 cdf interval:

$$f(x) = \Pr(\hat{Q}_{o,06} = Q_{o,06,L}) \quad (8)$$

$$= \frac{F_U(Q_{o,06}|\hat{D}_{o,07},w_{06}) - F(Q_{o,07}|D_{o,07},w_{07})}{F_U(Q_{o,06}|\hat{D}_{o,07},w_{06}) - F_L(Q_{o,06}|\hat{D}_{o,07},w_{06})}$$

$$\Pr(\hat{Q}_{o,06} = Q_{o,06,U}) = 1 - f(x) \quad (9)$$

5) Generate random number n . If $n \leq f(x)$, set count to lower bound 2006 count $Q_{o,06,L}$; otherwise set count to upper bound $Q_{o,06,U}$.

There are fewer count data recorded at very high demand values, and as a result the cumulative probability distributions of counts conditional on high demands are often based on small and incomplete data sets. To avoid reliance on probability distributions constructed using sparse data, all counts recorded with demands beyond the capacity threshold were combined into a single truncating probability distribution at the cut-off demand. For all simulated demands higher than that of the demand truncation point, this combined probability distribution is used for count simulation.

IV. DESCRIPTION OF DATA

The Aviation System Performance Metrics (ASPM) database is part of the Federal Aviation Administration's (FAA's) Operations and Performance Data system. Data from the "Download/Airport" section of the ASPM database was used for this analysis. The data includes hourly as well as quarter-hourly arrival and departure counts, demands, and visibility conditions (either visual (VFR) or instrument (IFR)

flight rules). The data is available for 77 major airports in the United States.

ASPM count data are based on individual aircraft landing and take-off times as supplied through Airline Service Quality Performance (ASQP) data or Enhanced Traffic Management System (ETMS) messages.

ASPM provides the perfect data set to construct the counterfactual scenarios described in the previous section; however, some particular characteristics of the ASPM demand data selected for this analysis must be noted. Firstly, the demand data used here is based on the updated flight plan just before a flight is due to take off at the origin airport; it does not reflect demand as defined by airline schedules. As a result, for flights arriving at a given airport, the delay calculated in this analysis includes all delays that occur between the filed flight plan take-off time (demand) and actual landing time (count), but does not include the delays between scheduled and flight plan take-off times (although this information can also be found in the ASPM dataset). For flights departing the airport, the delay calculated in this analysis includes the delay incurred between the time that the flight was scheduled to depart according to the flight plan, and the time that it actually does depart. As a result, the calculated delay will not include the effects of ground delay programs (GDP), the effects of air traffic management (ATM), plus other mechanisms that would cause a flight to deviate from its schedule. Secondly, the reported demand represents the total number of aircraft that were available for operation o (arrival or departure) in time interval t . An aircraft will count towards demand in each and every time interval starting in the one when it was first available to land/depart until the time interval when it is actually able to do so. As a result, the demand $D_o(t)$, reported in t includes the "new" demand $D'_o(t)$, plus the queued (unserved) aircraft from the previous time interval $[D_o(t-1) - Q_o(t-1)]$. $D'_o(t)$ for each 15-minute interval is easily calculated from the ASPM dataset, and is used for input to the simulation.

$$D'_o(t) = D_o(t) - [D_o(t-1) - Q_o(t-1)] \quad (10)$$

Where $D'_o(t)$ is the "new" demand for operation type o in time interval t

Δt is the length of one time interval

$D_o(t)$ is the total demand for operation o in time interval t

$Q_o(t-1)$ is the throughput for operation o in time interval $t-1$.

If an aircraft's demand and service times fall within the same or adjacent intervals, its delay is recorded to be zero. For instance, if time intervals are 15 minutes in length, an aircraft will not be counted towards delay if its demand and actual service times are, for instance, 1 minute and 14 minutes into the interval respectively. Also, ASPM counts will never exceed the total demand in any given time interval, meaning that operations which occur earlier than scheduled are not counted as negative delay or a delay savings.

Data from LGA, EWR, and JFK were obtained for May 1 through September 30 2006, and May 1 through September 30 2007. From the data, we can derive the following information about demands, throughput, and delay. Figure 5 displays cumulative arrival demands by hour averaged over all days from May 1 through September 30. It can be observed that the total daily demand (averaged over all days) at LGA and EWR has decreased (between 2% and 3%) from 2006 to 2007 while it has increased significantly (by approximately 18%) at JFK. As expected, departure demands exhibit very similar trends and as a result are not displayed here. Figure 6 displays the average arrival count recorded during VFR conditions plotted against demand. The average arrival count per demand was calculated by averaging all counts recorded at each demand level from 0 to 70+. Observe that the arrival counts match arrival demands up to a certain point, after which this trend stops as the facility cannot serve at the demanded rate any longer. After this peak count level, arrival counts remain steady or begin to decrease until the slope of the curve flattens out. Also beyond the peak, all demand cannot fully be served within the same time period any longer. The peak arrival count is the realized arrival capacity for a given airport [7]. Based on this simple yet reliable capacity estimation, Figure 6 suggests that the arrival capacities of all three airports have decreased from 2006 to 2007. One can also observe that higher arrival demands were reported at LGA and JFK in 2007, which suggests that there were longer queues, which in turn suggests that aircraft waited longer for service and therefore experienced greater delay in 2007. The same phenomenon, however, was not recorded at EWR. A similar analysis can be applied to the averaged departure counts in Figure 7, which implies that departure capacities have dropped at LGA and EWR but have increased at JFK from 2006 to 2007. However, much higher demands (and therefore queuing) were reported at JFK in 2007, which may be the result of increased demand and/or more severe demand peaking effects, as the data does not seem to suggest that capacity has decreased.

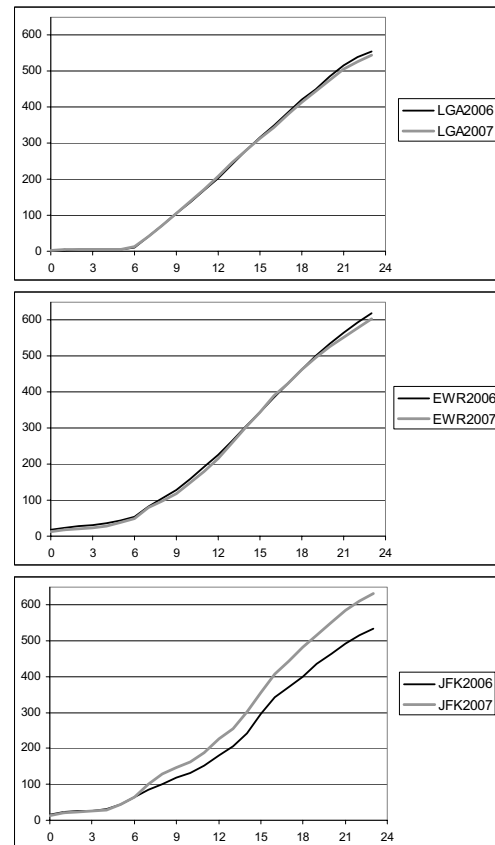


Figure 5. Cumulative Arrival Demands (Average Day of Ops), by Hour

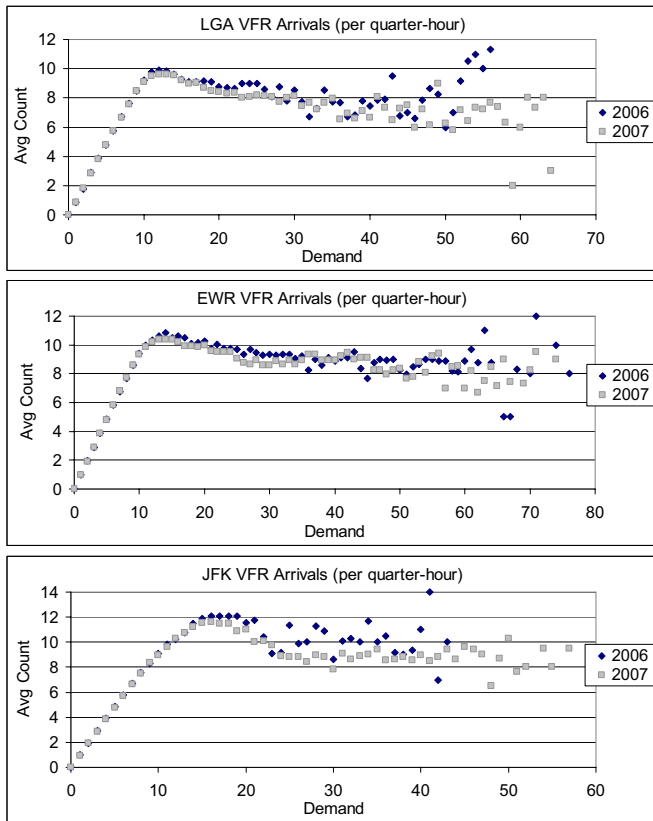


Figure 6. Average Arrival Counts vs. Demand, in VFR, by Quarter-Hour

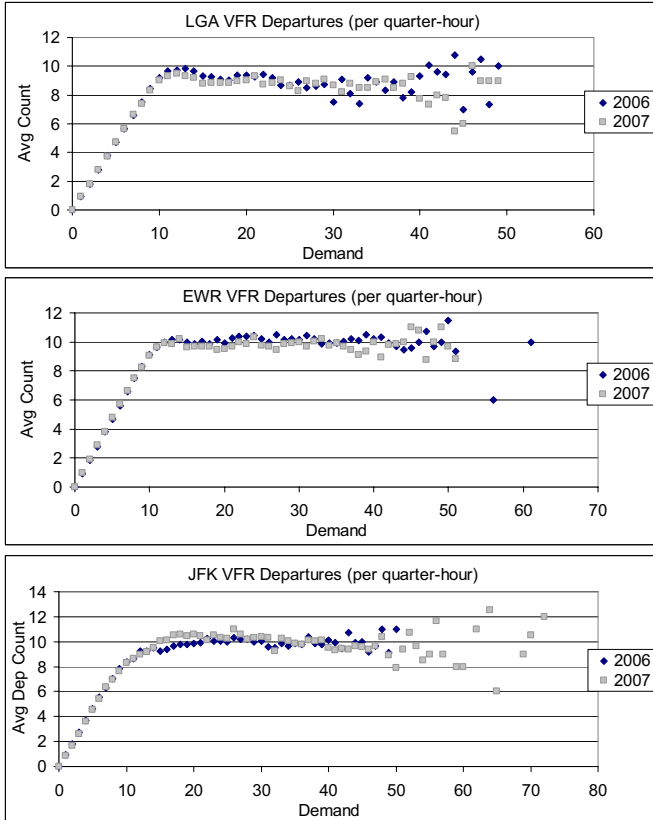


Figure 7. Average Departure Counts vs. Demand, in VFR, by Quarter-Hour

The average delay per flight was also calculated for arrival and departure operations at each airport from May through September of 2006 and 2007 as per Equation (7). Recall that delay is calculated against flight plan demand, and the data is tabulated in 15-minute intervals (such that $\Delta t=15$ min). The data set contains $T=14,688$ quarter-hour intervals.

The delay results are summarized in Table III.

TABLE III. AVERAGE DELAY PER FLIGHT, MAY-SEPT 2006 & 2007

		Average delay per flight (min)		Change (from 2006 to 2007)
		2006	2007	
LGA	Departure	8.56	10.72	+2.16
	Arrival	8.85	10.7	+1.85
EWR	Departure	11.53	9.95	-1.58
	Arrival	11.46	12.06	+0.60
JFK	Departure	12.06	14.38	+2.32
	Arrival	3.23	8.11	+4.88

The average delay per flight increased at both LGA and JFK between 2006 and 2007, and significantly so for JFK arrivals. Average delay has decreased by about 1.6 minutes per departing flight at EWR, and for arrival flights it has increased 0.6 minutes. These results from ASPM are consistent with the OPSNET data discussed previously.

Modeling the counterfactual scenarios involves recreating the structure of the ASPM demand and count data by assigning simulated 2007 demand and 2007 throughput values, and then calculating queue lengths and average delay in the same manner as was done for the data shown in Table III.

V. RESULTS

The simulation results are summarized in Table IV. The reported counterfactual delays are the average of 10 simulation runs for each scenario. The standard deviations of the 10 runs are also reported.

TABLE IV. DELAY RESULTS

	Average Delay per Flight (min)			Δ delay due to Δ demand (implies)	Δ delay due to Δ throughput (implies)	SD**
	2006	CF*	2007			
LGA						
Departure	8.56	6.49	10.72	-2.08	4.24	0.115
				<i>Demand has decreased</i>	<i>Throughput has decreased</i>	
Arrival	8.85	6.03	10.70	-2.82	4.67	0.097
				<i>Demand has decreased</i>	<i>Throughput has decreased</i>	
EWR						
Departure	11.53	5.78	9.95	-5.76	4.18	0.064
				<i>Demand has decreased</i>	<i>Throughput has decreased</i>	
Arrival	11.46	6.65	12.06	-4.81	5.41	0.097
				<i>Demand has decreased</i>	<i>Throughput has decreased</i>	
JFK						
Departure	12.06	19.09	14.38	7.03	-4.71	0.138
				<i>Demand has increased</i>	<i>Throughput has increased</i>	
Arrival	3.23	4.91	8.11	1.68	3.20	0.084
				<i>Demand has increased</i>	<i>Throughput has decreased</i>	

* Counterfactual, referring to scenario with 2007 demand and 2006 throughput
 ** Standard deviation of counterfactual delay, for 10 simulation runs made

TABLE V. DELAY RESULTS (COUNTERFACTUAL SCENARIO II)

	Average Delay per Flight (min)			Δ delay due to Δ demand (implies)	Δ delay due to Δ throughput (implies)	SD**
	2006	CF*	2007			
LGA						
Departure	8.56	13.08	10.72	4.52	-2.36	0.093
				<i>Throughput has decreased</i>	<i>Demand has decreased</i>	
Arrival	8.85	13.36	10.70	4.51	-2.66	0.287
				<i>Throughput has decreased</i>	<i>Demand has decreased</i>	
EWR						
Departure	11.53	20.33	9.95	8.80	-10.38	0.088
				<i>Throughput has decreased</i>	<i>Demand has decreased</i>	
Arrival	11.46	16.40	12.06	4.94	-4.34	0.351
				<i>Throughput has decreased</i>	<i>Demand has decreased</i>	
JFK						
Departure	12.06	9.10	14.38	-2.97	5.29	0.060
				<i>Throughput has increased</i>	<i>Demand has increased</i>	
Arrival	3.23	6.52	8.11	3.29	1.59	0.084
				<i>Throughput has decreased</i>	<i>Demand has increased</i>	

* Counterfactual, referring to scenario with 2006 demand and 2007 throughput
 ** Standard deviation of counterfactual delay, for 10 simulation runs made

The results are consistent with the trends seen in the OPSNET data, as well as the ASPM data presented in the previous section and used for this simulation. At both LGA and EWR, arrival and departure demand changes have results in decreases in arrival and departure delay, implying that demand has declined. In addition, delays attributed to changes in throughput have increased, which would imply that throughput has dropped as well. At JFK, arrival and departure delays have increased due to changes in demand, suggesting that demands have gone up (with the departure demands having caused relatively significant increases in delay). However, increases in departure throughput have caused departure delays to drop while arrival throughput may have decreased and caused a subsequent increase in arrival delay.

The counterfactual scenario can also be constructed by swapping the demand and throughput years and simulating 2006 demand with 2007 throughput; in other words, using the same procedure described above but with the years switched. In this case, the difference between the counterfactual and 2006 base year delays can attributed solely to changes in throughput, and the difference between the 2007 base year and counterfactual scenario delays to changes in demand. Table V contains the results of this simulation.

The delay trends in Tables IV and V are consistent with one another. It also appears that the magnitudes of the changes in delay are consistent between the two analyses at LGA, EWR arrivals, and JFK, although there is greater discrepancy in the departure results for EWR. Because demand, throughput and delay are not necessarily related linearly, the “direction” in which the counterfactual scenario is simulated could have a significant effect on the delay results (of Tables IV and V). However, the choice regarding which way to simulate the counterfactual scenario is arbitrary, and consequently the two sets of delay results may serve to validate the simulation process. The differences between the two sets of results for EWR departures may be due to other dependent effects not accounted for or readily apparent in the simulation process. Also, as demands increase, delays also increase at much faster rates; conversely, when demands are lower an increase in throughput can result in a significantly greater delay reduction [2]. This may account for the fact that the Table V results for EWR show much larger changes in delay between the two years than Table IV.

We can make a few inferences based on the results in Tables IV and V above. Firstly, of the three airports JFK has experienced the largest overall increase in delay due to changes in throughput and demand. In particular, a substantial growth in arrival and departure demands has contributed to the large increase in delay at JFK. Departure throughputs have not similarly increased to offset this rise in demand, while the problem in the arrival operations is further exacerbated by a decrease in throughput. Decreased throughput does not necessarily mean a drop in airport capacity. In fact, sources at

the FAA believe that fleet mix changes in 2007 at JFK led to higher minimum in-trail separations, which would certainly reduce throughput. It has also been suggested that New York airspace controllers had grown more conservative about aircraft separations due to safety concerns. This would support the findings of Figures 6 and 7, which suggest that capacities have generally decreased (except for JFK departures) between 2006 and 2007. The drop in demand at LGA and EWR (Figure 5) occurred alongside a drop in throughput, and generally resulted in an overall increase in delay at these airports.

VI. CONCLUSIONS

The New York airports experienced a very significant rise in delays over the summer of 2007 compared to previous periods, most specifically that of summer 2006. The purpose of this work was to estimate how much of this change in delay was due to demand changes and how much was due to throughput changes. Because demand and throughput change simultaneously, the purpose of this work was to quantify how changes in each contribute to a change in delay, and ultimately provide information about the causes of delay at one greater level of detail. To do this, an empirically driven simulation procedure was developed from classical queuing concepts, and applied to a case study of the three major New York area airports in summer 2006 and 2007. This procedure consists of a stochastic throughput function whose main advantages are that it uses non-parametric, empirically-based probability distributions and that capacity need not be estimated explicitly. The throughput function was used to recreate the structure of the ASPM data and construct the intermediate "counterfactual" scenario, by which the delay changes from 2006 to 2007 could be attributed to either demand or throughput.

The simulation results confirmed the OPSNET and ASPM data results. The counterfactual scenario was first constructed with 2007 demand and 2006 throughput. At both LGA and EWR, arrival and departure demand changes have results in decreases in arrival and departure delay, implying that demand has declined. In addition, delays attributed to changes in throughput have increased, which would imply that throughput has dropped as well. At JFK, arrival and departure delays have increased due to changes in demand, suggesting that demands have gone up. However, increases in departure throughput have caused departure delays to drop while arrival throughput may have decreased and caused a subsequent increase in arrival delay. The counterfactual scenario was also constructed with 2006 demand and 2007 throughput, and the results of this simulation served to validate the previous simulation results.

VII. FURTHER WORK

This procedure is a starting point from which we can further analyze and deconstruct the causes of operational delay at airports in terms of demand and throughput. However, knowing only the demand and throughput effects on delay has limited importance; it would be beneficial to identify factors other than flight rule conditions that influence demand and throughput. This could, in turn, be used to re-specify the throughput function to control for additional factors not yet included in the model. Phenomenon yet uncontrolled for might include fleet mix changes, and arrival/departure interaction effects (the model as of yet assumes arrivals & departures to be independent of one another). Another direction for future work is to base delay calculations on a demand scenario other than that of the flight plan, such as demand recorded at the time flights are scheduled by the airlines to arrive or depart. Using this, the effects of GDP as well as all the effects of ATM at origin airports could be incorporated into the analysis.

ACKNOWLEDGMENT

The authors would like to thank the Air Traffic Organization at the FAA, particularly Joe Post, Dan Murphy and Michael Wells, for sponsoring this work as well as providing invaluable insights and suggestions. We would also like to thank Dave Knorr at the FAA and Ken Wright at MITRE for their assistance.

REFERENCES

- [1] G.F. Newell, "Airport capacity and delays," *Transportation Science*, Vol. 13, no. 3, 1979.
- [2] Mark Hansen, "Post-deployment analysis of capacity and delay impacts of an airport enhancement: case of a new runway at Detroit," *Air Traffic Control Quarterly*, Vol. 12, pp. 339-365, 2004.
- [3] Ken Belson, "Study puts price tag on delays at airports," *New York Times*, December 2, 2007.
- [4] Dave Knorr, "Delay Metrics: Why are Delays Increasing?," Presentation given at the National Airspace System Performance Workshop, Asilomar, Pacific Grove, CA, 2007.
- [5] Ken Wright, "Pangs of New York: Reducing Delays at Kennedy, LaGuardia, and Newark Airports", panel member, Transportation Research Board Annual Meeting, Washington, D.C., 2008.
- [6] Carlos Daganzo, *Fundamentals of transportation and traffic operations*, 1st ed.. Oxford; New York: Pergamon, 1997.
- [7] Yu Zhang, Jasenka Rakas, and Mark Hansen, "Methodology for Estimating Airport Capacity and Throughput Performance," unpublished.

Propagation of Airspace Congestion. An Exploratory Correlation Analysis

Claus Gwiggner, Kota Kageyama, Sakae Nagaoka

Electronic Navigation Research Institute, 7-42-23 Jindaiji-higashi, Chofu
Tokyo, 182-0012, Japan
Tel: 0081-422-41-3184
email: [claus,kage,nagaoka]@enri.go.jp

Abstract—We analyze how large gaps between the planned and realized number of aircraft into flight sectors propagate through the European- and the Japanese Airspace. For this we analyze the sample cross-correlation matrix of the most congested part of the networks. Because of the motion of aircraft, gaps propagate to neighboring sectors, expecting positive correlation coefficients. The question in the analysis is whether there are unexpected coefficients. Such coefficients would be caused by traffic controllers or flow managers who compensate for strong gaps by re-routings or speed adjustments. Such strategies would often lead to negative correlation coefficients. Our results show that meaningful correlations appear on two levels: (i) locally, that is between a sector and its direct neighbors and (ii) globally on ‘traffic highways’, that is between sectors that are connected through a flight route with high traffic densities. This is true for both, the European- and the Japanese Airspace. Moreover, all correlations are positive and their time-lags correspond to the average travel times. No unexpected correlations have been found. We conclude that no systematic strategies to compensate strong delays are applied by controllers. The results are useful to justify predictive congestion models for future flow planning. They also give a first insight into how controllers deal with their workload, although a more detailed analysis is required to explore this topic.

Index Terms—Flow analysis, correlation analysis

I. INTRODUCTION

Airspace is divided into geographical regions, called sectors. A flight plan is a sequence $(S_1, t_1), \dots, (S_n, t_n)$ of sectors S_i and entry times t_i in the sector. Due to uncertainties (weather conditions, congestion etc.), aircraft can deviate from their flight plans. [BLHM05] classify the major sources of uncertainty as

- Demand uncertainty: flights fail to meet planned departure, arrival or en-route travel times. Contributing factors are mechanical problems, boarding passengers or weather conditions.
- Capacity uncertainty: airport and airspace throughput levels vary. Contributing factors are weather conditions and changes in flight sequences that disturb scheduled departure or arrival spacing.
- Flow control uncertainty: actions are taken by the traffic controllers in response to demand and capacity uncertainty. Examples are re-routing, re-sectorization and temporary capacity limitations. The human element of

decision making adds another layer of uncertainty to the whole system.

Deviations from flight plans lead to gaps between the planned and the real number of aircraft entering flight sectors. For example in the year 2004, 17.7 % of European flights departed- and 18.5 % arrived more than 15 min behind their schedule [EUR06].

Obviously, a gap between planned and real number of entries in a sector S in time slot t propagates to its neighboring sectors in slot $t+1$, because aircraft cannot stand still. On the other hand, pilots and air traffic controllers can compensate gaps by re-routing or speed adjustments of all aircraft.

In this article we analyze past flight data to see how such gaps propagate in reality through the airspace. Are there strategies of controllers to compensate the gaps successfully? We will look at (i) local propagation, that is propagation between a sector and its direct neighborhood and (ii) global propagation, i.e. between a sector and any other sector in the system. Based on such knowledge flow planning can be improved, because systematic gaps can be controlled, once their mechanisms are understood.

The article is divided into two parts and a conclusion: in the first part we explain the method and give some examples from literature. In the second part we report our results. We conclude with a critical comment and motivate future work.

II. METHOD AND RELATED WORK

We consider $Z_t = [Z_{1t}, Z_{2t}, \dots, Z_{mt}]'$, $t, Z_{it} \in \mathbf{Z}$ as a random process where Z_{it} represents the gaps between planned and realized number of aircraft entering sector i in time slot t . Our aim is to study the correlation structure of the process. Positive correlation between two sectors i and j in time slots t_1 and t_2 has the meaning that gaps above average in sector i and slot t_1 are associated with gaps above average in sector j and slot t_2 . As mentioned above, we expect such correlations between neighboring sectors. But we are more interested in unexpected correlations in the real traffic data. For example, take a sector with a

crossing of two routes. When a traffic peak on the first route is predicted to arrive at the sector some time ahead, the controllers could coordinate with sectors on the second route to re-route aircraft for compensation. Such a strategy would cause negative correlation between the sectors along the two routes. Likewise, vanishing correlation between two sectors, when conditioned on the value of other sectors, might help reveal network effects.

Related work from the ATM domain can be found in the analysis of flight data on a sector level: [WCGM03] and [WSZ⁺05] analyze uncertainties in sector demand. One of their observations is that flow control actions against congestion are visible in the data, in cases that the predicted peak counts are greater than some alert value displayed in the Enhanced Traffic Management System (ETMS). [RSWB06] analyze radar data to identify traffic flows in the U.S. airspace. They define a flow as a cluster of aircraft with similar trajectory properties. A trajectory is a high-dimensional vector of geographical components. They apply several clustering techniques to the data. But even after enhancing the data set with additional features (e.g. aircraft type), they conclude that none of the algorithms provides satisfactory results for practical purposes. A correlation analysis of sector data, as proposed in our article, has not been identified in literature review. This might be due to the known difficulties in the interpretation of auto- and cross correlation coefficients [Ken89], [Dig90]. On the other hand, correlation analysis is the first step in an analysis of multiple time series, as for example applied to highway traffic prediction in [KP03]. In what follows, we analyze the risk of misleading coefficients in our data before visualizing the most interesting correlation patterns. This is exploratory work with the aim to generate new hypotheses about the phenomenon.

A. Inference for cross-correlation matrices

In this part we define the sample correlation matrix function between multiple time series and derive bounds for the variability of its coefficients.

1) *Estimation*: We use the standard estimators of lag- k crosscorrelation

$$\hat{\rho}_{ij}(k) = \frac{\hat{\gamma}_{ij}(k)}{[\hat{\gamma}_{ii}(0)\hat{\gamma}_{jj}(0)]^{1/2}}$$

with sample crosscovariance elements

$$\hat{\gamma}_{ij}(k) = \frac{1}{n-k} \sum_{t=1}^{n-k} (Z_{it} - \bar{Z}_i)(Z_{j,t+k} - \bar{Z}_j)$$

where

$$\bar{Z}_i = \frac{1}{n} \sum_{t=1}^n Z_{it}$$

are the component-wise sample means of an observation consisting of n time slots.

These estimators are asymptotically normally distributed [Ken89]. Modifications exist to address issues of bias and

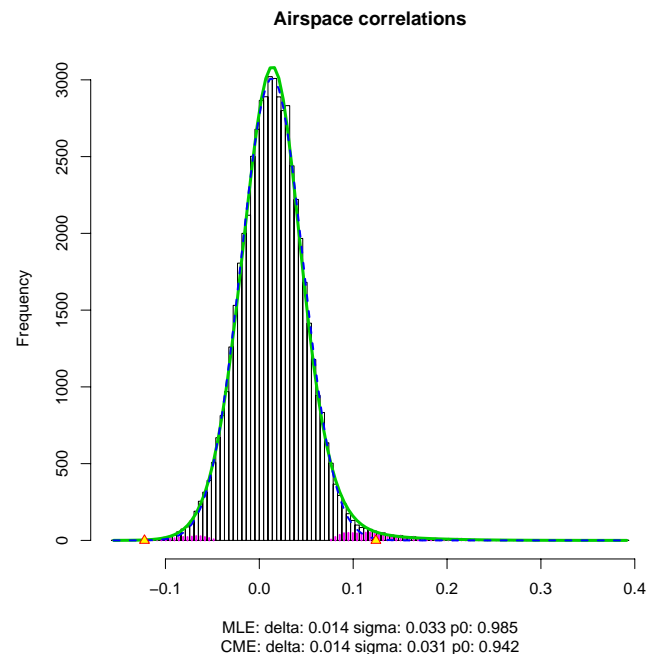


Figure 1. Histogram of estimated correlation coefficients. Bold: empirical distribution. Blue: empirical null distribution. Triangles: local false discovery rate. Data: Japanese Airspace.

high-dimensionality [Ken89], [LW04] and [SS05]. A disadvantage of the latter approaches is that the sample properties of their estimators are not known. Note also that the matrices do not have to be invertible for our study.

2) *Sample Variability*: Our objective is to decide whether the coefficients of the correlation matrix differ significantly from 0. For this, the variance of the sample correlations has to be known. For a large number n of independent observations the variance of a single sample correlation coefficient under the hypothesis that the true correlation is 0 is $\frac{1}{n-1}$ [Sap06]. There are two reasons why this result cannot be used directly in our analysis: (i) our observations are not independent and (ii) there is a large number of hypotheses to be evaluated.

a) *Bartlett*: When observations are dependent, a result from Bartlett gives insight into the problem [KSO83]. It shows that when the stationary series $Z_i(t), Z_j(t)$ are uncorrelated and estimated from a single realization

$$V[\hat{\rho}_{ij}(k)] = \frac{1}{n-k} \sum_{s=-\infty}^{\infty} \rho_{ii}(s)\rho_{jj}(s) \quad (1)$$

This means that even for large n , the variance of the sample correlations depends on all correlations of the original processes, which are generally unknown. Consequences are (i) a risk of ‘spurious’ correlations and (ii) that it is impossible to estimate these quantities directly from a finite sample. In practice, approximations are often used, for example by assuming that the individual series correspond to white noise (for example after pre-whitening).

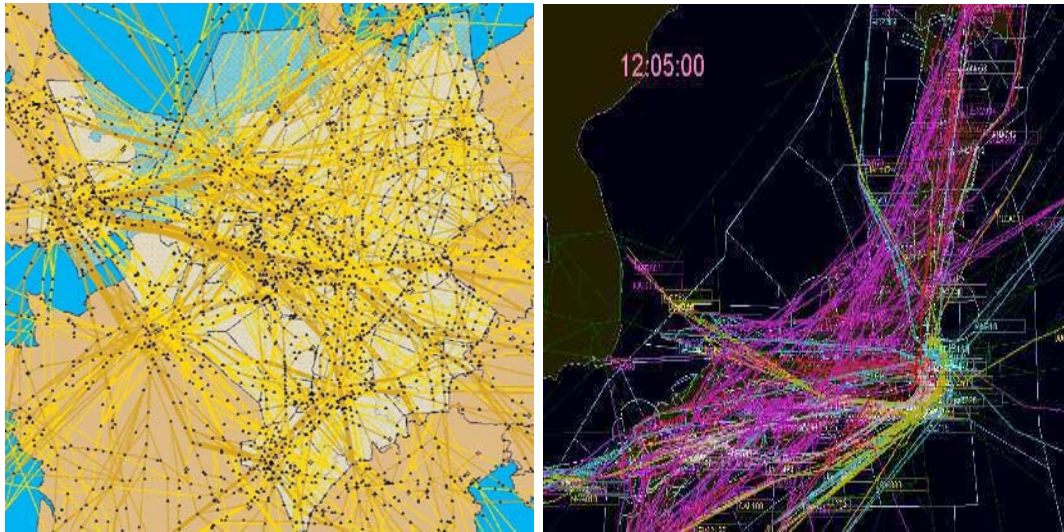


Figure 2. Airspaces. Left: European Central Airspace. Right: Japanese Airspace.

l_{max} \ ρ_{max}	0.01	0.05	0.1	0.2
10	1.002	1.05	1.20	1.8
75	1.015	1.375	2.50	7.0
100	1.02	1.50	3.00	9.0
150	1.03	1.75	4.00	13.0

Table I

VARIANCE INFLATION AS A FUNCTION OF SERIAL CORRELATION

To characterize the risk of spurious correlation in our instance, we calculate the variance inflation for several dependency structures compared to independent observations due to Bartlett’s formula. Looking ahead to Figure 4 we used scenarios with a low amount of constant dependency ρ_{max} up to time-slot l_{max} in order to obtain upper bounds for the inflation. The dependency structures are the following:

$$\rho_{ii}(s) = \rho_{jj}(s) = \begin{cases} \rho_{max} & s < l_{max} \\ 0 & else \end{cases}$$

Under this structure, equation 1 becomes

$$V(\hat{\rho}_{ij}(k)) < \frac{1}{n-k} \sum_{s=-\infty}^{\infty} \rho_{ii}^2(s) = \frac{1}{n-k} (1 + 2l_{max}\rho_{max}^2)$$

Table I shows $nV(\hat{\rho}_{ij}(1))$ for different values of ρ_{max} and l_{max} : For example, for a correlation of $\rho_{max} = 0.1$ up to lag $l_{max} = 10$, an inflation of 20% would occur. For stronger correlations, an explosion of the variance can be seen (bottom right part of the table). Again, looking ahead to Figure 4, we expect weak correlations in our series. We can expect 30 - 70 % increase of variance with respect to independent realizations.

b) *False discovery rates*: The second problem is that of the large number of coefficients to be evaluated. Classical hypothesis tests would expect a large number of rejections

by their very nature [Efr04]. [ETST01] proposes a heuristic method to identify a number of ‘interesting’ coefficients in large-scale testing contexts. They define the *local false discovery rate*

$$fdr(\hat{\rho}) \equiv f_0(\hat{\rho})/f(\hat{\rho})$$

where $f_0(\hat{\rho})$ is the density of uninteresting coefficients and $f(\hat{\rho})$ the density of all coefficients. *fdr* is the expected proportion of null coefficients in a selection of coefficients with value $\hat{\rho}$. Interesting coefficients are those with $fdr(\hat{\rho}) < c$, a threshold value, comparable in meaning with the significance level of classical tests.

Figure 1 shows the histogram of all $21 \times 21 \times 30 = 13230$ cross-correlation coefficients in our matrix for the Japanese Airspace (please see below for details on the selection of the 21 sectors). It has been estimated from 11 days of data, each consisting of 288 observation intervals. The bold line (green) is the empirical distribution, fitted by a polynomial of degree 3. The dotted blue line is the empirical null distribution, fitted by Efron’s method. It is a normal distribution with unknown variance. Both distributions look almost identical; small differences can be seen at the peak and $\hat{\rho} \sim 0.1$. The triangles mark the interval, outside which the computed $fdr < 0.2$. Finally, the pink bars represent the estimated mass of non-null coefficients. The majority of their mass lies inside the *fdr* interval. We obtain three results: (i), correlation coefficients above 0.12 can be regarded as interesting, (ii), the standard deviation of the empirical null distribution is 0.033, which is $\approx 86\%$ larger than the null variance for independent observations (for 11 days, each 288 observations). This is in agreement with the results from the previous paragraph. And (iii), a risk that interesting coefficients will be undetected exists.

To summarize, we analyzed how dependent observations

and a large number of variables affect statistical methods to infer significant correlation coefficients. The first approach showed that a variance inflation has to be expected and the second that there is risk of leaving interesting coefficients undetected. Both methods suggest a rather small critical value for interesting coefficients. At this point we remind that we wish to explore meaningful patterns of correlation rather than single coefficients. Subjective judgement may prove useful in this task.

III. RESULTS

We analyzed correlations in the European and in the Japanese airspace. The left part of Figure 2 shows the most congested part of the European Airspace. It comprises 31 sectors covering London, Zurich and Berlin, belonging to 9 control centers. The daily number of aircraft is about 8000 for this area. The yellow routes are from North to South and the brown ones from South to North. Between London and Frankfurt, one can see a bi-directional high density route. The Japanese airspace can be seen in the right part of Figure 2. Our area of interest contains 21 sectors, covering Fukuoka (south), Tokyo (center) and Sapporo (north). These sectors belong to 3 control centers. For these sectors, more than 85 % of the entry-times of the aircraft could be determined accurately. About 4000 aircraft use this part of the airspace every day. One can see high traffic routes from/to Tokyo (yellow, blue) as well an important number of over-flights (pink).

More formally, we consider the vector of random processes $GAP_t = PLN_t - REAL_t, t \in \mathbf{R}$, where the i th component $GAP_{i,t} = PLN_{i,t} - REAL_{i,t}$ represents the gaps between the planned and realized number of entries in sector i . The process is observed in 5 minutes time intervals, leading to 288 samples per day. For the European Airspace, 91 weekdays are available (Mon-Thu) in the summer period May, 13 - Sept. 29, 2004. For the Japanese Airspace, 11 days from August and November in the Year 2006 are available.

c) Time-Plots: Typical time plots of one component process $GAP_{i,t}$ can be seen in Figure 3. The top panel shows a sector from the European Airspace, the bottom shows an example from the Japanese Airspace. In both, the gaps fluctuate around 0, the variance looks constant during the day (7-19h). The marginal distributions of the processes turned out to be symmetric, as expected (not shown). In the following, we assume that the component processes are second-order stationary during the day.

d) Cross-correlation plots: We now analyze in more detail cross-correlations between local neighbors (local correlation) and between far lying sectors (non-local correlation). Figure 4 shows typical cross-correlation matrices. In the left panel, the 2x2-matrix from the two neighboring sectors T01 and T27 from the Japanese airspace are shown. The diagonal elements correspond to the autocorrelation functions (acf) up to lag 30, corresponding to 2h30. Both

Airspace	Type	avg	max
Europe	local	0.19	0.34
	non-local	0.16	0.24
Japan	local	0.24	0.28
	non-local	0.23	0.36

Table II
SUMMARY STATISTICS FOR CORRELATION COEFFICIENTS. TOP:
EUROPEAN AIRSPACE. BOTTOM: JAPANESE AIRSPACE.

Airspace	Type	# coeffs	lag-range
Europe	local	1.29	[-4, 3]
	non-local	1.09	[-6, 6]
Japan	local	4.3	[-4, 2]
	non-local	5.2	[-8, 8]

Table III
SUMMARY STATISTICS FOR CORRELATED SECTORS. TOP: EUROPEAN
AIRSPACE. BOTTOM: JAPANESE AIRSPACE.

show no peaks. The off-diagonal elements display the cross-correlations for positive lags in the upper diagonal $\rho(GAP_{i,t}, GAP_{j,t+k})$ and negative lags in the lower diagonal $\rho(GAP_{i,t}, GAP_{j,t-k})$. A peak at lag -3 has value 0.26. Its neighbors (lag -2 and -4) show still some higher value than the remaining ones. These three coefficients are the only interesting in the plot.

For more insight into correlation between far lying sectors, we analyze the two sectors EXH and EUY from European Airspace. They are separated by the two sectors EUF and EXE. Their correlation matrix function is plotted in the right panel of Figure 4. A decay of autocorrelation, starting from -0.1, can be seen. A peak in the cross-correlation is found at lag -5.

Table II summarizes the significant correlations of the full cross correlation matrices. In Europe, local correlations are on average 0.19 and have a maximum of 0.34 (columns 2, 3). The non-local correlations are on average 0.16 and have a maximum of 0.24. In Japan, the local correlations are on average 0.24 with a maximum of 0.28. And the non-local correlations are on average 0.23 and have a maximum of 0.36. All correlation coefficients are positive. Table III summarizes how two sectors are correlated. Of interest are the number of significant coefficients (at different lag values) and the time lag of these coefficients. In Europe, for locally correlated sectors, 77 % have exactly one significant coefficient, 19 % have two and 4 % three or four, leading to an average of 1.29 coefficients (column 2). In the Japanese Airspace, the average number is 4.3. For non-locally correlated European sectors, 91 % have exactly one and 9 % have two significant coefficients, averaging 1.09. The Japanese is higher again, with 5.2 significant coefficients per correlated sectors. Local correlations occur between lags -4 and 3, and non-local ones between lags -6 and 6 (column 3) in the European and between lags [-4,2] and [-8,8] in the Japanese Airspace. The

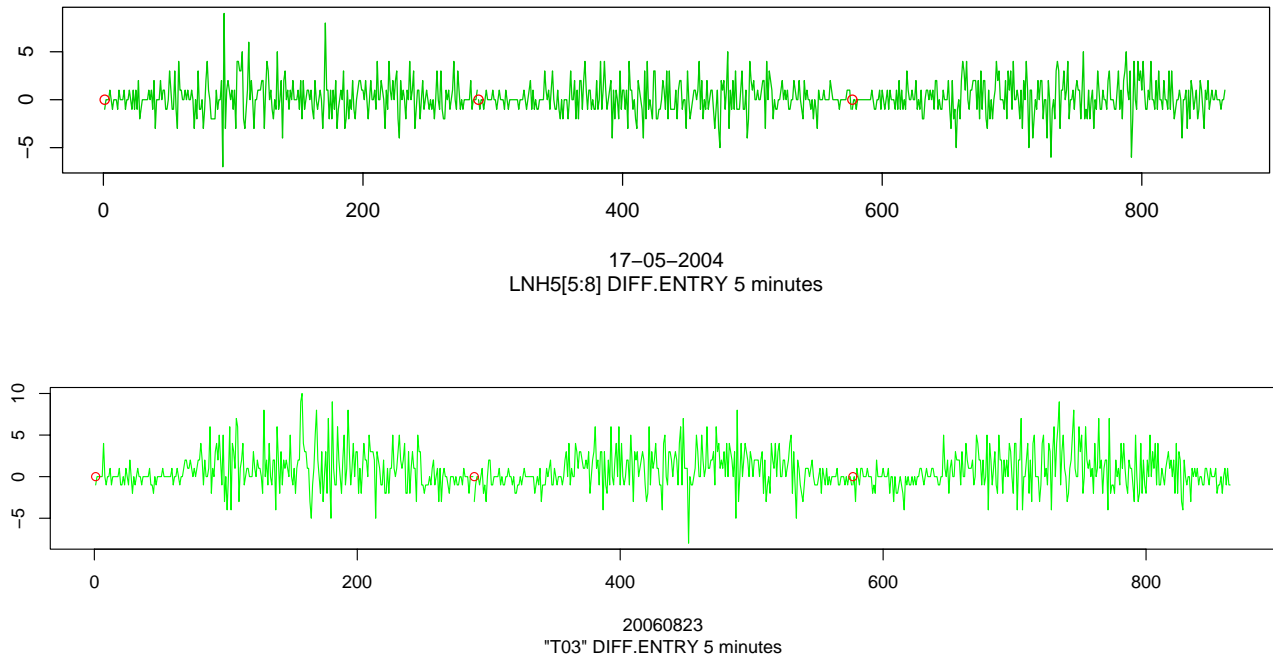


Figure 3. 3 successive week-days of gaps between planned and realized traffic at a sector entry, 5 minutes time-scale. Top: Sector from European Airspace. Bottom: Sector from Japanese Airspace. Daily repeating patterns. Constant mean around 0, constant variance over time (except night hours).

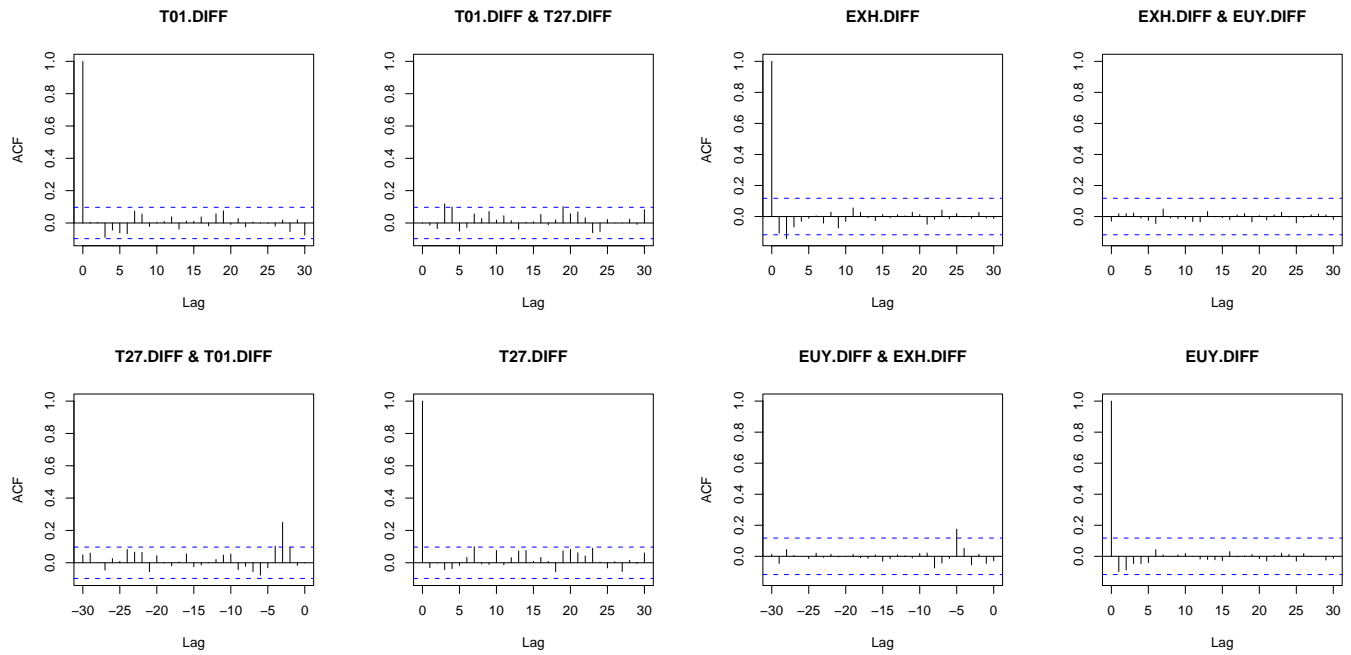


Figure 4. Cross-correlation matrices. Left: local neighbors. Right: non-local sectors

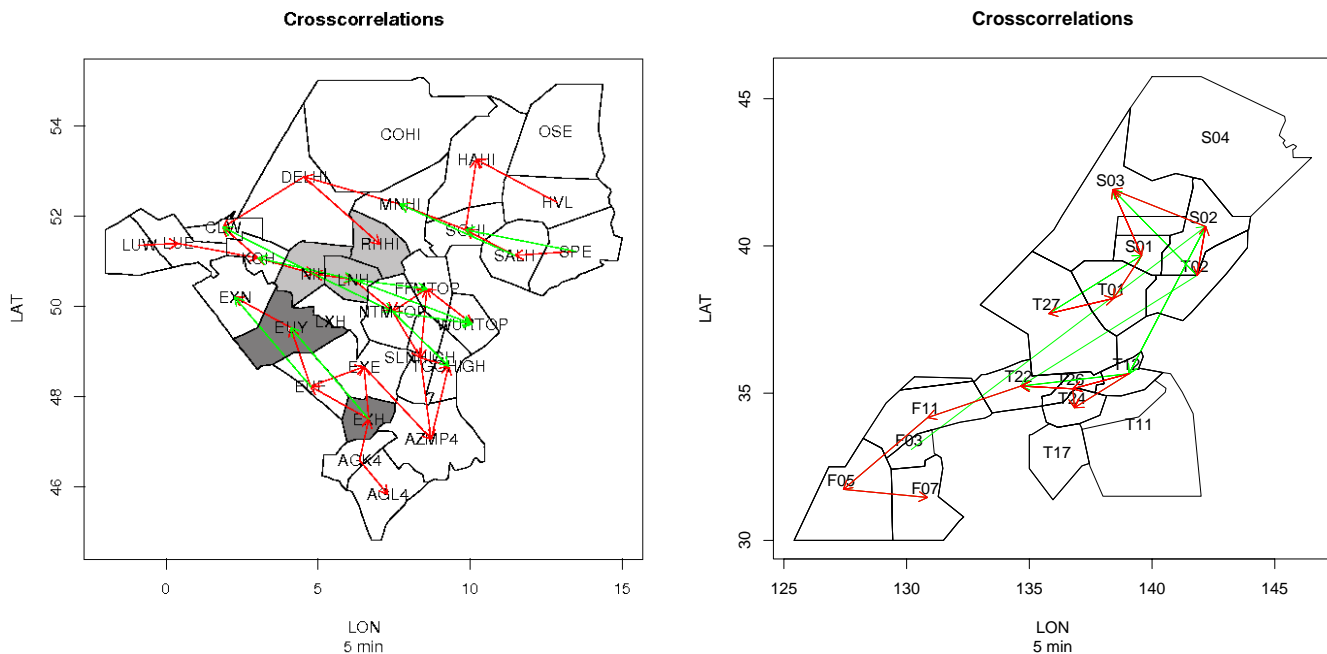


Figure 5. Visualization of the cross-correlation matrix. Left: Europe, Right: Japan.

higher average values in the Japanese Airspace have been analyzed further: there are generally many coefficients close to the critical value. This can be attributed to the higher sample variability as compared to the European data, because of the smaller sample size and because of the quality of the Japanese Airspace data [Gwi08].

The weak autocorrelation of the component processes and the sparse number of peaks in the crosscorrelation matrices suggest that the correlation structure in the system (i) does not contain spurious correlations because the component processes do not imply a severe variance inflation and (ii) has an intuitive explanation: all coefficients lie in the range of expectation since the traversal time for one sector lies between 6 and 10 minutes.

e) *Visualization of correlation matrix:* The correlation matrix functions for all 31 European and all 21 Japanese sectors were estimated up to lag $k = 30$, corresponding to 2.5 hours.

Figure 5 visualizes the results. An arrow between two sectors (i, j) represents a significant correlation at least one lag k . Positive and negative lags have opposite arrows. Local correlations are drawn in red. They reproduce almost the route network. For example, in the central flow (Frankfurt-London), they are bi-directional, whereas in the flow from Zurich to London, they are mono-directional. Non-local correlations are plotted in green. They reproduce only routes with high traffic densities. No correlations between two sectors that are not connected by a route are found.

IV. CONCLUSION AND FUTURE WORK

We analyzed how gaps between planned and realized traffic propagate through the European and the Japanese airspace. For this we did a correlation analysis for the most congested part of the systems. Because of the motion of aircraft, gaps propagate to neighboring sectors, expecting positive correlation coefficients. The question in the analysis was whether there are unexpected coefficients. Such coefficients would be caused by traffic controllers or flow managers who compensate for high gaps by re-routings or speed adjustments. Such strategies would often lead to negative correlation coefficients. We first analyzed the risk of obtaining misleading coefficients in a large correlation matrix. Then, we analyzed data from the European and Japanese airspace.

Our main results were:

- European and Japanese Airspace show similar patterns.
- significant cross-correlations appear on two levels: (i) locally, that is between a sector S and a direct neighbor and (ii) on high density routes, that is between two sectors S_1, S_2 that are connected through a flight route with high traffic densities.
- all correlations are positive.
- their lags correspond to the average traversal times.

No unexpected correlations have been found, and none of the correlations appears to be induced by the autocorrelation structure of a component process.

On the other hand one can argue that systematic re-routings would cause only weak correlations. Also, correlation assumes that the only source of covariation lies in the two variables under study. Indeed, the average strength of correlation was

0.2 in our data sets. This means that the non-existence for such strategies cannot be concluded; it can only be confirmed that such strategies currently show very weak effects in the counts of aircraft entering flight sectors. Such information is useful for demand prediction based on traffic densities: network-effects from far-lying sectors appear to have negligible effect. In order to get a deeper understanding of how controllers treat high workloads, a more specific model should be built. As a next step, inspiration for the construction of semi-empirical models (of conflict probabilities) can be found in the work of [Jar03]. This work is a step toward the identification of the mechanisms that lead to congestion in air traffic. Based on this, flow planning can be improved by taking into account the traffic predictions.

V. ACKNOWLEDGEMENTS

The first author likes to thank to Eurocontrol Experimental Center for a PhD grant under which the European data has been analyzed. He also wishes to thank to M. Pierre Collet for several discussions about correlation.

VI. VALIDATION

Figure 6 shows 4 scatter plots of variables in the system. The two upper ones are from the Japanese- the two lower ones from European Airspace. In each panel the bold line is the sample mean. It is reasonably linear. No other functional form of dependency is visible, neither. The first and third have significant coefficients of linear correlation. The second and fourth ones have not. Thus, linear correlation as a measure for dependence seems justified, even if the dependency between the variables is visibly weak.

REFERENCES

[BLHM05] M.O. Ball, D. Lovell, R. Hoffman, and A. Mukherjee. Response mechanisms for dynamic air traffic flow management. In *Proceedings of the 6th Europe-USA ATM Seminar. Baltimore, US.*, 2005.

[Dig90] P.J. Diggle. *Time Series: A Biostatistical Introduction*. Oxford University Press, 1990.

[Efr04] B. Efron. Large-scale simultaneous hypothesis testing: the choice of a null hypothesis. *Journal of the American Statistical Association*, 99/465:96–104, 2004.

[ETST01] B. Efron, R. Tibshirani, J. Storey, and V. Tusher. Empirical Bayes analysis of a microarray experiment. *Journal of the American Statistical Association*, 96:1151–1160, 2001.

[EUR06] EUROCONTROL. *Performance Review Report 2005*. EUROCONTROL, Brussels, Belgium, 2006.

[Gwi08] C. Gwiggner. Accuracy of regulated flight plan data. Technical report, ENRI, 2008.

[Jar03] M. R. Jardin. Toward real-time en route air traffic control optimization. *PhD Thesis. Stanford University*, 2003.

[Ken89] M.G. Kendall. *Time Series*. Oxford University Press, 3rd edition, 1989.

[KP03] Y. Kamarianakis and P. Prastacos. Forecasting traffic flow conditions in an urban network: Comparison of multivariate and univariate approaches. *Journal of the Transportation Research Board*, 1857:74–84, 2003.

[KSO83] M.G. Kendall, A. Stuart, and K. Ord. *The Advanced Theory of Statistics. Volume 3: Design and Analysis, and Time Series. 4th edition*. Charles Griffin & Company Ltd., London, 1983.

[LW04] O. Ledoit and M. Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88-2:365–411, 2004.

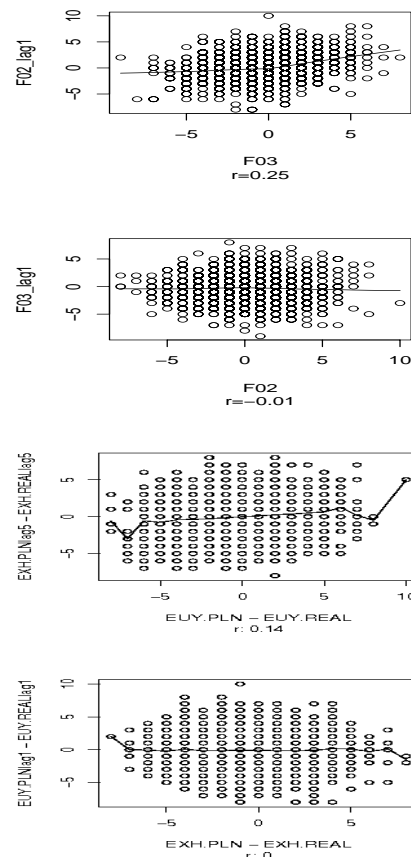


Figure 6. Scatterplots of gaps in one sector against time-lagged gaps in other sectors. First and third panel: significant coefficients (Japan/Europe), second and fourth: non-significant coefficients.

Smoothed traffic complexity metrics for airspace configuration schedules

David Gianazza, DSN, Toulouse, France
Email : {lastname}@recherche.enac.fr

Abstract—This paper is a continuation of previous research on optimal airspace configuration. It is expected to improve the predictability and the flexibility of the airspace management process by computing realistic predictions of the sectors opening schedules in *En-route* ATC centers. In previous papers, we selected relevant complexity metrics to predict the controllers workload, using neural networks trained on recorded airspace configurations. We also introduced new algorithms to build optimally balanced airspace configurations, exploring all possible combinations of elementary sectors.

As a result of this previous work, we were able to compute realistic schedules on a whole day of traffic, using complexity metrics that were computed from recorded radar tracks. The raw metrics, however, showed high variations in time which caused a "configuration switching" phenomenon. Although the number of control sectors in the computed schedule stayed globally close to the recorded number of sectors, the airspace was reconfigured much more often than in reality. The present paper shows how the input metrics can be smoothed in order to avoid this problem, and what may be the subsequent problems caused by the smoothing strategy.

INTRODUCTION

Over the years, and in a context of increasing air traffic demand, there has been a growing need to increase the capacity of the Air Traffic Management system. Improving the *predictability* of the system's response to the traffic demand is also a crucial issue, as it would allow a better use of the existing resources and an earlier anticipation of future congestions.

The work presented in this paper is the continuation of previous research on airspace configuration schedules ([1], [2], [3]) and air traffic complexity metrics ([4], [5]) previously led at the Global Optimization Laboratory (CENA/ENAC) and now continued within the Planification, Optimization, and Modeling team of DSN/DTI-R&D. The initial aim of this research is to compute realistic sectors opening schedules for *en-route* air traffic control centres, given an input traffic demand on a chosen day.

The current FMP/CFMU working method to build airspace configuration schedules relies on pre-defined sectorization scenarios, where the incoming traffic flows¹ are matched against the sector capacities² to detect

¹The metric used is the "incoming flow", also called "flight counts" or "traffic-volume" in some Eurocontrol documentations ([6]) or "traffic load" in the CFMU handbook. For a sector, it is the number of flights that will enter the sector within the next 60 minutes (or any other chosen period of time).

²The sector capacity is defined as a threshold value on the number of flights that may enter the sector in a chosen period of time.

potential overloads. Although it may prove effective in practice as it relies on the FMP/CFMU operators experience, this method is not grounded on a solid assessment of the actual controllers workload. Consequently, implementing any strategy to optimize the airspace schedule on this basis may lead to unexpected results (see [1]). Another drawback of the current method is that only a small subset of all possible airspace configurations is used.

In [3], new algorithms were proposed, using more relevant complexity metrics to assess the controllers workload, and exploring all possible combinations of elementary sectors to build optimal airspace configurations. As a result of this previous work, we were able to compute realistic airspace configuration schedules on a whole day of traffic, using raw complexity metrics computed from recorded radar tracks. The raw metrics, however, showed high variations in time which caused a "configuration switching" phenomenon. Although the number of control sectors in the computed schedule stayed globally close to the recorded number of sectors, the airspace was reconfigured much more often than in reality.

The present paper shows how the input metrics can be smoothed in order to avoid this problem. The next section first provides a short overview of the current research on airspace configuration and air traffic complexity. Section II describes the algorithms used to predict the sector status and to build airspace configurations, mainly focusing on the few improvements that were made since [3] was published. The experimental procedure applied to select the best smoothing parameters is described in section III. Results are provided in sections IV and V. Section VI concludes this paper.

I. OVERVIEW

A. Airspace configuration

Current research on airspace configuration is manifold and may deal with strategic airspace partitioning (see [7] and included references, [8]), pre-tactical sectors opening schedules ([9]), or tactical airspace management ([13]). In this paper, we are mainly concerned with pre-tactical airspace configuration schedules, although some of the proposed algorithms may also be used in tactical applications, provided the complexity metrics being used are relevant in that context.

The FMP/CFMU working method to build sectors opening schedules was shortly described in the introduction. Current research led by Eurocontrol proposes

short-term improvements of the Flow Management process, mainly by avoiding unnecessary regulations when building sectors opening schemes ([9],[6], [11]). One of the main concerns is the *network effect* observed in ATM regulations ([12]). These studies still use incoming flows and sector capacities, and a small number of pre-defined configurations.

In the United States, the main concern seems to be the dynamic adjustment of the airspace structure to the traffic flows reroutings caused by severe weather conditions. It is expected that more flexible boundaries would allow a more efficient use of airspace and increase the overall capacity. In [13], pre-defined scenarios of airspace sectorizations associated to traffic rerouting scenarios are proposed as a short-term improvement to the current practice.

A more dynamic resectorization with flexible boundaries is envisioned in future operational concepts ([14], [15], and some SESAR *Operational Improvement* steps). It is expected that moving the sector's boundaries in real-time to adapt to the traffic demand would increase the capacity and the efficiency of the ATM system. The actual capabilities and potential benefits of this new operational paradigm are still largely unknown at this early stage, however. There is also some concern that unlimited flexibility in the sectors boundaries would lead to a loss of situational awareness by the air traffic controllers (see discussion and literature review in [16]).

The work presented in this paper is more medium-term research, trying to improve the predictability and the flexibility of today's airspace management in Europe. The idea is to find the optimal combination of elementary (or modular) sectors that will provide the maximum capacity to a given input traffic, and balance the controllers workload as best as possible among the control sectors.

This airspace partitioning problem would be difficult to solve without choosing a heuristic if every combination of sectors was possible. The partitioning of the whole ATCC's airspace into control sectors is highly combinatorial ([2]), even with relatively few elementary sectors. Hopefully, the list of possible *control sectors* (either elementary or collapsed sectors) that can be operated in an air traffic control center is relatively small³, as not all combinations of elementary sectors are operationally valid⁴. So we may explore all valid airspace configurations, which may be built with operationally valid control sectors only, using classical tree search methods ([1], [3]).

These algorithms are applied to the prediction of airspace configuration schedules, optimally balancing the workload among the control sectors. Consequently, we need a way to assess the controller's workload, and it was proposed to use relevant air traffic complexity metrics to that purpose ([4], [5]).

³The list of control sectors is available from the ATCC's database

⁴One usually does not merge sectors which are not geographically connex, for example.

B. Air traffic complexity

A multitude of air traffic complexity metrics have been proposed in the literature (see [17] and [18] for a review), and many studies tried to correlate some of these metrics to the controllers workload, using various methods: linear ([19]) or logistic ([20]) regression, cross-sectional time series analysis ([21]), neural networks ([22]),... Many ways to quantify the controller's workload have also been tried: physical activity ([23], [21]), physiological indicators ([24], [25]), simulation models of the controller's tasks ([26], [27]), subjective ratings ([19], [22], [20]). The reader may refer to [4] for a discussion on these variables. Let us just say that, in addition to being subject to noise and biases⁵, most of the above dependent variables require relatively heavy experimental setups to collect the data, usually with the active participation of controllers. Databases are often small and might exhibit low variability, which may in turn harm the statistical relevance of the results.

In order to avoid some of these drawbacks, we proposed a new dependent variable for which a large amount of data is available from the ATCC databases, and which does reflect an operational reality. The basic idea, introduced in [29], is that the decisions to split (resp. merge) a sector are mostly taken when the controller is close to overload (resp. under-load). So the sector status (merged, operated, or split) is directly related to the controller's workload and may therefore provide an acceptable dependent variable. In [4] and [5], neural networks were trained on recorded patterns of metrics and sector statuses⁶ to select the most relevant metrics for our airspace configuration problem.

The proposed method allowed to select a subset of only 6 relevant indicators among the initial 28 chosen from [19], [22], [30], [31] and other sources. The airspace configuration schedules obtained with these metrics as input were quite realistic ([3]) when computed from recorded radar tracks. In this previous work, however, the input metrics were not smoothed, and a "configuration switching" phenomenon was observed. Let us now see, after a short description of the algorithms, if smoothed metrics provide better results.

II. ALGORITHMS

Our aim is to build a realistic schedule of the airspace configuration throughout the day. To that purpose, one needs first a correct assessment of the workload generated by the traffic throughput in a control sector, and second an algorithm exploring all possible airspace configurations to find out the optimal one, with respect to the workload balance over control sectors.

A. A neural network to predict the sector status

Neural networks are used to issue sector status probabilities for each control sector of a candidate configura-

⁵such as the subjective ratings recency effect denounced in [25], or raters errors in the case of "over-the shoulder workload ratings" [28]

⁶In our case, a pattern is a vector of complexity metrics measured at a time t in a given sector, together with the sector status (*merged*, *normal*, or *split*) that was recorded at this time.

tion. Beyond the similarities with the biological model, an artificial neural network may be viewed as a statistical processor, making probabilistic assumptions about data ([32]). A *training* set of *patterns* is used to determine a statistical model of the process which produced this data. Once correctly trained, the neural network uses this model to make predictions on new data. The reader may refer to [33] and [34] for an extensive presentation of neural networks for pattern recognition.

In our case, the neural network is trained on recorded airspace configurations, considering the actual status of each control sector : *merged* when the sector is collapsed with other sectors to form a larger sector (low workload), *normal* when the control sector is opened (normal workload), or *split* into smaller sectors operated separately (high workload)⁷. The input variables are the relevant complexity metrics, or any candidate subset of metrics, normalized by subtracting the mean value and by dividing by the standard deviation. The output of the neural network is a triple of sector status probabilities ($p_{merge}, p_{normal}, p_{split}$).

The network is unable to make complex recommendations such as to split the sector's volume in several parts and then to merge each of these parts with other sectors. It only recommends to merge the sector when the workload is low, or split it when the workload is high, or operate it normally when the workload is acceptable. As we are necessarily in one of the above three cases, the sum of the three probabilities $p_{merge}, p_{normal},$ and p_{split} is always 1.

More details on neural networks applied to sector status prediction, in the context of airspace configuration, can be found in previous works ([3]). How these networks were used to select the most relevant metrics is described in [4] and [5]. The same network's topology and training algorithms are used in the work presented here to select the most relevant smoothing strategy for the input metrics.

The software implementation is different, though. In previous works, the `nnet` R package developed by Pr. Ripley was used. As it is envisioned in a near future to try other types of neural networks, more suited to time series, some new software⁸ was developed. A backpropagation method⁹ and a BFGS¹⁰ quasi-Newton optimization method were implemented in Ocaml language. The same stopping parameters as in previous works with `nnet` were used.

⁷Irrelevant statuses, such as when a part of the initial sector is merged with one control sector, and the other part with another control sector, were discarded in the neural network's training.

⁸ANNiML (Artificial Neural Networks in ML) is written in Ocaml and should be made available soon, probably under GNU Lesser General Public License.

⁹Backpropagation of the output error through the network's layers allows to approximate the partial derivatives of the error function with respect to the weights

¹⁰BFGS (Broyden-Fletcher-Goldfarb-Shanno) is an iterative local optimization method, starting from an initial point (weights values in our case) and using an approximate hessian and the gradient of the objective function to find a local optimum. Note that different initial points may lead to different local optima.

B. Tree search algorithms for well-balanced sector configurations

As previously told, the neural network cannot issue complex recommendations on how to reconfigure several control sectors. A tree search algorithm was used to that purpose, exploring all possible combinations of elementary sectors, to find out the optimal one.

An optimal configuration is one for which the workload among the control sectors is balanced as best as possible, while using the less possible resources, and satisfying operational constraints such as a maximum number of available working positions for example.

Once again, we used the same algorithm as in [3] to compute optimal airspace configurations, with a few improvements that shall be detailed later in this section, and with the aim to study the influence of the smoothing strategy on the computed opening schedule.

Let us just describe the main features of this algorithm. Starting at time $t=0$ with a configuration where all elementary sectors are assigned to a single controller's working position, the situation is reconsidered every minute of the day, using the status probabilities ($p_{merge}, p_{normal}, p_{split}$) of each control sector in the current configuration to decide if the airspace should be reconfigured or not.

The decision criterion may be straightforward (taking the action corresponding to the highest probability), or it may propose to take an action only when the corresponding probability is close enough to 1, and when the difference between the two highest probabilities is sufficient. The first, straightforward, decision criterion was called D1 in [3], and the second was name D2, with decision parameters η (threshold on the difference between the two highest probabilities, for merging decisions), α (proximity of p_{merge} to 1) and β (proximity of p_{split} to 1). Figure 1 illustrates criterion D2, showing the evolution of the sector status probabilities just before a "split" decision, when p_{split} reaches $1 - \beta$.

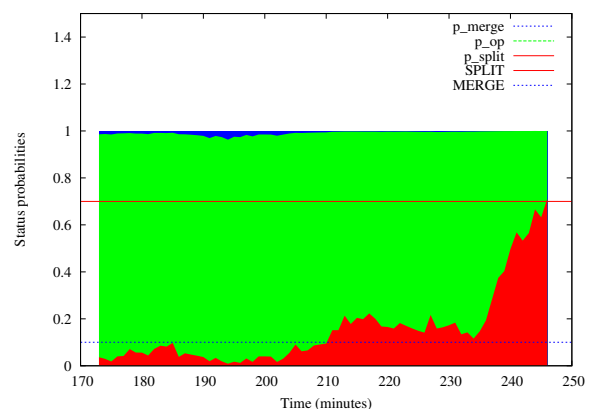


Fig. 1. Example of sector status probabilities in NGA sector (Brest ATCC) just before the algorithm decides to split the sector.

Once a decision to reconfigure some control sectors is taken, the corresponding elementary sectors are recombined, exploring all possible partitions of this set.

Some drawbacks of this local recombination method were highlighted in [3], for example in the case where the decision criterion triggers a "merge" action for two control sectors which are not neighbours. This is typically a case where the local recombination leads to no change, because the airspace should be reconfigured on a larger scale. A solution to this problem is to reconfigure the whole airspace in such cases. However, exploring exhaustively the whole tree of possible configurations by computing all of them becomes very rapidly computationally intensive even with a relatively small number of elementary sectors.

So the previous algorithm was improved as follows. Local recombinations are made as before when the control sectors that need to be reconfigured are geographically connected. If this is not the case, a full airspace reconfiguration is triggered, using a *Branch & bound* algorithm to explore all possible combinations. The detailed description of this algorithm will be the subject of a next publication, but the reader may refer to [1], [2], and [29] where a very similar *Branch & bound* algorithm is detailed.

A second improvement introduced in this paper is about the cost function allowing to compare the candidate airspace configurations. A more simple and more understandable cost function was designed, where the cost depends on the number of control sectors and the maximum probability in each category (*merge*, *normal*, *split*).

An "ideal" configuration should have $(p_{merge}, p_{man}, p_{split}) = (0, 1, 0)$ for all its control sectors. This is not always possible, so we need to take account of overloaded or underloaded sectors, and ill-balanced configurations. The cost of a configuration c , with a vector x of complexity metrics measured at time t is expressed as follows:

$$cost(c, x, t) = \underbrace{xx}_{k_1} \underbrace{xxx}_{k_2} \underbrace{xx}_{k_3} \underbrace{xxx}_{k_4} \underbrace{xx}_{k_5} \underbrace{xxx}_{k_6}$$

where we have assigned:

- k_1 digits to the number of overloaded sectors,
- k_2 digits to the maximum value of p_{split} among the overloaded sectors, where the probability is suitably scaled to the allowed number of digits,
- k_3 digits to the number of under-loaded sectors,
- k_4 digits to the maximum value of p_{merge} among the under-loaded sectors,
- k_5 digits to the number of normally loaded sectors,
- k_6 digits to the maximum value of $1 - p_{normal}$ among the normally loaded sectors,

With this cost, the first priority is to have the less possible overloaded sectors, and if there still remains some then the maximum probability p_{split} among these sectors should be as small as possible. The same explanation stands for underloaded sectors. For normally loaded sectors, we still want to use the less possible resources, but workload should be balanced as well as possible among the sectors. So the minimum value of p_{normal} among the normally loaded sectors should be

as high as possible. This is why we use the *maximum* of $1 - p_{normal}$ in the cost, so that minimizing this cost will lead to more desirable configurations.

III. EXPERIMENTAL PROCEDURE

Each complexity metric x_i may be smoothed by taking its average value over a period of time $[t - \delta_1, t + \delta_2]$, where we may try different values for δ_1 and δ_2 for each metric.

For now, the metrics are computed on past data (recorded radar tracks). In future applications, they may be computed either from simulated trajectories following flight plans, in the context of airspace configuration schedules, or from real-time radar tracks and trajectory predictions for tactical airspace management purposes. For real-time applications, one may prefer to smooth the metrics on a time window $[t - \delta, t]$, considering only the past positions of the aircraft. We decided to try this strategy first, which may be applied also to simulated trajectories for airspace schedules.

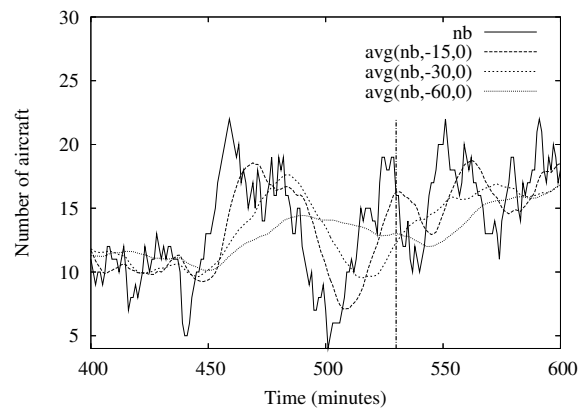


Fig. 2. Raw and smoothed number of aircraft in N sector (Brest ATCC). The splitting decision was taken at $t=530$ (vertical line).

As an illustration of the effect of smoothing on the metrics, figure 2 shows the number of aircraft within sector N (Brest ATCC), with different values of δ . We may notice the high variations in the raw aircraft count. The vertical line shows when the decision to split the sector into two smaller sectors was taken.

A. Testing different smoothing strategies

We would like to find out which combination of metrics and smoothing parameters is the best. This is a model selection problem. The main difficulties in model selection are the choice of a search strategy (how to explore the possible subsets of explanatory variables, knowing that the number of combinations is usually too large for an exhaustive search), and also the assessment of each model's performance (quality criterion, ability to generalize to fresh data).

In this paper, we will consider different values for the size of the smoothing window: 3, 5, 10, 15, 30, or 60 minutes. Ideally, we should make the same study as in [4], [5] but applied to the 27 complexity metrics with

all smoothing possibilities, which means 190 variables¹¹. The forward strategy that was used in previous works to explore different combinations of variables would take too much computation time, so it was decided to focus on the 6 most relevant variables found in our previous studies. These were the sector volume V , the number of aircraft within the sector Nb , the average vertical speed avg_vs , the incoming flows with time horizons of 15 minutes and 60 minutes (F_{15} , F_{60}), and the number of potential crossings with an angle greater than 20 degrees ($inter_hori$).

As a first approach, and keeping in mind that it is a fairly restrictive search strategy, it was decided to try different smoothing values, *applying the same smoothing window to all variables in the set of relevant raw metrics*. The reference set of variables is $REF = \{V, Nb, avg_vs, F_{60}, F_{15}, inter_hori\}$. The other combinations that were tested are **SM3**, **SM5**, **SM10**, **SM15**, **SM30**, **SM60**, which contain the same complexity metrics, smoothed respectively using time windows of 3, 5, 10, 15, 30, or 60 minutes.

B. Model selection and performance assesment

In our previous works ([4], [5]), the mean AIC¹² (averaged over the sample's size) was used to compare the performance of a given neural network on data samples of different sizes (a training set and a test set), and the mean BIC¹³ was used to compare different neural networks, trained on candidate subsets of complexity metrics (models of different sizes). In this paper, we will also use the mean BIC to compare the candidate models and assess the improvements provided by smoothing the metrics values, and the AIC to assess the generalization performance.

Once trained on past data, it important to check if the neural network also provides good predictions of the sector status when feeded with new inputs. This is generally done by splitting the initial data set in two samples: a train set and a test set. This *split-sample* (or hold-out) procedure is generally satisfying on large data samples, but may be prone to overfitting¹⁴ problems

¹¹27 metrics multiplied by 7 smoothing values (counting a zero value for the raw metrics), plus the sector volume.

¹²Akaike's "An Information Criterion" $AIC = 2\lambda - 2\ln(L)$, where λ is the number of unadjusted parameters of the model (i.e. the number of weights and biases of the network), and $\ln(L)$ is the log-likelihood error. When used for model selection with neural networks, AIC tends to overfit (see discussion in [34], p. 61), leading to select bigger models. The Schwartz's Bayesian Information Criterion is usually preferred.

¹³Schwartz's Bayesian Information Criterion $BIC = 2\lambda \ln(N) - 2\ln(L)$, where N is the size of the data sample. The BIC criterion gives a higher penalization than AIC to big models, but varies with the size of the data sample, so it may not be used to compare the performances of a neural network on samples of different sizes. Note that AIC and BIC are not absolute criteria: their evaluation is specific to the underlying "true" model, and only the relative differences in the criterion's value is useful.

¹⁴*Overfitting* occurs when the statistical model fits very well the data from which it was derived, but cannot generalize well on fresh data. The number of parameters in the model (network's weights for example) and few data samples may both cause overfitting problems. A neural network with too few weights may not be able to capture all the variations of the response to the input x , whereas a network with too many weights will more likely be subject to overfitting (see [32]).

on small samples. It was used in [4] and [5] with good results, but one may argue that the selected models may only be fit to the chosen train and test sets, although some tests on a second test set (another day of traffic) proved also satisfying.

So it was decided to apply a more sophisticated procedure, using first a *k-fold cross-validation* method for the model selection, and second a *split-sample* method (or hold-out validation) to assess the generalization performance of the best model. The initial data set is randomly split in two samples. The first one (*training set*) is again divided in k sub-samples and used for an iterative k -fold cross-validation allowing to select the best smoothing parameter. Then, the neural network is trained on the whole training set, and the generalization performance is checked on the *test set*.

In our case, we applied a 10-fold cross-validation, iteratively holding out one of the 10 sub-samples of the training set to assess the candidate model, and training the neural network on the 9 remaining sub-samples. The Schwartz's Bayesian Information Criterion (BIC) is computed on the sample that was not used to train the network. The BIC is averaged on the 10 runs for each model. The best model is found by comparing the average BIC.

Once we have found the best model, the neural network is trained on the whole training set (the 10 samples). The generalization performance of the trained network is assessed by comparing the AIC value found for the training set to the AIC of the test set. As the training method is an iterative local optimization (BFGS) which may fall into local optima depending on the chosen initial weights vector, ten training runs are made with different random values of the initial weights¹⁵.

C. Comparison of airspace configurations schedules

So far we have only detailed how to compare different statistical models allowing to predict the sector status from smoothed complexity metrics. Our final goal, however, is to build realistic airspace configuration schedules. So we also need to consider the influence of the smoothing strategy on the overall airspace configuration.

Ideally, the computed schedule should reproduce the actual configurations recorded that day. However, there is a high variability in the decisions made by control room managers on *how* to reconfigure the airspace, which comes in addition to the variability of decisions on *when* to reconfigure. We may hope that our sector status prediction could give an indication on when to trigger a reconfiguration and allow to build realistic configurations, but our algorithms may not compute exactly the same configuration as in reality.

We will assess the realism of the computed schedule by comparing the number of control sectors to the actual number of sectors that were opened that day. The Pearson's correlation coefficient may give an indication of the linear correlation between the computed and the

¹⁵Note that the ten runs of the cross-validation were also made with different random initial weights

real number of control sectors. However it may not be always reliable¹⁶ so we will also compute an *ad-hoc* "dissimilarity measure" which is the surface delimited by the two curves, divided by the surface of the real schedule. With this measure, two identical curves shall have a dissimilarity 0 if they are exactly superposed. In addition, we will also consider the number of recon-figurations throughout the day, which should be close enough to the real one.

So we don't have a unique quantified measure of similarity between airspace configuration schedules for now: the influence of the smoothing parameter on the opening schedule is assessed by considering both the number of control sectors and the number of configuration changes.

But before looking how smoothing the complexity metrics may change the overall airspace configuration schedule, let us show some results on the influence of the smoothing parameter on the prediction of the sector status.

IV. INFLUENCE OF SMOOTHED METRICS ON SECTOR STATUS PREDICTION

The results of the 10-fold cross-validation with different values of the smoothing window are presented in tables I, II, and III.

Set	mean BIC	BIC std dev
REF	1.163	2.7E - 2
SM3	1.156	3.0E - 02
SM5	1.141	2.9E - 02
SM10	1.117	2.4E - 02
SM15	1.114	2.4E - 02
SM30	1.059	2.6E - 02
SM60	1.046	3.5E - 02

TABLE I
MEAN BIC VALUES AND STANDARD DEVIATIONS FOR THE CROSS-VALIDATION

Table I shows the mean value and the standard deviation of the BIC criterion over the 10 runs of the cross-validation, for each candidate model. A somewhat surprising result is that SM60 (smoothing over 60 minutes) seems to provide the best results if we look only at the mean BIC. However, considering the standard deviation, it is not obvious that there is a true statistical difference between SM30 and SM60. Note also that the model with the lowest mean BIC is the one with the highest standard deviation.

Tables II and III show the mean correct classification rates and their standard deviations, over the ten runs, for all classes (Global) and also for each sector status class. Let us notice that the main improvement, when smoothing the input metrics, is made for the class corresponding to the normal domain of operation.

Let us now assess the generalization performance of the models. As explained in the previous section, the

¹⁶The correlation coefficient between two equal variables x and $y = x$ will be 1. Let us note however that this coefficient is not sufficient to actually measure how close we are to equality: the correlation coefficient between a variable x and another variable $y = x + d$, where d is a constant offset, will also be 1.

Set	Global	Merged	Normal	Split
REF	82.074%	88.353%	62.799%	90.322%
SM3	82.428%	88.451%	63.691%	90.587%
SM5	82.673%	88.545%	64.279%	90.75%
SM10	83.454%	89.401%	64.918%	91.548%
SM15	83.784%	89.365%	66.066%	91.699%
SM30	84.947%	89.910%	68.036%	93.147%
SM60	85.811%	90.698%	69.798%	93.242%

TABLE II
CORRECT CLASSIFICATION RATES

Set	Global	Merged	Normal	Split
REF	0.446%	0.616%	1.543%	0.572%
SM3	0.662%	0.586%	1.853%	0.641%
SM5	0.662%	0.852%	1.275%	0.556%
SM10	0.574%	0.714%	1.419%	0.384%
SM15	0.696%	0.771%	1.822%	0.636%
SM30	0.648%	0.568%	1.667%	0.794%
SM60	0.601%	1.091%	1.488%	0.794%

TABLE III
STANDARD DEVIATIONS OF THE CORRECT CLASSIFICATION RATES

neural network is trained again on the whole training set of patterns (instead of 9 sub-samples in the cross-validation). Ten runs were made with different random initial weights. The difference with cross-validation is that there are all made on the same training set.

Set	Training set		Test set	
	mean AIC	AIC std dev	mean AIC	AIC std dev
REF	0.765	1.2E - 02	0.781	1.5E - 02
SM3	0.750	1.3E - 02	0.757	1.3E - 02
SM5	0.743	1.6E - 02	0.751	1.5E - 02
SM10	0.733	2.1E - 02	0.744	2.3E - 02
SM15	0.710	2.2E - 02	0.727	2.2E - 02
SM30	0.681	1.5E - 02	0.697	1.3E - 02
SM60	0.644	1.6E - 02	0.662	1.8E - 02

TABLE IV
GENERALIZATION PERFORMANCE: MEAN AIC VALUES AND STANDARD DEVIATIONS FOR THE TRAINING SET AND THE TEST SET

Table IV shows the mean AIC values and the standard deviation, over the 9 best results out of the 10 runs¹⁷, for each smoothing parameter. The neural networks performances on the training and test sets are quite close, with any smoothing parameter. All models seem to generalize well, and show little differences in that respect (see figure 3).

V. INFLUENCE OF SMOOTHED METRICS ON OPENING SCHEDULES

The last section was dedicated to the influence of the smoothing strategy on the performance of the sector status prediction. Now, let us see how it modifies the resulting opening schedules, comparing the different models on a same day and for a chosen air traffic centre (Brest ATCC, 2003, June 1st). The same algorithms

¹⁷In two cases, it happened that the choice of the random initial weights and the training process relying on a local optimization led to significantly less performing networks. So it was decided to remove the ten percent less performing networks from the results.

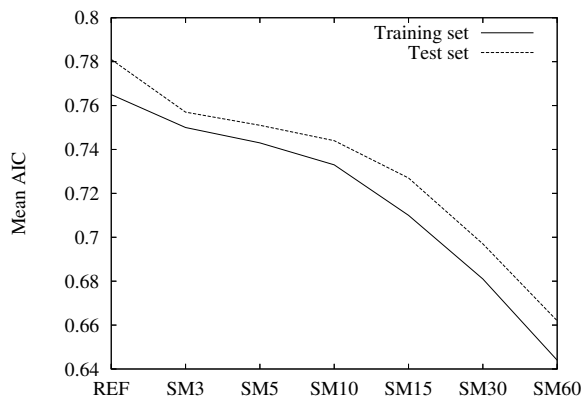


Fig. 3. Mean AIC for the training set and the test set.

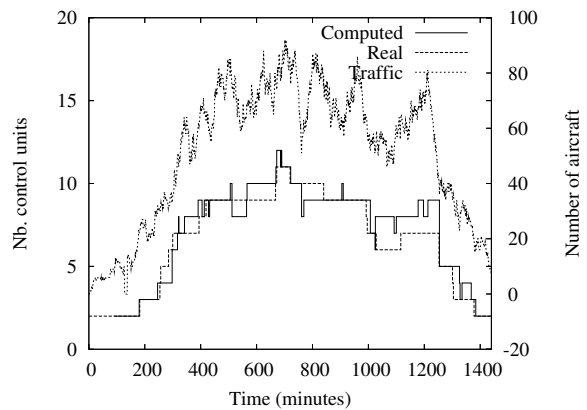


Fig. 4. Number of control sectors (computed schedule, real configurations), and traffic for Brest ATCC (2003, June 1st), with REF setup

were used for all models, with the same values for the split/merge decision parameters ($\eta = 0.2, \alpha = 0.1$, and $\beta = 0.3$).

Set	Correlation coeff.	Dissimilarity	Nb. config.
REF	0.9443	0.1169	101
SM3	0.9196	0.1489	202
SM5	0.9142	0.1510	179
SM10	0.9038	0.1783	125
SM15	0.9065	0.1681	140
SM30	0.9471	0.1094	34
SM60	0.9101	0.1426	23

TABLE V
CORRELATION COEFFICIENT AND NUMBER OF AIRSPACE CONFIGURATIONS FOR EACH MODEL

Table V shows the correlation coefficient, the dissimilarity measure, and the number of configurations for each model. All models show a good correlation, above 0.9 to the recorded number of control sectors. The number of reconfigurations is fairly high when smoothing on less than 15 minutes, showing a lot of "configuration switching", whereas SM30 and SM60 are much closer to the 28 airspace configurations that were actually used that day. Considering the dissimilarity measure and the number of configurations, SM30 seems to be the model that is most similar to reality. Let us now have a closer look at each computed schedule.

Figure 4 shows the reference situation, for Brest ATCC (2003, June 1st). The number of control sectors computed by our algorithm, using raw complexity metrics, can be compared to the actual number of control sectors that were opened, for each minute of this day. The evolution of the number of aircraft within the center is also displayed, above the two other curves. Let us remind that the number of aircraft is not sufficient to explain the number of control sectors, as other complexity metrics are also involved in the explanation of the sector status, and as the traffic load may not be equally dispatched among the sectors. It is still a good indication of the overall traffic load, however.

We may notice that, while the computed output stays globally close the recorded number of control sectors,

it also shows many variations around the actual curve, more or less following the traffic trends on that day. Notice the peak of traffic around 20:00 UTC (1200 minutes after 00:00), where the curve of the computed schedule apparently better follows the traffic trend than the actual configurations (we shall see later that this depends on the chosen smoothing parameter).

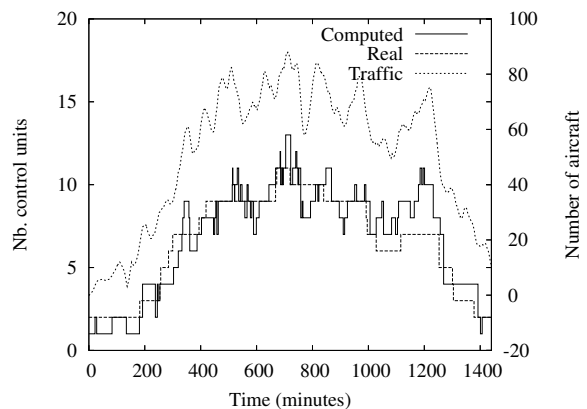


Fig. 5. Number of control sectors (computed schedule, real configurations), and traffic for Brest ATCC (2003, June 1st), with smoothed metrics (SM15)

Figures 5, 6 and 7 show the airspace schedule computed with smoothed metrics, using a smoothing window of 15, 30, or 60 minutes respectively. The traffic load's curve displayed on each figure shows the smoothed number of aircraft, using the smoothing window corresponding to each model.

At this point, when comparing figures 4, 5, 6, and 7, we may notice two phenomena which are not quantified by the measures of correlation and the number of reconfigurations. First of all, considering the peak of traffic around 20:00 UTC (1200 minutes after 00:00), we can see that the more you smooth the metrics, the less the computed number of control sectors reflects this peak of traffic. In fact, it becomes closer to the actual number of control sectors.

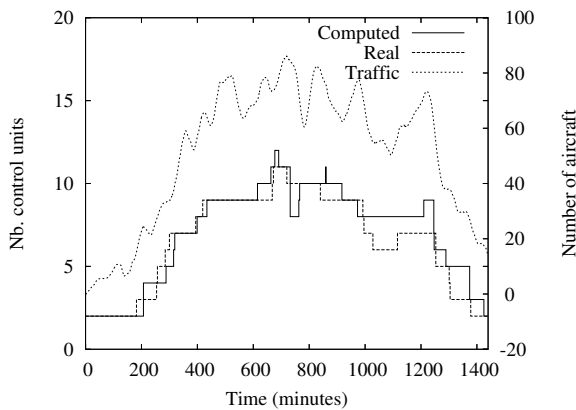


Fig. 6. Number of control sectors (computed schedule, real configurations), and traffic for Brest ATCC (2003, June 1st), with smoothed metrics (SM30)

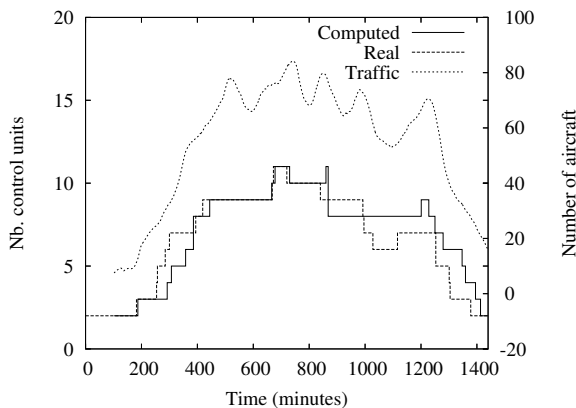


Fig. 7. Number of control sectors (computed schedule, real configurations), and traffic for Brest ATCC (2003, June 1st), with smoothed metrics (SM60)

The second conclusion that may be drawn from these figures is that smoothing the input metrics leads to delay the decisions to reconfigure the airspace. This is most visible on figure 7 (SM60) where the "climbing steps" corresponding to the split decisions in the morning and the "descending steps" of the merge decisions towards the end of the day are both on the right of the actual curve. In other words, the sector status prediction seems more performant *on average* when smoothing over 60 minutes¹⁸, but smoothing too much leads to take late split/merge decisions, thus delaying the moments at which the reconfigurations should be triggered.

All these experiments were made using the same decision parameters ($\eta = 0.2$, $\alpha = 0.1$, and $\beta = 0.3$) for all models. These parameters also have an influence on the moment at which reconfigurations are triggered. Some other parameter values were tried ($\eta = 0$, $\alpha = 0.5$, and $\beta = 0.5$), with the aim to improve the reactivity of the reconfiguration algorithm. For SM60, the reconfigurations were triggered slightly

¹⁸Although it was not such a clear-cut in deciding which of SM30 or SM60 was the best model, in table I

earlier but still the same phenomenon was observed, and the number of configurations increased to 38 configurations. Other trials were made, mixing metrics smoothed over 60 minutes and metrics smoothed over 10 minutes, with similar results.

So, smoothing the metrics over 15 minutes or less allows a higher reactivity to the traffic variations, but with much more reconfigurations than observed in real life. Among the models that were tested, SM30 (smoothing the input metrics over 30 minutes) seems the best compromise, considering the performance of the sector status prediction, but also the realism of the computed airspace configuration schedule. It seems to better capture the moments at which the reconfigurations are triggered, than when smoothing over 60 minutes.

VI. CONCLUSION AND PERSPECTIVES

The opening schedule computed with metrics smoothed over 30 minutes showed a number of reconfigurations close to reality, and with a number of control sectors well correlated to the actual configurations. It seems the best compromise among the models tested so far with the chosen neural network topology.

In a pre-tactical context, smoothing over relatively long periods of time may have positive consequences. The model should be more robust to uncertainties on aircraft trajectories when the complexity metrics will be computed from flight plans instead of past radar tracks.

In regard to the instant workload of a controller operating a sector at a time t , this smoothing strategy seems too drastic and may lead to miss the exact moments at which reconfigurations should be triggered, if this model was to be used for tactical purposes in a dynamic airspace management tool. An explanation is that only snapshots of the traffic situation – i.e. metrics values measured at time t – were used to predict the sector status. We may expect better results by considering the input metrics as time series, and by using recurrent neural networks instead of simple feed-forward networks. Provided this approach proves successful, the airspace configuration algorithms may prove useful for tactical purposes: flow managers may issue *what-if* requests and get some feedback on the resulting sectorization and workload balance among the control sectors.

Further works shall address both issues: improve the statistical model by using time series and recurrent networks to better capture the instant workload, and test the current model on simulated traffic, using flight plans as inputs, in order to predict the airspace opening schedule for the next day. Other smoothing strategies may also be tried, with different smoothing parameters for each metric for example, or with smoothing intervals centered on the current time.

REFERENCES

- [1] D. Gianazza and J. M. Alliot. Optimization of air traffic control sector configurations using tree search methods and genetic algorithms. In *Proceedings of the 21st Digital Avionics Systems Conference*, 2002.

- [2] D. Gianazza, J. M. Alliot, and G. Granger. Optimal combinations of air traffic control sectors using classical and stochastic methods. In *Proceedings of the 2002 International Conference on Artificial Intelligence*, 2002.
- [3] D. Gianazza. Airspace configuration using air traffic complexity metrics. In *7th USA/Europe Seminar on Air Traffic Management Research and Development*, 2007. best paper of "Dynamic Airspace Configuration" track.
- [4] D. Gianazza and K. Guittet. Evaluation of air traffic complexity metrics using neural networks and sector status. In *Proceedings of the 2nd International Conference on Research in Air Transportation*. ICRAT, 2006.
- [5] D. Gianazza and K. Guittet. Selection and evaluation of air traffic complexity metrics. In *Proceedings of the 25th Digital Avionics Systems Conference*. DASC, 2006.
- [6] C. Verlhac and S. Manchon. Improved configuration optimizer methodology to use a decision support tool. Technical note EEC Note No. 10/05, Eurocontrol Project NCD-F-FM, June 2005.
- [7] A. Klein. An efficient method for airspace analysis and partitioning based on equalized traffic mass. In *Proceedings of the 6th USA/Europe Air Traffic Management R & D Seminar*, 2005.
- [8] C.E. Bichot and N. Durand. A tool to design functional airspace blocks. In *7th USA/Europe Seminar on Air Traffic Management Research and Development*, 2007.
- [9] C. Verlhac and S. Manchon. Optimization of opening schemes. In *Proceedings of the fourth USA/Europe Air Traffic Management R&D Seminar*, 2001.
- [10] P. Flener, J. Pearson, M. Agren, C. Garcia-Avello, M. Celiktin, and S. Dissing. Air-traffic complexity resolution in multi-sector planning using constraint programming. In *7th USA/Europe Seminar on Air Traffic Management Research and Development*, 2007.
- [11] C. Verlhac, A. Schweitzer, E. Dumont, and S. Manchon. Improved configuration optimizer technical documentation. Technical note EEC Note No. 11/05, Eurocontrol Project NCD-F-FM, June 2005.
- [12] Brankica Pestic Le Foll. Network effect: A possible model to highlight interdependencies between flow management regulations. Master's thesis, Faculty of Transport and Traffic Engineering of Belgrade, 2006.
- [13] A. Klein, P. Kopardekar, M. D. Rodgers, and H. Kaing. Airspace playbook: Dynamic airspace reallocation coordinated with the national severe weather playbook. In *Proceedings of the 7th AIAA Aviation Technology, Integration and Operations Conference*, 2007.
- [14] Global air traffic management operational concept. Technical report, International Civil Aviation Organization, 2005.
- [15] H. Swenson, R. Barhydt, and M. Landis. Next Generation Air Transportation System (NGATS) Air Traffic Management (ATM)-Airspace Project. Technical report, National Aeronautics and Space Administration, 2006.
- [16] E. S. Stein, P. S. Della Rocco, and R. L. Sollenberger. Dynamic resectorization in air traffic control: A human factors perspective. Technical report DOT/FAA/TC-TN06/19, Atlantic City International Airport, NJ: Federal Aviation Administration William J. Hughes Technical Center, 2006.
- [17] Cognitive complexity in air traffic control, a literature review. Technical report, Eurocontrol experimental centre, 2004.
- [18] P. Kopardekar. Dynamic density: A review of proposed variables. FAA WJHTC internal document. overall conclusions and recommendations, Federal Aviation Administration, 2000.
- [19] P. Kopardekar and S. Magyarits. Measurement and prediction of dynamic density. In *Proceedings of the 5th USA/Europe Air Traffic Management R & D Seminar*, 2003.
- [20] A. J. Masalonis, M. B. Callahan, and C. R. Wanke. Dynamic density and complexity metrics for realtime traffic flow management. In *Proceedings of the 5th USA/Europe Air Traffic Management R & D Seminar*, 2003.
- [21] A. Majumdar, W. Y. Ochieng, G. McAuley, J.M. Lenzi, and C. Lepadetu. The factors affecting airspace capacity in europe: A framework methodology based on cross sectional time-series analysis using simulated controller workload data. In *Proceedings of the 6th USA/Europe Air Traffic Management R & D Seminar*, 2005.
- [22] G.B. Chatterji and B. Sridhar. Measures for air traffic controller workload prediction. In *Proceedings of the First AIAA Aircraft Technology, Integration, and Operations Forum*, 2001.
- [23] I. V. Laudeman, S. G. Shelden, R. Branstrom, and C. L. Brasil. Dynamic density: An air traffic management metric. Technical report, 1999.
- [24] J.H. Crump. Review of stress in air traffic control: Its measurement and effects. *Aviation, Space and Environmental Medicine*, 1979.
- [25] P. Averty, S. Athènes, C. Collet, and A. Dittmar. Evaluating a new index of mental workload in real ATC situation using psychological measures. Note CENA NR02-763, CENA, 2002.
- [26] A. Yousefi, G. L. Donohue, and K. M. Qureshi. Investigation of en route metrics for model validation and airspace design using the total airport and airspace modeler (TAAM). In *Proceedings of the fifth USA/Europe Air Traffic Management R&D Seminar*, 2003.
- [27] ACT-540 NAS Advanced Concepts Branch. An evaluation of dynamic density metrics using RAMS. Technical report (draft) DOT/FAA/CT-TN, Federal Aviation Administration, April 2001.
- [28] C. Mannings, S. Mill, C. Fox, E. Pfeleiderer, and H. Mogilka. The relationship between air traffic control events and measures of controller taskload and workload. In *Proceedings of the 4th Air Traffic Management Research & Development Seminar*, 2001.
- [29] D. Gianazza. *Optimisation des flux de trafic aérien*. PhD thesis, Institut National Polytechnique de Toulouse, 2004.
- [30] D. Delahaye and S. Puechmorel. Air traffic complexity: towards intrinsic metrics. In *Proceedings of the third USA/Europe Air Traffic Management R & D Seminar*, 2000.
- [31] P. Averty. Conflict perception by ATCS admits doubt but not inconsistency. In *Proceedings of the 6th Air Traffic Management Research & Development Seminar*, 2005.
- [32] M. I. Jordan and C. Bishop. *Neural Networks*. CRC Press, 1997.
- [33] C. M. Bishop. *Neural networks for pattern recognition*. Oxford University Press, 1996. ISBN: 0-198-53864-2.
- [34] B. D. Ripley. *Pattern recognition and neural networks*. Cambridge University Press, 1996. ISBN: 0-521-46086-7.

KEYWORDS

Air traffic complexity, airspace configuration, neural networks.

BIOGRAPHY

David Gianazza is currently researcher in the Planning, Optimization, and Modelling team¹⁹ of the DSN/DTI R&D domain. He received his engineer's degrees (IEEAC in 1986, IAC in 1996) from the french civil aviation academy (ENAC) and his M.Sc. (1996) and Ph.D. (2004) in computer science from the "Institut National Polytechnique de Toulouse" (INPT). A more detailed biography may be found here: www.recherche.enac.fr/~gianazza/

¹⁹The POM team was formerly part of the Global Optimization Laboratory LOG CENA/ENAC

Network Restructuring Models for Improved ATS Forecasts

Daniel A. DeLaurentis, Tatsuya Kotegawa, Aaron Sengstacken

School of Aeronautics and Astronautics

Purdue University

West Lafayette, IN 47907 USA

ddelaure@purdue.edu, tat@purdue.edu, asengsta@purdue.edu

Abstract— Current air traffic forecast methods employed by the FAA function under the assumption that the flight route network will not change, that is, no new flight routes will be added and no existing flight routes will be removed. However, in reality the competitive nature of the airline industry is such that new routes are routinely added between cities possessing significant passenger demand while other city-pairs are removed. This paper investigates models for forecasting network reconfiguration that exploit knowledge of network structure in the Air Transportation System (ATS), with the goal of improving overall forecast that drives policy and infrastructure enhancement decision-making.

Keywords-forecast; network theory; air traffic

I. INTRODUCTION

In order to synthesize long term plans for new technology, infrastructure improvements, policy enhancements, and regulations for the Air Transportation System (ATS), an understanding of air traffic dynamics is needed (i.e., determining how, when and where would air traffic arise or shift in the future). To meet this need, the FAA Air Traffic Organization (ATO) Office of Performance Analysis and Strategy (PAS) produces air traffic forecasts to project future demand, identify operational shortfalls, determine workforce requirements, and estimate the benefits of future investments. In the current forecast algorithm, the projected schedules are based upon the assumption that the future route network structure will be the same as the current network structure. That is, no new direct service routes are added between cities and, thus, the existing airline hub airports will continue to operate as hub airports.

However, the flight service route network structure is likely to change over time. The competitive nature of the airline industry is such that new direct routes are routinely added between cities with significant passenger demand and routes are also removed when demand dwindles. In addition, the location and number of airline hubs are not fixed; within the past several years, two major hubs have been eliminated (St. Louis and Pittsburgh), one airline hub opened and subsequently closed (Washington Dulles International Airport), and several other hubs were substantially

restructured. Looking further, scenarios are now taking shape in which environmentally-inspired imperatives may significantly modify the feasible sets of operations and network reconfiguration states. Overall, in order to enhance the ATS forecast precision, a better understanding of restructuring dynamics is required. *Motivated by this goal, research described in this paper is focused on investigating several models for forecasting the mechanism of network restructuring, in particular the aspect of new flight service route formation.* Families of parameters that describe the network topology are used as predictor variables in these models.

The remainder of the paper is organized as follows. After an introduction to network theory and some examples of its use in previous efforts for analyzing the ATS (Section II), Section III describes the data source and assumptions for all analysis. Detailed explanation of the three forecast algorithms developed up to date, along with key implications will follow in Section IV. Section V summarizes the interim results from these forecast algorithms.

II. NETWORK THEORY

A. Background

Multiple networks subsist in the overall ATS; the primary ones are summarized in Table I. The transport network topology was analyzed in the present study in which airports (nodes) are interconnected by flight routes (links). Modern Network Theory (also known as Network Science)^{1,2} has produced powerful results from multiple domains (e.g. physics, information, social science, biology) in recent years concerning how real world networks evolve. Some researchers have begun to explore application for analyzing air transportation networks. Guimera et al analyzed the worldwide air transportation network topology and computed measures which characterized the relative importance of cities/airports.³ Bonnefoy and Hansman⁴ used a plot of the weighted degree distribution for light jet operations to understand the capability of airports to attract the use of Very Light Jets (VLJs). A significant body of work exists in the

TABLE I. MULTIPLE, INTERACTING NETWORKS IN THE ATS

Network	Node (N) & Link(L)	Time Scale
Demand	N : Homes/Business L : Demand for Trips	Months/Years
Mobility	N : Origin/Destination L : Actual PAX trips	Days/Weeks
Transport	N: Airports L: Flight Routes	Days/Weeks
Operator	N: Aircraft / Crew L: Mission	Hours
Infrastructure	N: Waypoints and Airports L: Flight Routes	Months

TABLE II. DEFINITIONS FOR SELECTED NETWORK MEASURES

Parameter	Symbol	Description
Node	N/A	Airport
Node Degree	k_i	Number of flight routes existing at node i
Node Weight	w_i	Amount of operations associated with node i
Link Weight	r_{ij}	Amount of operations between node i and j
Clustering Coefficient	C_i	Measure of local cohesiveness for a node. Higher C_i implies that it is more likely an alternate connection path exists when a existing link fails
Eigenvector Centrality	x_i	A centrality measure of a node determined by its own and neighbors' degree. In the transport network, the importance of one airport is determined not only by its own number of routes supported, but also the number of routes and traffic level of airports with which it directly connects (an airport with high eigenvector centrality is likely to be very busy itself and also connected to other busy airports)
Population*	pop_i	Population within a 50 mile radius of node i

related domain of operations research on the design of optimal networks for particular instances and applications (e.g. schedule for an airline). However, these approaches generally do not pursue insight into the underlying structure of networks,

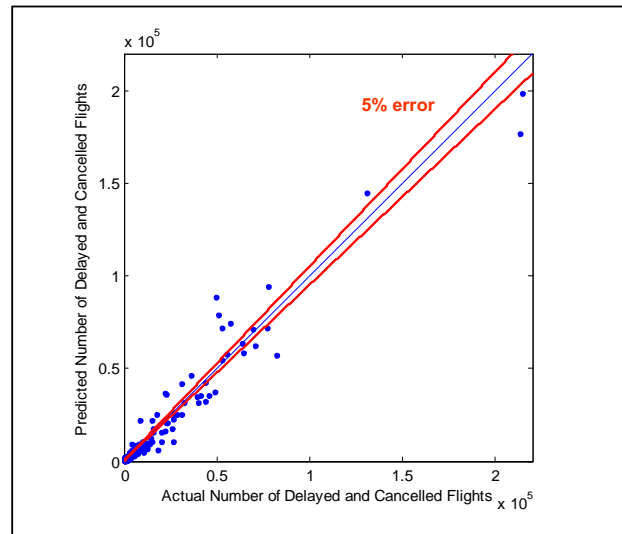


Figure 1. Prediction validation for 2004 delayed operations regression model (each data point is an airport)

the role this structure plays in future designs, nor the interplay between networks from multiple domains. Examination of the ATS using network theory at the national level and assessment of associated analysis models and techniques as a framework to provide both insight into ATS structure and a useful systems analysis has been a topic for our work⁵. The forecast of service route restructuring presented in this paper is one example application. Table II summarizes key network theory parameters that will be discussed and utilized for the remainder of this paper. More details can be found in [6].

The manner in which some of these parameters translate into real world performance and operations metrics is also topic of ongoing research⁴. One example of such a mapping is depicted in Eq. (1). This expression is a multivariate regression model for predicting the number of delayed operations for an airport using its degree, clustering coefficient, eigenvector centrality, degree weight and surrounding population as predictor variables.

$$\sqrt{\text{Delayed Ops per Year}} = 0.01928 + 0.147k_i + 0.02606C_i + 0.56722x_i + 0.20758w_i + 0.07462pop_i. \quad (1)$$

All variables are normalized using the corresponding maximum value, and the model produces a good coefficient of determination ($R^2 = 0.95$). The graph shown in Figure 1 displays the comparison between the actual and predicted number of delayed operations (for airports that registered at least one delay) for the 2004 ATS. A 5% error interval is also included. Eigenvector centrality and degree compose the majority of the regression model (significantly high F-values and Type II Sum of Squares compared to the other variables in the model). The number of expected delays can be forecasted using Eq. (1), but in order to reduce traffic congestion more attention should be placed on airports with not only high degree but also higher eigenvector centrality since these two

variables are anticipated to be the main source of operation delays. The primary implication for utilizing network theory as an ATS analysis tool, then, is that these measures can be efficient indicators of network operational performance. Also, focusing on the high-level characteristic of the ATS network generates deeper understanding on the nature of the ATS without being overwhelmed by its complexity.

III. DATA SOURCE AND ASSUMPTIONS

The primary research conducted under this study follows a similar approach to the delay regression model presented in the previous section. The objective is to determine if network theory parameters can be utilized to identify unconnected city-pairs that are most likely to connect in the future. The data used for this study was obtained from Air Carrier Statistics database family maintained by the U.S. Bureau of Transportation Statistics⁷. In particular, the Form 41 T- 100 Domestic segment (All US Carriers) database was used to construct the network studied. The BTS monitors 2627 total airports; however, the ATS network analyzed in this study was restricted to airports that had at least one cumulative commercial flight since 1990. This criterion reduces the network size to 887 nodes (airports). Several different measures are available for use in defining a link, such as the number of passengers, available seats, flights scheduled or actually performed. Since the transport network was explored in this research, a link was constituted by performed passenger flights per year between airports. Each flight route was required to have a minimum of 24 annual flights to be defined as a link in order to filter out any spontaneous, irregular flights that may bring ‘noise’ to the network analysis. To further simplify the analysis, the ATS network was assumed to be undirected and the number of arrival and departure operations were simply added together to compute the w_i and r_{ij} .

The source of the ATS network evolution can be broken down into four basic categories—flight route addition due to network expansion or reconfiguration, and flight route removal due to network contraction or reconfiguration. Reconfiguration refers to the ‘re-wiring’ of links within a

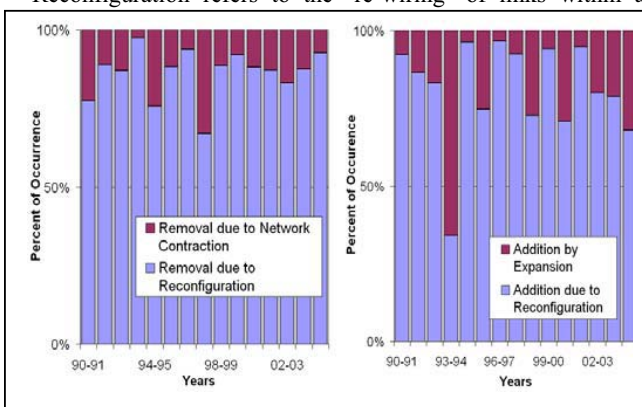


Figure 2. Variation in source of ATS network topology evolution

defined set of nodes; no new nodes are added and no pre-existing nodes are removed to create or destroy a link. Network contraction and expansion are opposite to reconfiguration, with links either being created by connecting to newly developed nodes or removed by detaching nodes from an existing set. Figure 2 illustrates the morphing of the ATS network categorized in these four evolution sets—it can be seen that the vast majority of the flight routes removed and created are a result of network reconfiguration. Thus, all forecast model and results described in the following section will only examine the mechanism of flight route construction due to reconfiguration. Investigation of the mechanism for the other three evolution categories (flight route removal due to network contraction / reconfiguration and flight route addition due to expansion) can be done relatively easily by supplying historical data sets to the algorithm that corresponds to the evolution category of interest.

IV. ROUTE CONSTRUCTION FORECAST ALGORITHMS

Three prototype forecast algorithms were created, compared and contrasted a) the logistic regression model, b) fitness function model and c) the artificial neural network approach. In this paper, the logistic regression model is discussed in detail. A brief summary for each approach is listed below.

Logistic regression is a statistical method to train a probability curve for event occurrence based on historical data input. The event for which the occurrence probability is calculated will be the construction of a new flight route between unconnected city-pairs and the inputs will be the parametric characteristics of the flight route. The iteratively-reweighted least squares (IRLS) method was utilized as the algorithm to fit the regression model with historical data.

Fitness function model is a network growth logic which operates under principles of the scale free network model where nodes with higher importance, or fitness value, are granted a higher probability to construct a new link. The initial composition of the function that computes nodal fitness projects growth that favors highly connected nodes (a hub-and-spoke type growth) that is typical in the ATS today. However, the fitness function can be modified to investigate the efficacy of various types of network growth mechanisms corresponding to a mix of different business models.

The **Artificial Neural Network (ANN)** is composed of a set of interconnected neurons that mimic human brain activity in attempting to develop optimal input-output mappings for prediction. Though some underlying fundamentals are similar to logistic regression, the ANN usually has higher precision. One drawback is that the relationship between input and output remains to be a ‘black box’—it cannot be expressed in terms of explicit equations as is typical in conventional statistical models. Also, due to the higher computational requirements of the ANN algorithm, the network to be analyzed via ANN must be kept relatively small.

A. Logistic Regression Model

A new flight route in the network context represents a new pathway between unconnected node pairs. The characteristic of new routes can be described by observing the traits of the airport pairs that create the route. The traits of the airport pair can be captured from two perspectives: a) by examining the list of parameters for each of the airports and b) by examining the relative difference of parameters between the airports. A record of network parameters (referred to as [parameter] list) for each airport involved in a new route indicates the type of airports that are most likely to be involved in a new connection. On the other hand, a record of parameter difference, or *deviation*, between the airport pairs produces a pair-wise measure that may be better able to characterize new connection formation. In particular, the type of connection can be categorized into either homogenous (connection between ‘large-large’ airports or ‘small-small’ airports) or heterogeneous (connection between ‘small-large’ airports).

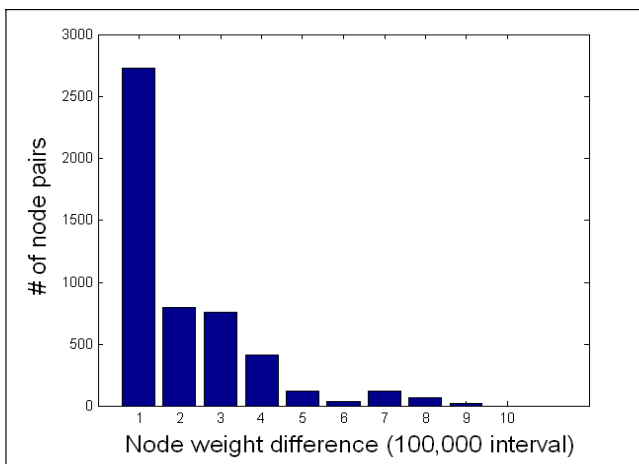


Figure 3. Node weight list distribution for new flight routes established in the ATS network between 1990 and 2005

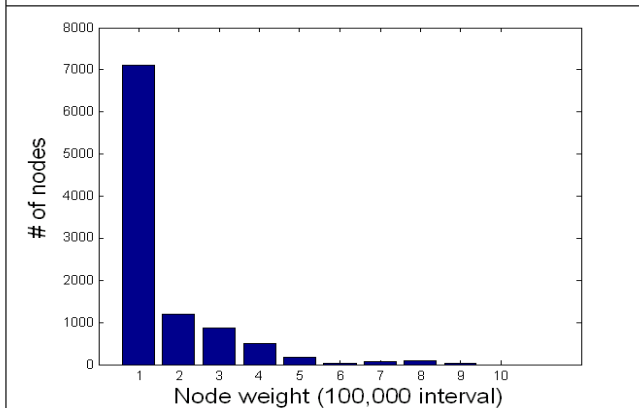


Figure 4. Node weight deviation divergence distribution for new flight routes established in the ATS network between 1990 and 2005

By combining insights from the parameter list and parameter deviation traits of the airport pairs that construct a new connection, the patterns that facilitate new flight routes can be extracted. The histograms in Figure 3 and 4 illustrate the distribution of node weight list and deviation for airports that formed new flight routes in the ATS network between 1990 and 2005. Figure 3 shows that most of the nodes involved in flight route constructions had relatively low traffic (between 1 and 100,000 annual operations), and Figure 4 shows that the difference (deviation) in the traffic of nodes involved in new links was mostly homogenous. The implication is that most new flight routes are established between airports that have lower traffic. A similar exercise was carried out for the remainder of the network parameters listed in Table II.

Parametric data are fed into the logistic regression model via design matrix X which ultimately gives node pairs that follow such trends higher likelihood of connection. Design matrix X is structured as shown in Eq. (2) for which all network theory variables in Table II are included, along with the distance information between node i and j . The second column of X , r_{ij} , signifies the occurrence of a new flight route construction for node i and j between observation years. If a new route is established between i and j a ‘1’ is placed in r_{ij} and if not a ‘0’ is placed.

$$X = \begin{bmatrix} 1 & r_{ij} & k_i & k_j & abs(k_i - k_j) & w_i & \dots \\ 1 & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \quad (2)$$

Based on the design matrix input, the regression model computes the variable parameter estimates using the standard iteratively-reweighted least squares (IRLS) algorithm and feeds the estimates into Eq. (3) which computes the probability of an unconnected node pair ij will construct a new flight route.

$$P_{connect,ij} = \frac{1}{1 + e^{-\hat{x}X_2,ij}} \quad (3)$$

X_2 in Eq. (2) is a matrix that contains the network parameter and parameter deviation information structured identically to X , except X_2 only includes data for unconnected node pairs. The design matrix X contains information for all connected and non-connected node pairs for probability curve training purposes. After Eq. (2) has been computed, $P_{connect,ij}$ is compared to a random number ($rand$) between 0 and 1 and the algorithm predicts a new flight route construction between node i and j if $P_{connect,ij} > rand$.

B. Accuracy Measures

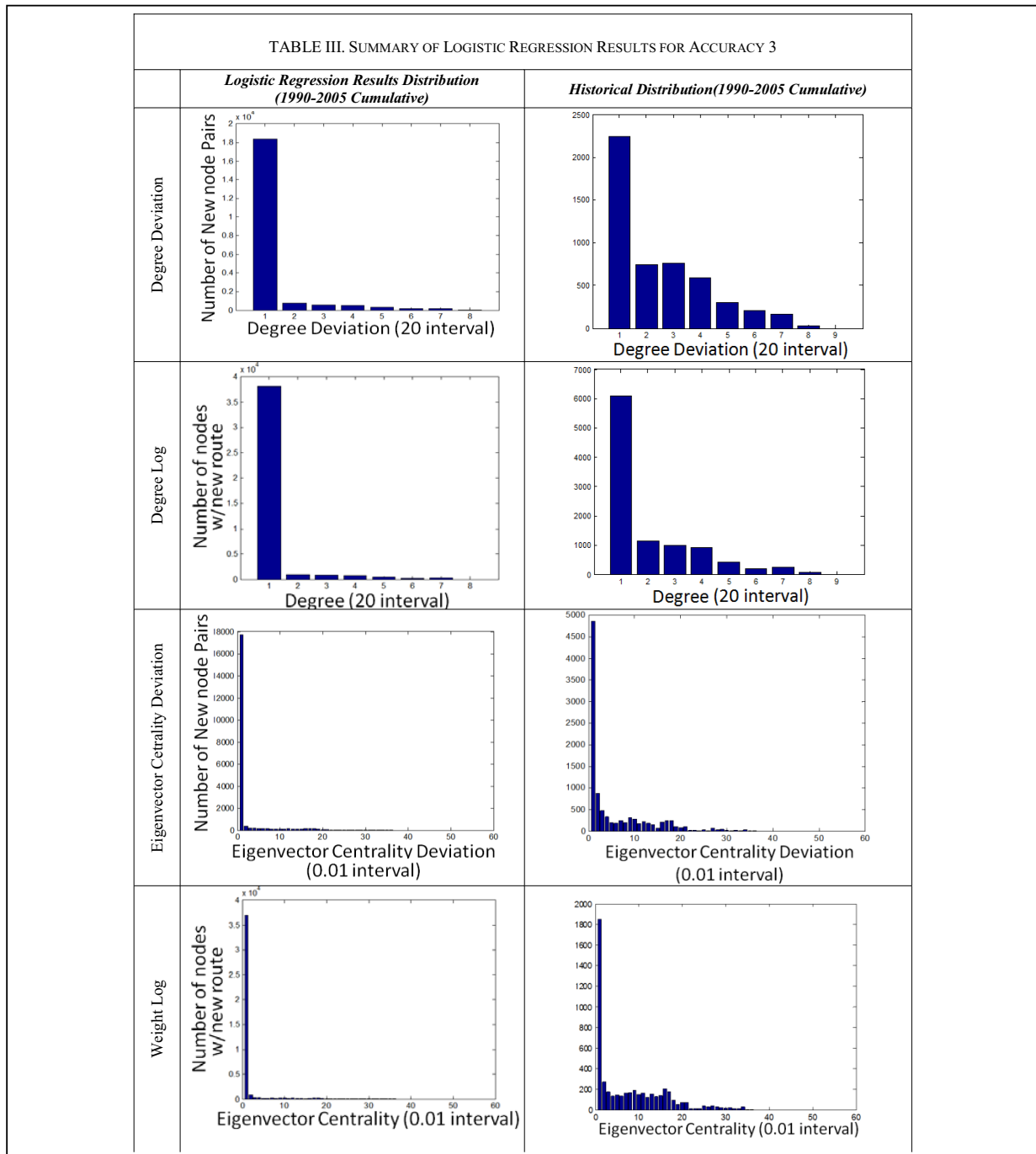
Three accuracy measures are employed to assess the forecast precision.

$$Accuracy\ 1 = \frac{\text{number of correctly predicted routes}}{\text{total number of predicted routes}} \quad (3)$$

$$Accuracy\ 2 = \frac{\text{number of correctly predicted routes}}{\text{number of actual new routes}} \quad (4)$$

Accuracy 1 shown in Eq. (3) was used to check how many new routes the forecast algorithm was predicting in order to obtain the correct new route. If the algorithm is predicting thousands of new routes to acquire only few correct new routes, accuracy 1 will be very low. On the other hand, accuracy 2, shown in Eq. (4) simply describes how many of the predicted new routes were correct, with respect to the number of actual new routes.

Accuracy 3 is a special type of accuracy measure which the coherence in distribution of characteristic trends for new links between the data and forecast model is examined. This is done by comparing the node parameter list and divergence histogram curve from the data and forecast algorithm, such as those seen in Figure 3 and Figure 4. The goal of employing accuracy 3 is to make sure that the forecast methods are predicting the future ATS network in the ‘right direction’; a formal equation to describe accuracy 3 currently does not exist.



C. Logistic Regression Results and Discussion

Results for an iteration of the logistic regression model are shown on Table IV; output for each year is an average over 10 runs. Inputs to the model consist of all the network theory parameters and deviation values for variables listed on Table II as well as the distance between airport pairs. Forecasts are done on a year-over-year basis; that is, parameters from only the previous year are utilized for the forecast.

The 'Correctly Predicted' column indicates the total number of correctly predicted routes for that year while the 'Total Predicted' column indicates the total number of predicted routes from the forecast algorithm. The logistic regression model has relatively high Accuracy 2 but low Accuracy 1 across all years, indicating that the algorithm can correctly forecast a significant number of new routes but does so by forecasting many additional routes in the process.

Accuracy 3 outcomes for both list and deviation distributions for degree and eigenvector centrality are shown in Table III. The first important finding was that the distributions produced by the logistic regression differ from those observed from historical data. In particular, the logistic regression allocates too much preference for connection to airports and airport pairs with small valued network parameters. This result, however, does not mask a second important finding from these results: across both parameter list and deviation distributions [for nodes with new connections], new routes were being established primarily between 'small' airports, whether defining small by degree or centrality significance. Appropriately, the logistic regression model distributes higher probability to establish connections between these small airports...it just distributes too much importance to these. Owing to the fact that the current ATS network is dominated by hub-and-spoke style architectures⁸, there exist many small, spoke airports and very few large, hub airports. Since there are more small nodes in the network, many small-

TABLE IV. SUMMARY OF LOGISTIC REGRESSION RESULTS

Year	Correctly Predicted	Total Predictions	Accuracy1	Accuracy 2
1990	N/A	N/A	N/A	N/A
1991	83.7	1217.6	0.0700	0.4314
1992	88.7	1427.6	0.0624	0.4264
1993	68.8	1065.1	0.0658	0.4145
1994	80.0	1365.8	0.0594	0.4645
1995	146.7	1395.8	0.1077	0.3976
1996	82.6	1409.8	0.0596	0.4325
1997	78.2	1489.0	0.0535	0.3476
1998	62.7	1177.6	0.0552	0.3968
1999	71.3	1488.4	0.0488	0.4006
2000	113.6	1980.2	0.0580	0.3381
2001	52.0	1286.3	0.0410	0.3824
2002	388.8	2328.9	0.1676	0.2878
2003	128.2	1123.5	0.1159	0.2728
2004	120.5	1043.1	0.1157	0.2953
2005	104.4	1088.1	0.0970	0.2806
Average	111.3	1392.5	0.0785	0.3713

to-small airport pairs arise as candidates for flight route construction. Abundant small-to-small airport connection candidates coupled with the forecast model favoring small-to-small airport connections from historical trends results in significant number of over-predictions for small-to small airport connections. This conclusion is the message conveyed from simultaneous consideration of Accuracy 1 and 3 metrics.

D. Brief Introduction and Analysis of the Artificial Neural Network

The Artificial Neural Network (ANN) is composed of a set of interconnected neurons that mimic human brain activity. Through supervised back-propagation training techniques, an ANN is able to achieve desired input-output mapping by adjusting the weights associated with each neuronal connection in the network. While the basic concept underlying ANN is similar to that of the logistic regression, the ANN usually has higher accuracy due to its higher degrees of freedom. However, the relationship between input and output for a trained ANN remains difficult to describe, unlike the logistic regression model. Also, the size of the network that can be analyzed with an ANN was restricted to a smaller size (~250 nodes) than for the logistic regression model due to the computational intensity of the ANN training algorithm.

The ANN approach proceeded via a feed-forward, fully-connected network algorithm⁹. After training the ANN with historical data, it was used to predict connections between two airport nodes. To capture the ATS network dynamics, the airport metrics for the previous three years were used at the input neuron layer resulting in an input layer of 63 neurons, a hidden layer consisting of 126 neurons, and a single output neuron. The input neurons represent two airport nodes, the hidden layer neurons used a *tan-sig* activation function, and the output neuron used a *log-sig* activation function. The single output neuron indicated the connectivity between the two airport nodes—1 for connected, 0 unconnected. The training data consisted of 50% of the historical data, while 25% was used for testing and 25% for validation. Once again, it should be noted that, for research reported here, the ANN was used to forecast only a subset of the ATS, mainly due to current computational limitations. In particular, historical data from the American Airlines (composed of routes operated by American Airlines, American Eagle and Executive Airline) and Southwest Airlines Transport Networks were employed to evaluate the accuracy of the ANN algorithm.

The trained ANN had extremely high accuracy rates in predicting new flight routes, with a minimum value of 70% for both Southwest and American Airlines Transport Networks. The ATS network used for the ANN forecast algorithm was abbreviated to 224 nodes (recalling that the logistic regression model considers 887 nodes). The 224 nodes included in the ANN training were the most active nodes in the ATS, excluding smaller, inactive airport nodes. Results for the two airline network forecasts along with translation to Accuracy 1 and 2 are shown below in Table V and VI.

TABLE V. TRAINED ANN RESULTS (SOUTHWEST AIRLINES NETWORK)

		Historical Data	
		Connect	Disconnect
Network Simulation	Connect	2850	1104
	Disconnect	738	380706
Accuracy 1 = 72.08%			
Accuracy 2 = 79.43%			

TABLE VI. TRAINED ANN RESULTS (AMERICAN AIRLINES NETWORK)

		Historical Data	
		Connect	Disconnect
Network Simulation	Connect	7291	2788
	Disconnect	2962	372357
Accuracy 1 = 72.33%			
Accuracy 2 = 71.11%			

The results displayed in Tables V and VI are separated into four cells. The sum of rows in the table describes the forecast results by the ANN, and the sum of columns describes the actual status of the unconnected node pairs. For example, in the American Airlines results (Table VI), the ANN forecasted a total of (7,291+2,788) 10,079 new flight routes (city pairs). Out of this total number of predicted new routes, in actuality 7,291 formed connections as determined from the historical data while 2,788 were disconnected (i.e., 'false alarms'). Similarly, the ANN forecasted that (2,962+372,357) 375,319 node pairs would remain disconnected but in actuality 2,962 out of these 375,319 made a connection. The overall accuracy results of the ANN are impressive when compared to the logistic regression model; however, it is difficult to extract any insights on the ATS evolution mechanism itself, since the relationships inside the trained ANN do not relate directly to the meaning of the input data (it is just an optimal prediction configuration). It is noted here again that the network size was significantly reduced in the ANN case.

E. Brief Introduction and Analysis of the Fitness Function Method

The fitness function model is a network growth logic which operates under the fundamentals of scale-free network model⁸. In this type of growth mechanism and network model, nodes with higher importance, or fitness value, are granted a higher probability to participate in a new link. The procedure begins by reading in the network topology from the previous year. For each node in the network, a fitness value was calculated through a specific functional composition of several nodal metrics listed in Table II. The initial functional composition used in the research was simply a ratio of individual nodal parameter of airports and the network sum of that parameter. For example, if a particular node has $k=10$ and the total k for the entire network is 100, its fitness function will be $10/100 = 0.1$. This type of fitness function projects growth that favors highly connected and important nodes (a hub-and-spoke type growth) that is typical in the ATS today.

However, the fitness function can be modified to allow various types of network growth mechanisms corresponding to a different mix of business models that might emerge in the future. This ability to tailor scenarios in an explicit manner dealing directly with service provider behavior is an attractive advantage of this approach. Subsequent to the fitness calculation, a pair-wise fitness was calculated for each node pair, and this was used to determine a probability of linking for all unconnected node pairs in the network. Links are added to the topology based on those pairs with high link probability (under some randomness).

Unlike the logistic regression model and the ANN approach, for which historical trends were directly projected to forecasting, the fitness function algorithm employs insights from growth models developed from the network science domain. Various combinations of network parameters (summarized in Table II) were investigated for the fitness function to determine which combination best suited the forecasting task. The fitness function that combines distance, degree, eigenvector centrality and nodal weight produced the forecast with highest accuracy. Results for an iteration of this fitness function model are shown on Table VII, noting once again that output for each year is an average over 10 runs. In comparison with the logistic regression model, the fitness function model produces poor results in the form of Accuracy 1 and 2. The problem of 'over-forecasting' was not resolved.

Surprisingly, however, the fitness function has improved Accuracy 3 results especially in the parameter list histograms (not displayed). The fitness function seems to develop the correct traits for choosing the nodes that develop new routes, but the specific prediction of 'which nodes' is relatively low perhaps due to the large pool of new connection candidates (there are approximately 4 million unconnected node pairs to choose from!). Even though Accuracy 1 and 2 for the fitness function approach may be lower than the ANN or logistic

TABLE VII. SUMMARY OF LOGISTIC REGRESSION RESULTS

Year	Correctly Predicted	Total Predictions	Accuracy1	Accuracy 2
1990	N/A	N/A	N/A	N/A
1991	36.9	1217.6	0.0700	0.4314
1992	45.3	1427.6	0.0624	0.4264
1993	28.5	1065.1	0.0658	0.4145
1994	31.1	1365.8	0.0594	0.4645
1995	75.3	1395.8	0.1077	0.3976
1996	33.6	1409.8	0.0596	0.4325
1997	46.7	1489.0	0.0535	0.3476
1998	30.9	1177.6	0.0552	0.3968
1999	29.8	1488.4	0.0488	0.4006
2000	51.6	1980.2	0.0580	0.3381
2001	23.3	1286.3	0.0410	0.3824
2002	139.5	2328.9	0.1676	0.2875
2003	54.7	1123.5	0.1159	0.2728
2004	45.4	1043.1	0.1157	0.2953
2005	48.1	1088.1	0.0970	0.2806
Average	111.3	1392.5	0.0785	0.3713

regression, the results appear sensible for an algorithm that does not depend on historical trends. This latter fact makes it difficult to judge the fitness function model performance comparative to the other two models described in this paper that purely utilizes historical trends as input. In addition, models that are independent of historical trends have a distinct advantage over models that are dependent—it is more likely that the forecast accuracy can be maintained even if the characteristic of the ATS is significantly shifted from the nature of past trends. The logistic regression and ANN model will be able to forecast the ATS at higher accuracy levels if and only if the ATS continues to evolve in the direction it has been evolving. However, if new policy, technology or operation methods that revolutionizes the ATS are introduced, the logistic and ANN will require new historical data to accumulate before further accurate forecast can be made. With algorithms like the fitness function model, change in ATS characteristics can be readily introduced by appropriately adjusting the fitness function calculation. Combinations of the models, therefore, also seem like a promising avenue for further research.

V. CONCLUSION AND FUTURE WORK

Current air traffic forecast methods employed at the FAA function under the assumption that the flight route network will not change, that is, no new flight routes will be added and no existing flight routes will be removed. In reality, the competitive nature of the airline industry and the potential need for new policies relating to the environment are such that new routes are routinely added between cities possessing significant passenger demand and other city-pairs are removed.

Research performed under this project and described in this paper explored means to understand network reconfiguration dynamics in the ATS. In particular, the aim was to expand the capabilities of the existing ATS forecast methods developed by the FAA, ultimately leading to improved decision-support in maintaining and enhancing the ATS. Employing network theory variables and concepts as a foundation to characterize the network of flight service routes in the ATS, three families of models were developed and tested: a) Logistic regression, b) a network topology based fitness function method, and c) an artificial neural network (ANN) algorithm. Results indicate that each has merit under differing accuracy metrics and each has methodological drawbacks. Advantages and disadvantages were documented. Overall, the logistic regression appears to capture more likely new city pairs, though in an inefficient manner as compared to the fitness function model. The ANN has superb prediction capabilities but was only tested on a sub-set of the network data due to computational and time constraints of this short duration study.

There still is much room for expansion in the current ATS forecast capabilities described in this paper. First, means are

available to increase all accuracy measures for each forecast algorithm. Some proposed methods to meet this goal include the implementation of more accurate and precise data of the ATS (i.e. ETMS instead of BTS), parsing the ATS network into sub-networks such as specific aircraft class or service provider, and removing variables deemed insignificant to the model or causing high multicollinearity. Forecasting based on multiple previous years for the logistic regression and fitness function model may also increase accuracy. Eventually, multiple forecast methods may be merged to go beyond the limit of individual methods. Second, enhanced ability to implement future scenarios will greatly improve the value of this research. All forecast methods are essentially based under an assumption in which the future ATS will grow in the way it has in the past. However, this is not true. New types of airline services, emergence of innovative technologies as well as new regulations and policies will impact the future state of network configurations; each of these may also drastically change the fundamental principles of operation of the ATS. In order to anticipate the effect for some of these ground-breaking factors in the forecast algorithms, a better understanding and mapping of the ATS is required. Finally, combining the best of the algorithms for predicting new city pairs with the FAA's current forecast method (based largely on the FRATAR algorithm) constitutes the most immediate next step.

ACKNOWLEDGMENT

The authors acknowledge sponsorship of this research by the FAA Air Traffic Organization Office of Performance Analysis and Strategy. The contributions of Donald Fry, En-pei Han and Sricharan Ayyalasomayajula at Purdue University are also acknowledged and appreciated.

REFERENCES

- [1] Newman, M., "The structure and function of complex networks," *SIAM Review*, Vol. 45, 2003, pp. 167-256.
- [2] Albert, R., Barabási, A.-L., "Statistical mechanics of complex networks," *Rev. Mod. Phys.* Vol. 74, pp. 47-97 2002.
- [3] Guimera, R., Mossa, S., Turtchi, A., Amaral, L., "The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles," *Proc. Nat. Acad. Sci.*, Vol. 102, 2005, pp. 7794-7799.
- [4] Bonnefoy, P., Hansman, R., "Potential Impacts of Very Light Jets in the National Airspace System," *Journal of Aircraft*, Vol. 44, No. 4, 2007, pp. 1318-1326.
- [5] DeLaurentis, D., Han, E-P., Kotegawa, T., "Network-Theoretic Approach for Analyzing Connectivity in Air Transportation Capacity Networks," accepted for publication in *AIAA Journal of Aircraft*.
- [6] DeLaurentis, D., Kotegawa, T., "Establishment of a Network-based Simulation of Future Air Transportation Concepts," AIAA 2006-7719, Sep., 2006
- [7] U.S. Bureau of Transportation Statistics (BTS), www.bts.gov; [accessed numerous times over study period]
- [8] Conway S., "Scale-free Networks and Commercial Air Carrier Transportation in the United States," Proceedings of the 24th Congress of the International Council of the Aeronautical Sciences (ICAS), Yokohama, Japan, 2004.
- [9] Tsoukalas, L. H. and R. E. Uhrig (1997). *Fuzzy and Neural Approaches in Engineering*, John Wiley and Sons, Inc.

Resource Allocation in Flow-Constrained Areas with Stochastic Termination Times

Moein Ganji¹⁴, Alex Nguyen¹³, David Lovell¹⁴, Michael Ball¹²

University of Maryland, College Park, MD 20742

¹ Institute for Systems Research

² Robert H. Smith School of Business

³ Department of Electrical and Computer Engineering

⁴ Department of Civil and Environmental Engineering

Abstract

In this paper we formulate an optimization problem for the assignment of dispositions to flights whose preferred flight plans pass through a flow-constrained area. For each flight, the disposition can be either to depart as scheduled but via a secondary route that avoids the flow-constrained area, or to use the originally intended route but to depart with a controlled departure time and accompanying ground delay. We anticipate that the capacity through the flow-constrained area will increase at some future time once the weather activity clears. The model is a two-stage stochastic program that represents the time of this capacity windfall as a random variable, and determines expected costs given a second-stage decision, conditioning on that time. The goal is to minimize the expected cost over the entire distribution of possible capacity increase times.

I. INTRODUCTION

A flow-constrained area (FCA) is a region of the national airspace system (NAS) where a capacity-demand imbalance is expected due to some unexpected condition such as adverse weather, security concerns, special-use airspace, or others. Flow-constrained areas might be drawn as polygons in a two-dimensional space, although in practice they are usually represented by a single straight line, functioning as a cordon.

When an FCA has been defined, it is then often the case that an airspace flow program (AFP) is invoked by the Federal Aviation Administration (FAA). An AFP is a traffic management initiative (TMI) issued by the FAA to resolve the anticipated capacity-demand imbalance associated with the FCA. It is the goal of this paper to develop a method by which, given the aggregate data described here, specific orders for individual flights can be developed for a single FCA that a) maximize the utilization of the constrained airspace, b) prevent the capacity of the FCA from being exceeded, and c) achieve a system-wide delay minimization objective. We recognize that this model cannot be directly applied to AFP planning as it does not address issues related to the manner in which the FAA and the flight operators collaborate in reaching

a final decision regarding each flight. Our goal here is to develop relevant stochastic optimization models. We intend to address issues related to collaborative decision making (CDM) in later papers.

II RELATED RESEARCH

The research most closely related to this paper has to do with airport ground holding. Much work has been done in addressing the airport ground holding problem, including the development of stochastic integer programming models, [1], [4], [5], [6], [11]. However, there is still much active research in the development of models for managing flights through congested areas of en route airspace under weather uncertainty.

In [10], the rerouting of a single aircraft to avoid multiple storms and minimize the expected delay was examined. In this model, the weather uncertainty was treated as a two-state Markov chain, with the weather being stationary in location and either existing or not existing at each phase in time. A dynamic programming approach was used to solve the routing of the aircraft through a gridded airspace, and the aircraft was allowed to hedge by taking a path towards a storm with the possibility that the storm may resolve by the time the aircraft arrived. The focus of the work was on finding the optimal geometrical flight path of the aircraft, and not on allocation of time slots through the weather area. Follow-on work expanded to modeling multiple aircraft with multiple states of weather and attempted to consider capacity and separation constraints at the storms.[9][8]

Initial steps at a concept of operations that describes the terminology, process, and technologies required to increase the effectiveness of uncertain weather information and the use of a probabilistic decision tree to model the state space of the weather scenarios was provided in [1]. Making use of this framework is a model recently proposed that uses a decision-tree approach with two-stage stochastic linear programming with recourse to apportion flows of aircraft over multiple routing options in the presence of uncertain weather [3]. In

the model, an initial decision is made to assign flights to various paths to hedge against imperfect knowledge of weather conditions, and the decision is later revised using deterministic weather information at staging nodes on these network paths that are close enough to the weather that the upcoming weather activity is assumed known with perfect knowledge. Since this is a linear programming model, only continuous proportions of traffic flow can be obtained at an aggregate level, and not decisions on which individual flights should be sent and when they should arrive at the weather. In [7], a stochastic integer programming model is developed based on the use of scenario trees to address combined ground delay-rerouting strategies in response to en route weather events. While this model is conceptually more general than ours, by developing a more structured approach we hope to develop a more scalable model.

Recently, a Ration-by-Distance (RBD) method was proposed as an alternative to the Ration-by-Schedule (RBS) method currently used for Ground Delay Programs (GDPs) that maximizes expected throughput into an airport and minimizes total delay if the GDP cancels earlier than anticipated [3]. This approach considers probabilities of scenarios of GDP cancellation times and assigns a greater proportion of delays to shorter-haul flights such that when the GDP clears and all flights are allowed to depart unrestricted, the aircraft are in such a position that the expected total delay can be minimized. While this problem was applied to GDPs, the principles of a probabilistic clearing time where there is a sudden increase in capacity and making initial decisions such that the aircraft are positioned to take the most advantage of the clearing is similar to our problem.

III. MODEL

A. Model inputs

Our base model inputs consist of information about the FCA, which is consistent with the information used in AFP planning.

- Location of the FCA
- Nominal (good weather) capacity of the FCA
- Reduced FCA (bad weather) capacity of the FCA
- Start time of the AFP
- Planned end time of the AFP

From a list of scheduled flights and their flight plans, we determine the set of flights whose paths cross the FCA, and who therefore would be subject to departure time and/or route controls under an AFP. We also require a set of alternate routes for each flight (see Figure 1). The alternate route for each flight should be dependent on the geometry of the FCA and the origin-destination pair it serves. These most likely would be submitted by carriers in response to an AFP; for the purposes of this paper it is assumed they are submitted

exogenously, although for testing purposes it was necessary to synthesize some alternate routes.

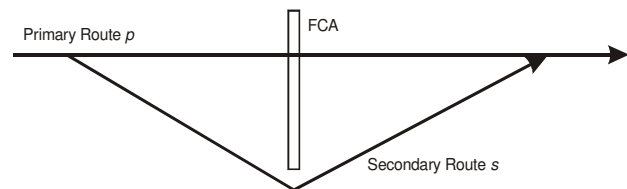


Figure 1. Primary route through the FCA and secondary route bypassing FCA.

B. Controls

In order not to exceed the (reduced) FCA capacity, each flight will be assigned one of two dispositions in the initial plan reacting to the FCA:

1. The flight is assigned to its primary route, with a controlled departure time that is no earlier than its scheduled departure time. Given an estimate of en route time, this is tantamount to an appointment (i.e., a slot) at the FCA boundary. Some flights might be important enough that they depart on time, the AFP notwithstanding. Other flights might be assigned some ground delay.
2. The flight is assigned to its secondary route, and is assumed to depart at its scheduled departure time.

This limited set of conceived actions imposes several important assumptions:

- We do not consider airborne holding as a metering mechanism to synchronize a flight on its primary route with its slot time at the FCA
- We assume that any necessary number of flights can be assigned to their secondary routes without exceeding any capacity constraints in other parts of the airspace. (In fact, our model can easily be extended to handle such “other” capacity constraints but in this initial version we do not include them.)
- We assume that, when the weather clears, the FCA capacity increases immediately (“in one step”), back to some higher capacity.
- The random variable is the time at which the FCA capacity increases back to a higher value. We assume that perfect knowledge of the realization of this random variable is not gained until the scenario actually occurs, and so no recourse can be taken until the scenario is realized.

C. Scenarios and future responses

The outputs of this model are:

1. An initial plan that designates whether a flight is assigned to its primary route or secondary route; for those assigned to their primary route an amount of ground delay (possibly zero) is assigned.
2. A recourse action for each flight under each possible early clearance time.

We model the time at which the weather clears (i.e. FAC capacity increases) as a discrete random variable, with some exogenous distribution. For any realization of the capacity increase time, the flights in question will be in some particular configuration as specified in the initial plan. Some will have departed, either on their primary or secondary routes, some will already have completed their journeys, and some will still be at their departure airports.

Flights that were originally assigned to their primary route and that have already taken off will be assumed to continue with that plan. For any such flight, the primary route is assumed to be best, so no recourse action is necessary.

We now consider flights originally assigned to their primary route that have not yet taken off. We need not consider transferring them to their secondary routes, because if that were a good idea in the improved capacity situation, it would also have been a good idea in the initial plan. Thus, the only possible change in disposition for these flights involves potentially moving their controlled departure time, i.e. reducing their assigned ground delay. Constraints are required to define the range of times the flight can arrive at the FCA boundary based on the required en route time and the time the recourse action is taken (clearance time). We also explicitly enforce the restriction that under such situations the assigned ground delay (or equivalently arrival time at FCA) cannot be increased.

All other flights not yet considered were originally assigned to their secondary routes, with departure times as originally scheduled. These secondary routes avoid the FCA somehow. Under the FCA capacity windfall, some of those flights may now have an opportunity to use the FCA. If a flight has not yet taken off, and it is decided that it can use the FCA, the lowest cost way to do this is to re-assign it back to their primary route, with some controlled departure time no earlier than their scheduled departure time. If, on the other hand, the flight has already taken off, then the only mechanism to allow it the use of the FCA is a hybrid route that includes that portion (and perhaps more) of the secondary route already flown, plus a deviation that traverses the FCA and presumably rejoins the primary route at some point after the FCA (see Figure 2). A flight that is already en route via its secondary route may or may not prefer such a hybrid path, depending on the difference in cost (time, fuel, etc.) between doing that and continuing on its secondary route. There may be many possible hybrid routes, and perhaps only a limited set of those would be acceptable to carriers and air traffic control (ATC).

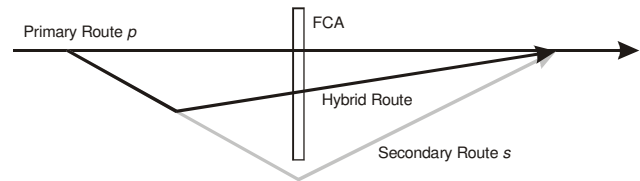


Figure 2. Reverting from secondary route back to primary route through FCA.

For each possible value of the capacity windfall time, we determine the expected locations of all affected flights at that time, and also what would be the best change in disposition, if any, for each of those flights according to a system performance metric. With this information, we can compute the conditional cost associated with flying these flights under that realization of the stochastic event. Ultimately, then, the goal of the optimization problem is to minimize the expected total cost, given these conditional costs and their probabilities.

D. Model development

We start by defining the discrete lattice on which time will be represented. We assume there is an index set $\{1, \dots, T\}$ of size T that demarcates equally spaced time slots, each of duration Δt . Each of these represents a possible appointment time window at the FCA. The nominal capacity of the FCA should be specified in terms of the maximum number of flights permissible during one of these time windows. The number of time slots T then depends directly on Δt and the total duration of an AFP, perhaps inflated to allow for ending times later than the original estimate. The reference time $t=1$ can be chosen as the earliest scheduled departure time of all of the affected flights. The actual time indicated by the index t is then $(t-1/2)\Delta t$.

The flights affected by the FCA can be determined from the filed flight plans for that day, minus known cancellations and re-routes at the time the AFP is invoked. These flights are indexed according to the set $\{1, \dots, F\}$. In the rest of the paper, any specific reference to a time period t and flight f assumes that $t \in \{1, 2, \dots, T\}$ and $f \in \{1, \dots, F\}$.

1) First stage (Initial Plan)

There are two sets of assignment variables that are related to decisions about the dispositions of flights. One set represents the initial plan, which are the decisions provided by the model that will be enacted immediately once the model is run and the AFP is declared. The second set represents conditional decisions (recourse actions) based on the random variable representing the time at which the capacity windfall takes place, which we do not know at the time of the execution of this optimization problem, but that we condition for when determining the best initial plan.

For the initial plan, we define the following set of binary decision variables:

$$x_{f,t}^p = \begin{cases} 1, & \text{if flight } f \text{ uses its primary route and} \\ & \text{has an appointment time } t \text{ at the FCA} \\ 0, & \text{otherwise} \end{cases}$$

$$x_f^s = \begin{cases} 1, & \text{if flight } f \text{ is assigned to its secondary} \\ & \text{route} \\ 0, & \text{otherwise} \end{cases}$$

Every flight f needs to have an assigned disposition under the initial plan, thus:

$$\sum_t x_{f,t}^p + x_f^s = 1 \quad \forall f \quad (1)$$

We require that any flight that is assigned to its primary route cannot be given an appointment slot at the FCA that is earlier than its scheduled departure time plus the expected en route time required to arrive at the FCA. If $E_f \Delta t$ represents the en route time (from its origin to the FCA) for flight f , and $D_f \Delta t$ is the scheduled departure time for flight f , then:

$$\sum_{t=1}^{D_f + E_f} x_{f,t}^p = 0 \quad \forall f \quad (2)$$

This construction requires that en route times and scheduled departure times are represented on the same discrete lattice as the FCA appointment times.

No similar constraint is applied to flights assigned to their secondary routes under the initial plan, because they are not metered at any point and hence are expected to depart at their originally scheduled departure time. There is no provision in the model for a flight to depart early, despite the fact that the secondary route takes more time than the primary route (since, subject to minor variations, airlines do not allow flights to take off before their scheduled departure times).

It might be the case that for a particular flight f , there is a latest slot time l_f at the FCA that the carrier who owns that flight would be willing to accept. Slots later than l_f can be prevented via the following constraint:

$$\sum_{t=l_f+1}^T x_{f,t}^p = 0 \quad (3)$$

For any flight for which l_f is not explicitly provided, l_f is an effective time beyond which it would never make sense not to choose the secondary route.

The initial constrained capacity (maximum number of flights) for time window t can now be defined as C_t^0 and the constraint to enforce it is:

$$\sum_f x_{f,t}^p \leq C_t^0 \quad \forall t \quad (4)$$

2) Second Stage (Revised Plan)

The variables and constraints defined so far represent the first stage of the stochastic program. It is assumed that these decisions will be enacted deterministically immediately after the FCA is declared. Next, we describe the second stage of the stochastic program – those variables that represent the conditional decisions we expect *would be* made if any of a number of possible capacity windfall times happens to come true in the future. We model the time slot at which this occurs as a discrete random variable with domain Ω and probability mass function

$$f_U(u) = \Pr\{U = u\} \quad \forall u \in \Omega$$

Under a capacity windfall, a flight that was originally assigned to its primary route with a controlled departure time might still be given the same general disposition, although its departure time could be moved earlier if that were beneficial to the system goal. We let

$$y_{f,t}^p | u = \begin{cases} 1, & \text{if at the time } U = u \text{ of the capacity windfall,} \\ & \text{flight } f \text{ is assigned to its primary route with} \\ & \text{appointment slot } t \text{ at the FCA} \\ 0, & \text{otherwise} \end{cases}$$

We will (shortly) introduce other variables for the other possible second stage flight dispositions, and we will require that all flights be assigned a disposition under every possible realization of the stochastic event U . For now, we proceed by obviating values of $y_{f,t}^p | u$ that would either be physically infeasible or politically imprudent. Later, structural constraints plus pressure from the objective function will lead to the best possible selection of second stage dispositions for all flights.

First, it is impossible to assign a flight to a slot that would require it to depart before its scheduled departure time:

$$y_{f,t}^p | u = x_{f,t}^p \quad \forall f, u, \quad \forall t \in \{1, \dots, D_f + E_f\} \quad (5)$$

This constraint works with constraint (2) to achieve the required result.

Given the timing U of the capacity windfall, some flights may already have taken off. If they did so via their primary route (with a controlled departure time), then their second stage disposition should match that of the first stage:

$$y_{f,t}^p | u = x_{f,t}^p \quad \forall f, u, \quad \forall t \in \{1, \dots, u + E_f\} \quad (6)$$

A closer look at constraint (6) reveals that it also satisfies an important requirement for flights that have not yet taken off. For any particular flight f and given the capacity windfall time u , the collection of primary stage variables $\{x_{f,t}^p\}_{t=1}^{t=u+E_f}$ will either contain one at exactly one position or it will consist entirely of zeros. In the former case, this means that the flight has already taken off, and that situation has been dealt with. In the latter case, this is indicative of the fact that these slot times are infeasible. Thus, even for flights that have not yet

taken off, constraints (2) and (6) insure that they will not be assigned, in the second stage, to their primary routes with slot times that they cannot achieve.

Looking at constraints (5) and (6), it is clear that they can be combined:

$$y_{f,t}^p | u = x_{f,t}^p \quad \forall f, u, \quad \forall t \in \{1, \dots, \max(u, D_f) + E_f\} \quad (7)$$

On the other hand, for flights that already took off via their secondary routes (and therefore at their scheduled departure times), the only possible second stage dispositions are secondary or hybrid routes, so assignments to primary routes for these flights must be prevented:

$$\sum_t y_{f,t}^p | u \leq 1 - x_f^s \quad \forall u, \forall f \ni D_f < u \quad (8)$$

In addition, we will not allow a flight whose controlled departure time is being moved in the face of a capacity windfall to be worse off than it was before this event materialized:

$$y_{f,t}^p | u \leq \sum_{q \geq t} x_{f,q}^p + x_f^s \quad \forall u, f, t \quad (9)$$

Notice that we want to allow for the possibility that flights originally assigned to their secondary routes can revert, under the appropriate circumstances and if the optimization decides this is best, to their primary route if they have not already taken off, which is why the variable x_f^s appears in constraint (9).

For flights that were originally assigned to the secondary route, the increased capacity at the FCA might allow some of these flights to pass through the FCA and thus improve their flight path by returning to the primary route at some point after the FCA or continuing directly to the destination. For a flight that has not yet departed, one could choose to have the same structure apply, but the portions of the total flight path spent on the secondary and reverting routes would then have to have length zero. In this paper, as will be shown later, we use a different approach. We define the second-stage decision variables for this choice as follows:

$$y_{f,t}^h | u = \begin{cases} 1, & \text{if flight } f \text{ was originally assigned to its} \\ & \text{secondary route, but under capacity} \\ & \text{clearing time } u \text{ has been assigned an} \\ & \text{FCA appointment slot } t \\ 0, & \text{otherwise} \end{cases}$$

This decision can only be reached for flights that were originally assigned to their secondary routes:

$$y_{f,t}^h | u \leq x_f^s \quad \forall u, f, t \quad (10)$$

However, it should be obvious that the objective function will enforce this behavior implicitly.

The flights in question will be on their secondary routes and diverting onto a hybrid route that passes through the FCA. We need to impose constraints that insure that these flights are

only assigned to FCA time slots they can feasibly reach. If a flight diverts from its secondary route to its hybrid route at time t^d there will be an earliest time it can reach the FCA. Figure 3 illustrates the geometry used to compute the parameter used by our model:

$t_{r,t}^d$ is the time at which f must divert from its secondary route so as to use its hybrid route that arrives at the FCA at time t .

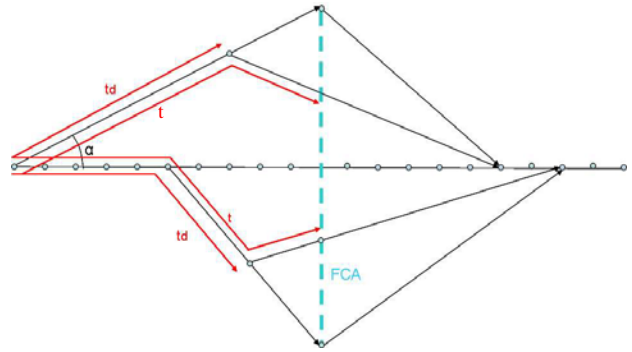


Figure 3. Possible diversion from secondary to hybrid route.

The following constraint prevents a flight from diverting to its hybrid route before the weather is actually cleared.

$$y_{f,t}^h | u = 0 \quad \forall f, u, \quad \forall t | t_{f,t}^d \leq u \quad (11)$$

In addition, the time slot assignment cannot be later than the latest time for which it would be reasonable to accept an assignment at the FCA considering the geometry of its secondary route:

$$y_{f,t}^h | u = 0 \quad \forall f, u, \quad \forall t > l_f \quad (12)$$

The final option possible, if the capacity increase occurs at time u , is to leave a flight that was originally assigned to its secondary route on that route, with the same (scheduled) departure time:

$$y_f^s | u = \begin{cases} 1, & \text{if flight } f \text{ was originally assigned to its} \\ & \text{secondary route, and if, under AFP stop} \\ & \text{time } u, \text{ that decision remains unchanged} \\ 0 & \text{otherwise} \end{cases}$$

Practically speaking, it would never make sense to assign a flight to its secondary route under the recourse if it had not also been given the same assignment in the initial plan. It might seem, therefore, that the following constraint is necessary:

$$y_f^s | u \leq x_f^s \quad \forall u, f \quad (13)$$

However, it should be obvious that the objective function will enforce this behavior implicitly. If it was cost effective to assign a flight to its secondary route under the recourse, it would be cost effective to do so under the initial plan.

Constraints 10 and 13 can be combined into a single constraint:

$$y_{f,t}^h |u + y_f^s |u \leq x_f^s \quad \forall u, f, t \quad (14)$$

It would be possible, given the constraints developed so far, to assign a flight to a hybrid route that essentially reverts to the primary route immediately. In other words, this would be an assignment that is tantamount to taking off on the primary route at the scheduled departure time, which is a more logical way to interpret this outcome. Therefore we introduce the following constraint to enforce this behavior:

$$y_{f,D_f+E_f}^h |u = 0 \quad \forall f, u \quad (15)$$

For each time scenario u , every flight f must be assigned to one of these dispositions. Furthermore, if the disposition involves being scheduled into a slot appointment at the FCA, no more than one slot can be assigned to a given flight. Given that the decision variables are required to be binary, the following constraint addresses both of these concerns:

$$\sum_t y_{f,t}^p |u + \sum_t y_{f,t}^h |u + y_f^s |u = 1 \quad \forall u, f \quad (16)$$

For any value $U=u$, there will be a new capacity profile $C^u(t)$ that agrees with $C^0(t)$ up to time $t=u$, but represents an increase in capacity beyond that point. For example, if $C^0(t)$ had been a constant vector, then $C^u(t)$ could be a step function that makes a jump at time $t=u$. On the other hand, if $C^0(t)$ had been a periodic 0-1 function, then $C^u(t)$ might just have an increased duty cycle after time $t=u$. Figure 4 shows examples of both of these extremes. A wide variety of profiles for $C^u(t)$ are possible; the only real requirements are that it agree with $C^0(t)$ prior to time $t=u$, and that after that time, it supports a higher rate of flow than that was possible under the initial plan. The capacity constraint under the scenario $U=u$ can now be written as:

$$\sum_f y_{f,t}^p |u + \sum_f y_{f,t}^h |u \leq C_t^u \quad \forall u, t \quad (17)$$

3) Objective Function

Since our model involves the specification of decisions that are conditioned random events, the objective function will be an expected value. To emphasize the paradigm of creating a plan (our initial plan) together with contingency plans (our recourse actions), we represent the objective function as the sum of the deterministic cost of the initial plan minus the expected savings from recourse actions.

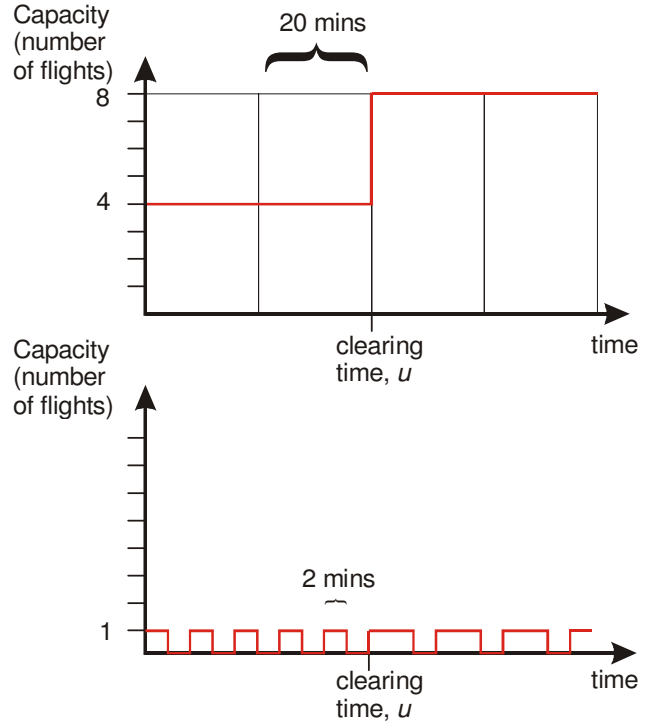


Figure 4. Capacity functions with high (top) and low (bottom) throughput before and after clearing.

Therefore the objective function can thus be represented as:

$$\text{Min} \left[C(X) - \sum_u P_u S(Y_u) \right] \quad (18)$$

Or more precisely:

$$\text{Min} \quad Z = z^1 + z^2 - \sum_u P_u (z_u^3 + z_u^4) \quad (19)$$

Where,

$$z^1 = \sum_f \sum_t c_{f,t}^p x_{f,t}^p \quad (20)$$

$$z^2 = \sum_f c_f^s x_f^s \quad (21)$$

$$z_u^3 = z^1 - \sum_f \sum_t c_{f,t}^p y_{f,t}^p |u + \sum_f \sum_t c_f^s s_{f,t}^p |u \quad (22)$$

$$z_u^4 = \sum_f \sum_t sv_{f,t}^h y_{f,t}^h |u \quad (23)$$

where

$c_{f,t}^p$ is the cost of assigning flight f to its primary route so that it arrives at the FCA at time t .

c_f^s is the cost of assigning flight f to its secondary route.

$sv_{f,t}^h$ is the savings incurred if flight f starts out on its secondary route but reverts to a hybrid route that arrives at the FCA at time t .

$s_{f,t}^p$ is a dummy binary variable that works as an indicator. It takes value of one when a flight initially assigned to its

secondary route is assigned back to its primary route under revised plan.

So;

$$s_{f,t}^p = \text{Min}(x_{f,t}^s, y_{f,t}^p) \tag{24}$$

IV. COMPUTATIONAL EXPERIMENT

We conducted a computational experiment to give some preliminary evidence as to the computational feasibility of the model and its impact on decision making. We now describe the problem data. Flights, their routes and alternate routes were generated artificially based on the airspace geometry given in Figure 5. There were three types of flights:

Short haul: length - 60 min: origin-to-FCA – 30 min, FCA-to-destination – 30 min; reroute angle – $\arctan(30/30)$.

Medium haul: length – 180 min: origin-to-FCA – 90 min, FCA-to-destination – 90 min; reroute angle – $\arctan(30/90)$.

Long haul: length – 300 min: origin-to-FCA – 150 min, FCA-to-destination – 150 min; reroute angle – $\arctan(30/150)$.

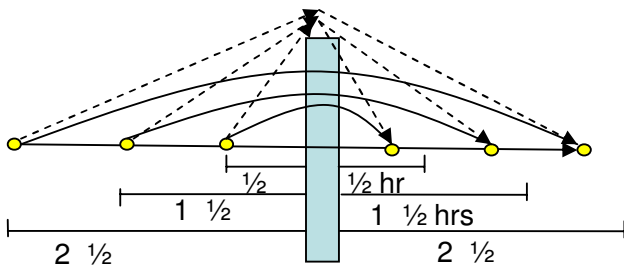


Figure 5: Airspace Geometry for Flight Generation

There were $F=200$ flights with one flight departing every 1 minute and departures alternating among the three flight types. First flight departed at 2:00 PM. There were $T=200$ time slots; each slot had a width of $\Delta t=2$ minutes. Initially, the FCA had restricted capacity of 1 flight per every three time slots (10 flights per hour). In all cases, the FCA cleared by 7:00 PM so that capacity rose to 4 flights per time slot (120 flights per hour). There were four possible early clearance times: 3:00 PM, 4:20 PM, 5:30 PM, each occurring with probability 0.25. In event of early clearance, slot capacity rose from 1/3 to 2 flights for slots between the clearance time and 7:00 PM.

Three cases were run:

All Options: This was the complete model as defined in the paper.

Reroute but No Recourse: In this case, the reroute option did not include the possibility of recourse, i.e. the $y_{f,t}^h$ variables are all fixed at zero. This corresponds to a decision making scenario where the possibility of rerouting after departure is not taken into account.

No Reroute: In the case, no rerouting is allowed. This corresponds to a decision making scenario under which the problem is solved only using ground delays.

Table 1 gives the results of our experiments. The costs of various solution components as well as the total expected cost are given. Note that the “All Options” scenario produced substantial cost savings over the other cases (particularly the No Reroute case). The fact that the initial plan costs (cost of the assigned ground delay $C(x_p)$ and cost of complete reroutes $C(x_r)$) changed significantly among the cases shows that taking the various recourse options into account can substantially alter the initial plan.

Also note that running times are given. A 2.8 GHz Intel® Pentium® based computer was used with 1.99 GB of RAM. The IP solver used was XPress MP® vers 2007B.

	All Options			Reroute but No Recourse			No Reroute		
air/ground delay cost	4			4			4		
Objective Function	2379.4			2622.8			6808		
C(xp)	187			455			11381		
C(xr)	764.7			722.0			0		
U-Clearance time	3:00 PM	4:20 PM	5:40 PM	3:00 PM	4:20 PM	5:40 PM	3:00 PM	4:20 PM	5:40 PM
Probability of clearance	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
SV(yip)	61	114	0	345	364	110	10117	6247	1928
SV(yh)+SV(xs->yp)	595.1	207.8	19.7	446.6	68.7	0.0	0.0	0.0	0.0
GD(u)	126	73	187	110	91	345	1264	5134	9453
AD(u)	169.6	556.9	745.0	275.4	653.3	722.0	0.0	0.0	0.0
C(u)	804.4	2300.6	3166.9	1211.6	2704.0	3232.8	1264.0	5134.0	9453.0
E[cost]	2379.4			2622.8			6808.0		
E[GD]	143.3			250.3			6808.0		
E[AD]	559.0			593.1			0.0		
n(xp=1)	65			73			200		
n(yp=1)	148	77	65	150	84	73	200	200	200
n(xs=1)	135			127			0		
n(yh=1)	41	55	10	0	0	0	0	0	0
n(ys=1)	11	68	125	50	116	127	0	0	0
n(xs->yp)	83	12	0	77	11	0	0	0	0
Running Time (sec)	283			226			428		

Table 1: Computational Results

V. CONCLUSIONS AND FUTURE WORK

In this paper, we have defined the basics of a stochastic optimization model for simultaneously making ground delay and reroute decisions in response to en route airspace congestion. We have also given the results of an initial computational experiment. Future steps should include more computational experiments and model refinements aimed at improving the computational performance of the integer program and at exploring the changes in airspace planning the model provides. We anticipate the need to provide many refinements and extensions to this model to better address practical problem solving. Further, another vital direction is the development of strategies necessary to embed this model within CDM processes necessary for the delivery of practical air traffic flow management solutions.

VI. ACKNOWLEDGEMENTS

The authors gratefully acknowledge the support of the National Aeronautics and Space Administration Airspace Systems Program under ARMD NRA: NNH06ZNH001.

References

- [1] M.O. Ball, R. Hoffman, A. Odoni and R. Rifkin, "A Stochastic Integer Program with Dual Network Structure and its Application to the Ground-Holding Problem", *Operations Research*, 51,167-171, 2003
- [2] G. Davidson and J. Krozel, "Strategic Traffic Flow Management Concept of Operations", Aviation Technology, Integration and Operations (ATIO), 2004.
- [3] R. Hoffman, J. Krozel, G. Davidson and D. Kierstead, "Probabilistic Scenario-Based Event Planning for Traffic Flow Management". AIAA Guidance, Navigation, and Control Conference, August 2007.
- [4] R. Hoffman, M. O. Ball and A. Mukherjee, "Ration-by-Distance With Equity Guarantees: A New Approach to Ground Delay Program Planning and Control", Proceedings of the USA/Europe Air Traffic Management R&D Seminar, ATM 2007.
- [5] B. Kotnyek and O. Richetta, "Equitable Models for the Stochastic Ground-Holding Problem Under Collaborative Decision Making", *Transportation Science*, 40, 133-146, 2006.
- [6] A. Mukherjee and M. Hansen, "A dynamic stochastic model for the single airport ground holding problem", *Transportation Science*, 41 444-456, 2007.
- [7] A. Mukherjee and M. Hansen, "A dynamic rerouting model for the air traffic flow management", *Transportation Research - Part B*, in press.
- [8] A. Nilim and L. El Ghaoui, "Algorithms for Air Traffic Flow Management Under Stochastic Environments," IEEE American Control Conference, 2004.
- [9] A. Nilim, L. El Ghaoui and V. Duong, "Multi-Aircraft Routing and Traffic Flow Management under Uncertainty", Proc. ATM 2003, Budapest, Hungary, 2003.
- [10] A. Nilim, L. El-Ghaoui, V. Duong and Hansen, M., "Trajectory-Based Air Traffic Management (TB-ATM) Under Weather Uncertainty," Proceedings of the 4rd USA/Europe Air Traffic Management R&D Seminar, Dec. 2001.
- [11] O. Richetta and A. R. Odoni, Solving optimally the static ground-holding problem in air traffic control. *Transportation Science* 27, 228-238, 1993.

Trajectory Prediction : a Functional Regression Approach

S. Puechmorel

ENAC Dept. MI LMA

7, Avenue Edouard Belin

31055 TOULOUSE FRANCE

email: puechmor@recherche.enac.fr

D. Delahaye

ENAC Dept. MI LMA

7, Avenue Edouard Belin

31055 TOULOUSE FRANCE

email: delahaye@recherche.enac.fr

L. Boussouf

INSA

157, Avenue de Rangueil

31400 TOULOUSE FRANCE

email: lboussou@etud.insa-toulouse.fr

Index Terms—Functional regression, Functional data, Trajectory Prediction

Abstract—Accurate trajectory prediction is an important issue for decision support tools in the field of ATM. This paper presents a new approach that considers trajectories as points in a functional space. By finding an expansion of observed trajectories on a suitable basis and truncating the expansion to a finite number of terms, standard regression algorithms can be used. Within this framework, full segments of trajectories can be forecasted up to 10-15 minutes.

I. INTRODUCTION

Functional data analysis is an active branch of statistics in which relevant objects are mappings belonging to a well defined space, most of the time an Hilbert space. It has been proven very efficient for problems where preserving the functional nature of data is of great importance : curves classification, functional dependence learning and similar problems [15]. In the recent literature an increasing attention has been paid to linear functional regression [2],[6] and some of its generalization [23],[14]. In this setting, either a scalar value or a mapping (the *response*), possibly contaminated by an independent measure noise is assumed to be linearly dependent on a mapping (the *predictor*). In the functional framework, the equivalent of the slope coefficient in the classical finite dimensional linear model is a kernel function that has to be estimated. Solving the associated least square problem leads to the well known Wiener-Hopf equation that generally admits no unique solution. One of the main issue in functional regression is thus to add some extra assumption on the regressor kernel so that the original ill-posed problem can be solved. On the other hand, in the field of ATM, the need for accurate trajectory predictor has appeared as a prerequisite for Decision Support Tools (DST). Air traffic management research and development has produced a substantial collection of decision support tools that provide automated conflict detection and resolution [4], [1], [22], trial planning [10], controller advisories for metering and sequencing [20], [3], traffic load forecasting [11], [9], weather impact assessment [8], [19], [5]. The ability to properly forecast future aircraft trajectories is central in many of those decision support tools. As a result, trajectory prediction (TP) and the treatment of trajectory prediction uncertainty continue as active areas of research and

development (eg [17], [21], [12], [16], [18]). In this paper we will present an innovative approach based on functional regression for solving the short to mid-term trajectory prediction (TP) problem. Long-term prediction is yet beyond the scope of this study, but considering a database of trajectories and taking into account intents of aircraft through flight plans may allow an extension of our methodology to encompass it. The first part of the paper will be devoted to a short compendium of available trajectory goodness-of-fit metrics, then the main results on functional regression will be exposed, with the potential applications and improvements of existing algorithms for the specific trajectory prediction problem.

II. TRAJECTORY PREDICTION METRICS

When an aircraft flies from a city A to a city B, it has to be managed by air traffic controllers in order to avoid collisions with others aircraft. Everyday, about 8000 aircraft fly in the French airspace, inducing a huge amount of control workload. Such a workload, is then spread by the mean of the airspace sectoring. The airspace is divided into geometrical sectors, each of them being assigned to a controller team. When a conflict between two (or more) aircraft is detected, the controller changes their routes (heading, speed, altitude) in order to keep a minimum distance between them during the crossing. All flying aircraft are then monitored during their navigation and so from the departure till the destination.

When a controller observes its traffic on the radar screen, he tries to identify convergent aircraft which may be in conflict in a near future, in order to apply manoeuvres that will separate them. The problem is to estimate where the aircraft will be located in this near future (5-10-20 minutes); this process is call trajectory prediction. This prediction may be also very useful in order to estimate the workload level in control sector to prevent over capacity event. As a mater of fact, it is very useful to estimate when an aircraft will enter a sector in order to compute the associated sector workload and to apply regulation if necessary. When a sector is expected to be overloaded, the aircraft involved in such a process will be speeded up or slow down by the controller in order to adapt the demand to the actual capacity as much as possible.

The trajectory prediction depends mainly on the residual noise after filtering which are the weight of the aircraft, the

temperature and the wind. The residual noise is integrated with time with a growing covariance matrix indicating that the estimated position is less and less accurate. The weight of the aircraft is relevant in the flight dynamics model but is still a raw data. The engines of aircraft are sensitive to the air temperature and such a data is very useful to model the trust of the aircraft but it is also very difficult to measure on real time. Finally, the wind influences strongly the cinematic of the aircraft and limits also the trajectory prediction. Based on the available accuracy, the actual limit of the trajectory prediction is about 15 minutes for the conflict detection. It means that after 15 minutes the uncertainty is so big that the estimated position is no more useful for such application.

One of the issues in trajectory prediction is to measure how accurately a model will fit to a target trajectory. Unfortunately, many different metrics can be proposed, each of them focusing on a specific aspect of accuracy. Most of the time, the proposed metrics fall into one of these categories [13] :

- Time coincidence. The time difference between a predicted event and a real event is used as a measure of TP accuracy. Time coincidence is relevant in applications where synchronizing is important, like sequencing traffic, or when the DST uses time information to instruct controller about the order in which actions have to be taken.
- Spatial coincidence. Similar to the previous one except that spatial distance at specified time (or more generally at events that can be predicted with the knowledge of aircraft positions up to a given time) between the model and the real aircraft is computed. Spatial coincidence can be refined by further splitting into altitude and horizontal error. Furthermore, for some applications, mainly conflict predictors and/or solvers, spatial difference is projected onto a vector normal to the real trajectory (cross-track error) and onto a vector tangent to the real trajectory (along-track error).
- 4D coincidence. Trajectories are considered to be 4D curves, and distance between such curves is computed. Most of the metrics derived for spatial coincidence can be extended to the 4D setting, with the benefit of including a kind of time coincidence, thus generalizing in some sense the previous two aspects.
- Morphological similarity. Different in nature from the previous metrics, an intrinsic distance between trajectories considered as curves in a 3D space can be derived from Riemannian geometry. Since only the shape of the trajectory is taken into account, this metric is relevant mainly for trajectory design tools.

Except for the last one, all those basic metrics can be integrated along trajectories to produce a mean value indicator (the classical L^2 distance is for example obtained by integrating the standard spatial coincidence metric over time). Within the frame of functional regression, the standard choice is to consider L^2 distance as goodness-of-fit measure. In the following, we will use this spatial coincidence metric along

with a specific 4D distance. Investigation of different kind of accuracy evaluations is planned for future work.

III. FUNCTIONAL REGRESSION

A. The functional nature of the trajectory prediction problem

An aircraft trajectory is by definition a mapping from a time interval $[a, b]$ to the space \mathbb{R}^3 (sometimes, it is convenient to add speed, so that the resulting expanded state space is \mathbb{R}^6). Such a trajectory is indeed the observed result of a complex evolution process that involves flight dynamics, external actions (pilot, ATC) and atmospheric factors (wind, temperature ...). The complete description of the trajectory using all these factors is generally not possible, because many influencing factors are unknown (aircraft mass, local wind, etc ...), so a less accurate but tractable model has to be chosen. For the purpose of short term prediction, a linear controlled model is accurate enough. The main assumption made is that the derivative of acceleration is zero (in a weak sense, since in most models commands are piecewise constant functions). Based on this observation, we will focus on trajectories belong to the Sobolev space of square integrable mappings with square integrable derivatives (in distributional sense) up to order 3. From now, we will assume that all trajectories belong to this space.

B. Linear regression with a functional predictor

The linear functional regression problem can be stated as follows :

- The predictor and the response are square integrable mappings from respective compact time intervals $[a, b]$, $[c, d]$ to \mathbb{R}^n (resp. \mathbb{R}^m).
- The data set consists of pairs $(X_i, Y_i)_{i=1 \dots N}$ of predictor/response. It is assumed that the X_i, Y_i are sample trajectories of two underlying smooth Hilbert random processes (for a general account on these processes, see [7]), with unknown smooth mean μ_X, μ_Y and covariance kernels B_X, B_Y .
- The functional linear model on the predictor X has the general form :

$$\hat{Y} : t \mapsto f(t) + \int_{[a,b]} K(s,t)X(s)ds$$

with $f : [c, d] \rightarrow \mathbb{R}^n$ a smooth square integrable mapping and $K : [a, d] \times [c, d] \rightarrow M_{m,n}(\mathbb{R})$ a smooth square integrable (m, n) -matrix valued kernel.

- The solution of the functional regression problem is the optimal couple (f, K) that minimize the mean square error between Y and \hat{Y} .

Most of the time, the related literature on the subject addresses the problem with $n = 1$, that is for real valued trajectories. In our setting, this clearly means that it enforces the fact that the x, y, z components of a trajectory can be treated as independents scalar valued mappings. Although we will see later that functional regression models satisfying some invariance properties must fall into this category, there is no reason to limit ourselves to kernels with values in the

set of diagonal matrices. It worth notice too that the basic theory makes the assumption that trajectories are continuously observed, which is clearly not the case of ATC data. The reference [23] extends the least square criterion to irregularly spaced samples on predictor and response, providing us with the right framework for trajectory prediction applications and will be the starting point of our work. The solution of the functional regression problem is known to satisfy a Wiener-Hopf equation :

$$E[XY] = \int_{[a,b]} K(s,t)B_x(s,t)ds \quad (1)$$

Unfortunately, this equation has generally not a unique solution. Furthermore, solving 1 from sampled trajectories yields an even more ill-posed problem. Solutions to this problem, mainly by using regularization, have been proposed in [15]. Most of the time, an expansion of the predictor on an Hilbert basis is used to solve the functional regression (note that at some point of the process, all expansion must be truncated in order to obtain something computable). Several choices exist for such a basis. In many papers it is recommended to use Karhunen-Loeve expansion, with the eigenfunctions of the covariance operator as basis :

$$\psi_i(t) = \lambda_i \int_{[a,b]} B_X(s,t)\psi_i(s)ds$$

with λ_i the eigenvalue corresponding to ψ_i . Since the covariance and mean functions are unknown, there are to be estimated from the data. The procedure used in [23] is to use a weighted sum kernel approximation. A complete treatment can be found in [24]. It has to be noted that this particular choice is essentially heuristic : since the basis functions depend only on the predictor and not on the response, it cannot be guaranteed that the first q eigenfunctions, associated to the q largest eigenvalues, are the q most predictive (the integer q represents the truncation index used when solving the functional regression problem).

C. Solving the functional regression

The framework in this part will be the one chosen in [23]. Under the general assumptions of the previous section, we will further assume that the data set consists in a finite number of sparsely sampled predictor/response pairs. Let X_i (resp. Y_i) be the realization of predictor process X (resp. response process Y) corresponding to observation i in the data set. Let M_i (resp. N_i) be the number of samples available for this observation and let $X_{i,j}$, $j = 1 \dots M_i$ (resp. $Y_{i,j}$, $j = 1 \dots N_i$) be the actual samples along trajectories X_i (resp. Y_i) with corresponding sample times $\tau_{i,j}$ (resp. $\tilde{\tau}_{i,j}$). The number of samples M_i, N_i and the sampling times are assumed to be random variables independent from the processes X, Y . As mentioned before, the first step towards solving the regression problem is to find an expansion of the predictor and the response on respective (infinite but countable) basis, respectively $(\phi_i)_{i \in \mathbb{N}}$ and $(\psi_i)_{i \in \mathbb{N}}$. A widely used procedure in the field of functional data analysis is to find a smooth approximation

to covariance function of X (resp. Y) then find estimators of the eigenvalues/eigenvectors of the covariance operators (such procedure is known as Functional Principal Component Analysis FPCA). As mentioned before, it is not guaranteed that this will result in an optimal representation for regression purpose, but it has proved quite efficient and robust in practice. However, finding the eigenvalue/eigenvectors from the empirical covariance function obtained from the measurements is quite a lengthy process. First of all, a smooth estimator of mean and covariance function has to be obtained. A first approach is to use a local linear smoother. Given a kernel K and a bandwidth parameter h , the local linear smoother for the (t, X) scatterplot is obtained by minimization over a, b of :

$$\sum_{i=1}^n \sum_{j=1}^{M_i} K\left(\frac{t - \tau_{ij}}{h}\right) (X_{ij} - a - b(t - \tau_{ij}))^2$$

with t being a fixed time. The optimal values a, b obtained for a fixed t give a local linear model, so that the estimated mean at time t is $\hat{\mu}_X(t) = a$. For covariance estimation, the procedure is roughly the same, but instead of considering the samples X_{ij} of X , the empirical covariance function :

$$C'_X(\tau_{i,j_1}, \tau_{i,j_2}) = (X_{i,j_1} - \hat{\mu}_X(\tau_{i,j_1}))(X_{i,j_2} - \hat{\mu}_X(\tau_{i,j_2}))$$

is used. The corresponding local linear model is bivariate and corresponds to the minimum over a, b, c of :

$$\sum_{i=1}^n \sum_{j=1}^{M_i} \sum_{k=1, k \neq j}^{M_i} K\left(\frac{t - \tau_{ij}}{h}, \frac{s - \tau_{ik}}{h}\right) (C'_X(\tau_{i,j}, \tau_{i,k}) - a - b(t - \tau_{i,j}) - c(s - \tau_{i,k}))^2 \quad (2)$$

The estimated covariance $\hat{C}_X(t, s)$ at time t, s is thus the term a in the previous expression. Several packages exist for such estimation. For our purpose, an ad-hoc algorithm has been developed : the general principle will be described below.

It must be noted that local linear smoother is not the only usable procedure for fitting a smooth curve or surface to scatterplot data. Smoothing splines seem to be a good choice too. The difference in terms of performance between those two approaches has not been investigated yet, but at first sight there is no evidence why the first one will be better : it has been chosen in our application only because of its availability and some asymptotic approximation results.

Next step is to estimate eigenvalues and eigenvectors of the covariance function by solving the functional equation :

$$\int_{[a,b]} \hat{C}_X(t,s)\psi(s)ds = \lambda\psi(t)$$

Several numerical procedures can be found in the numerical analysis literature for solving such problems. We have applied a Nyström method on a regular grid for the trajectory application (see details in the next section). The result is two finite sets of pairs eigenvalues/eigenvectors $(\lambda_i, \psi_i)_{i=1 \dots P}$ and $(\mu_i, \theta_i)_{i=1 \dots Q}$ for X and Y respectively. The number of representing functions (the P and Q integers) has to be chosen

on either a leave-one-out or an AIC : the later gives good results in the trajectory case.

Given the eigenvalues/eigenfunctions pairs for X and Y respectively, it is possible to compute an estimate of the covariance of X and Y components respectively :

$$\hat{\sigma}_{XY}(i, j) = \int_{[c,d]} \int_{[a,b]} \theta_j(s)\psi_i(t)\hat{C}_X(t, s)dtds$$

The resulting optimal kernel solving (approximately) the functional regression problem is then obtained as :

$$K(t, s) = \sum_{i=1}^P \sum_{j=1}^Q \frac{\hat{\sigma}_{XY}(i, j)}{\lambda_i} \psi_i(t)\theta_j(s)$$

IV. APPLICATION TO TRAJECTORY PREDICTION

Applying functional regression to trajectories implies :

- Extending all previous estimators to vector valued ones (thus replacing the covariance function by a 3×3 matrix valued function).
- Find the right predictor and response.

A. Principal components in the vector case

Recall that the chosen representation basis is obtained by functional principal component analysis. For trajectory prediction purpose, all random processes have values in \mathbb{R}^3 , so that canonical procedures have to be extended. Estimation of mean and covariance functions can be used readily since the proposed local linear estimator extends componentwise to the 3-dimensional case. It should be noted however that computing the mean function involve 3 times more computation than for the scalar case and computational task is scaled by a 6 factor for the covariance (due to the symmetry of the covariance matrix). An important step in the design of a linear smoother is the choice of weighting kernel and bandwidth. The problem has been addressed in the field of non parametric statistics and it is known that the kernel has less influence than the bandwidth. The Epanechnikov kernel :

$$K_e(t) = \frac{3}{4}(1-t^2)1_{[-1,1]}(t)$$

has some interesting optimality properties and is easy to compute. Another choice is the Gaussian kernel :

$$K_g(t) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{t^2}{2})$$

For very fast computation, it is still possible to use a uniform kernel :

$$K_u(t) = \frac{1}{2}1_{[-1,1]}(t)$$

Since the data set is usually large (around 1000 trajectories sampled at 10s), a compactly supported kernel in the local linear smoother allows a reduced computational load and an complexity mostly independent of the number of samples in a trajectory. The Gaussian kernel is not compactly supported, but decreases very fast at infinity so that practically it can be set to 0 outside an compact interval. The optimal bandwidth can be found in the limit of large samples. If the kernel is

K , and the target function f is supported in the interval $[a, b]$ then the asymptotically optimal bandwidth is :

$$h^5 = \frac{\sigma^2(b-a)^2 \int K^2(x)dx}{N (\int x^2 K(x)dx)^2 \int_a^b f''^2(x)dx}$$

with σ^2 the variance of the noise, N the number of samples. Since f is unknown, one has to estimate both σ^2 and f'' . Since f has to be estimated, it is clear that only some rough guess on plausible value can be done. A first approach is simply to increase the order of the model as :

$$\sum_{i=1}^n \sum_{j=1}^{M_i} K\left(\frac{t-\tau_{ij}}{h}\right) (X_{ij} - a - b(t-\tau_{ij}) - c(t-\tau_{ij})^2)^2$$

the c coefficient will then give an estimate of $f''(\tau_{ij})/2$. This method is appealing for trajectory modelling since curvature can be obtained directly from c . However, experiments on real data show that little is gained since the second order model itself has a bandwidth that must be set heuristically. A second approach is simply to compute using a finite difference operator. It works surprisingly well for such a naive approach, probably because noise is not dominant in ATM data. Assuming that the bandwidth is known, solving for the parameters of the local linear model can be done easily. For a given vector or matrix valued function X sampled at times τ_{ij} with values A_{ij} , the mean value at time t can be obtained using the following formula :

$$\hat{X}(t) = -dm_X + ef_X$$

with :

$$d = \sum_{i=1}^n \sum_{j=1}^{N_i} K_e\left(\frac{t-\tau_{ij}}{h}\right) (t-\tau_{ij})^2$$

$$e = \sum_{i=1}^n \sum_{j=1}^{N_i} K_e\left(\frac{t-\tau_{ij}}{h}\right) (t-\tau_{ij})$$

and :

$$m_X = \sum_{i=1}^n \sum_{j=1}^{N_i} K_e\left(\frac{t-\tau_{ij}}{h}\right) X_{ij}$$

$$f_X = \sum_{i=1}^n \sum_{j=1}^{N_i} K_e\left(\frac{t-\tau_{ij}}{h}\right) (t-\tau_{ij})X_{ij}$$

Since K_e is compactly supported (or approximately such), the inner sum in the previous expression will only involve a number of terms bounded by a constant so that the computational task for computing $\hat{X}(t)$ at a given t scales linearly with the number of trajectories in data set but independent from the size of the trajectories themselves. Once the covariance functions have been obtained for both the predictor and the response, the eigenvalues/eigenfunctions are computed by solving an homogeneous Fredholm equation of the second kind :

$$\int_{[a,b]} C(s, t)\phi(s)ds = \lambda\phi(t)$$

with C the covariance function of interest. Several numerical methods can be found in the literature for such a problem. In our setting, the Nyström method based on Gauss-Legendre quadrature formula has proved very well fitted. The method proceeds by approximating the integral operator as :

$$\sum_{i=1}^n w_i C(s_i, t) \phi(s_i)$$

with w_i and s_i respectively the weights and the abscissa of the Gauss-Legendre quadrature approximation. For n points, the abscissa are the roots of the n -th Legendre polynomial P_n while the weights are computed as :

$$w_i = \frac{2}{(1 - s_i^2) P'_i(s_i)^2}$$

with the Legendre polynomials P_i evaluated by the recursion :

$$(i + 1)P_{i+1} = (2i + 1)xP_i - iP_{i-1}$$

with conventionally $P_{-1} = 0, P_0 = 1$. The Fredholm equation of the second kind is approximated by the following eigenvalue problem :

$$\sum_{i=1}^n w_i C(s_i, s_j) \phi(s_i) = \lambda \phi(s_j), j = 1 \dots n$$

The result is a set of eigenvectors $\phi(s_j, i), j = 1 \dots n, i = 1 \dots n$ and eigenvalues λ_i . The approximation to the eigenfunction of the original covariance operator associated to eigenvalue λ_i is then :

$$\phi_i(t) = \frac{1}{\lambda_i} \sum_{i=1}^n w_i C(s_i, t) \phi(s_i)$$

(the case $\lambda_i = 0$ will not occur in our application). Since our original processes X, Y take their values in \mathbb{R}^3 , all previous equations are to be taken as vector ones. The main consequence is that the approximate eigenvalue has to be solved with a full $3n \times 3n$ system. Fredholm equation solving is quite a critical step in the overall algorithm, so attention has been paid to its accuracy. In fact, Gaussian quadrature is not the only procedure that can be used, but any quadrature formula based on samples will work. A comparison has been made between low accuracy rectangle method for approximating the integral and the complete Nyström algorithm with Gauss-Legendre quadrature. For that purpose, the test set has been obtained by generating trajectories of a simple random process (namely a square root function with an additive gaussian noise and a random scaling).

The error obtained with the two methods on the test set is, as a function of the number of eigenfunctions : Nyström algorithm appears to be more accurate than rectangle quadrature, at least for small number of eigenfunctions. However, method tends to yield ill conditioned matrices when a large number of eigenfunctions is required : this phenomenon explains why Nyström is outperformed by rectangle quadrature in such cases. The difference between the two procedures remains anyway quit low, indicating that a high order quadrature formula is not of great importance for our purpose.

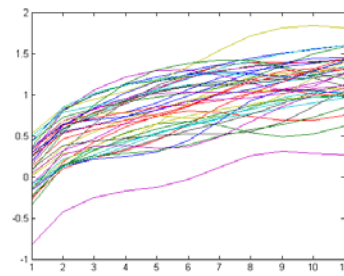


Fig. 1. Test functions for Fredholm equation

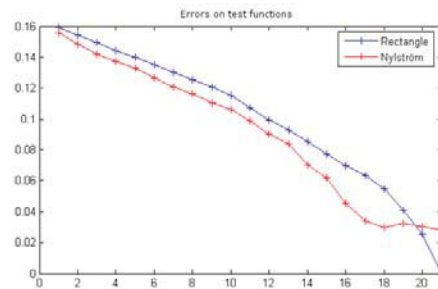


Fig. 2. Error comparison

B. Predictor

Finding the right predictor is a critical task in applying functional regression. For trajectory prediction purpose, it is natural to consider a part of the observed trajectory as the predictor, and a part of the future trajectory as target. The learning database has thus been chosen by selecting homogeneous segments of 20 radar plots from a day of traffic, then for each segment cut into two 10 plots pieces. The first piece will be used as predictor and second one as target. A total of 3200 trajectories has been considered, with a final database of 100 segments. In the traffic, a test database with the same number of segments and similar characteristics has been selected too. Since the random process associated to trajectories has no reason to be stationary, the Karhunen-Loeve basis is a priori different for the predictor and the target. These two basis will be denoted respectively by $(\phi_i)_{i \in \mathbb{N}}$ and $(\psi_i)_{i \in \mathbb{N}}$. Let $(X_k, Y_k)_k$ be the k -th sample from the learning base (that is X_k is the first half on segment k while Y_k is the second half), the regression problem is to find an optimal K such that :

$$\sum_{k=1}^N \int \|Y_k(t) - \int K(t, s) X_k(s) ds\|^2 dt$$

is minimal. The kernel K can be expressed using basis $(\phi_i)_{i \in \mathbb{N}}, (\psi_i)_{i \in \mathbb{N}}$ as :

$$K(t, s) = \sum_i \sum_j K_{ij} \psi_i(t) \phi_j(s)$$

Using orthogonality of the Karhunen-Loeve eigenfunctions, the original problem is reduced to find optimal an optimal

sequence (K_{ij}) minimizing :

$$\sum_{k=1}^N \int \left\| \sum_i \psi_i(t) \left(b_{ik} - \sum_j K_{ij} a_{jk} \right) \right\|^2 dt$$

assuming that the expansions of X_k, Y_k are :

$$X_k(s) = \sum_j a_{jk} \phi_j(s), Y_k(t) = \sum_i b_{ik} \psi_i(t)$$

using the orthonormality of $(\psi_i)_{i \in \mathbb{N}}$, the minimization problem reduces further to :

$$\min_{(K_{ij})} \sum_{k=1}^N \sum_i \left(b_{ik} - \sum_j K_{ij} a_{jk} \right)^2$$

In practice, expansions are truncated to a fixed rank. Let P be the corresponding integer. The approximate finite dimensional minimization problem is :

$$\min_{(K_{ij})} \sum_{k=1}^N \sum_i^P \left(b_{ik} - \sum_j^P K_{ij} a_{jk} \right)^2$$

which is nothing but a linear least mean square problem that can be solved with the help of normal equations or QR factorization. The result of prediction for first the test functions (same as in previous sections) is summarized below :

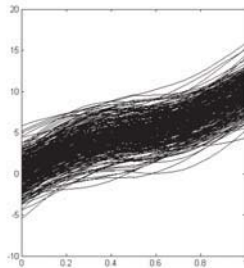


Fig. 3. Test functions for prediction

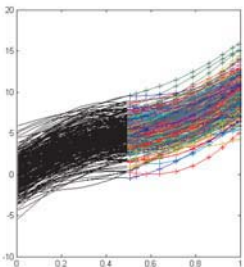


Fig. 4. Predicted second half of trajectories

In the case of real traffic, the obtained results are : The relative error for in the horizontal plane is kept low and increases as expected with prediction time. In the case of

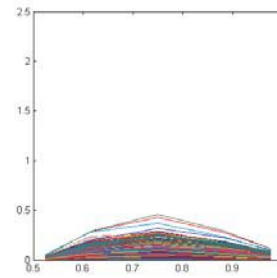


Fig. 5. Relative prediction error

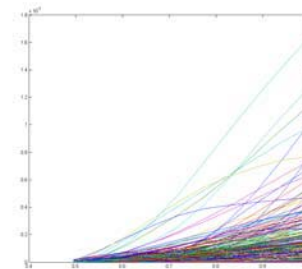


Fig. 6. Relative prediction error (x-y plane)

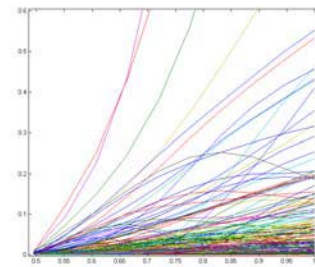


Fig. 7. Relative prediction error (z component)

z component, relative error is high, but measurements are discrete (flight levels), thus there is an intrinsic noise coming from quantization. For this component, prediction is indeed built-in with smoothing : a comparison with a smoothed trajectory yields much lower error.

C. Clustered regression

From now, only the case of trajectories originating from the same underlying stochastic process has been considered. While several operational situations fall in this category (especially when the predictor consists of small parts of trajectories), it is known that this assumption is false when applied to large areas of the airspace. To deal with this problem, it may be necessary to introduce a cluster regression. The data set of trajectories is partitioned into homogeneous classes (clusters) based on a relative distance criterion (most of the time, a L^2 norm or a Sobolev norm is used). Once the clustering has been done,

functional regressors are computed cluster by cluster, assuming that all trajectories in the same cluster are close enough to be modeled as samples of the same stochastic process. When applying functional regression to a new trajectory, the closest cluster is chosen, then the corresponding regressor is used. This way of applying functional regression has proved to be very efficient for inhomogeneous 1D curves ; however, application to 3D trajectories has not been done yet.

V. CONCLUSION AND FUTURE WORK

The functional regression is a promising approach on simulated situations. Besides of producing an estimation of the future positions of aircraft in the short to mid term prediction range, it is possible to derive confidence regions for the actual position, thus yielding a better control on the quality of the produced solution. The computational task involved is heavy, but has to be done only once : as soon as the Karhunen-Loeve has been produced, it can be used at low cost. The future work will be first experiments on selected learning data sets so that functional regression can be compared with other methods. A second aspect will be to introduce other kind of decompositions (namely wavelets and curvlets basis) that are known to perform much like Karhunen-Loeve but at a much lower computational cost and to investigate the cluster functional regression. Furthermore, a functional regression software adapted to trajectory prediction is currently under development.

REFERENCES

- [1] D.J. Brudnicki and A.L. McFarland, *User request evaluation tool (uret) conflict probe performance and benefits assessment*, FAA/Eurocontrol ATM Seminar Saclay France, 1997.
- [2] H. Cardot, F. Ferraty, and P. Sarda, *Functional linear model*, Statist. Probab. Letters **45** (2003), 11–22.
- [3] R. Coppenbarger, R. Lanier, D. Sweet, and S. Dorsky, *Design and development of the enroute descent advisor (eda) for conflict-free arrival metering*, AIAA-2004-4875 AIAA GNC Conference Providence RI, 2004.
- [4] H. Erzberger, R.A. Paielli, D.R. Isaacson, and M.M. Eshowl, *Conflict detection in the presence of prediction error*, FAA/Eurocontrol ATM Seminar Saclay France, 1997.
- [5] J. Evans and al, *Reducing severe weather delays in congested airspace with weather support for tactical air traffic management*, FAA/Eurocontrol ATM Seminar BUdapest Hungary, 2003.
- [6] F. Ferraty and P. Vieu, *Functional nonparametric statistics : A double infinite dimensional framework*, North-Holland, 2003.
- [7] I.I. Gikhman and A.V. Skorokhod, *Introduction to the theory of random processes*, Dover, 1996.
- [8] D.B. Kirk and al, *Problem analysis resolution and ranking (parr) development and assessment*, FAA/Eurocontrol ATM Seminar Santa Fe NM, 2004.
- [9] A. Masalonis and al, *Using probabilistic demand prediction for traffic flow management decision support*, AIAA-2004-5231 AIAA GNC Conference Providence RI, 2004.
- [10] B.D. McNally, R.E. Bach, and W. Chan, *Field test evaluation of the ctas conflict prediction and trial planning capability*, AIAA-1998-4480 AIAA GNC Conference Boston MA, 1998.
- [11] C. Meckiff, R. Chone, and J.P. Nicolaon, *The tactical load smoother for multi-sector planning*, FAA/Eurocontrol ATM Seminar Orlando FL, 1998.
- [12] S. Mondoloni, S.M. Paglone, and S. Green, *rajectory modeling accuracy for air traffic management decision support tools*, ICAS Congress Toronto ON, 2002.
- [13] S. Mondoloni, Swierstra S., and M. Paglione, *Assessing trajectory prediction performane – metrics definition*, 24th Digital and Avionics Systems Conference, 2005.
- [14] H. G. Müller and U. Stadtmüller, *Generalized functional linear models*, Annals of statistics **33** (2005), no. 2, 774–805.
- [15] J. Ramsay and B. W. Silverman, *Functional data analysis*, Springer-Verlag, 1997.
- [16] H.F. Ryan, M. Paglione, and S. Green, *Review of trajectory accuracy methodology and comparison of error measure metrics*, AIAA-2004-4787 AIAA GNC Conference Providence RI, 2004.
- [17] R. Slattery and Y. Zhao, *Trajectory synthesis for air traffic automation*, AIAA Journal Guidance Control and Dynamics, vol. 20, 1997, pp. 232–238.
- [18] S.Mondoloni and I. Bayraktutar, *mpact of factors, conditions and metrics on trajectory prediction accuracy*.
- [19] V. Sud and al, *Air traffic flow management collaborative routing coordination tools*, AIAA-2001-4112 AIAA GNC Conference Montreal PQ, 2001.
- [20] H.N Swensen and al, *Design and operational evaluation of the traffic management advisor at the forth worth air route traffic control center*, FAA/Eurocontrol ATM Seminar Saclay France, 1997.
- [21] S. Swierstra and S. Green, *Common trajectory prediction capability for decision support tools*, FAA/Eurocontrol ATM Seminar BUdapest Hungary, 2003.
- [22] A. Vink, *Eatchip medium term conflict detection: Part 1 eatchip context*, FAA/Eurocontrol ATM Seminar Saclay France, 1997.
- [23] F. Yao, H. G. Müller, and J. L. Wang, *Functional linear regression for longitudinal data*, Annals of statistics **33** (2005), no. 6, 2873–2903.
- [24] ———, *Functional linear regression for sparse longitudinal data*, Journal of Amer. Statist.Assoc. **100** (2005), 577–590.

The robust flight level assignment problem

Olivier Klopfenstein^(1,2)

(1) France Telecom *R&D*
38-40 rue du Général-Leclerc
92794 Issy-les-Moulineaux Cedex 9, France
olivier.klopfenstein@orange-ftgroup.com

Dritan Nace⁽²⁾

(2) Laboratoire Heudiasyc UMR CNRS 6599
Université de Technologie de Compiègne
60205 Compiègne Cedex, France
nace@utc.fr

Abstract—This paper studies the robust flight-level assignment problem. Our goal is reducing the cost (and more specifically the delay) induced by airspace congestion through an appropriated flight level assignment (FLA) taking account of uncertainties. We investigate a robust optimization framework inspired by Bertsimas and Sim work for linear programs and propose appropriate models for the robust flight level assignment problem.

Keywords: Flight-level assignment, Robustness, Linear programming.

I. INTRODUCTION

Alleviating delays caused by airspace congestion is, and will continue to be, critical to the operation of the European air traffic control system. Two kinds of congestion can be identified corresponding to two different areas of airspace: terminal congestion (around airports) and en-route congestion (between airports). We will focus on congestion in the airspace rather than at airports, and we are interested here in a specific direction involving flight-level optimization with respect to a given traffic demand and given routes. In other words, our goal is to reduce the additional cost (including delay) induced by conflict resolution procedures through a better assignment of flight levels. Indeed, in case of en-route conflicts, some aircraft has to be rerouted, which produces some delay. This delay can be reduced through increasing the speed of aircraft, which yields additional energetic cost. We typically consider here the energetic cost due to conflict resolution, called simply conflict or energetic cost in the remainder of the paper. With respect to the FLA problem, we restrict ourselves to only three possible levels for each flight. Despite this restriction, the problem remains highly combinatorial due to the large number of simultaneous flights. The flight level assignment problem is shown to be NP-complete in the strong sense [2], which makes it hard to solve at optimum even for reduced size instances. The problem becomes rather more difficult when involving the uncertainties in ATM. More precisely, an important question that we raise in this paper is how to include the potential en-route conflicts associated with each aircraft in the model and take into account the uncertainties related to it. All this leads to the robust flight-level assignment problem and the associated mathematical model, which is the main focus of this work.

A. Related works

Optimization problems in ATM have been widely studied, and we do not intend to mention all of them. We prefer

to focus on some work related somehow to the flight-level assignment problem. Let us first cite Bertsimas and Stock [7], [8] who have looked at the Air Traffic Flow Management Rerouting Problem (TFMRP), considering simultaneously the time and the route assignment problem through a deterministic approach. First in [7], they handle the Air Traffic Flow Management Problem (TFMP) with En-route Capacities, and then in [8] they show how to optimally control aircraft by rerouting, delaying, or adjusting the speeds of the aircraft in the ATC system to avoid airspace regions with reduced capacities due to weather conditions. Delahaye and Odoni in [11], study the problem of airspace congestion from the stochastic optimization point of view and propose a genetic algorithm. Barnier and Brisset (see [10]), consider the problem of level assignment while using an ideal sector-less environment. The main idea is to allocate different flight levels to intersecting routes in order to avoid conflicts. A straight line between an origin and destination pair represents the path of a flow of flights between these two airports; in other words, only direct routes are considered. Then, if two flows are in conflict, they must be routed on two different levels. The problem becomes a graph coloring one: given a graph with a set of vertices and a set of edges, the problem is to color the edges such that any two intersecting edges (not at their extreme vertices) have two different colors, and the number of colors used is the lowest possible. Some other research on this problem, also based on the graph coloring problem is presented in Letrouit's thesis (see [16]). The route assignment problem here is handled by several tasks. The first task is minimizing the number of required levels when assigning each route to a level from the beginning to the end of a flight, and the second task is the distribution of routes among N levels in order to minimize the number of intersections between the routes having the same level. More recently, Constans et al. (see [19]) have studied the problem from the angle of aircraft speed modification. They propose minimizing conflict risks by dynamically imposing feasible modifications on the speeds of the aircraft. Doan et al. (see [9]) have presented a deterministic model intended to optimize route and flight-level assignment in a trajectory-based ATM environment. The aim of the latter study is to address the problem of airspace congestion, and in particular to reduce the number of potential en-route conflicts. This work was the starting point for the study presented here.

Let us recall that our goal is reducing the cost induced by

potential en-route conflicts. An important question is then how modeling the induced cost given conflict probabilities. Clearly, the cost induced by an en-route conflict greatly depends on the conflict resolution methods and for a majority of cases, the delay is asymmetrically distributed to the involved aircraft. Due to uncertainties, it is not possible to determine in advance if some conflict will occur and which aircraft will be delayed. All this justifies the need for a robust flight level assignment as depicted below.

B. Paper organization

This paper is organized as follows. After this introduction, in Section II we recall briefly works on robust linear programming (LP). In Section III, we focus on the FLA problem. We present binary linear programming models and in Section IV we discuss its robust versions. Section V is devoted to some numerical results, including remarks on implementation and data estimation. Some concluding notes are provided in Section VI.

II. THE ROBUST BINARY LINEAR PROGRAMMING PROBLEMS

Robust optimization is one of the common approaches to take account of uncertainties in optimization problems. We refer to [18] for a survey in the context of combinatorial optimization. The usual goal of robust optimization is to find the best solution which remains feasible for a whole set of possible events. The main criticism for robust models is the so-called over-conservatism: the obtained solution will be feasible for all the possible events, regardless their occurrence probability. In practice, the worst case may impose a large cost, while being highly improbable. To remedy this disadvantage, some works have proposed to relax this worst case condition [4], [6]. As a result, the solution computed may be feasible for most of the events, but not all of them.

This is the spirit of the robust model proposed by Bertsimas and Sim [6], where the feasibility degree of a solution can be controlled. An important advantage of this model is to be easily used also with integer variables (see e.g. [5]). Indeed, the initial integer linear program (without uncertainty) is transformed into another robust integer linear program. A similar model has been proposed in [15]. The main interest of this latter approach lies in the existence of an efficient solution heuristic. Hence, it can be used on large integer linear problems.

III. THE FLIGHT LEVEL ASSIGNMENT PROBLEM

In this section we present LP based approaches for both deterministic and robust variants of the FLA problem.

A. An LP model for the FLA problem

Notation:

- L denotes the set of possible flight-levels l . We denote with L_i the set of preferred flight levels associated with flight i .

- The set of flights is noted with F . F^l groups all flights allowed to fly to level l .
- x_i^l : binary variable $(0, 1)$, takes value 1 when the flight i , fly on level l and 0 otherwise.
- b_i^l : gives the profit associated with flight i when flying on level l .
- p_{ij} : gives the cost penalty associated with aircraft i when resolving a potential conflict with aircraft j . When dealing with the robust variant, it will denote a random variable associated with the additional cost that an aircraft can have due to some potential conflict.
- P_i^l : gives the admissible cumulated cost for a given flight i and level l .
- S_i^l : gives the set of flights j having a potential conflict with flight i at level l .

Given the above notation, an LP model associated with the FLA problem denoted with P is as follows:

$$\max \sum_{i \in F, l \in L} b_i^l x_i^l \quad (1)$$

$$\sum_{j \in S_i^l} p_{ij} x_j^l \leq M_i^l (1 - x_i^l) + P_i^l, \quad i \in F, l \in L_i, \quad (2)$$

$$\sum_{l \in L_i} x_i^l = 1, \quad i \in F, \quad (3)$$

$$x_i^l \in \{0, 1\} \quad i \in F, l \in L_i, \quad (4)$$

where M_i^l gives a sufficient large value, (for instance $M_i^l = \sum_{j \in S_i^l} p_{ij}$). Without loss of generality, we assume that for the given P_i^l values there exists a feasible solution for problem P .

The above model is a binary integer LP problem involving a large number of constraints and variables, which makes it hard to be solved by exact methods. From mathematical point of view, it can be seen as a specific case of a multi-dimensional multiple-choice knapsack problem. We will provide in the following the framework of an approximated method for the FLA problem. The main idea behind this method is considering the assignment problem separately for each level. There are two main bricks: the first one, **Step 0**, is devoted to maximize the number of flights assigned to their preferred level. Thus, we solve a reduced problem involving only flights with their preferred level and next, fix all assigned flights. The second brick is concerned with the remaining flights. This problem, called P^l , is slightly different to P_i^l as we use as constants (i.e., $x_i = 1$) for all already assigned flights, and we add all other flights concerned with this level that are not yet assigned. Both these problems are different to P as we do not need to force any flight to be assigned to some level, as described below. It provides the main block to construct the solution approach outlined below:

Approximated flight level assignment (ApproxFLA)

- Step 0:** Proceed with robust flight assignment separately for each level (Solve problem P^l);
Fix the level for flights already assigned.
- Step 1:** Proceed with robust flight assignment separately for each level (Solve problem P^l);
Fix the level for flights already assigned.
- Step 2:** If all flights are assigned, STOP.
Else, increase the admissible cost for each non assigned flight and return to **Step 1**.

The key element of the method is the procedure of flight assignment (P^l and P^l) associated with a given level of **Step 0** and **Step 1**. We will focus only on P^l used in **Step 0**. Before detailing the mathematical formulation, let give some precision on the notation. As there is no need to distinguish flight levels, the binary variable x_i^l is now replaced by x_i , and as before it takes value 1 when the flight i flies on level l and 0 otherwise. Respectively b_i^l and P_i^l are now replaced by b_i and P_i . For sake of simplicity we will allow ourselves to use the same notation for F^l as in P , but here it groups only flights having l as their preferred level. Notice also that the order of level examination would have an impact on the obtained solution. We propose to start with the most loaded levels. Problem P^l follows:

$$\max \sum_{i \in F^l} b_i x_i \quad (5)$$

$$\sum_{j \in S_i^l} p_{ij} x_j + M_i x_i \leq M_i + P_i, \quad i \in F^l, \quad (6)$$

$$x_i \in \{0, 1\} \quad i \in F^l. \quad (7)$$

The above model has an interesting structure as it corresponds to a *simple* multi-dimensional knapsack problem. The problem P^l can be written in a similar way except that some constraints of type $x_i = 1$ are added and F^l groups all concerned flights.

IV. MODELING AND SOLVING THE ROBUST FLA PROBLEM

Assuming separate probability conditions, the robust version of the FLA can be formulated with probability constraints as follows:

$$\max \sum_{i \in F^l, l \in L} b_i x_i$$

$$Pr\left(\sum_{j \in S_i^l} p_{ij} x_j + M_i x_i \leq M_i + P_i\right) \geq 1 - \epsilon, \quad \forall i \in F^l, l \in L.$$

Following the Bertsimas and Sim work, we can deduce the robust variant of the above ILP problem. This yields still another ILP problem, which is at least as difficult as the standard problem. All this justifies heading to approximated methods to deal with it: we will make use of the framework approximated method given above for the FLA problem,

except that we consider the robust variant of the P^l problem (called RP^l), instead.

The key element of the method is the procedure of robust flight assignment RP^l associated with a given level of **Step 0**. Let us deduce first its robust variant.

A. Modeling the RP^l problem

In the P^l we have assumed that p_{ij} are some given constants expressing the potential cost induced by some conflict involving aircraft i and j . In the following, we assume that p_{ij} are random variables that take values in an interval data already estimated. This assumption leads us to the robust version of the P^l problem, that is RP^l , and subsequently to the robust variant of the FLA problem. Naturally, some way to take into consideration the uncertainties is not to allow all en-route potential conflicts to count for the total cost estimation. We have thus a robust version of the FLA problem in the sense that for a given aircraft only a part of potential en-route conflicts are assumed to occur and expected to generate additional costs. First, let us precise the assumptions related to the robust problem RP^l and the ways used to introduce the uncertainty in the model. Following the Bertsimas and Sim works on this area, it seems natural to model the uncertainty by introducing a protection coefficient, which gives the maximum number of conflicts that can occur for a given flight. In our model we do not make use of conflict probabilities in a direct way but consider their consequences, that is the corresponding additional costs. These potential costs are modeled by intervals $[0, \bar{p}_{ij}]$, with $\bar{p}_{ij} > 0$. Resolving a conflict that involves a pair of aircraft, yields delay and hence an additional cost, non necessarily symmetrically distributed among involved aircraft. This statement leads us to the following assumption: any flight i will experience (most probably) a reduced number of potential conflicts during his time flight, (which yields additional costs to the involved aircraft) and this number (Γ_i) varies in $[0, |S_i^l|]$. Then, we are interested in "best" solutions that remain feasible for any scenario with at most Γ_i coefficients taking the worst value \bar{p}_{ij} . Such a solution is obtained through the following program:

$$\max \sum_{i \in F^l} b_i x_i$$

$$\sum_{j \in S} \bar{p}_{ij} x_j + M_i x_i \leq M_i + P_i, \quad i \in F^l, S \subseteq S_i^l : |S| = \Gamma_i,$$

$$x_i \in \{0, 1\} \quad i \in F^l.$$

The above program contains a large number of constraints and it is hard to solve at optimum. However, it has been shown by Bertsimas and Sim that it can be modeled through an ILP (Integer Linear Programming). The latter program, provided below, still contains a large number of constraints and variables and remains hard to be solved by exact methods.

Following the work of Bertsimas and Sim, the robust variant of problem P_l with respect to a given vector Γ , denoted RP_{Γ} , is as follows:

$$\begin{aligned} & \max \sum_{i \in F^l} b_i x_i \\ \Gamma_i z_i + \sum_{j \in S_i^l} \delta_{ij} y_{ij} + M_i x_i & \leq M_i + P_i, \quad i \in F^l, \\ z_i + \delta_{ij} y_{ij} & \geq \delta_{ij} x_j \quad i \in F^l, j \in S_i^l \\ x_i \in \{0, 1\}, z_i \geq 0, y_{ij} \geq 0 & \quad i \in F^l, j \in S_i^l. \end{aligned}$$

Hence, we have opted to use another (alternative) robust model of above, following the one introduced in [15]. The following model uses a parameter vector $\gamma \in [0, 1]^{|F^l|}$ instead of the vector Γ :

$$\begin{aligned} & \max \sum_{i \in F^l} b_i x_i \\ M_i x_i + \min \left\{ \sum_{j \in S_i^l} \bar{p}_{ij} x_j, \gamma_i \cdot \sum_{j \in S_i^l} \bar{p}_{ij} \right\} & \leq M_i + P_i \quad i \in F^l, \\ x_i \in \{0, 1\}, \quad i \in F^l. & \end{aligned}$$

The above model is denoted below $RP_{l\gamma}$. This formulation can be simplified a lot. Let us focus on the robust constraint i . Either we consider the worst case (maximum conflict induced costs), or we have a constraint: $M_i x_i + \gamma_i \cdot \sum_{j \in S_i^l} \bar{p}_{ij} \leq M_i + P_i$. In this latter case, two sub-cases occur: when $\gamma_i \cdot \sum_{j \in S_i^l} \bar{p}_{ij} > P_i$, then $x_i = 0$; when $\gamma_i \cdot \sum_{j \in S_i^l} \bar{p}_{ij} \leq P_i$, we have a dummy constraint which can be ignored.

Hence, this robust model leads to three different configurations:

- either $x_i = 0$: the flight i does not use level l ;
- or $x_i = 1$ and no constraint is associated to flight i : this means that flight i uses level l with zero conflict costs;
- or $x_i = 1$ and the worst case is taken into account: the flight i uses level l with maximal conflict costs.

These three cases are in fact summarized in the two following ones:

- either flight i has zero conflict costs;
- or flight i is associated maximal conflict costs.

Hence, the analysis of the above robust model leads to a new one, which is very simple. Let $I_c \subseteq F^l$ be a subset of flights:

$$\begin{aligned} & \max \sum_{i \in F^l} b_i x_i \\ M_i x_i + \sum_{j \in S_i^l} \bar{p}_{ij} x_j & \leq M_i + P_i \quad i \in I_c, \\ x_i \in \{0, 1\}, \quad i \in F^l. & \end{aligned}$$

The parameter enabling to tune robustness is the subset I_c , and we denote the problem by $RP^l(I_c)$.

B. Solving the RP^l problem

In the precedent section we have described how an instance of the robust FLA problem can be modeled by ILP. Let recall that we are interested in robust solutions that remain feasible

in a large part of scenarios, that is, which has a high enough feasibility probability. Obviously, if we take $\gamma_i = 1$, for all i in F^l (which gives $I_c = F^l$), we obtain feasible solution for all scenarios. One idea is to start with $I_c = \phi$ and to make it grow gradually until a solution with the desired feasibility probability is achieved. Such ideas have already been exploited in [14], [15]. The algorithm can be depicted as follows:

A fast heuristic approach for solving RP^l

- Step 0:** Set $I_c = \phi$.
 Select an index $i \in F^l$ such that:
 $i = \arg \min \{P_i - E[\sum_{j \in S_i^l} p_{ij}]\}$
 Set $I_c \leftarrow I_c \cup \{i\}$.
- Step 1:** Solve $RP^l(I_c)$.
 Let \bar{x} be the solution found.
- Step 2:** If feasibility probability of \bar{x} is high enough, STOP.
 Else, select an index $i \in F^l \setminus I_c$ such that:
 $i = \arg \min \{P_i - E[\sum_{j \in S_i^l} p_{ij} \bar{x}_j]\}$
 Set $I_c \leftarrow I_c \cup \{i\}$; Return to **Step 1**.
-

As it can be seen from the algorithm, during **Step 0** we look for a *strongly constrained* constraint to introduce in I_c . The solution will admit all flights in this level except the flight i or a few constraining the selected flight i (it depends on the associated benefits). At this stage, the above solution is most probably not feasible and we need to pursue with other steps in order to further constrain the set of flights to be assigned at this level. An immediate way to accelerate the algorithm is to introduce at step 0 in I_c a larger number of constraints. For more details on the general framework of the algorithm and a deeper study on its theoretical properties, we refer to [15].

Notice also that the above algorithm doesn't ensure the optimality of the obtained solution. An important element of the resolution scheme given above is measuring the probability of the obtained solution. There are two ways to estimate the feasibility probability associated with some solution.

1) *First method:* The main idea behind the first method is using the Hoeffding's inequality [12], which is a result in probability theory that gives an upper bound on the probability for the sum of random variables to deviate from its expected value. This yields general results but it could be pertinent since variables p_{ij} can be assumed independents in our model. Let us recall first this fundamental result (see [12] for details):

Let X_1, \dots, X_n be independent random variables. Assume that the X_i are almost surely bounded; that is, assume for $1 \leq i \leq n$ that $Pr(X_i \in [a_i, b_i]) = 1$. Let be $S = \sum_{i=1}^n X_i$ and $E[S]$ its expected value. Then, we have the inequality

$$Pr(S - E[S] \geq nt) \leq \exp\left(-\frac{2n^2 t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right),$$

which is valid for positive values of t .

To apply this result to our problem, we first need to compute the expected value for each random variable. For doing this, let us try to express these variables in a more formal way.

Recall that the random variable p_{ij} corresponds to the cost induced by some resolution conflict procedure. Then, it de-

pend to two factors: first, if some conflict is occurring, and second, the resolution procedure engaged by the air traffic controllers. Hence, the first event is an en-route conflict¹ modeled below with a random variable c_{ij} , which follows a Bernoulli distribution ($Pr(c_{ij} = 1) = q_{ij}$, and $Pr(c_{ij} = 0) = 1 - q_{ij}$). In case of conflict, the cost induced to the involved aircraft, represented by random variables p'_{ij} , is assumed identically distributed in the interval $[0, \bar{p}_{ij}]$. We have: $p_{ij} = p'_{ij} \cdot c_{ij}$. Then, $E[p_{ij}] = E[p'_{ij} \cdot c_{ij}]$. Since the random variables p'_{ij} and c_{ij} are stochastically independent: $E[p_{ij}] = E[p'_{ij}] \cdot E[c_{ij}] = \frac{\bar{p}_{ij}}{2} q_{ij}$. Clearly for a given vector x , we obtain $E[p_{ij}x_j] = \frac{\bar{p}_{ij}}{2} q_{ij}x_j$.

Considering that the FLA problem has separated probability constraints, we need to ensure that for each constraint the following probability condition is satisfied:

$$Pr(\sum_{j \in S_i^l} p_{ij}x_j + M_i x_i \leq M_i + P_i) \geq 1 - \epsilon. \text{ As}$$

$$Pr(\sum_{j \in S_i^l} p_{ij}x_j + M_i x_i \leq M_i + P_i) \geq$$

$$Pr(\sum_{j \in S_i^l} p_{ij}x_j \leq P_i),$$

we restrict ourselves in ensuring that

$$Pr(\sum_{j \in S_i^l} p_{ij}x_j \leq P_i) \geq 1 - \epsilon \text{ for all } x_i = 1.$$

Applying the Hoeffding's inequality we have:

$$Pr(\sum_{j \in S_i^l} p_{ij}x_j \geq P_i) =$$

$$Pr(\sum_{j \in S_i^l} p_{ij}x_j - E[\sum_{j \in S_i^l} p_{ij}x_j] \geq P_i - E[\sum_{j \in S_i^l} p_{ij}x_j]) \leq$$

$$\exp\left(-\frac{2(P_i - \sum_{j \in S_i^l} \frac{\bar{p}_{ij}}{2} q_{ij}x_j)^2}{\sum_{j \in S_i^l} \bar{p}_{ij}^2 x_j}\right) = \epsilon'_i.$$

Notice that the interval of values for variable $p_{ij}x_j$ is given by $[0, \bar{p}_{ij}]$, which explains the above formula. Clearly, we have reached a feasible robust solution x when for all i in F^l with $x_i = 1$, we have $\epsilon'_i \leq \epsilon$.

The method is attractive and does not need restrictive probability conditions but it could lead to costly solutions as the probability bounds are quite general and could be weak. To remedy this, a natural idea is to use Monte-Carlo simulation.

Remark. It is possible to formulate an ILP model for computing a robust solution with the desired feasibility probability. For this, we need to combine the search for some robust solution x with some additional conditions that a feasible solution must satisfy. As shown in [15], it yields an ILP model involving additional variables and constraints.

2) *Second method:* The second way to handle the feasibility probability computation is using Monte-Carlo simulation. The main idea behind is simulating the departures times for all flights, simulating next the most convenient resolution en-route procedure, and estimating the induced cost. Once all coefficients of the model estimated, we check the feasibility of our solution for the given scenario. We repeat this a large number of times, and deduce the feasibility probability associated with the robust solution.

V. NUMERICAL TESTS

Our approach for the robust FLA problem is implemented in C++ using CPLEX 10.0. Let us give some details on the

¹The conflict probability associated with a pair of aircraft can be computed following the method given in [3].

TABLE I
TEST INSTANCE

Network	Number of Flights	Used Airports	Used WayPoints
NET_FR	1377	134	769

implementation approach: we start by considering levels one by one, from the most loaded to the least one. For each level, we start with a set of a reduced number of flights. More precisely, we initiate the RP^l problem with about 5% of concerned flights. We choose the most constrained ones, that is in increasing order of $\{P_i - E[\sum_{j \in S_i^l} p_{ij}]\}$ values. Further iterations could be necessary to ensure the probability feasibility of the obtained solution. Hence, for a given solution x , we add in the RP^l problem a few new flights in decreasing order of $\epsilon'_i (> \epsilon)$ values.

For our tests we use collected data on departure and arrival times, aircraft type, velocities, trajectory crossing angle and flight levels for a set of flights. Next, for each flight we will compute the en-route conflict probability following the guidelines given in [3]. The test data corresponds to French air traffic of August 12th 1999. Table I presents the characteristics of test data. All the tests were run on a machine with the following configuration: Windows XP, 1 processors Pentium 4 2.4GHz, 1 Gb of RAM.

At this stage, the first difficulty encountered when implementing the model, is concerned with providing the right parameters \bar{p}_{ij} and P_i . Indeed, the best choice would be to estimate the interval $[0, \bar{p}_{ij}]$ as a function of crossing angle, and type of aircraft, and last, estimate the P_i^l as a few percent of the energetic cost of the flight. In this first series of tests, essentially because of lack of data and time, we have set the same unitary cost for all conflicts. Thus, we have set $P_i = \max \alpha, c * (duration - 1) + \alpha$, where duration gives the flight duration, c and α are both constants. For instance, for any flight with duration less than 1 hour, we have fixed P_i to $\alpha = 3$, while for the others we also take into account the duration of the flight according to the above formula with $c = 0.1$. Our goal is to measure the impact of robustness on the number of flights assigned to their preferred level comparing to these that have to be changed. We have also varied the level of robustness parameter ϵ . To measure the feasibility of the solution, we have used the Hoeffding's formula. In table II are shown some results obtained with the above parameters for three different values of ϵ , which gives the allowed infeasibility probability. The second column, ("Number of changed levels") gives the number of flights not assigned to their preferred levels and accommodated to adjacent levels because of en-route conflicts. The last column ("Gap Robust/Deterministic") gives the percentage of additional flights assigned to their preferred levels thanks to robustness in comparison with the deterministic model. These results show that when using the robust model we can have some increase in the capacity of accommodating flights in their preferred levels with very high probability feasibility comparing to the standard problem when considering the worst case. This latter case is computed by

TABLE II
NUMERICAL RESULTS

ϵ	Number of changed levels	Gap Robust/Deterministic
0.05	170	6.5%
0.10	163	10.4%
0.15	139	23.6%
0.20	135	25.8%
0.25	126	30.7%

adding all constraints corresponding to flights in the problem. We do not need at all to do any computation on the probability feasibility: the obtained solution is feasible for any case and all constraints are satisfied. For our deterministic problem using the above set of parameters we have obtained 182 flights not assigned to their preferred levels.

In our computations, most of remained flights not assigned to their preferred levels are accommodated to adjacent levels, while for some of them we have needed to increment their cumulated allowed cost as indicated in **Step 2** of the *ApproxFLA* Algorithm described in Section III.

Indeed, we expected to have a larger difference between the deterministic and the probabilistic model. We believe that this is because of using the Hoeffding bound which is somehow weak. To remedy this, two directions need to be followed: first, using a better parameterizing of the model, and next switching to Monte-Carlo simulations, better suited to this kind of problems. This work is in progress.

VI. CONCLUDING REMARKS

In this paper we have provided a mathematical model for the robust FLA problem. We have first discussed the model following the Bertsimas and Sim [5] approach and focus on a second one inspired from [15], for which an approximated tractable iterative approach is available. We have adapted this later work in the context of ATM for solving the robust FLA problem. This work is a first stage to achieve a thorough study on the robust flight level assignment problem. As remarked above, the obtained results rises the problem of how parameterizing the model. Another point, in addition to those shown above, is related with the assumption of considering only en-route conflicts between aircraft flying horizontally in the same level. We are actually thinking in considering air conflicts that involve crossing aircraft flying on different levels, for instance when one of them is climbing or descending. This assumption will also allow a better modeling of the problem and can contribute in avoiding the above limitations of the robust model. Further investigations are needed.

REFERENCES

- [1] Ahuja R.K., Magnanti T.L. Orlin J. B., *Network Flows: Theory, Algorithms and Applications*, Prentice Hall, 1993.
- [2] Bashllari A., Nace D., Carlier J., *A note on the flight level assignment problem*, InoWorkshop, 4-6 december 2007.
- [3] Bashllari, A. Kaciroti, N. Nace, D. Fundo, A. *Conflict Probability Estimations Based on Geometrical and Bayesian Approaches* 10th International IEEE Conference on Intelligent Transportation Systems, (ITSC'07), Seattle, Washington, USA, October, 2007.
- [4] Ben-Tal A. and Nemirovski A., *Robust solutions of Linear Programming problems contaminated with uncertain data*, Math. Program. (Ser. A), 88 (2000), pp. 411-424.
- [5] Bertsimas D. and Sim M., *Robust discrete optimization and network flows*, Math. Program. (Ser. B), 98 (2003), pp. 49-71.
- [6] Bertsimas D. and Sim M., *The Price of Robustness*, Operations Research, 52-1 (2004), pp. 35-53.
- [7] Bertsimas D. and Stock Patterson, S., *The air traffic flow management problem with en-route capacities*, Operations Research, Vol. 46, pp. 406-422, 1998.
- [8] Bertsimas D. and Stock Patterson, S., *The traffic flow management rerouting problem in Air Traffic Control: A dynamic Network Flow Approach*, Transportation Science 2000 INFORMS, Vol. 34 No. 3, pp. 239-255, 2000.
- [9] Doan, N-L., Duong V., Nace, D. *The Air Route Network Design Problem*, RIVF'2004 Conference, Hanoi, February 2004.
- [10] Barnier N. and P. Brisset, *Graph Coloring for Air Traffic Flow Management*, Proceedings CPAIOR'02, pp. 1-15, 2002.
- [11] Delahaye D. and Odoni, A., *Airspace Congestion Smoothing by Stochastic Optimization*, Evolutionary Programming VI, pp. 163-176, 1997.
- [12] Hoeffding W., *Probability inequalities for sums of bounded random variables*, Journal of the American Statistical Association 58 (301): 1330, 1963.
- [13] Kellerer, H., Pferschy, U., Pisinger, D., *Knapsack Problems*, Springer (2004).
- [14] Klopfenstein, O., and Nace, D., A robust approach to the chance-constrained knapsack problem, to appear in Operations Research Letters.
- [15] O. Klopfenstein, Tractable algorithms for chance-constrained combinatorial problems, submitted. Available online on www.optimization-online.org.
- [16] Letrouit V., *Optimisation du Rseau des Routes Ariennes en Europe*, PhD Dissertation, INPG, 1998.
- [17] Martello S., and Toth P., *Knapsack Problems: Algorithms and Computer Implementations*, Wiley (1990).
- [18] Y. Nikulin, Robustness in Combinatorial Optimization and Scheduling Theory: An Annotated Bibliography, available on www.optimization-online.org (2004).
- [19] Constans S., Fontaine B., Fondacci R, *Minimizing Potential Conflict Quantity with Speed Control*, ICRA 2006, Belgrade.

Stochastic Airspace Demand for Strategic Traffic Flow Management

Jeff Henderson and Antonio Trani

The Charles E. Via, Jr. Department of Civil and Environmental Engineering
Virginia Tech
Blacksburg, VA
jmhender@vt.edu

Abstract— In this paper we consider the problem of predicting the demand for en route airspace sectors considering uncertain flight departure time and en route conditions. Flight, airport, and airline conditions that lead to greater variance in departure time prediction errors are examined and used to develop kernel-smoothed empirical probability density functions for flight departure time predictions. The structure of the departure time prediction errors is found to vary across the departing airport type. A similar analysis is performed for the en route airspace to characterize the random component of airspace sector traversal time. Variance of en route sector traversal times is found to increase for shorter duration planned sector traversal times. A method that combines these sources of uncertainty is presented and applied to two days of historical traffic conditions for east coast U.S. airspace sectors. Results of this analysis indicate that the mean absolute prediction error of the airspace demand can be reduced by 20% when using the probabilistic method as compared to a deterministic procedure. Similarly, standard deviation of the error in airspace demand is reduced by 23 to 25% also indicating a reduced spread in the demand estimation.

Keywords— en route; airspace; traffic flow management; demand; probabilistic

I. INTRODUCTION

In 2006 aircraft operating in the National Airspace System (NAS) experienced in excess of five hundred thousand aircraft hours of airborne delay [1]. The number and duration of delays are expected to worsen during a projected growth of 47.5 million to 67.7 million flights operating under instrument flight rules (IFR) from 2006 to 2017 respectively [2]. A combination of improved traffic flow management practices and an increase in airspace capacity would be required to mitigate these expected delays. The Next Generation Air Transportation System (NextGen) program is one such current initiative [3].

The focus of this work is to demonstrate how stochastic models can support en route traffic flow management decision-making under uncertainty. Current traffic flow management practice is based on the deterministic Enhanced Traffic Management System (ETMS) and the experience of air traffic controllers and managers [4]. ETMS provides forecasts of airport departures and arrivals, sector entry and exits, airway entry and exits, and waypoint crossings [5]. The

drawback of these forecasts is the inability to consider a range of potential scenarios so that traffic flow managers must be more conservative in their decision-making. Conversely, the deterministic forecast may under represent the congestion potential during volatile conditions, such as severe convective weather, leading to capacity overload.

There is a body of work in the recent literature focusing on quantifying and modeling stochastic elements of the NAS en route airspace. The estimated time of departure is the single largest source of uncertainty for flights that have not departed from the origin airport [6]. The work on pre-departure uncertainty has focused on quantifying variance and confidence bounds under various weather and flight-specific attributes at a range of look-ahead times [7-9].

The prediction of departure time is one component in the estimation of en route airspace sector demand. Meyn details a method to estimate sector and airport demand from arrival probability distributions and sector traversal time [10]. Only a single source of uncertainty is modeled at an unspecified look-ahead time. Mueller et al. note that departure time, wind forecasting errors, deviations from the flight plan, and aircraft performance and weight uncertainty can lead to errors in sector demand prediction [11]. The climb phase of flight, especially step climbs that are mandated by air traffic controllers in congested airspace, is identified as the source of the highest trajectory performance prediction errors with empirical results presented.

Flow models are another proposed approach to improve the estimation of sector demand by considering air traffic demand at a high level. Many of the current models are deterministic though well-suited to metering traffic flows to an arrival fix at a busy airport [12-15]. Probabilistic versions have also been developed but at the more macroscopic center level [16]. An attempt to establish a relationship between planned and observed sector counts is discussed in [17].

The following sections describe extensions to current stochastic airspace demand models to include pre-departure uncertainty, en route traversal uncertainty, and route uncertainty. A method to combine departure time uncertainty and en route traversal time uncertainty is presented and applied to one day of historical airspace conditions to quantify the benefit of a probabilistic approach.

II. DEPARTURE UNCERTAINTY

Previous work in the area of sector demand estimation has noted several factors that lead to errors in sector demand [6-8]. The work presented here will focus on pre-departure, sector traversal, and route uncertainty. Pre-departure uncertainty is the difference between the proposed wheels-off time at the origin airport and the measured departure time as recorded in ETMS. The ETMS system does not provide the most accurate historical prediction of wheels-off time, however errors of a few minutes are considered negligible in the context of this analysis.

The quantity of interest is the departure prediction error and not deviation from the schedule so the lateness of a flight is not what is being measured. The following is a partial list of factors that can result in poor estimations of departure time: aircraft arriving late from a previous leg, unavailable gates from the previous leg, crew arriving late, aircraft servicing, de-icing operations, runway direction reversals, taxiway availability, etc.

The procedure to calculate departure uncertainty begins by collecting all relevant messages from the ETMS historical data including the flight schedule (FS), flight plan (FZ), flight amendment (AF), control departure time (CTRL), and flight cancellation (RZ) messages. The analysis days for this study are shown in Table I from which 1,238,730 departure observations are extracted. Information from the previous day is also used to obtain full flight plan and schedule information. Gate push-back times are obtained from the FAA Airline Service Quality Performance (ASQP) database [18]. Definitions for departure time are shown in Table II.

Messages are then sorted by time of entry into the ETMS system. For each message a modeled departure time may be recorded at 0, 15, 30, 60, and 120 minute look-ahead times. A modeled departure time is not recorded if a more recent message is received prior to one of the look-ahead times.

TABLE I. STUDY ANALYSIS DAYS.

Day	Year(s)	Day	Year(s)
February 19	2000-2005	September 26	2000-2004
May 10	2000-2005	October 23	2004
June 11	2004-2005	December 1	2001-2002
June 27	2004	December 3	2000
July 14	2005	November 28	2004
July 27	2000-2004	November 30	2003

TABLE II. DEPARTURE TIME DEFINITIONS.

Notation	Definition
ETMS modeled departure time	Gate pushback time + ETMS modeled taxi time
ETMS modeled taxi time	Moving average of last five taxi times for that flight
Wheels-off time	Estimated runway-off time for flight from ETMS message
Departure time prediction error	Wheels-off time – ETMS modeled departure time
Look-ahead time	ETMS modeled departure time – current time

For example if a flight plan message is received at 0200Z with a modeled departure time of 0330Z, then a subsequent flight amendment is received at 0250Z then only an error observation corresponding to a 60 minute look-ahead time is recorded for the 0200Z message.

The analysis proceeds by attempting to find structural variance in the prediction error data. Exploratory analysis strongly suggests the existence of non-constant variance across variables, otherwise known as heteroscedasticity. A modified least squares regression procedure is used since errors are non-normal and right-skewed even under a logarithmic transformation. Another method that accounts for non-constant variance is the class of generalized autoregressive conditional heteroscedasticity (GARCH) models most suitable to time series analysis but with limited applicability to this problem [19].

The modified regression procedure is as follows. The regression form of **Error! Reference source not found.** shows a response variable Y to be a function of two independent random variables X and ϵ and a coefficient matrix β .

$$Y = \beta X + \epsilon \tag{1}$$

If the variance is constant then $\epsilon \sim N(0, \sigma^2)$ is independent of X. This model is extended by allowing the variance to be a function of X as shown in [20]. An exponential link function is used in this analysis but others may be substituted.

$$\epsilon \sim N(0, \sigma^2 \lambda(X)) \tag{2}$$

$$\lambda(X) = \exp(\Gamma X) \tag{3}$$

This type of regression on the variance does not eliminate the non-normality problem but it does allow an investigation into the conditions that lead to larger variance. A total of 26 explanatory variables are considered representing flight, airline, airport, and weather conditions.

The model is calibrated using SAS [21] with coefficients for the 14 selected variables presented in Table III. The exponential of the coefficients is also shown since the coefficients must be transformed back and used as a multiplier effect. All variables are significant at the 5% level though the test for statistical significance is somewhat questionable in this case. A logarithmic transform of the error response at a 30 minute look-ahead time (E_{LAT30}) is used to better approximate normality .

$$Y = \log(E_{LAT30} + 60 \text{ minutes}) \tag{4}$$

Since the error could not be completely transformed to normality a categorical analysis of the distribution of errors is considered based on the results of the regression analysis. The first of the groupings uses the departing airport type of the flight. A box-and-whisker diagram of the errors (Fig. 1) shows the 25th percentile, median, and 75th percentile of the errors as a box. The difference between the 75th percentile and the 25th percentile is known as the interquartile range

(IQR) and is used to mark the largest and smallest observations that are “valid” as whiskers. A data point is

TABLE III. REGRESSION COEFFICIENTS FOR MEAN AND VARIANCE OF DEPARTURE TIME PREDICTION ERROR AT A 30 MINUTE LOOK-AHEAD TIME.

	Coefficient for Mean Estimate			Coefficient for Variance Estimate		
	β	$exp(\beta)$	S.E. ^a	Γ	$exp(\Gamma)$	S.E. ^a
Intercept	4.175	65.040	0.0009	0.277	1.319	0.0005
Depart from a large hub	0.014	1.014	0.0009	-0.019	0.981	0.0051
Depart from a medium hub	-0.004	0.996	0.0011	0.038	1.039	0.0060
Depart from a small hub	-0.064	0.938	0.0013	-0.303	0.739	0.0059
Depart from a non-hub	-0.004	0.996	0.0011	-0.170	0.844	0.0066
Departing airport operating under instrument conditions (IMC)	0	1		0.284	1.328	0.0025
A large carrier (top 25 by operations) departing from a large hub airport	0	1		-0.367	0.693	0.0030
A large carrier (top 25 by operations) departing from a medium hub airport	0	1		-0.681	0.506	0.0048
A large carrier (top 25 by operations) departing from a non-hub airport	0	1		-0.048	0.953	0.0088
A large carrier (top 25 by operations) departing from a foreign airport	0	1		-0.313	0.731	0.0061
A small carrier (not top 25 by operations) departing from a small airport and arriving to a large hub	0	1		0.379	1.461	0.0048
If flight plan is amended	0	1		0.100	1.105	0.0018
If convection is forecasted to impact this flight (origin airport, en route, destination airport)	0	1		0.047	1.048	0.0022
If flight has been cancelled and reactivated	0	1		0.198	1.219	0.0039
If flight has been both amended and cancelled and reactivated	0	1		0.408	1.504	0.0052
Sample size	1,130,874					
Log Likelihood	-83,918.2					

^a Standard Error

considered valid if it is less than 1.5(IQR) from the box. Outliers are indicated by a “+”. The notable characteristics of the diagram are that the median is relatively constant between airport types, all the distributions are right-skewed (positively skewed), variance increases as airport size decreases, and there are numerous outliers for each distribution, which is the reason for the solid red line. Large variance for smaller airports may seem counter-intuitive but the type of airline operating at these airports has an impact.

A categorical grouping that includes factors in addition to departing airport size is shown in Fig. 2 with a corresponding box-and-whisker plot in Fig. 3. The clustering procedure covers all cases and the order is generally as follows: amendment, cancelled, forecasted convection, airport type, if the carrier is one of the top 25 carriers by operations, and arriving airport size. Clusters are ranked by mean error then by variance so that cluster 1 has the lowest mean and variance while cluster 10 has the highest mean and variance. The highest variance is for flights that have amendments or that have been cancelled and reactivated. By separating smaller carriers from larger carriers this analysis shows that larger carriers have lower variance than smaller carriers and smaller airports have lower variance than larger airports when corrected for carrier type. However, since smaller carriers dominate smaller airports we get the results shown in Fig. 1.

An attempt is made to generalize the errors to a probability distribution. However, since the error is right-skewed and peaked around 0 the standard distributions are poor approximations (Fig. 4). Histograms are constructed and compared to the lognormal for each of the two groupings considered here: airport type and clustered data. For each of these histograms the lognormal approximations are significantly different from the observed empirical distribution.

In kernel smoothing a probability mass, such as a normal or other symmetric density function, is placed at each data point. The equations to place the probability mass are straightforward. Begin by specifying a kernel that satisfies (5). In this case the standard normal distribution is chosen $K(x) \sim N(0,1)$. The density at each value is estimated by summing all kernels as detailed in where n is the number

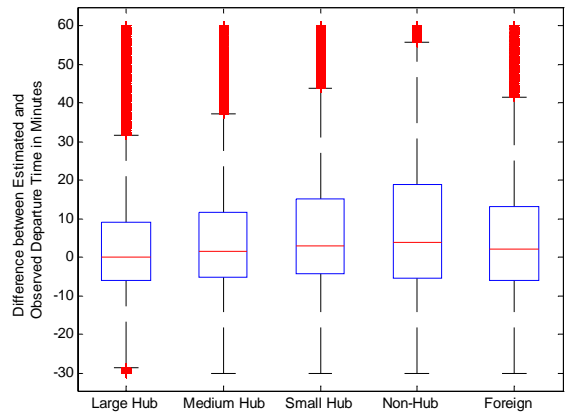


Figure 1. Box-and-whisker diagram for departure time prediction errors at 30 minute look-ahead times by departing airport type.

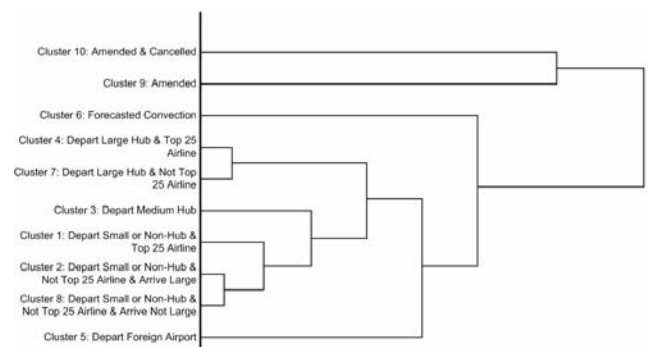


Figure 2. Tree diagram for clustering.

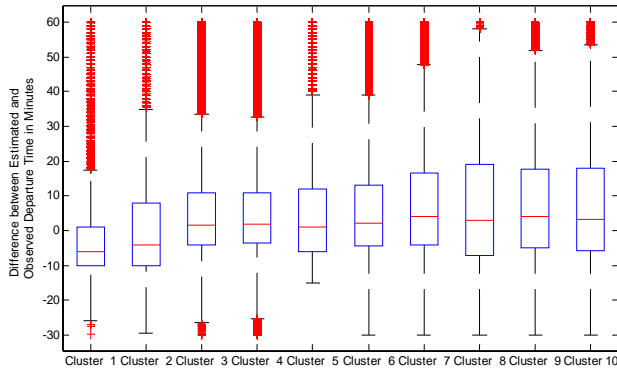


Figure 3. Box-and-whisker diagram for departure time prediction errors at 30 minute look-ahead times by cluster.

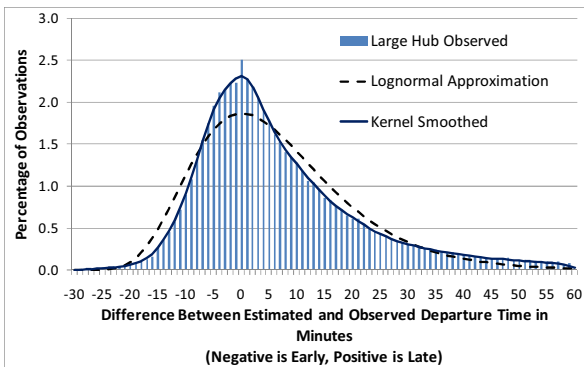


Figure 4. Histogram of departure time prediction errors at a 30 minute look-ahead time frame for large hub airports for 30 days in 2000-2005. A lognormal approximation and kernel smoothed distribution are also displayed illustrating a poor and good fit respectively.

of samples and h is the bandwidth. The optimal bandwidth parameter is a function of sample size and variation of the distribution. Larger sample sizes permit a smaller bandwidth while larger variances require increased bandwidth. The kernel smoothing parameter estimation results are excluded for brevity.

$$\int K(x)dx = 1 \tag{5}$$

$$\hat{f}(x; h) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \tag{6}$$

III. SECTOR TRAVERSAL TIME VARIATION

Another source of randomness in the estimation of sector demand is the traversal time through the sector. Controller actions such as: speed changes, vectoring, issuing a holding pattern, or the clearance of a more direct route may cause the flight to spend more or less time in the sector than planned. The approach used compares the planned sector flight time obtained through simulation to the observed sector traversal time from the processed ETMS radar track data (TZ). The scope of the analysis includes the following east coast air route traffic control centers (ARTCCs): Chicago (ZAU),

Indianapolis (ZID), Atlanta (ZTL), Jacksonville (ZJX), Miami (ZMA), Washington (ZDC), Cleveland (ZOB), New York (ZNY), and Boston (ZBW).

To obtain planned sector traversal times the most recent flight plan or amendment before the actual departure is extracted from ETMS data. The flight plan data is converted into a format suitable for the RAMS Plus airspace simulation software [22]. Other information including aircraft performance, airport locations, navigational aids (NAVAIDs), fixes, airways, standard terminal arrivals (STARs), and departure paths are also converted to the RAMS format. Aircraft performance uses EUROCONTROL’s Base of Aircraft DATA (BADA) [23] which is different from the ETMS system aircraft performance models [5]. The largest source of uncertainty in aircraft performance modeling is the prediction of aircraft weight. In this analysis we assume a nominal, or average, weight for each flight based on the three aircraft mass categories contained in BADA: low, nominal, and high. Each of the flight plans are then simulated to obtain the time of sector entry, time of sector exit, and sector traversal time. The air traffic controller functionality of RAMS is turned off so there is no conflict resolution for flights predicted to violate minimum separation standards.

The ratio of observed sector traversal time to planned sector traversal time, which is obtained from processing the RAMS output files, is examined for structure. A plot of the standard deviation of the ratio of observed to planned sector traversal times by planned sector traversal time and observed airspace density (Fig. 5) shows that the planned traversal time through the sector has a larger effect than the observed airspace density. This does not mean to suggest that airspace density has no effect since sectors with shorter traversal times are typically more congested than those with longer traversal times. The assertion here is that flown traversal time is mostly impacted by planned time for a flight to cross a sector.

Based on this observation a series of kernel-smoothed densities are developed for planned traversal times (t_p) as follows: $\{t_p | 0 < t_p \leq 4 \text{ minutes}\}$, $\{t_p | 4 < t_p \leq 8 \text{ minutes}\}$, $\{t_p | 8 < t_p \leq 12 \text{ minutes}\}$, $\{t_p | 12 < t_p \leq 16 \text{ minutes}\}$, $\{t_p | 16 \text{ minutes} < t_p\}$. A sample kernel-smoothed ratio for the 4 to 8 minute planned traversal time interval is shown in Fig. 6.

Alternatively, an error distribution that considers the relative difference between the observed and planned traversal times is also considered but not selected (i.e. error distribution = observed traversal time – planned traversal time). Due to the difference between the high and low range, e.g. 4 to 8 minutes in Fig. 6, a negative sector traversal time may be implied from the resulting error distribution. The ratio distribution is more appropriate in this case since traversal times are always positive and relative to the planned traversal time.

IV. SECTOR HIT RATE

The last source of uncertainty considered in this analysis is the sector hit rate which is defined as the rate at which the planned sectors for a flight plan match the observed or flown

sequence of sectors. It is a combined consideration of the route and altitude forecast accuracy. Consider ordered sets of planned sectors P and flown sectors F . A hit is defined if the planned and flown sectors match and is then used to calculate the overall hit rate. Note that it is possible for a flight to enter the same sector more than once so by this definition the hit rate is restricted to be ≤ 1 .

$$Hit\ Rate = \frac{|PNF|}{|P|} \tag{7}$$

The simulation and playback results from the sector traversal analysis in RAMS are also used to calculate the hit rate. The results of the hit rate analysis show an overall average hit rate of 73%. Conditions that lead to re-routing such as severe weather and airspace congestion were not included here. Further work that could find a relationship to predict sector hit rate probabilities under various conditions would be beneficial.

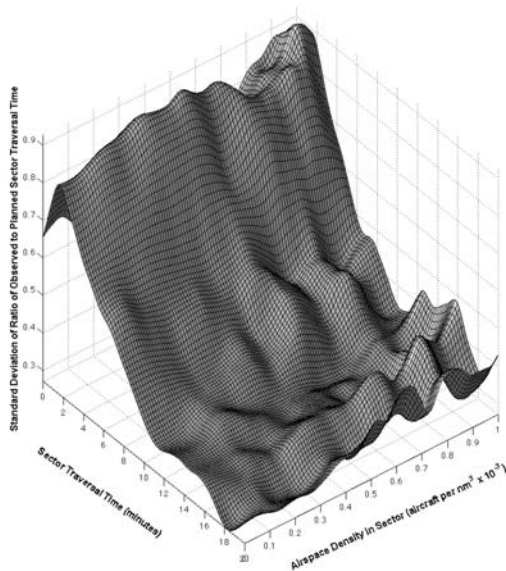


Figure 5. Standard deviation of ratio of observed to planned sector traversal time by planned sector traversal time and observed airspace density in the sector.

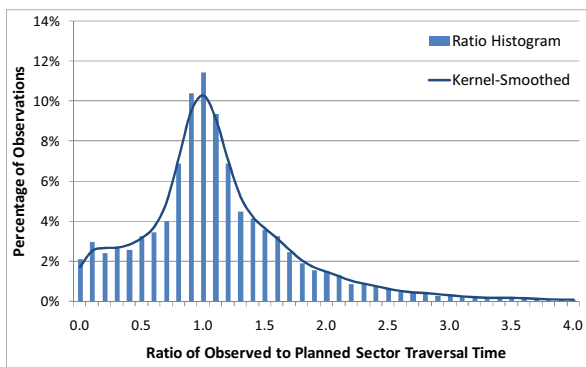


Figure 6. Ratio of observed to planned sector traversal time for a planned traversal time of 4 to 8 minutes.

V. PROBABILISTIC SECTOR DEMAND

We now use the departure uncertainty and sector traversal variability models to develop sector demand at 10, 30, and 60 minute look-ahead times to sector entry. We are interested in the probability for a given flight to occupy a sector as a function of time (Fig. 7). The resulting distribution is not a probability density function since the area under the curve does not equal 1. This distribution would then be used in the calculation of sector demand by time period.

The equations in this section represent the application of standard statistical methods, such as the convolution theorem [24], and conventions used in calculations. The following nomenclature is used throughout this section:

- $\hat{d}_i(t)$ = Expected count, or demand, for sector i , during time period t
- $f_i^{departure}(\cdot)$ = Distribution of errors in predicting a flights departure time as calculated in Section II for the flight traversing the sector at position i under consideration
- $f_i^{enroute}(\cdot)$ = Distribution of flight traversal time through sector at position i considering the stochastic en route component
- $f_i^{total}(\cdot)$ = Probabilistic demand distribution for sector at position i considering both departure and en route sources of randomness
- $f_k^{ratio}(\cdot)$ = Distribution of ratio r for sector k as calculated in Section III
- F = Set $\{s_1^{flown}, \dots, s_m^{flown}\}$ of sectors traversed using the flight's flown trajectory
- i, j = Position indices where position i is the first sector after the departing airport and positions m, n are the last sector before the arrival airport
- k = Sector index for ratio distribution $f_k^{ratio}(\cdot)$
- $LAT^{departure}$ = Look-ahead time to departure (wheels-off)
- LAT^{sector} = Look-ahead time to sector entry
- m, n = Number of sectors that a flight crosses when following the flown (m) or planned (n) trajectory
- $p_0(t)$ = Probability that the flight under consideration arrives to a sector i during time period t
- $p_1(t)$ = Probability that the flight under consideration does not arrive to a sector i during time period t

- P = Set $\{s_1^{planned}, \dots, s_n^{planned}\}$ of sectors in the flight plan
- r = Ratio of observed $t_j^{traversal}$ to planned $\hat{t}_i^{traversal}$ traversal times
- $s_i^{planned}$ = Sector in position i in the flight plan
- s_j^{flown} = Sector at position j in the set of flown sectors F
- \hat{t}_i^{entry} = Planned time of entry into a sector at position i
- \hat{t}_i^{LAT} = Planned time of entry into a sector at position i at look-ahead time LAT
- $\hat{t}_i^{traversal}$ = Planned traversal time through a sector at position i
- t_j^{entry} = Flown, or observed, time of entry into a sector at position j
- $t_j^{traversal}$ = Flown, or observed, traversal time through a sector at position j
- t_k^T = Variable used to convert from a ratio distribution to a relative error distribution

The analysis of historical ETMS data presented is sector based so to generate demand each sector in a flight plan is examined. To start we consider a single flight, its associated flight plan, and one of the sectors that the flight traverses when it follows its flight plan.

There are two cases to be considered for demand prediction for en route flights. There are additional considerations for flights that have not departed that are discussed later in this section. In the first case we find a flown sector that satisfies the conditions listed in (8-10). The first of these conditions is that the flown sector must also be included in the set of planned sectors. As shown in Section IV there are cases where the flown sector does not appear in the flight plan. The second condition ensures that the flown sector is at least the look-ahead time away from the planned sector. The third condition specifies that there is no closer flown sector.

So if a sector is found that satisfies the three conditions in (8-10) an improved estimate of the estimated sector entry time is calculated. Otherwise, for the second case where

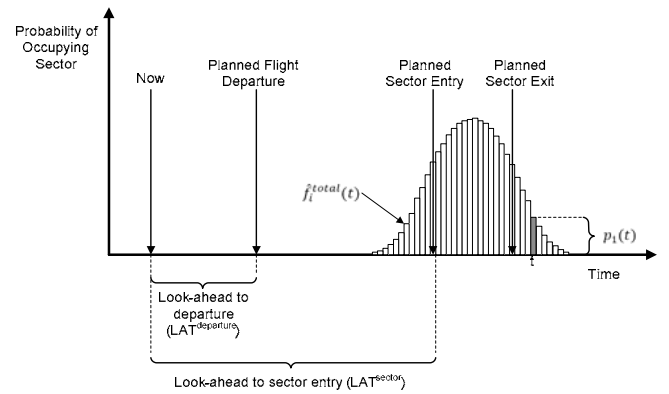


Figure 7. Distribution of probabilistic sector demand.

no sector is found that satisfies (8-10) the uncorrected planned time of entry into a sector is used which is the second condition in .

$$s_j^{flown} \in P \tag{8}$$

$$\hat{t}_i^{entry} - t_j^{entry} \geq LAT^{sector} \tag{9}$$

$$\{s_k^{flown} \in F \mid [LAT]^{sector} \leq \hat{t}_i^{entry} - t_k^{entry} \dots \tag{10}$$

$$\hat{t}_i^{LAT} = \begin{cases} \hat{t}_i^{entry} + (t_j^{entry} - \hat{t}_j^{entry}), & \text{if } \exists s_j \\ \hat{t}_i^{entry}, & \text{otherwise} \end{cases} \tag{11}$$

The next step is to determine the set of sectors for which traversal time ratio distributions will be considered and included in the analysis. If a flown sector is found that satisfies (8-10) then all sectors after and including the flown sector are included, otherwise all sectors are used to update the uncertainty distribution starting from the origin airport (12). Each of these sectors is matched with an appropriate ratio distribution that is described in Section III and categorized by the ratio of observed to planned sector traversal times (13). Since we are interested in the time relative to the corrected sector entry time calculated in we convert the basis of the distribution in . Planned traversal times are subtracted for all sectors excluding the planned sector under consideration so that all distributions are error distributions except the planned sector under consideration. For the planned sector under consideration the expected traversal time is included in the distribution to achieve a correct demand value.

$$j = \begin{cases} j, & \text{if } \exists s_j^{flown} \\ 1, & \text{otherwise} \end{cases} \tag{12}$$

$$f_k^{ratio}(r; \hat{t}_k^{traversal}), \quad \forall k = j, \dots, i \tag{13}$$

$$t_k^T = \begin{cases} (r - 1)\hat{t}_k^{traversal}, & \text{if } j \leq k < i \\ r\hat{t}_k^{traversal}, & \text{if } k = i \end{cases} \tag{14}$$

$\forall k = j, \dots, i$

The distributions are summed by the standard convolution (i.e. the * operator) method of taking the

discrete Fast Fourier Transform \mathcal{F} of each of the distributions, performing an element-by-element multiplication, then transforming back using the Inverse Fast Fourier Transform \mathcal{F}^{-1} [24]. The general case is shown in for distributions $f(\cdot)$ and $g(\cdot)$ and in for the distributions considered here. If a sector is found satisfying (8-10) then the result of is the density function relative to the corrected sector entry time to be added to the sector demand. Otherwise the departure time prediction error is also considered.

If the departure time prediction error needs to be considered then the look-ahead time for the departure uncertainty is calculated using the look-ahead time to the sector, the planned entry time into the sector, and the planned flight departure . The departure look-ahead time is rounded up to one of the available departure look-ahead

$$(f * g) = \mathcal{F}^{-1}[\mathcal{F}(f) \cdot \mathcal{F}(g)] \tag{15}$$

$$f_i^{enroute}(t) = \left[\left(f_j^{ratio} \left(t_j^T \right) \dots * f_{j+1}^{ratio} \left(t_{j+1}^T \right) \right) \dots \right] \tag{16}$$

times of {0, 15, 30, 60, 120 minutes} and used to select a departure uncertainty distribution. The departure uncertainty distribution is combined with the en route uncertainty distribution to arrive at a total uncertainty distribution .

$$LAT^{departure} = LAT^{sector} - \left(\hat{t}_i^{entry} - \hat{t}^{depart} \right) \tag{17}$$

$$f_i^{total}(t) = \left[f_i^{departure}(t) * f_i^{enroute}(t) \right] \tag{18}$$

To estimate the demand for the planned sector under consideration during any time period a summation of the discrete total error distributions is performed . If a distribution of demand for a sector is required rather than just the expected count then a discrete probability density function is constructed for each flight consisting of the probability that the flight arrives during a time period p_1 or does not arrive p_0 . A series of convolution operators for each flight similar to that shown in may be used to derive a distribution of sector counts for the purpose of obtaining confidence bounds.

$$\hat{d}_i(t) = \sum_{\forall \text{ flights}} f_i^{total}(t) \tag{19}$$

$$p_1(t) = P(N = 1) = f_i^{total}(t) \tag{20}$$

$$p_0(t) = P(N = 0) = 1 - p_1 \tag{21}$$

The method presented in this section implicitly assumes statistical independence for the departure and en route error distributions. Though this assertion is not strictly true it does allow for efficient demand uncertainty calculations. Methods that consider the covariance between the sector-based uncertainty distributions would also need to be computationally efficient to be useful for strategic traffic flow management.

VI. PERFORMANCE OF PROBABILISTIC SECTOR DEMAND MODEL

The historical traffic conditions on the date of August 29, 2005 is used to compare the performance of the probabilistic model for sector demand to a deterministic model at 10, 30, and 60 minute look-ahead times to sector entry in 1-minute intervals. Recall from the departure uncertainty section that two groupings are considered: one based on departing airport type and one based on a clustering that considers additional factors. Overall comparisons are made by considering the standard deviation of the demand prediction error and the mean absolute value of the prediction error (Table IV). The standard deviation of the error is reduced by 25% and the prediction error reduced by 20% when using the probabilistic methods as compared to the deterministic method. Results indicate that the cluster grouping method that considers additional factors offers little improvement on the method that only considers airport type in the departure uncertainty. Both methods also consider the en route random component as described in . A histogram detailing the distribution of the prediction error at a 30 minute look-ahead time to sector entry is shown in Fig. 8. This deterministic to probabilistic comparison is challenged by the fact that deterministic errors are discrete whereas the probabilistic errors may take on any real value.

Analysis of a second day of traffic data is performed for the date of July 27, 2005. The mean absolute prediction error for sector demand is reduced by 20% and the standard deviation by 23%, similar to the first day analysis results.

VII. CONCLUSIONS AND FUTURE WORK

In this paper departure time and en route sources of uncertainty are quantified. Airport size and sector traversal time are key indicators of the level of uncertainty expected for a flight. A probabilistic method is presented that combines airport and en route sources of uncertainty to produce improved estimates for sector demand. These more robust sector demand estimates have the potential to more efficiently use available airspace and identify volatile conditions that lead to higher controller workload. The probabilistic method is validated using historical traffic conditions for airspace sectors on the east coast of the U.S. for two days. Results indicate that the probabilistic method has the potential to reduce the standard deviation of the prediction error by 23 to 25% and the mean absolute prediction error by 20%. The sector hit rate, which is the rate that the planned sectors match the observed sectors, is a significant source of uncertainty for developing airspace sector demand estimates. Future work that can predict changes to the hit rate would be useful in improving sector demand estimates. Other future work could include identifying structure in the departure time and en route travel time error distributions.

VIII. ACKNOWLEDGEMENTS

This work was partially funded by Virginia Tech's Dean of Engineering Fellowship and by the FAA NAS Strategy Simulator (NSS) project. The results presented here are the

views of the authors and not those of the FAA or any other federal agency.

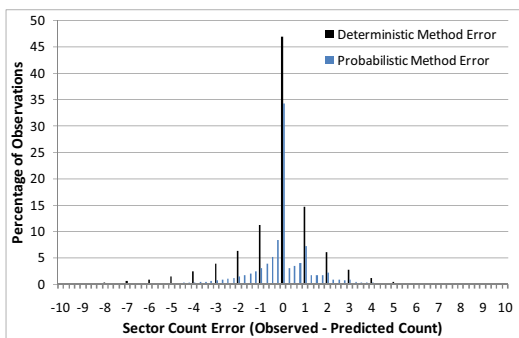


Figure 8. Histogram comparison of the distribution of sector count errors at a 30 minute look-ahead time on August 29, 2005 for deterministic and probabilistic (airport grouping) methods.

TABLE IV. COMPARISON OF DETERMINISTIC AND PROBABILISTIC METHODS FOR SECTOR COUNTS USING DATA FROM AUGUST 29, 2005.

	Deterministic Demand	Probabilistic Demand (Airport Grouping)	Probabilistic Demand (Cluster Grouping)
S.D. ^a 10 min. ^c	1.909	1.487	1.485
S.D. ^a 30 min. ^c	1.971	1.499	1.497
S.D. ^a 60 min. ^c	2.003	1.508	1.507
M.A.P.E. ^b 10 min. ^c	1.109	0.900	0.899
M.A.P.E. ^b 30 min. ^c	1.144	0.909	0.907
M.A.P.E. ^b 60 min. ^c	1.159	0.914	0.913

^a Standard deviation of the prediction error.

^b Mean absolute value of the prediction error.

^c Look-ahead time in minutes

REFERENCES

[1] FAA, "Aviation System Performance Metrics (A.S.P.M.)." vol. 2007 Washington, D.C., 2007.

[2] FAA, "FAA Aerospace Forecasts: Fiscal Years 2006-2017," in *31st Annual FAA Aviation Forecast Conference*, 2006.

[3] Joint Planning and Development Office, "Next Generation Air Transportation System: Integrated Plan," 2004.

[4] C. Wanke and D. Greenbaum, "Incremental, Probabilistic Decision Making for En Route Traffic Management," in *7th U.S.A./Europe Air Traffic Management R&D Seminar* Barcelona, Spain: EUROCONTROL/FAA, 2007.

[5] Volpe National Transportation Systems Center, "Enhanced Traffic Management System (ETMS) Reference Manual Version 7.5," U.S. Department of Transportation VNTSC-DTS56-TMS-004, November 2002 2002.

[6] C. Wanke, D. Greenbaum, S. Mulgund, and L. Song, "Modeling Traffic Prediction Uncertainty for Traffic Management Decision Support," in *AIAA*

Guidance, Navigation, and Control Conference and Exhibit Providence, Rhode Island, 2004.

[7] K.S. Lindsay, D.P. Greenbaum, and C.R. Wanke, "Predeparture Uncertainty and Prediction Performance in Collaborative Routing Coordination Tools," *Journal of Guidance, Control, and Dynamics*, vol. 28, pp. 1178-1186, 2005.

[8] J. Krozel, D. Rosman, and S.R. Grabbe, "Analysis of En-Route Sector Demand Error Sources," in *AIAA Guidance, Navigation, and Control Conference* Monterey, California, 2002.

[9] S.C. Mohleji and N. Tene, "Minimizing departure prediction uncertainties for efficient RNP aircraft operations at major airports," in *6th AIAA Aviation Technology, Integration, and Operations Conference* Wichita, Kansas, 2006.

[10] L.A. Meyn, "Probabilistic Methods for Air Traffic Demand Forecasting," in *AIAA Guidance, Navigation, and Control Conference* Monterey, California, 2002.

[11] K.T. Mueller, J.A. Sorensen, and G.J. Couluris, "Strategic Aircraft Trajectory Prediction Uncertainty and Statistical Sector Traffic Load Modeling," in *AIAA Guidance, Navigation, and Control Conference* Monterey, California, 2002.

[12] P.K. Menon, G.D. Sweriduk, and K.D. Bilimoria, "New Approach for Modeling, Analysis, and Control of Air Traffic Flow," *Journal of Guidance, Control, and Dynamics*, vol. 27, pp. 737-744, 2004.

[13] A.M. Bayen, P. Grieder, G. Meyer, and C.J. Tomlin, "Lagrangian Delay Predictive Model for Sector-Based Air Traffic Flow," *Journal of Guidance, Control, and Dynamics*, vol. 28, pp. 1015-1026, 2005.

[14] D. Dugail, E. Feron, and K.D. Bilimoria, "Conflict-Free Conformance to En Route Flow Rate Constraints," in *AIAA Guidance, Navigation, and Control Conference* Monterey, California, 2002.

[15] P.K. Menon, G.D. Sweriduk, and K.D. Bilimoria, "Computer-Aided Eulerian Air Traffic Flow Modeling and Predictive Control," *Journal of Guidance, Control, and Dynamics*, vol. 29, pp. 12-19, 2006.

[16] B. Sridhar, T. Soni, K. Sheth, and G. B. Chatterji, "Aggregate Flow Model for Air-Traffic Management," *Journal of Guidance, Control, and Dynamics*, vol. 29, pp. 992-997, 2006.

[17] C. Gwiggner and V. Duong, "Average, Uncertainties and Interpretation in Flow Planning," in *2nd International Conference on Research in Air Transportation (ICRAT)* Belgrade, Serbia, 2006.

[18] FAA, "Airline Service Quality Performance (A.S.Q.P.)," Washington, D.C., 2006.

[19] R. S. Tsay, *Analysis of Financial Time Series*. Hoboken, N.J.: John Wiley & Sons, Inc., 2005.

- [20] W. H. Greene, *Econometric analysis*, 5 ed. Upper Saddle River, N.J.: Prentice Hall, 2003.
- [21] "S.A.S. 9.1.3 Help and Documentation," SAS Institute Inc., 2004.
- [22] "RAMS Plus Data Manual," I.S.A. Software Ltd., 2004.
- [23] EUROCONTROL, "Aircraft Performance Summary Tables for the Base of Aircraft Data (B.A.D.A.) Revision 3.5," European Organisation for the Safety of Air Navigation, Bretigny-sur-Orge, France EEC Note No. 09/03, 2003.
- [24] D. C. Champeney, *A handbook of Fourier theorems*. New York, N.Y.: Cambridge University Press, 1987.

MODELING THE OPERATIONAL IMPACT OF AIR TRAFFIC CONTROL AUTOMATION TOOLS:

A Case Study of Traffic Management Advisor

Megan Smirti, Mark Hansen
University of California at Berkeley
NEXTOR
Berkeley, CA, USA
msmirti@berkeley.edu

Xing Chen
CSSI, Inc
Arlington, VA, USA

Abstract— Traffic Management Advisor (TMA) is a decision support tool developed to assist Traffic Management Units (TMU) in metering and sequencing arrival traffic. This study examines the use and impact of TMA during its early stages of deployment at Chicago Center (ZAU). Determining impacts of use presents a methodological challenge because usage may depend on weather and traffic conditions, possibly leading to spurious results if simple with/without comparisons are made. In an effort to isolate the impact of TMA, this study therefore employs an alternate method. A preliminary understanding of TMA use is established through summary statistics. This enables the development and use of detailed statistical models to isolate the impact of TMA at ZAU. We find evidence through these detailed models that TMA use increased capacity in specific conditions and capacity variability was reduced in all scenarios. A simulation of these results on delay at Chicago O'Hare International Airport (ORD) showed that TMA use can decrease delay by 33%.

Keywords: Air Traffic Management, Capacity, Traffic Management Advisor

I. INTRODUCTION

The Federal Aviation Administration (FAA) developed the Free Flight Phase 1 (FFP1) program with the goal of automating certain functions of air traffic control to improve performance of the National Airspace System (NAS). The FFP1 program established metrics used to evaluate system deployments, which assisted the FAA in performing tests and evaluations before undertaking widespread deployment of the tools. Tools analyzed in recent years include User Request Evaluation Tool (URET) and Traffic Management Advisor (TMA), which is the focus of this paper. TMA is part of a suite of tools that was planned to increase the efficiency of flight operations in all five domains of the NAS [1].

As discussed by Hansen [2], air traffic control system evaluations present a unique challenge. Because the NAS is affected by many diverse factors, such as weather and demand, isolating the impact of a specific air traffic control enhancement is complicated. The challenge is even more difficult during early stages of deployment when the tool is

used only in selected time periods, which may be different from non-use periods in some systematic ways. In this study, to isolate the impact of TMA on airport operational capacity, we extend an econometric modeling method developed in [3] that considers capacity as a random variable. Our work contributes to the development of consistent and credible evaluation methods for automation tools, which will become increasingly important as NAS modernization proceeds.

Section II of this paper provides background on TMA, describes its functionality, and discusses previous benefit studies. Section III introduces summary statistics to aid in understanding how TMA is used, and describes the data used in the analysis. Section IV defines an econometric model used to determine the impacts of TMA implementation and presents estimation results. Section V isolates the capacity and variance of capacity effects of TMA to determine a change in delay from TMA use. Section VI concludes the research with discussion and recommendations.

II. TMA BACKGROUND

The role of TMA is to coordinate the transition between center and control airspace for arrivals. TMA was designed for decision support for the metering position of the Traffic Management Coordinators (TMC). However, as discussed by Bolic [4], the adaptation, or actual use instead of intended use, of systems developed for air traffic controllers (ATC) and traffic management coordinators (TMC) often diverges from the intended purpose. For example, at Los Angeles center, TMA was initially used to display traffic in a larger area than was previously available [2]. This increased "shared situational awareness" generated considerable operational benefit even when the decision support functionality was not in use.

TMA began initial daily use (IDU) at ZAU in June 2005. Adaptation also took place at ZAU, as TMA was used exclusively to facilitate the release of internal departures – those bound for an airport within the same Air Route Traffic Control Center (ARTCC) airspace. The TMA display screen is well suited to this function because of a detailed arrival

schedule for the major airports in the Chicago TRACON—ORD and Chicago Midway.

Implementation at ZAU followed the successful implementation of TMA at eight ARTCCs, with the first implementation in 1996 at Fort Worth. Later implementations were supported by studies finding benefits from TMA implementations at Fort Worth and other centers. These benefit studies relied on before and after analysis, including summary statistics and regression modeling. Two examples of such studies are below.

A. TMA at Minneapolis Center (ZMP)

Through a comparative analysis of airport acceptance rates (AAR) before and after TMA deployment, the FFP1 program office determined that TMA increased AAR at ZMP. A regression analysis was then performed to isolate the impact of TMA. By defining AAR as a function of TMA, metrological condition, and runway interaction, it was found that the increase in the AAR mean was not statistically significant. This regression treated TMA as a dummy variable which was set to 1 to signify a time period after TMA was deployed.

A similar study was performed regarding the total operations rate, or the sum of the airport acceptance and airport departure rates. This analysis found a statistically significant increase in the operations rate after TMA was deployed. It was concluded that optimized arrivals flows under TMA allowed the controllers to release more aircraft [5].

B. TMA at Los Angeles Center (ZLA)

The impact of TMA on internal release departures to LAX from other airports within ZLA was examined after the June 2001 TMA implementation. Similar to ZAU, TMA allowed the Traffic Management Unit (TMU) at ZLA to optimize the release of these departures by fitting them in to the arrival stream without causing delays. By calculating the mean delay before and after the deployment of TMA, it was found that both gate and airborne delay decreased after TMA deployment. It was concluded that because other airports experienced increases in gate and airborne delay for the same time period, TMA was able to reduce delay at LAX [6]. This study did not include a regression analysis and did not consider other factors which could have contributed to a decrease in delay, such as changes in demand.

III. EXPLORATORY TMA ANALYSIS

For the purpose of modeling the impact of TMA on airport runway capacity, the operational impact at Chicago O'Hare International Airport (ORD) was chosen for case study. Data were collected for the study period of July 2005, immediately after IDU of TMA, to mid-March 2006.¹ Data were gathered from the FAA's Aviation System Performance Metrics (ASPM) database. The "Airport Efficiency" portion of this database provides variables on quarterly-hour arrival and departure count and "demand" at ORD, which will be explored in greater detail in Section IV. Each entry includes corresponding information about the meteorological condition

¹ The period from December 19 to 25 was excluded, because schedules and operations are substantially changed by large volumes of holiday travel.

(MC), other weather related information, and runway configuration.

A TMA usage log was collected from ZAU to match the periods in ASPM with the periods when TMA was explicitly being used by the TMCs. During the study period, TMA was powered on and available for use from 6AM to 8PM daily. However, TMA was referred to sporadically by the TMCs; the times when TMA was assisting TMCs was recorded in a usage log [7]. To combine these data with ASPM data, time stamps on each of the data sets were matched.

A. TMA Use at ZAU

The following summarizes TMA usage data with the goal of gaining a general understanding of the factors affecting use of TMA during the study period. Discussions with TMCs, managers, and consultants supporting TMA implementation at ZAU revealed the policies and procedures affecting TMA use was sporadic; therefore, this study will focus on TMA usage periods rather than before and after TMA deployment periods. To determine the best model formulation, correlations between TMA use, meteorological conditions, and runway configuration are explored.

1) Meteorological Conditions

Table I summarizes TMA use in terms of visibility conditions at ORD. The three meteorological conditions classified are visual meteorological conditions (VMC), marginal visual meteorological conditions (MVMC), and instrumental meteorological conditions (IMC) [8]. Each quarter-hour data entry in ASPM is identified as either VMC or IMC. We further subdivided VMC into MVMC and "full" VMC, based on visibility criteria defined in [8].

TABLE I. CEILING AND VISIBILITY AVERAGES, BY METEOROLOGICAL CONDITIONS AND TMA USE

	IMC		MVMC		VMC	
	OFF	ON	OFF	ON	OFF	ON
Ceiling (100's Ft)	8.95	16.02	19.96	27.64	12.92	13.2
Visibility (statute mi)	3.14	2.06	7.93	7.94	9.57	9.71
no of obs. with TMA	54		165		1254	
total no of obs.	1694		2445		19362	

From Table I it can be seen that during the study period there were very few observations of TMA use under IMC. Out of the 1473 periods that TMA was used, only 3.67% (54 periods) were during IMC. For those few periods when TMA was used during IMC, it was typically during high ceiling conditions. The average ceiling condition under IMC and TMA use was almost double that of the average ceiling condition under IMC with no TMA use. Conditions under MVMC and VMC when TMA was and was not in use are more similar, although under MVMC the ceiling is considerably higher when TMA is in use.

2) Runway Configurations

Chicago O'Hare International Airport has 6 active runways in 3 pairs of parallel runways. There are a number of possible runway configurations at ORD for arrivals and departures that can be used at any given time. The five most frequently used runway configurations for arrivals and departures are shown Table II, along with the proportion of time each is used and proportion of TMA use. The configuration 4R, 9L, 9R | 4L, 9L, 32L, 32R is known as the default configuration for VMC and MVMC.

TABLE II. FIVE MOST COMMON RUNWAY CONFIGURATIONS AT ORD

Configuration	% of Periods Configuration Used	% of periods TMA Used
22R, 27L, 27R 22L, 32L, 32R	40.19	6.64
4R, 9L, 9R 4L, 9L, 32L, 32R	36.22	6.95
22R, 27L 22L, 32L, 32R	6.96	10.49
14R, 22L, 22R 9L, 22L, 27L	13.07	3.39
9R, 14L, 14R 4L, 9L, 22L	3.56	1.61

During the study period, TMA was "certified" on two runway configurations: 22R, 27L, 27R | 22L, 32L, 32R and 4R, 9L, 9R | 4L, 9L, 32L, 32R (referred to as configuration 1 and 2, respectively). This means that for these configurations TMA predicts the time when flights reach ORD entry fixes with sufficient accuracy. TMA was most likely used for these configurations and for 22R, 27L | 22L, 32L, 32R. This two arrival runway configuration was favored for two possible reasons. First it is very similar to the certified configuration 1. Second, TMCs noted that TMA did not "recognize" the third runway in configuration 22R, 27L, 27R | 22L, 32L, 32R when scheduling internal departures, a problem that did not arise when just two arrival runways were in use.

IV. ECONOMETRIC MODELING OF CAPACITY UNDER TMA

The following section introduces the econometric modeling technique used to model and determine the impact of TMA on operational capacity at ORD. This technique is based on the model developed by Hansen [3] to determine the capacity impact of new runway development.

A. Count and Demand Data Analysis

To accurately determine the capacity impact of TMA, the operations rate (operation count per unit time) is compared with operation demand per unit time. The data are divided into two groups based on TMA use; data for periods when TMA was in use are separated from data collected when TMA was not in use.

Data from ASPM were used for this analysis. The variable arrival (departure) count in the ASPM database indicates the number of arrivals (departures) in a time period (defined as a 15 minute interval). The variable arrival (departure) demand represents the number of aircraft scheduled to arrive (depart) in a specific time period. While demand for an operation often

leads to that operation occurring, scenarios exist where the arrival (departure) demand exceeds the arrival (departure) capacity, or the maximum number of aircraft that can perform the operation in a given period. In this case, some aircraft will be queued. Aircraft counting toward the demand in a given period that do not actually arrive (depart) in that period are counted toward demand in the subsequent period. Thus the difference between count and demand in a given period is essentially the size of the queue at the end of that period.

To measure demand, ASPM determines the expected arrival time of an aircraft by adding the en-route time to the wheels-off time. An arrival in a time period before the calculated time is counted towards the demand in the earlier period in which it arrives; an arrival at the calculated time is counted toward the demand for that period; and an arrival after the calculated time is counted toward the demand in all time periods between the calculated arrival and the actual arrival time. Departure demand is calculated similarly, based on the actual pushback time plus an airport-specific unimpeded taxi time, or when a flight is subject to a ground delay program (GDP), the estimated time when the flight will be cleared for departure under the GDP.

The model developed for this study will use the data to determine the change in capacity for arrivals only due to TMA use. A model is constructed which treats capacity as a random variable, by calculating capacity as a function whose distribution depends on weather, runway configuration, demand, and TMA use. This methodology uses statistical procedures that estimate the relationship between these factors and capacity.

To isolate the impact of TMA, the capacity function includes a dummy variable which is set to 1 if TMA is in use in time period t , and it is set to zero otherwise. The parameter of primary importance is the coefficient on the dummy variable representing TMA use. This parameter is the contribution to capacity of TMA. If the coefficient is negative, it can be concluded that TMA reduces capacity; if it is positive, it can be concluded that TMA increases capacity. This coefficient for operation type O (where O = arrivals only for this study) will be termed β_0 .

The example in Fig. 1 depicts β_0 . The solid curve is a sample probability distribution of runway capacity. The second dashed curve is a sample probability distribution for runway capacity when TMA is in use, but when other conditions (weather, etc.) are similar. The difference in the mean values of these curves, represented by the curve peaks, is β_0 . Fig. 1 depicts a case when TMA use affects only the mean of the capacity distribution. TMA use may also affect the variance of the capacity distribution by consistently feeding traffic to the airport at a more consistent rate. Both effects are considered below in section B.

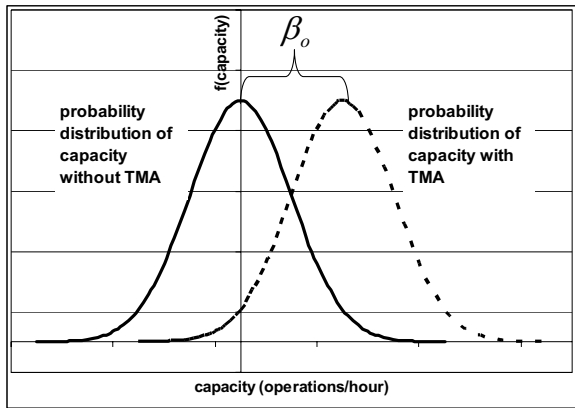


Figure 1. Depiction of β_0 , the contribution of TMA to capacity.

B. Operational Impact: Censored Regression Model

The model to be used in this section is a censored regression, or tobit, model, which measures the difference in capacity due to TMA use. A censored regression model is appropriate because it is impossible for a count value to exceed a demand value. Throughput, or runway operations per unit time, is therefore censored by demand.

The tobit model formulation is below. The model will calculate the capacity based on the known operation demand and the known operation count. To isolate the impact of runway configuration and meteorological condition, there are separate models for each configuration and condition. We estimated the model for 4 different data sets. Models were estimated for VMC and MVMC and for runway configurations 1 (22R, 27L, 27R | 22L, 32L, 32R) and 2 (4R, 9L, 9R | 4L, 9L, 32L, 32R). Each model considers capacity as a function of demand, windspeed, and TMA use. Each model also captures the variance of capacity, and analyzes the impact of TMA on this variance.

The model specification is below.

$$\begin{aligned}
 Q_o(t) &= \min(D_o(t), C_o(t)) \\
 C_o(t) &= \alpha_o + \beta_o A(t) + \gamma_o W(t) + \tau_o D_o(t) + \varepsilon_o
 \end{aligned}
 \tag{1}$$

Where:

$Q_o(t)$	is the count for operation of type o (either arrivals or departures) in 15-minute time period t;
$D_o(t)$	is the demand for operations of type o in time period t;
$C_o(t)$	is the ORD capacity for operations of type o in time period t;
$A(t)$	is equal to 1 if TMA is in use in time period t and 0 otherwise;
$W(t)$	is the windspeed in time period t;
ε_o	is a stochastic error term, assumed to be IID normal with mean 0 and variance $\sigma_o^2 + \rho_o A(t)$;
$\alpha_o, \beta_o,$ $\gamma_o, \tau_o,$ σ_o^2, ρ_o	are parameters to be estimated.

The model is estimated using a maximum likelihood method, which will find the parameters that best fit the data. Mainly, we are interested in β_0 and ρ_0 , the effects of TMA on the mean and the variance of the capacity distribution. The detailed model estimation technique is discussed in great depth by Hansen [3].

1) Illustration of Censored Regression Model Results

For illustrative purposes, the full model results for one data set will be described in detail. We chose the model for VMC conditions and runway configuration 1 for this illustration. Estimation results appear in Table III.

TABLE III. CENSORED REGRESSION MODEL RESULTS

Parameter	Symbol	Estimate (Standard Error) T-Statistic
Intercept	α_o	26.164 (0.333) 78.517
Effect of TMA on capacity	β_o	1.720 (0.412) 4.17214
Effect of Windspeed	γ_o	-0.201 (0.026) -7.797
Effect of Demand	τ_o	0.000 (0.000) -0.055
Variance	σ_o^2	5.994 (0.101) 59.338
Effect of TMA on Capacity Variance	ρ_o	-1.267 (0.310) -4.089

The model results show that the baseline quarter-hour capacity for arrivals at ORD is 26.164 arrivals, which is the equivalent of 104.656 arrivals per hour. This is very close to the benchmarked 100 arrivals per hour determined by the FAA [9]. The results also show that when TMA is being used by the TMCs, arrival capacity is increased by 1.720 arrivals per quarter hour, or 6.880 arrivals per hour. This is equivalent to a 6.6% capacity increase. The results show that windspeed decreases arrival capacity by -.201 arrivals per quarter hour, and that demand has no impact on capacity. The estimated variance is 5.994 arrivals per quarter hour squared, which is decreased by -1.267 when TMA is in use. All parameters except demand are significant at the 0.05 level (denoted by the boldface type).

2) Model Results for the Impact of TMA on Arrival Capacity and Variance of Capacity

The impact of TMA on the capacity mean, measured by β_0 , and capacity variance, ρ_0 , for the four sets of MC and runway configuration are shown in Table IV.

TABLE IV. THE EFFECT OF TMA ON CAPACITY MEAN AND CAPACITY VARIANCE

MC & RW Configuration	β_o Values (Standard Error) T-Statistic	ρ_o Values (Standard Error) T-Statistic
VMC, RW 1	1.720 (0.412) 4.172	-1.267 (0.310) -2.976
VMC, RW 2	.302 (.357) .846	-1.892 (.345) -5.479
MVMC, RW 1	.822 (.914) .899	-1.554 (.642) -2.420
MVMC, RW 2	-1.318 (.961) -1.372	-1.784 (.718) -2.485

The β_o values are not significant in three of four meteorological conditions and runway configuration cases. Under VMC and runway configuration 1, capacity mean is significantly higher due to TMA. There is a possible “self-selection” bias in this case because it represents favorable conditions, which could encourage TMA use.

The ρ_o values indicate the estimated change in capacity variance when TMA is in use. The results suggest that arrival capacity variance did decline when TMA was in use. We also note that these results are consistent with the FFP1 LAX study [6], which found less dispersion between arrival counts and throughput after TMA was implemented.

If TMA usage at ZAU did in fact reduce arrival capacity variance, this would have an important benefit. It would reduce delay, because a negative capacity deviation is more likely to have an adverse effect than is positive deviation to have a beneficial effect. In many cases, positive deviations cannot be fully exploited because there is insufficient demand. While a negative deviation can also be inconsequential, it is more likely to contribute to a queue going into the next period.

The following section explores how the use of TMA can affect delay due to its capacity and variance impacts.

V. DELAY IMPACT ESTIMATION

To illustrate the potential of TMA use to save minutes of flight delay, a simulation was employed. The operational count if TMA was in use 100% of the time was simulated and compared with operational count if TMA had never been in use during the study period. To further isolate the capacity and variance effects of TMA, two potential operational count scenarios were calculated: one with the capacity effect of TMA calculated alone ($\rho_o=0$), and another with both the capacity and capacity variance effect. Operational demand was kept constant over all scenarios to fully illustrate the delay changes due to TMA.

A. Delay Calculation without TMA

Using demand and count data for all quarter hour periods at ORD collected for January 2006, a cumulative count curve was constructed. A cumulative curve in this case is a plot of

cumulative operational count on the y-axis and time on the x-axis. In the first period, cumulative operational count (n_1) is equal to the count of operations in period one (n_1'). In the second period, cumulative operational count (n_2) is the count in period two (n_2'), plus the count in period one (n_1'). Therefore the cumulative operational count in period two is $n_2=n_1+n_2'$. The count in period three is $n_3=n_2+n_3'$, and so on for all remaining periods. Cumulative demand is determined similarly.

The horizontal distance between any two points on the curves is equal to the wait time in queue that an operation (arrival) was delayed. The area between the two curves is the delay in flight-minutes for the time period of study.

To illustrate how this method can be used to determine the delay savings potential of TMA, the study period of January 6, 2006 from 13:15-21:15 was chosen. The first step was to construct the curves of cumulative demand and cumulative count in the “without TMA” scenario for this period. These curves can be seen in Fig. 2.

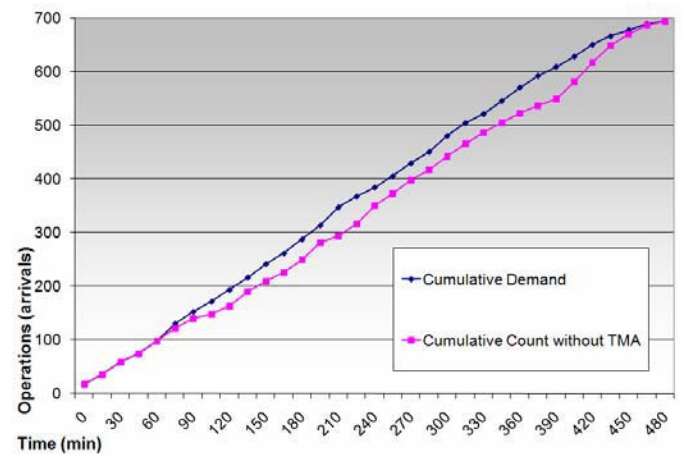


Figure 2. Cumulative Demand and Count: Without TMA Scenario

The area between the two curves, or the study period delay in flight-hours, is equal to 225.9 flight-hours.

B. Delay Calculation with TMA

To simulate and isolate the capacity effect and the variance of capacity effect of TMA, cumulative curves were constructed for the two scenarios. The estimated parameters of the capacity function from (1) were used to calculate the new capacity. The parameters of the best fit models are in Table V.

TABLE V. CAPACITY ESTIMATION EQUATION PARAMETERS

	Capacity Mean				Capacity Variance	
	α	β_0	γ	τ	σ^2	ρ_0
VMC, 1	26.164	1.720	-0.201	0.000	5.994	-1.267
VMC, 2	21.696	0.302	0.114	0.112	6.333	-1.892
MVMC, 1	28.151	0.822	-0.560	-0.021	6.028	-1.554
MVMC, 2	25.949	-1.318	-0.126	-0.054	5.740	-1.784

The following sections describe how the capacity effect and the variance of capacity effect were determined.

1) Simulation of TMA Capacity Effect

To isolate the effect on capacity of TMA, capacity was calculated as a function of the parameters in Table V depending on the MC and runway configuration. Capacity in each period was assumed to be normally distributed with mean

$$\mu_C = \alpha_0 + \beta_0 A(t) + \gamma_0 W(t) + \tau_0 D_0(t) \tag{2}$$

and variance $\sigma^2 + \rho_0$, where $\rho_0 = 0$. Capacities for each quarter hour period in the study period were then drawn from this distribution. Next, as in (1), the operational count was calculated as the minimum of the capacity and the operational demand. The unserved operations in any period were added to the operational demand of the next period.

The simulated cumulative operational count curve represents the operational count that would have been achieved if TMA was in use during the entire study period, but only the capacity effect of TMA was realized. The cumulative count of operations with the TMA capacity effect is shown below in Fig. 3, along with the cumulative count without TMA and the cumulative demand.

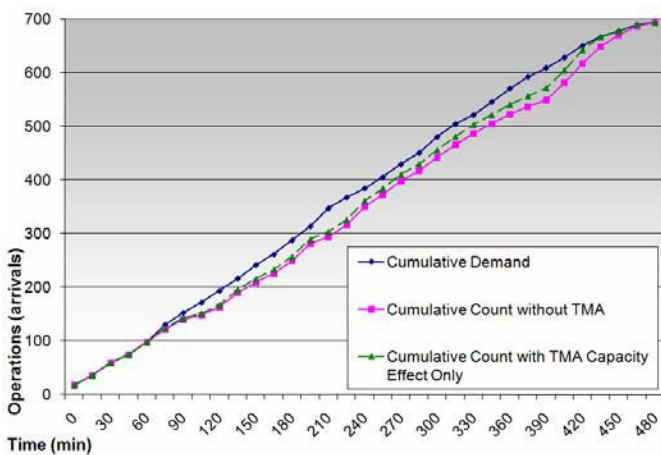


Figure 3. Cumulative Demand and Count: TMA Capacity Effect Only Scenario

The delay calculated for the TMA capacity effect only scenario was 147.7 flight-hours which is a delay savings of 78.1 flight-hours over the scenario when TMA is never in use.

2) Simulation of TMA Variance of Capacity Effect

To simulate the variance of capacity effect, the capacity effect along with the variance of capacity effect was calculated. The same method was used as for the TMA capacity effect only scenario. Capacity was assumed to be normally distributed with mean μ_C and variance $\sigma^2 + \rho_0$, where ρ_0 is the associated value for each MC and runway configuration from Table V. The cumulative operational count for the TMA capacity and capacity variance scenario can be seen in Fig. 4.

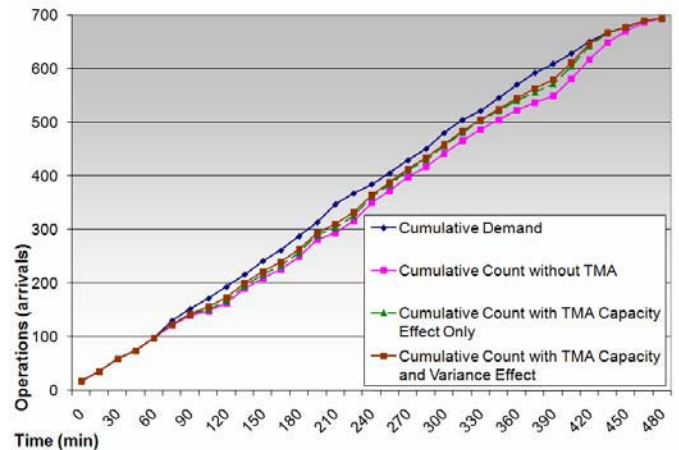


Figure 4. Cumulative Demand and Count: TMA Capacity Variance Effect

The delay for the TMA capacity and capacity variance effect was 121.0 flight-hours, which is a savings of 26.7 flight-hours as compared with the TMA capacity effect only scenario and an overall delay savings of 104.9 flight-hours.

Using the same method for the entire month of January 2006, if TMA had been in use 100% of the time, TMA would have saved 750 flight-hours of delay for arrivals compared to the “without TMA” scenario. Of these 750 flight-hours, 500 flight-hours of savings were due to capacity effect, and 250 flight-hours of savings were due to variance effect. This finding generalizes to a savings in delay of 9,000 flight-hours per year and about 10 seconds per flight.

VI. CONCLUSIONS

This study found that the use of TMA for releasing internal departures appears to have decreased capacity variance and in some cases increased capacity mean. Using the model results, it was found that increased use of TMA could lead to decreased delay of about 10 seconds per flight.

Additionally, we have furthered the use of censored regression applied to ASPM data as an evaluation method for ATM tools. In particular, we have shown how this method can be used to investigate the effect of new tools on the variance of capacity as well as its mean. In our particular case, we find that TMA use, even though it was restricted to releasing internal departures, had a measurable impact on arrival capacity variance at ORD.

Further study is necessary to assess the impact of TMA when it is used for time based metering. Time based metering went into effect in June 2007 at ZAU, and could decrease the variance in capacity by allowing controllers to effectively

manage capacity especially during high traffic periods. Understanding the impact of TMA on capacity and capacity variance due to time based metering, and comparing these findings with those in this study, would provide insight into the benefits of TMA when it is employed for its full range of uses rather than used only for more limited, adapted purposes.

REFERENCES

- [1] United States General Accounting Office. (2001). National Airspace System: Free Flight tools show promise, but implementation challenges remain. Available at: <http://www.gao.gov/new.items/d01932.pdf> Accessed on January 4, 2007.
- [2] Hansen, M., Mukherjee, A., Knorr, D., and Howell, D. Effect of T-TMA on capacity and delay at Los Angeles International Airport, *Transportation Research Record 1788*, 2002, pp. 43-48.
- [3] Hansen, M. (2004). Post-deployment analysis of capacity and delay impacts of an airport enhancement: Case of a new runway at Detroit. *Air Traffic Control Quarterly*, 12 (4).
- [4] Bolic, T. and Hansen, M. User Request Evaluation Tool (URET) adoption and adaptation; Three center case study, in *Proceedings of 4th USA/Europe Air Traffic Management R&D Seminar*, 2005.
- [5] Free Flight Program, Performance metrics results to date. June 2001 Report. Received from FAA.
- [6] Free Flight Program, Performance metrics results to date. June 2003 Report. Received from FAA.
- [7] Michalak, D. Traffic Management Lead at Chicago Center. Personal Interview. December 1, 2006.
- [8] de Neufville, R. and Odoni, A. 2003 Airport systems planning, design and management. McGraw-Hill.
- [9] Federal Aviation Administration. Airport capacity benchmark report 2004: Chicago O'Hare. Available at: <http://www.faa.gov/events/benchmarks/>. Accessed on January 4, 2007.

Track 2

CNS/ATM

Limitations of Subliminal Control in Air Traffic Management

Georgios Chaloulos
Automatic Control Laboratory
Department of Electrical Engineering
ETH Zürich
Zürich, Switzerland
Email: chaloulos@control.ee.ethz.ch

John Lygeros
Automatic Control Laboratory
Department of Electrical Engineering
ETH Zürich
Zürich, Switzerland
Email: lygeros@control.ee.ethz.ch

Abstract—An air traffic control concept under the name of *Subliminal Control* has been introduced. In this approach, an automated system, commanding minor speed adjustments imperceptible by the Air Traffic Controller, tries to keep the Air Traffic Controller's risk perception low, emulating a "lucky traffic". In this paper, we investigate the limits of this air traffic control approach. We test a proposed subliminal controller against several encounter geometries for level flights. A stochastic environment using wind forecast uncertainties is used for this purpose. The results demonstrate the cases where subliminal control can potentially reduce the workload of the ATC.

I. INTRODUCTION

The current Air Traffic Management (ATM) system is to a large extent based on a rigidly structured airspace and a mostly human-operated system architecture [1], [2]. For the separation assurance between aircraft, Air Traffic Controllers (ATC) have to make decisions under a highly uncertain and complex environment. To do so, they have to estimate the future positions of aircraft and intervene whenever they perceive a high risk of loss of separation. It is obvious that the projected traffic increase [3], [4] demands an increase in the number of aircraft per sector. This will result in more stress on the ATC. To alleviate some of this workload, several potential solutions have been proposed, including conflict detection and resolution algorithms (for a thorough overview and classification of the literature, the reader is referred to [5]).

An alternative solution was proposed in [6] under the name *Subliminal Control*. The premise is that minor speed adjustments, commanded by an automated system running in parallel with the ATC, can convert a potentially conflicting situation into a "lucky traffic" for the ATC, in the sense, that the trajectories turn out to ensure safe separation at an early stage, reducing the ATC's monitoring workload. These speed adjustments have to be as small as possible in order to remain imperceptible by the ATC. In this approach the human is still kept in the loop, and automation is introduced in a user-friendly way. Crück and Lygeros in [7] presented a mathematical framework for subliminal control, while in [8], a hybrid dynamical game is proposed in which the control has to minimize a cost representing the risk perceived by air-traffic controllers despite the uncertainty of trajectory prediction.

In this paper, we investigate the limits of subliminal control method for Air Traffic Control. Subliminal control is tested against several conflict encounters under stochastic environment, due to the presence of wind forecast uncertainty.

The paper is organized as follows: Section II briefly introduces the subliminal control concept, Section III describes the modeling of the risk perception of the ATC, Section IV discusses the flight simulation model, Section V presents the simulations results of this study and Section VI states the conclusions of this work.

II. SUBLIMINAL CONTROL

The main idea of subliminal control is to turn ATC's uncertainty about traffic evolution into an advantage. It has been shown that small adjustments of speeds commanded early enough can prevent a large percentage of conflicts [9]. Here we consider speed resets small enough to be within the uncertainty margin of the ATC (and hence, in principle, imperceptible). Results from the experiments of the European project ERASMUS [10] indicate that speed variations up to 12% may go unnoticed by the ATC.

For the subliminal control concept, instead of detecting conflicts and then resolving them, the problem considered is:

- 1) Predict the risk the ATC will perceive in the near future, when faced with a given traffic situation.
- 2) Reduce (if necessary) the risk perception by applying unnoticeable speed changes.

The function we use to compute the risk perception is described in Section III. Given this function, we assume that the automated system can predict the traffic with sufficient accuracy for a time horizon significantly longer than the ATC's "prediction horizon", i.e. the time ahead the ATC can foresee a dangerous situation. Then the task of the system will be to minimize the risk perception along all possible set of aircraft trajectories.

It should be emphasized here, that in our subliminal control setting, the system consists of two separate models: the model of the ATC, representing the risk perception at each time step and the aircraft/environment model, which is used for an accurate trajectory prediction. Thus, the model of the ATC's risk perception is used for the computation of the cost and the

aircraft/environment model for computing the control commands. Furthermore, the ATC's model introduces additional constraints for the maximum (in number and magnitude) speed change commands.

The system operates in a dynamic environment. Aircraft may enter or leave the sector of interest, deviate from their expected trajectory, etc. Due to the high uncertainties involved, it would be unrealistic to compute an optimal control valid for a very long time. Instead, we employ a receding horizon control approach, solving at each time step a finite horizon optimal control problem and re-compute a new optimal control law at each time step (or possibly whenever an event takes place, e.g. a probability of high risk perception is high).

III. RISK PERCEPTION OF THE ATC

The risk perception model we use is along the lines of the one proposed in [8]. The risk function for a given traffic situation is defined as:

$$V_{\text{ATC},n} : X_1 \times X_2 \times \cdots \times X_n \longrightarrow [0, 7],$$

where X_i denotes the state of aircraft i and n is the total number of aircraft in the traffic situation. The range of values $[0, 7]$ has been set in reference to the experimental setting in [11].

In the general situation of n aircraft, if a pair of them generates a high risk, then the whole situation will be perceived as a high risk situation. Of course, as the number of aircraft increases, the situation becomes more complex, influencing the risk perception of the ATC. Thus, we set for the risk function:

$$V_{\text{ATC},n}(X_1, \dots, X_n) = \lambda(n) \max_{i,j,i \neq j} V_{\text{ATC},2}(X_i, X_j),$$

where $\lambda(n)$ is a complexity coefficient associated with a n -aircraft situation. Since there is very few relevant data available to validate this approach, we do not pursue this aspect any further and concentrate on pairwise risk perception $V_{\text{ATC},2}(X_i, X_j)$.

The risk function for 2 aircraft is defined for the planar case as follows [8]:

$$V_{\text{ATC},2}(X_1, X_2) = \frac{b}{\max\{\text{Sep}(X_1, X_2) + cT_{go}(X_1, X_2), d\Delta\}},$$

where $\text{Sep}(X_1, X_2)$ is the minimum separation between the aircraft that the ATC expects to happen in the worst case inside his prediction horizon. T_{go} is the time at which this minimum separation occurs, Δ is the minimum prescribed separation (for the situation to remain conflict free) and b, c, d are design parameters.

Since ATC do not have a very good perception of the speeds of the aircraft, we assume they extrapolate trajectories using constant estimated speeds \hat{V}_1, \hat{V}_2 and an uncertainty margin α . The model that we use to represent their trajectory prediction process is

$$(S_{\text{ATC}}) \quad \begin{cases} \dot{x}_1(t) \in (1 + [-\alpha, \alpha])\hat{V}_1 \\ \dot{x}_2(t) \in (1 + [-\alpha, \alpha])\hat{V}_2 \end{cases} \quad (1)$$

with initial conditions $x_1(0), x_2(0)$ (the current aircraft positions).

IV. AIRCRAFT/FLIGHT ENVIRONMENT MODEL

We use the model developed in [12] to perform the simulations. This model allows one to capture multiple flights taking place at the same time. Each flight has an associated flight plan, aircraft dynamics and a Flight Management System (FMS). The evolution of flights is affected by the wind speed. The wind speed is modeled as a sum of a nominal and a stochastic part. The stochastic component is assumed to be correlated in time and space, i.e. the wind experienced by each aircraft at a given time is correlated to the wind experienced by all other aircraft at the same time and the wind experienced by all aircraft at earlier times [13]. The authors have shown in [14] that ignoring this correlation structure can result in high conflict probability estimation errors, when simulating more than one aircraft. Therefore, the evolutions of different flights are coupled to one another through the wind model.

The model is stochastic (because of the wind uncertainty) and hybrid, since it comprises both continuous and discrete dynamics; the former arise from the aircraft's physical motion, while the latter from the flight plans and the FMS.

A. Aircraft dynamics

The aircraft is modeled using a Point Mass Model (PMM), based on the Base of Aircraft Data (BADA) database [15]. The continuous dynamics for the aircraft motion are extensively described in [12]. Apart from the continuous dynamics, discrete dynamics also arise in our model, mainly because of the FMS and the flight plan.

The flight plan consists of a sequence of way-points $\{O(i)\}_{i=0}^M$, in three dimensions, $O(i) \in \mathbb{R}^3$. The sequence of the way-points defines a sequence of straight lines joining each way-point to the next. In our experiments, Requested Time of Arrival (RTA) for each way point is not implemented. As a result, the aircraft only corrects cross track deviations from the reference path, while along track errors are ignored. This assumption reflects what is known as a 3D FMS, which is currently the standard for most commercial aircraft.

The FMS can be thought of as a controller, which, by measuring the state and using it together with the flight plan, determines the values for the inputs. The control is to some extent continuous, but some parameters and set points of the controllers depend on the discrete dynamics of the FMS [12].

B. Stochastic environment

The stochasticity of our model arises because of uncertainty about the wind velocity. The wind velocity is modeled as a sum of two terms: a deterministic (nominal) component, representing the meteorological predictions available to ATC and a stochastic component, representing inaccuracy and uncertainty in these predictions. Since the meteorological predictions are known and available to the ATC before a flight takes place, the flight plans are adjusted taking them into consideration. Thus, the way the nominal wind affects aircraft trajectories is deterministic and known a priori. For simplicity reasons, we set the deterministic part of the wind to zero.

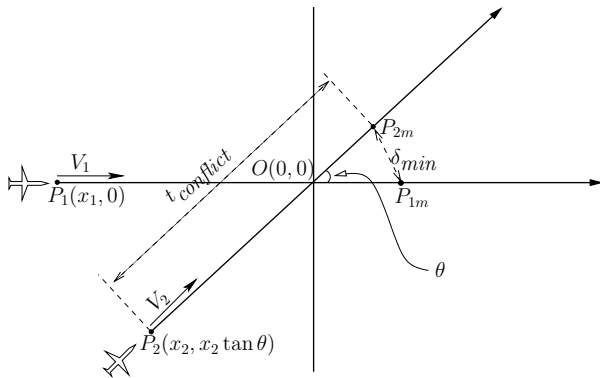


Fig. 1. Conflict scenario

The stochastic wind component is modeled as a random field $w : \mathbb{R} \times \mathbb{R}^3 \rightarrow \mathbb{R}^3$, where $w(t, P)$ represents the wind at point $P \in \mathbb{R}^3$ and at time $t \in \mathbb{R}$. We assume that $w(t, P)$ is a Gaussian random variable with zero mean. Recall that the wind experienced by each aircraft at a given time is correlated to the wind experienced by all other aircraft at the same time and the wind experienced by all aircraft at earlier times [13]. As discussed in [14], [16], this correlation structure cannot be ignored for accuracy reasons when simulating more than one aircraft. A detailed procedure for extracting wind samples with given spatio-temporal correlation can be found in [12].

V. SIMULATION RESULTS

Since subliminal control involves only speed alterations for the conflicting aircraft, it is reasonable to restrict ourselves in level flight scenarios. We consider two aircraft flying level at the same altitude, in straight lines, at constant airspeeds (see Fig. 1) without applying subliminal control. In the absence of a wind field, the minimum distance the two aircraft approach each other is denoted δ_{min} and the time this event occurs $t_{conflict}$ (time to minimum separation).

We construct flight plans to code this encounter geometry that intersect at $O(0, 0)$. $P_1(t) = (x_1(t), 0)$ and $P_2(t) = (x_2(t), x_2(t) \tan \theta)$ denote the positions of the aircraft at time t . For the simulation purposes, we use the nominal speed for an Airbus 321 cruising at 33000ft, which is 454knots [15].

We use four different values for the minimum separation δ_{min} : 0nm (where a mid-air collision would happen), 5nm, 10nm and 15nm. Three different crossing angles $\theta = (45^\circ, 90^\circ, 135^\circ)$ and 25 different values for nominal time to minimum separation $t_{conflict} = (1, 2, \dots, 25)$ minutes are considered. Even though nominally the aircraft would follow exactly their flight plans, uncertainty in aircraft motion forces them to a different minimum separation at a different time.

Concerning the risk perception model, we use $\alpha = 0.1$, $b = 49$, $c = 0$, $d = 1.4$ and $\Delta = 5$ nm. All distances are expressed in nautical miles. For prediction horizon, we assume that the ATC can predict up to 8 minutes ahead. We then say we have a *high risk* situation if $V_{ATC,2}(X_1, X_2) = 7$, a *medium risk* situation if $3.5 \leq V_{ATC,2}(X_1, X_2) < 7$, and a *low risk* situation otherwise.

To investigate the limits of the subliminal control, in all simulations, we apply the maximum speed change not perceived by the ATC (i.e. -12% or +6%) as early as possible (i.e. in the beginning of the simulation). Thus, we will try to determine how soon before an incident a speed change command has to be sent to the FMS of the aircraft. The aircraft FMS is assumed to immediately accept and implement the command. Since our system is stochastic, we perform Monte Carlo simulations to estimate the risk perceived by the ATC and the conflict probability of the aircraft by performing 1000 simulations and computing the fraction of them that enters conflict. By the term conflict we define a situation where two aircraft violate required minimum separation standards, in our case 5nm.

A. Simulations for $\delta_{min} = 0$ nm

Simulation results are shown in Figures 2-5. Figure 2 shows the probability of conflict for the three different crossing angles as a function of the time to minimum separation $t_{conflict}$. Solid lines correspond to $\theta = 45^\circ$, dotted lines to $\theta = 90^\circ$ and dashed lines to $\theta = 135^\circ$. The simulations where no speed changes are sent to the aircraft are plotted with blue color, while red color corresponds to speed change sent to only one aircraft and the simulations with speed clearances sent to both aircraft are plotted in green. One can observe that subliminal control is a good technique to solve potential conflicts up to 15 minutes before the time they would appear, sending speed clearances to both aircraft.

Figures 3-5 show the levels of the perceived risk of the ATC for the three crossing angles. Solid lines correspond to a high risk perception by the ATC (i.e. cases when the ATC would issue a conflict resolution command), dotted lines correspond to a medium risk perception (i.e. cases when the ATC would monitor the situation closely, waiting to see if it evolves into a high risk or a low risk situation) and dashed lines correspond to low risk situations (i.e. cases when the ATC would not expect the situation to evolve into an unsafe one). We observe that even though no conflict actually occurs, the risk perception of the ATC is low only in the case of $\theta = 45^\circ$ and if the subliminal controller issues speed clearance commands to both aircraft 25 minutes ahead of the expected time of the conflict. In all other cases, ATC's risk perception cannot be kept low applying subliminal control.

B. Simulations for $\delta_{min} = 5$ nm

Figures 6-9 illustrate the results for the simulations. This time, the conflict resolution can be easily handled by the subliminal controller, as changing the speed of only one aircraft is enough to resolve any conflict, even if the speed command is issued as late as only a minute before the conflict. The algorithm though is unable to keep the ATC confident that the traffic will not evolve into a conflict, unless the speed change command is issued (to both aircraft) no later than 23 minutes before the expected time of minimum separation. Thus, no flexibility for optimization between different possible solutions of the subliminal controller is left, since the controller has to

command the largest speed changes that are allowed to both aircraft.

C. Simulations for $\delta_{min} = 10\text{nm}$

This case is quite different from the previous ones, since the nominal minimum distance between the aircraft is adequate to almost ensure that no conflict will occur except for the case when aircraft are very far away and the remaining uncertainty is big (Figure 10). As expected, subliminal control guarantees in this case, too, no conflict between the aircraft. Figures 11-13 show that if a speed adjustment is sent to both aircraft at least 17 minutes before the expected occurrence of the minimum separation, the ATC will not perceive the situation as potentially dangerous. It is still required for both aircraft to adjust their speeds accordingly to avoid a medium risk situation, that would keep the ATC busy monitoring the situation, but on the other hand, an early enough decision can leave a small window for optimization depending on each aircraft's priorities (i.e. small speed adjustment vs. late speed adjustment).

D. Simulations for $\delta_{min} = 15\text{nm}$

As in the previous case, conflict avoidance is ensured in all cases, even when no speed control is applied to the aircraft (see Figure 14). This is not the case for the risk perception of the ATC though (see Figures 15-17), since a high risk perception is only avoided when the expected time to minimum separation is 5 minutes or less, which reduces the ATC uncertainty window. The risk perception can be kept low however, even by applying only one speed change, provided that it is applied around 17 minutes ahead of the expected time of the minimum separation. If both aircraft adjust their speeds, the commands can be issued just 8 minutes before the expected time of the minimum separation, leaving a big margin for an optimization depending on the aircraft's priorities.

VI. CONCLUSIONS

We have investigated the potential of the use of subliminal control to alleviate ATC's workload and monitoring of some potentially dangerous encounters. The results clearly indicate that, depending on the geometry, subliminal control can reduce the workload of the ATCs monitoring situations. Those can instead be solved early enough with minor speed adjustments, keeping the risk perception low. In all cases, care needs to be taken to ensure maneuvers remain subliminal. The accuracy of the trajectory prediction tools is also important, since more accurate tools would allow the application of subliminal control over longer horizons. As envisioned by the ERASMUS concept [10], a potential solution for this could be air-based trajectory prediction tool, that avoids radar measurement errors, and down-links the information to the ATC.

ACKNOWLEDGMENTS

This research is supported by the European Commission under the project ERASMUS, FP6-TREN-518276.

REFERENCES

- [1] C. Tomlin, "Hybrid control of air traffic management system," Ph.D. dissertation, Univ. California, Berkeley, 1998.
- [2] M. Nolan, *Fundamentals of Air Traffic Control*, 3rd ed., Belmont, CA: Wadsworth, 1998.
- [3] European Commission, "European transport policy for 2010: time to decide," 2001. [Online]. Available: http://www.europa.eu.int/comm/energy_transport/en/lb_en.html
- [4] Eurocontrol Air Traffic Statistics and Forecast Service, *Forecast of Annual Number of IFR Flights (2004-2010)*, February 2004. [Online]. Available: <http://www.eurocontrol.int/statfor/forecasts/index.html>
- [5] J. Kuchar and L. Yang, "A review of conflict detection and resolution methods," *IEEE Transactions on Intelligent Transportation Systems*, vol. 1, no. 4, pp. 179-189, 2000.
- [6] J. Villiers, "Automatisation du contrôle de la circulation aérienne - projet "ERASMUS" une voie originale pour mieux utiliser l'espace aérien," Institut du Transport Aérien - ITA, Tech. Rep. Volume 58, 2004.
- [7] E. Crück and J. Lygeros, "A Mathematical Framework for Subliminal Air Traffic Control," in *AIAA Guidance, Navigation and Control Conference and Exhibit*, Hilton Head Island, South Carolina, Aug. 2007.
- [8] —, "Hybrid modeling for the evaluation of risk perception by air-traffic controllers," in *European Control Conference*, July 2007.
- [9] J.-M. Alliot, N. Durand, and G. Granger, "A statistical analysis of the influence of vertical and ground speed errors on conflict probe," in *4th Air Traffic Management Research & Development Seminar*, Santa Fe, 2001.
- [10] E. Commission, "ERASMUS - En Route Air Traffic Soft Management Ultimate System." [Online]. Available: <http://www.atm-erasmus.com>
- [11] P. Avrty, K. Guittet, and P. Lezaud, "Perception of risks of conflict by air traffic controllers: the CREED project," DGAC/DSNA/DTI, Tech. Rep., July 2006. [Online]. Available: <http://www.dsna-dti.aviation-civile.gouv.fr/actualites/revues/revue03/03pgarticle2/03art2.htm>
- [12] I. Lympopoulos, A. Lecchini, W. Glover, J. Maciejowski, and J. Lygeros, "A stochastic hybrid model for air traffic management processes," Department of Engineering, Cambridge University, Tech. Rep. CUED/F-INFENG/TR.572, February 2007.
- [13] R. Cole, C. Richard, S. Kim, and D. Bailey, "An assessment of the 60 km rapid update cycle (RUC) with near real-time aircraft reports," MIT Lincoln Laboratory, Tech. Rep. NASA/A-1, July 15, 1998.
- [14] G. Chaloulos and J. Lygeros, "Effect of wind correlation on aircraft conflict probability," *AIAA Journal of Guidance, Control and Dynamics*, vol. 30, no. 6, pp. 1742-1752, 2007.
- [15] Eurocontrol Experimental Centre, *User Manual for the Base of Aircraft Data (BADA)*, 2004. [Online]. Available: <http://www.eurocontrol.fr/projects/bada/>
- [16] G. Chaloulos and J. Lygeros, "Wind Uncertainty Correlation and Aircraft Conflict Detection based on RUC-1 Forecasts," in *AIAA Guidance, Navigation and Control Conference and Exhibit*, Hilton Head Island, South Carolina, Aug. 2007.

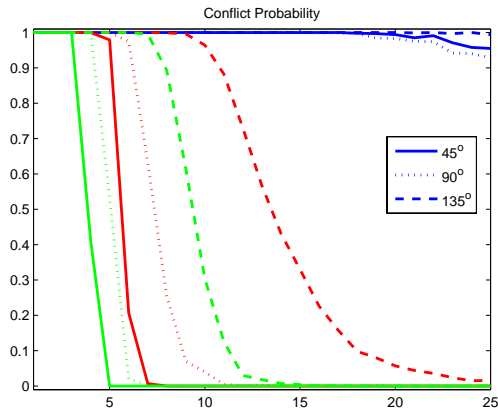


Fig. 2. Conflict Probability for $\delta_{min} = 0nm$

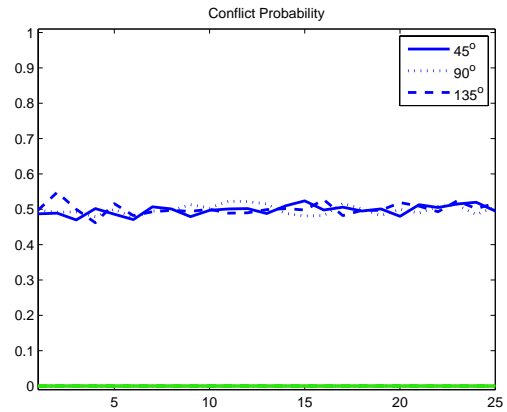


Fig. 6. Conflict Probability for $\delta_{min} = 5nm$

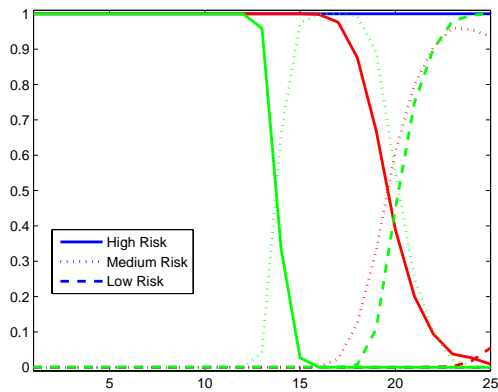


Fig. 3. Risk perceived by the ATC for $\theta = 45^\circ$

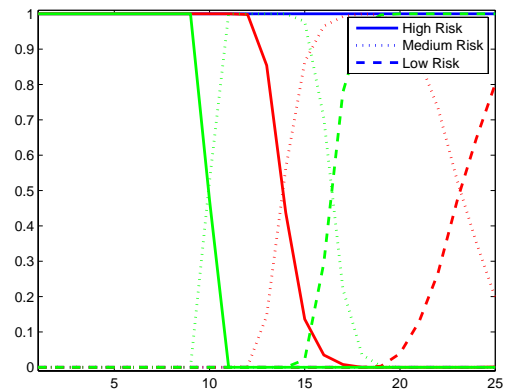


Fig. 7. Risk perceived by the ATC for $\theta = 45^\circ$

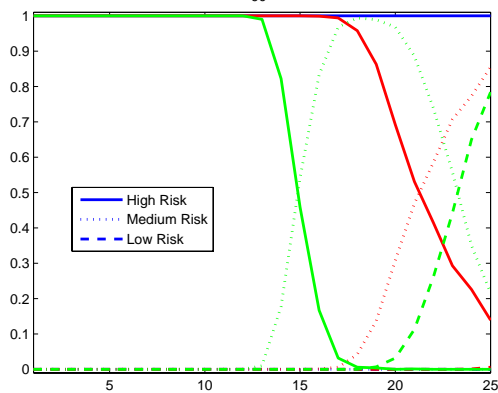


Fig. 4. Risk perceived by the ATC for $\theta = 90^\circ$

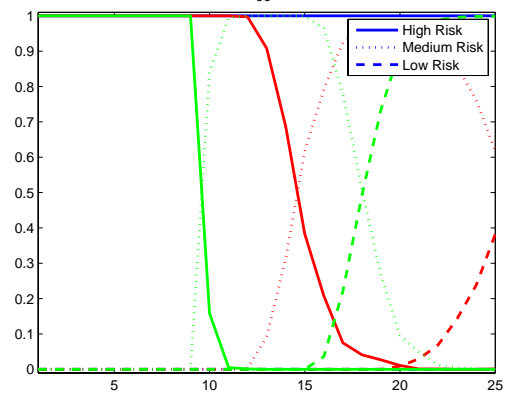


Fig. 8. Risk perceived by the ATC for $\theta = 90^\circ$

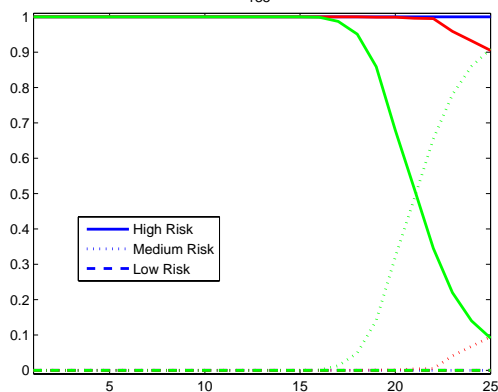


Fig. 5. Risk perceived by the ATC for $\theta = 135^\circ$

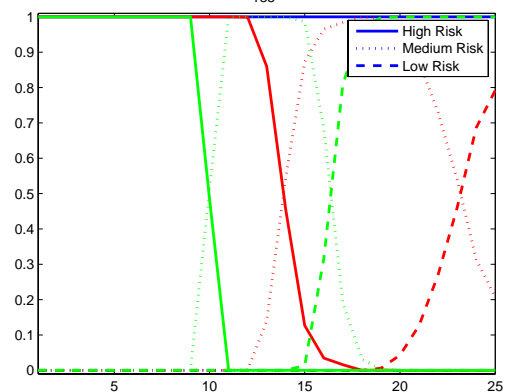


Fig. 9. Risk perceived by the ATC for $\theta = 135^\circ$

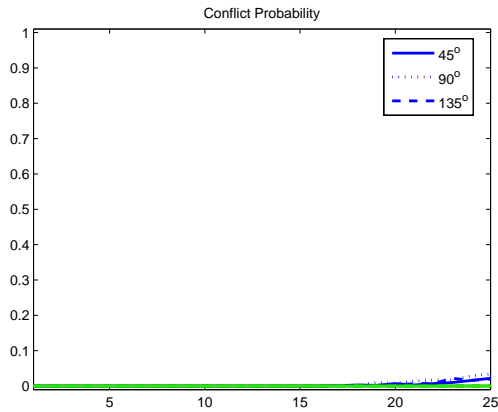


Fig. 10. Conflict Probability for $\delta_{min} = 10nm$

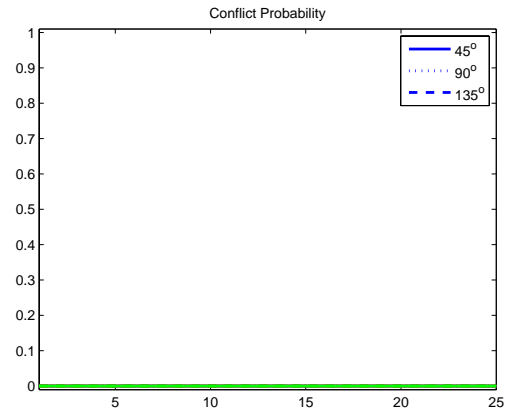


Fig. 14. Conflict Probability for $\delta_{min} = 15nm$

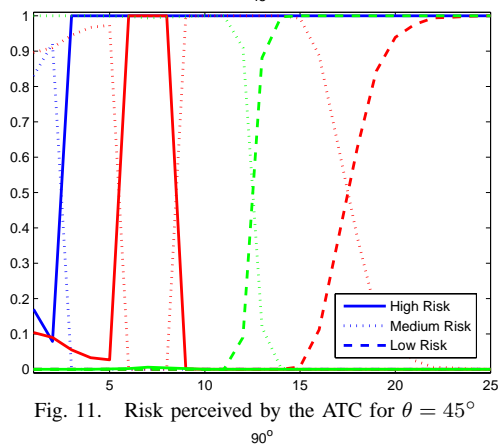


Fig. 11. Risk perceived by the ATC for $\theta = 45^\circ$

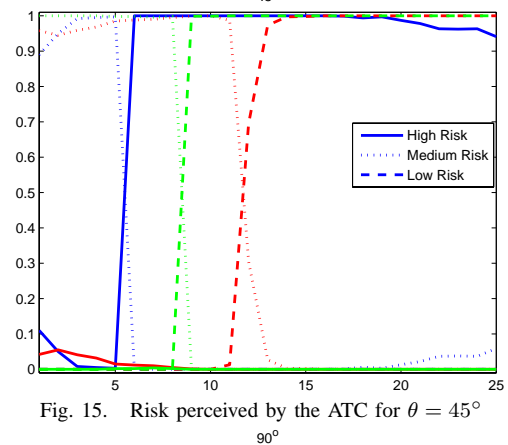


Fig. 15. Risk perceived by the ATC for $\theta = 45^\circ$

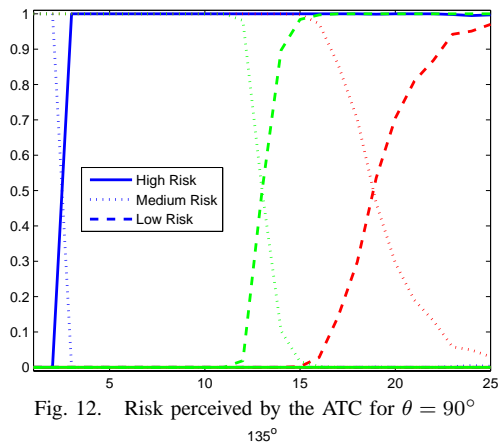


Fig. 12. Risk perceived by the ATC for $\theta = 90^\circ$

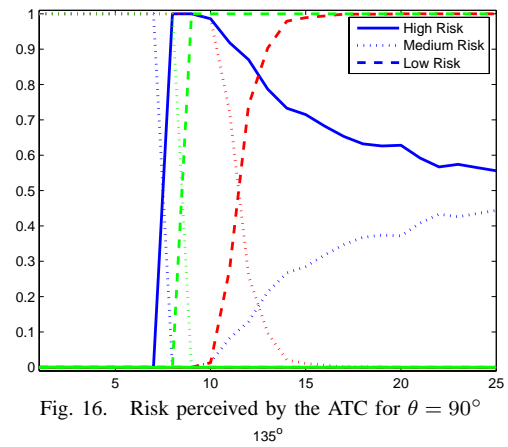


Fig. 16. Risk perceived by the ATC for $\theta = 90^\circ$

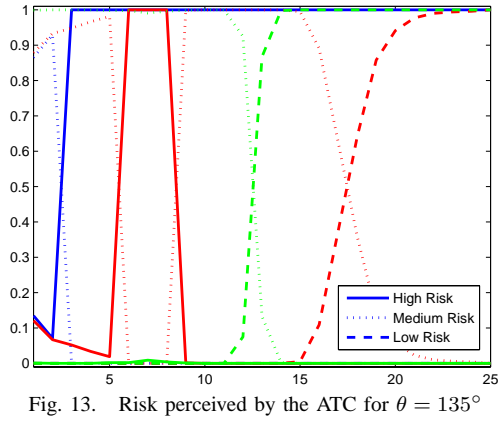


Fig. 13. Risk perceived by the ATC for $\theta = 135^\circ$

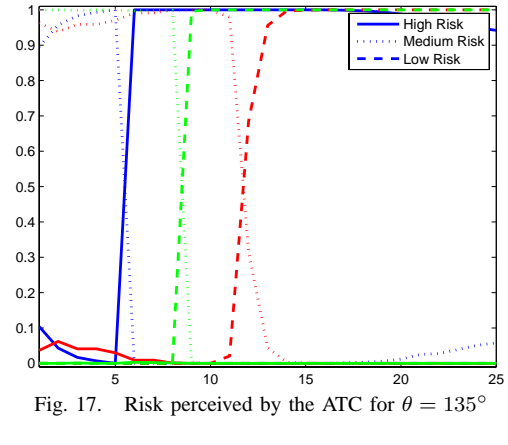


Fig. 17. Risk perceived by the ATC for $\theta = 135^\circ$

Distribution of Longitudinal Speed Prediction Error of ADS-C System

Masato Fujita

Air Traffic Management Department
Electronic Navigation Research Institute
Chofu, JAPAN
m-fujita@enri.go.jp

Abstract— The number of aircraft flying in oceanic airspaces is growing. To accommodate the traffic growth, the reduction of separation minimum for Automatic Dependent Surveillance – Contract (ADS-C) aircraft is required. However, the reduction of the separation minimum increases the collision risk of aircraft and the safety assessment prior to the reduction is expected. The probability distribution model of the longitudinal speed prediction error is a key parameter of the collision risk formula for the longitudinal separation minimum under ADS-C. In this paper, the empirical distribution of the longitudinal speed prediction error of aircraft in North Pacific routes is provided. Using Peak over Threshold (POT) technique, we found the distribution model which is appropriate for the risk estimation.

Keywords—component; Automatic Dependent Surveillance – Contract (ADS-C), Longitudinal Speed Prediction Error, Peak Over Threshold, Collision Risk

I. INTRODUCTION

NOPAC (North PACific) route system (Fig.1) is the most congested oceanic ATS route system in Fukuoka FIR. The number of aircraft flying NOPAC route system is growing. To accommodate the traffic growth, the reduction of separation minima is expected. The 50NM longitudinal separation minimum for ADS-C (Automatic Dependent Surveillance - Contract) aircraft has been implemented sequentially beginning from R220 and R580. In near future, the 30NM longitudinal separation minimum will be implemented.

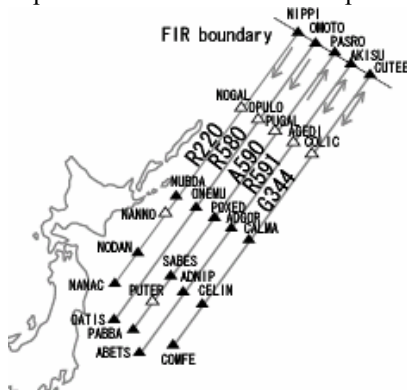


Figure 1. NOPAC route system

An aircraft under ADS-C circumstance transmits their position periodically. In Fukuoka FIR, the reporting interval is 1600 seconds in usual and 320 seconds in the case of strategic lateral offset. Under radar surveillance, position information is available in the order of seconds. Since the position information is rarely available under ADS-C circumstance, the prediction of trajectory is indispensable for surveillance.

For the quantitative estimation of mid-air collision risk of aircraft, the expected number of fatal accidents is often utilized as a risk indicator. It is called a collision risk. When the estimated collision risk does not exceed the target level of safety which is determined prior to the safety assessment, the situation is considered to be safe. When the separation minima are reduced, the collision risk increases. Hence, the safety assessment is required to confirm that the airspaces remain safe even under the reduced separation minima. (See [1].)

An aircraft pair collides if and only if they overlap in longitudinal, lateral and vertical dimension. Hence, the longitudinal overlap probability which is the probability that a pair of aircraft overlap in the longitudinal dimension should be estimated in the safety assessment of reduced longitudinal separation minimum. It is calculated using the probability distribution of aircraft position error due to aircraft navigation capability and the probability distribution of speed prediction error which causes from the position prediction performance of on-board systems and the interpolation performance of ground systems. (See [2] and [3].)

This paper gives the empirical distribution of longitudinal speed prediction error and the curve fitted to the empirical distribution applying POT (Peak over Threshold) technique in Extreme Value Theory. (See [4].)

II. CONCEPT OF ADS-C

Under ADS-C circumstance, ground stations transmit a message which tells the required type of downlink messages and the frequency of downlinks. It is called a contract message. An ADS-C aircraft downlink the required messages automatically as it is indicated in the contract messages. Downlink is executed periodically (periodic report), when the

event (lateral deviation, vertical rate change, waypoint change, altitude range change) occurs (event report), or a one-time-only report required by ATC (demand report). Downlink ADS messages are classified into basic messages and the other optional messages. A basic message contains stamped time, current position etc. Optional messages, for instance, provide the ground speed and direction at the reporting time, the location of the next waypoint and its estimated time of arrival and the predicted position at some future time instance. A ground ATC system, which is called ODP (Oceanic Data Processing system) in Japan, interpolates (and extrapolates) to predict the aircraft position from the optional messages till it receives next report.

In Japanese system, predicted route group messages, intermediate projected intended group messages and fixed projected intent group messages are utilized for the prediction of aircraft position at the ground system. The first message provides the location of the next waypoint over which the aircraft is passing and the estimated time of arrival. The third gives the predicted position at some future instance. Japanese system requires ADS aircraft to send the predicted position 37 minutes later. When an aircraft intends to change its speed or direction within 37 minutes, intermediate project intended group messages are coupled to inform when and where the speed and direction are changed.

III. DERIVATION OF EMPRICAL MODEL

A. Identification of ADS-C messages of aircraft on NOPAC

An ADS message includes a position report and the predicted position of aircraft. However, it does not contain the route name on which the aircraft intended to fly. To identify which route each aircraft flies on, FDPS (Flight Data Processing System) data set was utilized.

All flights in Fukuoka FIR are saved in FDPS data with their call signs, aircraft types, the departure and destination airports, the original flight plans, the waypoints over which the aircraft flew, the time instance when aircraft flew over the waypoints etc. All flights of NOPAC routes are identified by FDPS data set.

The ADS-C and ATS Facilities Notification (AFN) data set during September 1st 2005 to August 31st 2006 in the format of [5], [6] and [7] were provided by Kobe Aeronautical Satellite Center (the data on November 18th 2005 and from January 17th 2006 to February 9th 2006 could not be collected). The FDPS data set in the same period was provided by ATM Center.

The AFN procedure enables an ATS facility to become aware of an aircraft's data link capability and provides an exchange of address information. AFN messages are transmitted when an aircraft enters into a region where a data link service is provided by a service provider and when an aircraft is placed under the control of an adjacent ATS facility. All ADS-C messages transmitted in one flight are wedged by

AFN messages and ADS-C disconnect messages in chronological order. Since the AFN message contains the aircraft registration number, for every ADS-C message, the registration number of aircraft which transmits the ADS-C message is identified.

In many cases, the registration number of aircraft which was utilized in a flight is included in FDPS data. The corresponding ADS-C messages were identified using the registration number as the search key. However, in the case where no registration number of an ADS aircraft flying NOPAC is saved in FDPS data, we found the corresponding ADS messages manually with the help of self-developed GUI. (Fig. 2)

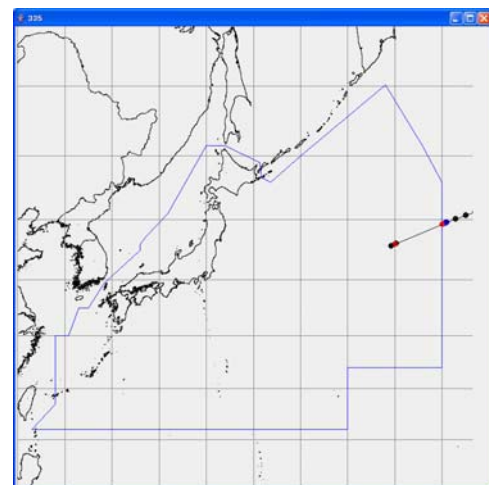


Figure 2. Display of GUI
(Black dots means the periodic reports and colored dots means the event-driven position reports)

B. Definition of Longitudinal Speed Prediction Errors

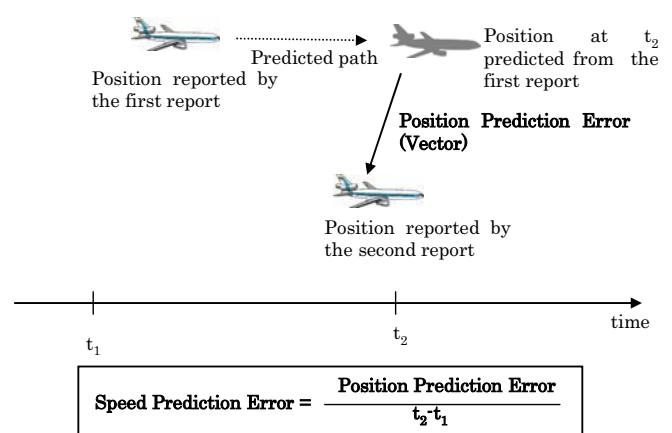


Figure 3. Definition of Speed Prediction Error

Consider successive two ADS-C messages transmitted by a single aircraft. Let t_1 be the time instance when the first message was transmitted and let t_2 be the time instance when the second message was transmitted. The position prediction error is defined as the difference of the reported position at t_2

We introduce how to find the coordinate of ‘the foot C of perpendicular from A on the route.’ Let H be the foot of perpendicular on the plane W. The coordinate of the point H is given by

$$(x_H, y_H, z_H) = \left(x_A - a \frac{ax_A + by_A + cz_A}{a^2 + b^2 + c^2}, y_A - b \frac{ax_A + by_A + cz_A}{a^2 + b^2 + c^2}, z_A - c \frac{ax_A + by_A + cz_A}{a^2 + b^2 + c^2} \right) \tag{8}$$

Hence the coordinate of the point C is given by

$$(x_C, y_C, z_C) = \left(\frac{Rx_H}{\sqrt{x_H^2 + y_H^2 + z_H^2}}, \frac{Ry_H}{\sqrt{x_H^2 + y_H^2 + z_H^2}}, \frac{Rz_H}{\sqrt{x_H^2 + y_H^2 + z_H^2}} \right) \tag{9}$$

We can find the coordinate of ‘the foot D of perpendicular from B on the route’ in the same way. We define the longitudinal speed prediction error by

$$\Delta P_x = d_s(C, D) \tag{10}$$

if the point C and D line up in the traveling direction of the aircraft, otherwise, it is defined by

$$\Delta P_x = -d_s(C, D) \tag{11}$$

D. Results

We study the longitudinal speed prediction error in the case where aircraft fly ‘straight and at a constant speed.’ When the event report is transmitted, the aircraft assumes to change its speed, heading or its vertical speed. Hence we only consider the successive ADS reports such that both of them are assumed to be periodic reports. (The basic group report following the contract message is assumed to be a demand report, when the contract message for a demand message is transmitted.)

Fig. 6 shows the time interval of successive periodic reports of aircraft flying in the NOPAC route system. Since the reporting time interval indicated in the contract message is 320 sec and 1600 sec, there are peaks at 6 min and 27 min.

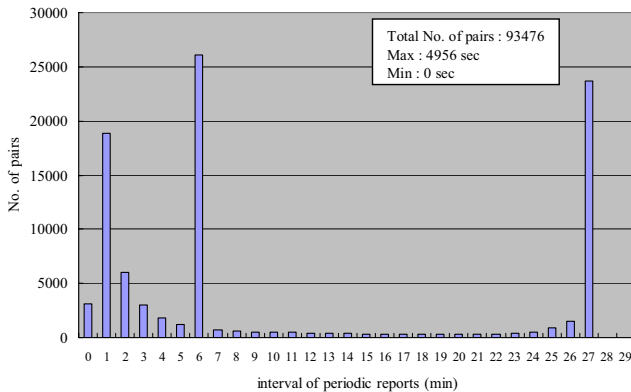


Figure 6. Distribution of intervals of periodic reports

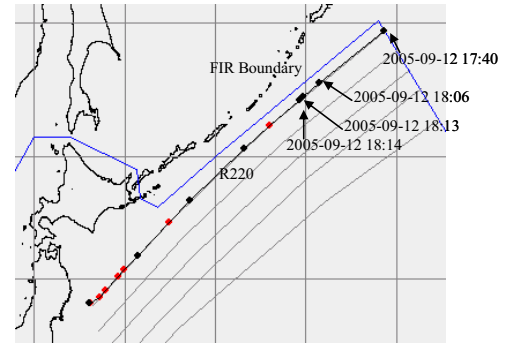


Figure 7. Trajectory of an aircraft with large longitudinal speed prediction error

In the rest of this paper, we only consider the periodic report pairs whose time interval is greater than 26 min and less than or equal to 27 min (right peak of Fig. 6). A few ADS-C reports in this data set are not coupled with fixed intent group. In the case where the estimated time of arrival at the next waypoint described in the predicted route group message is close to the stamped time of the basic message, the longitudinal prediction error is sometimes large in magnitude.

Fig. 7 shows the trajectory of an aircraft flying on R220 in the NOPAC route system. The dots show the position reported via ADS-C. The longitudinal speed prediction error of reports which were transmitted at 17:40 and 18:06 was -668 (knots). It turned out that the basic group transmitted at 17:40 is not coupled with ‘fixed projected intent group’ and estimated time of arrival given in ‘predicted route group’ is 10 sec later from the stamped time. The ‘predicted route group’ reports that the next waypoint is in the east of the reported position in spite of the westbound aircraft. It seems that the aircraft flies by the waypoint; however, the on-board system does not update the next waypoint. Hence the ground system possibly considers that the aircraft is flying in the opposite direction. If a fixed projected intent group is coupled in this case, a system might misunderstand the aircraft heading in a short period. However, the system makes an appropriate prediction based on the fixed projected intent group after a few seconds.

0.034% of basic reports are not coupled with ‘fixed projected intent group’ and 0.017% of basic reports are coupled with neither ‘fixed projected intent group’ nor ‘predicted route group’.

Fig. 8 shows the empirical distribution of the longitudinal speed prediction errors of periodic report pairs which are coupled with both fixed projected intent group and predicted route group and whose reporting time interval is greater than 26 min and less than or equal to 27 min. There are no incredibly large longitudinal speed prediction errors any more.

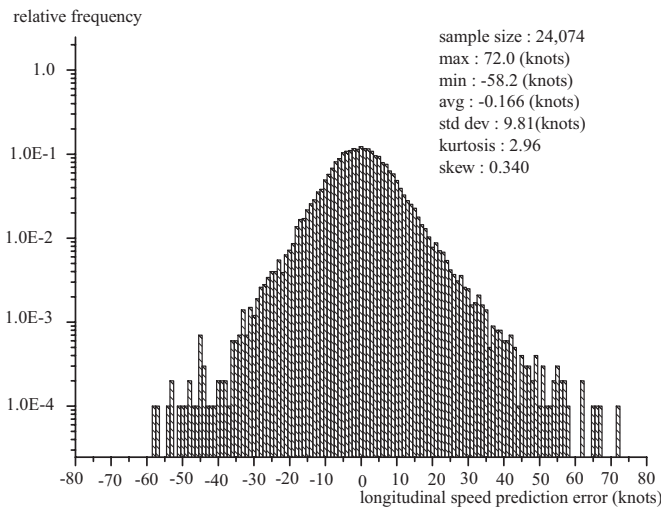


Figure 8. Empirical distribution of longitudinal speed prediction errors (Data omitted, NOPAC)

IV. DISTRIBUTION MODEL

On first sight, the empirical distribution given in Fig. 8 follows a normal distribution. Fig 9 shows the QQ-plot (quantile-quantile plot) for normal distribution. If the empirical distribution follows a normal distribution, the dots in Fig. 9 are on the red straight line. When ‘sample quantiles’ is larger than -20 and smaller than 20, the dots seem to be on the straight line, however, it is not the case for the data set outside of [-20, 20].

The average and standard deviation of restriction of the empirical distribution on [-20,20] are -0.1142 and 7.757, respectively. The histogram in Fig. 10 shows the empirical distribution and the graph of the probability density function of the normal distribution with average = -0.1142 and standard deviation = 7.757. Fig.10 shows that this normal distribution fits the empirical distribution well.

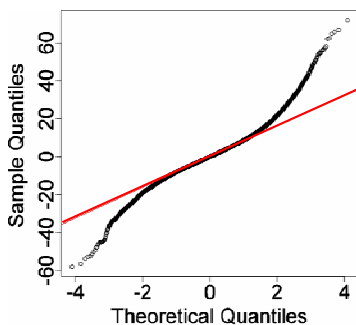


Figure 9. QQ-Plot of Fig.8 for normal distribution

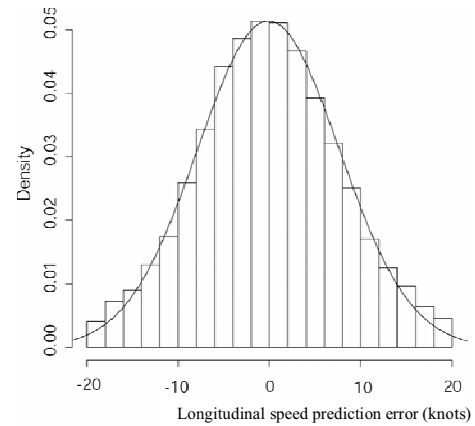


Figure 10. QQ-Plot of Fig.8 for normal distribution

The author applied POT technique to the data set. Extreme value theory claims that the conditional probability $\Pr\{Y < y | Y > u\}$ of distributions satisfying some technical assumptions approximately follows a generalized Pareto distribution when u is large enough. (More precisely, for any distribution which is in the domain of attraction, $\Pr\{Y < y | Y > u\}$ weakly converge to the generalized Pareto distributions as $u \rightarrow \infty$.) The cumulative distribution function of a generalized Pareto distribution is given by

$$H(y) = 1 - \left(1 + \xi \frac{y}{\sigma}\right)^{-1/\xi}, \quad 1 + \xi y / \sigma > 0. \quad (12)$$

When the shape parameter $\xi < 0$, the generalized Pareto distributions are Beta distributions. ($0 < y < \sigma / \xi$) In $\xi = 0$, they are exponential distributions and they are Pareto distribution in the case where $\xi > 0$. (See [4] and other related papers for more detail.)

The author analyzed the both-side tails of Fig. 8 using POT technique. The R-package extRemes [8] is utilized for the analysis. (R is a free statistical software for data analysis.) For the right tail, we set the threshold $u = 20$ considering the stability of estimated shape parameter ξ and scale parameter σ . The number of excesses of thresholds is 687 (2.85% of the whole data set). By maximum likelihood method, we found $\xi = 0.0426$ and 95% confidence interval is [-0.03419, 0.13204]. $\sigma = 7.63$ and its standard error is 0.4342. Fig. 11 and Fig. 12 show the QQ-plot and the density plot of this model, respectively. Since almost all dots are on the diagonal line in the QQ-plot diagram, the generalized Pareto distribution fits the right tail of the empirical distribution well. The fact $\xi = 0.0426$ suggests that the right tail is slightly thicker than an exponential distribution.

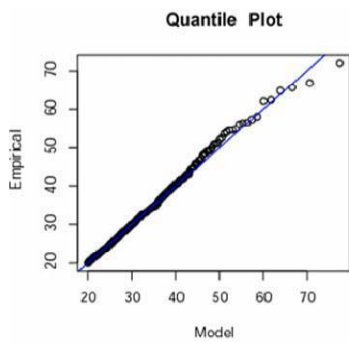


Figure 11. QQ-Plot of generalized Pareto distribution for the right tail of Fig. 8

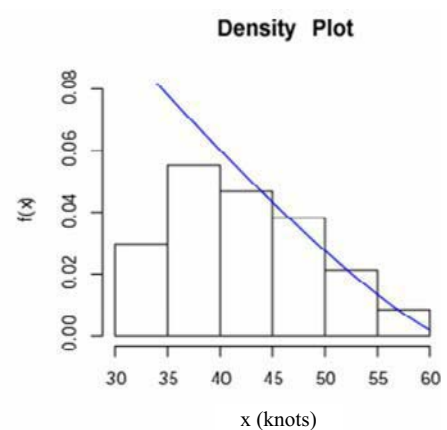


Figure 14. Density plot of generalized Pareto distribution for the left tail of Fig. 8

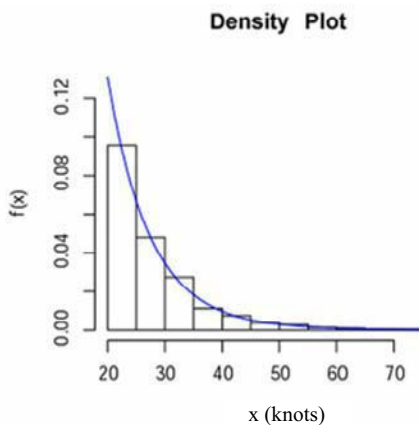


Figure 12. Density plot of generalized Pareto distribution for the right tail of Fig. 8

Under the assumption that both left and right tails follow the same distribution, we applied POT technique to find the shape of the tail. In this case, we have only to analyze the tail of the absolute value of empirical data. We set the threshold $u = 20$. The number of excesses of thresholds is 1161 (5.26%). $\xi = 0.0386$ with 95% confidence interval $[-0.01962, 0.10373]$. $\sigma = 7.093$ and its standard error is 0.3048. Fig. 15 and Fig. 16 show the QQ-plot and the density plot of this model, respectively. The generalized Pareto distribution fits the tail of the empirical distribution well judging from QQ-plot diagram.

The same analysis was conducted for the left tail. We set the threshold $u = 34$. The number of excesses of thresholds is 47 (0.195%). $\xi = -0.4484$ with 95% confidence interval $[-0.61369, -0.08794]$. $\sigma = 12.22$ and its standard error is 2.374. Fig. 13 and Fig. 14 show the QQ-plot and the density plot of this model, respectively. Because of small amount of data set, the estimated parameter has large standard deviation and some dots are apart from the diagonal line in QQ-plot diagram. Hence we cannot apply POT technique to determine the shape of left tail. One way to find the shape of right tail is to assume that both left and right tails follow the exactly same distribution.

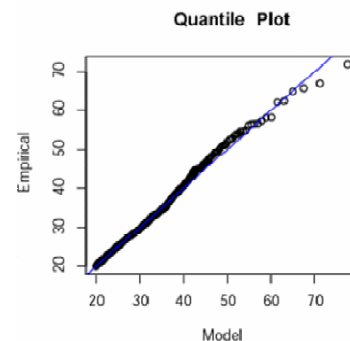


Figure 15. QQ-Plot of generalized Pareto distribution for the both tails of Fig. 8

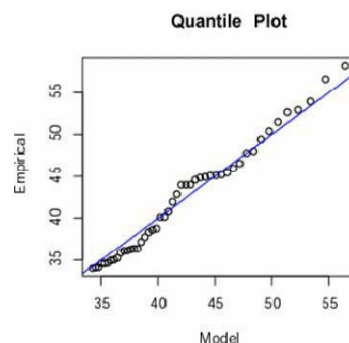


Figure 13. QQ-Plot of generalized Pareto distribution for the left tail of Fig. 8

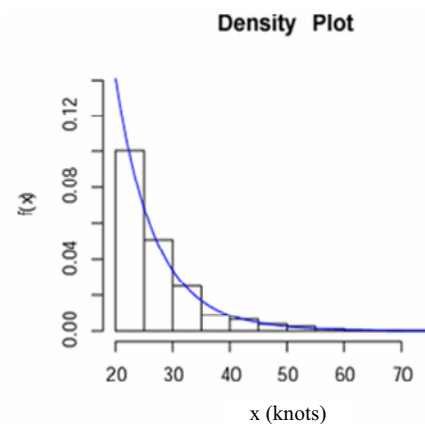


Figure 16. Density plot of generalized Pareto distribution for the both tails of Fig. 8

of Fig. 8

V. CONCLUSION

Let $\Phi(x)$ be the standard normal cumulative distribution function, namely,

$$\Phi(x) = \int_{-\infty}^x \frac{\exp(-u^2/2)}{\sqrt{2\pi}} du \quad (13)$$

Then the cumulative distribution function $F(x)$ of longitudinal speed prediction errors, x in knots, is given by

$$F(x) = \begin{cases} \frac{5.26/2}{100} \times \left(1 + 0.0386 \frac{-x-20}{7.093}\right)^{-1/0.0386} & x \leq -20 \\ \frac{2.63}{100} + \frac{94.74}{100} \times \frac{\Phi\left(\frac{x-(-0.1142)}{7.757}\right) - \Phi\left(\frac{-20-(-0.1142)}{7.757}\right)}{\Phi\left(\frac{20-(-0.1142)}{7.757}\right) - \Phi\left(\frac{-20-(-0.1142)}{7.757}\right)} & -20 < x < 20 \\ \left(1 - \frac{5.26/2}{100}\right) + \frac{5.26/2}{100} \times \left[1 - \left(1 + 0.0386 \frac{x-20}{7.093}\right)^{-1/0.0386}\right] & x \geq 20 \end{cases} \quad (14)$$

The simplified form is given by

$$F(x) = \begin{cases} 2.63 \times 10^{-2} \times \left(1 - 5.44 \times 10^{-3} \times (x+20)\right)^{-25.9067} & x \leq -20 \\ 2.63 \times 10^{-2} + 0.9569 \times \left(\Phi\left(\frac{x-(-0.1142)}{7.757}\right) - 5.18 \times 10^{-3}\right) & -20 < x < 20 \\ 0.9737 + 2.63 \times 10^{-2} \times \left(1 - \left(1 + 5.44 \times 10^{-3} \times (x-20)\right)^{-25.9067}\right) & x \geq 20 \end{cases} \quad (15)$$

The analysis of both-side tails suggests that the tail of the empirical distribution of longitudinal speed prediction errors follows a Pareto distribution which has slightly thicker tails than exponential distributions. It is hard to calculate the longitudinal overlap probability if the tail follows a Pareto distribution. Even if the tail of empirical distribution is thicker than the exponential distribution, the shape parameter ξ is so small that we may assume that the tail follows an exponential distribution in many cases.

We can assume that $|\text{longitudinal speed prediction error}|_{[20,\infty)} - 20$ follows an exponential distribution whose probability density function is $\exp(-x/\lambda)/\lambda$. Here $|\text{longitudinal speed prediction error}|_{[20,\infty)}$ denotes the restriction of the absolute value of longitudinal speed prediction error on $[20,\infty)$. The maximum likelihood estimator of λ is the average of the empirical data set of $|\text{longitudinal speed prediction error}|_{[20,\infty)} - 20$. It is 7.439. Hence we get the following probability density function of a model of longitudinal speed prediction errors.

$$f(x) = \begin{cases} \left(1 - \frac{5.26}{100}\right) \times \frac{\exp(-(x+0.1142)^2 / (2 * 7.757^2))}{\sqrt{2\pi} * 7.757} & |x| < 20 \\ 2.63 \times 10^{-2} \times \frac{\exp(-(|x|-20)/7.439)}{7.439} & |x| \geq 20 \end{cases} \quad (16)$$

This paper first reviews the basic concept of ADS-C and summarizes the methodology to find the longitudinal speed prediction errors of ADS-C.

A few ADS-C reports, which are not coupled with the fixed intent group and the estimated time of arrival at the next waypoint is close to the stamped time, have large longitudinal prediction errors in magnitude. (Max. 668 knots)

Fig. 8 shows the empirical distribution of the longitudinal speed prediction errors of periodic report pairs which are coupled with both fixed projected intent group and predicted route group and whose reporting time interval is greater than 26 min and less than or equal to 27 min. By QQ-plot, it turns out that this distribution on $[-20, 20]$ follows the normal distribution whose average is -0.1142 and whose standard deviation is 7.757 . POT (Peak over Threshold) technique of Extreme Value Theory was applied to find the shape of the tail of Fig. 8. The tails (outside -20 and 20) follow the generalized Pareto distribution whose shape parameter $\xi = 0.0386$ and the scale parameter $\sigma = 7.093$. Equation (15) gives the explicit description of the cumulative distribution function.

The density function of the distribution in (16) is also given under the assumption that the tails follow an exponential distribution.

ACKNOWLEDGMENT

The author thanks Japan Civil Aviation Bureau, ATM Center, Kobe Aeronautical Satellite Center and Mr. Sasamoto for the provision of information and data sets.

REFERENCES

- [1] Air Traffic Service, Annex 11 to Convention on International Civil Aviation Section 2.26
- [2] Anderson, D. A Collision Risk Model Based On Reliability Theory That Allows For Unequal Navigational Accuracy, ICAO SASP-WG/WHL/7 WP/20 REVISED, May 2005.
- [3] Fujita, M. Nagaoka, S and Amai, O. Safety Assessment prior to Implementation of 50NM Longitudinal Separation Minimum in R220 and R580, ICAO SASP-WG/WHL/9 WP/14, May 2006.
- [4] Coles, S. An Introduction to Statistical Modeling of Extreme Values, Springer Verlag London, 2001.
- [5] Automatic Dependent Surveillance (ADS), ARINC Characteristic 745-2, June, 1993.
- [6] Minimum Operational Performance Standards for Airborne Automatic Dependent Surveillance (ADS) Equipment, RTCA DO-212 October, 1992.
- [7] ATS Data Link Applications over ACARS Air-Ground Network, ARINC Characteristic 622-4, October, 2001.
- [8] <http://www.assessment.ucar.edu/toolkit/>

Three-Degree Decelerating Approaches in Arrival Streams

Continuous Descent Approaches in High Traffic Density

Arjen de Leege

Faculty of Aerospace Engineering
Delft University of Technology
Delft, The Netherlands
arjen.deleege@gmail.com

Alexander in't Veld

Faculty of Aerospace Engineering
Delft University of Technology
Delft, The Netherlands
a.c.intveld@tudelft.nl

Max Mulder

Faculty of Aerospace Engineering
Delft University of Technology
Delft, The Netherlands
m.mulder@tudelft.nl

René van Paassen

Faculty of Aerospace Engineering
Delft University of Technology
Delft, The Netherlands
m.m.vanpaassen@tudelft.nl

Abstract—Self-spacing is a solution for the runway capacity reduction that is intertwined with the use of continuous descent approaches in the current air traffic management system to reduce aircraft noise. In case of self-spacing the separation task is transferred from the air traffic controller to the pilot. The Three-Degree Decelerating Approach (TDDA) can be executed in a distance- or time-based self-spacing environment while yielding a noise reduction. A fast-time simulation tool has been developed to simulate arrival streams of different aircraft types executing the TDDA in both self-spacing scenarios under actual wind conditions. The tool was used to quantify the performance differences between distance- and time-based self-spacing in terms of capacity, noise reduction, and loss of separation. In the time-based scenario no effects of preceding aircraft on trailing aircraft could be identified. However, an increase in separation with a negative effect on the airport capacity in order to assure safe separation was required. In the distance-based self-spacing scenario a slow-down effect was observed that led to a decrease in the noise reduction towards the end of the arrival stream. This was solved by altering the initial separation between aircraft in the arrival stream. In the distance-based self-spacing scenario no negative effect on the runway capacity or safety has been identified.

Index Terms -- Continuous Descent Approach, capacity, self-spacing

I. INTRODUCTION

To accommodate the forecasted further growth of aviation without increasing the noise impact measures must be taken [1-3]. Promising procedures are Continuous Descent Approaches (CDAs) but are infeasible in the current air traffic management system because of the negative effect on the runway capacity. During the approach Air Traffic Control

(ATC) issues speed, altitude, and heading instructions to keep aircraft safely separated. During a CDA ATC can no longer give instructions; otherwise the aircraft are not able to follow their optimum descent path. Moreover it is unknown what the descent paths of the aircraft will be. The aircraft performance, pilot control strategy, and wind condition significantly affect the descent path. [3][4] Therefore Air Traffic Control (ATC) introduces additional spacing between aircraft to assure that the separation minima are respected, though at the cost of runway capacity. The capacity reduction prevents the CDA from being introduced at a large scale at the major airports in the world to reduce the noise nuisance in the vicinity of the airport.

The Three-Degree Decelerating Approach (TDDA) is a CDA capable of realizing a significant noise reduction. The procedure lies within the boundaries of present approach procedure limitations and can be implemented in a short term [4-6]. Major difference with the current ATM system and the key to application of a CDA without a drop in the runway capacity is the use of self-spacing. The spacing task is transferred from the air traffic controller to the pilot. The maneuverability of an aircraft while executing a CDA is limited and largely driven by the aircraft performance, wind conditions, and the control strategy of the pilot. The aircraft performance information is readily available in the cockpit rather than on the ground. A pilot can plan and execute, with the help of onboard systems, a CDA to remain safely separated and exploit the noise reduction potential of the CDA [5][6]. Previous research focused on the design of the procedure, and the required systems [4-6]. This paper discusses the feasibility of implementing the TDDA at high

traffic density airports in a distance- or time-based self-spacing environment. It also introduces intent-based trajectory predictions to prevent transient motions from occurring in an arrival stream of aircraft. As will be discussed later, the slinky effect can only occur when relying on distance-based self-spacing.

Section II addresses the TDDA procedure in a distance-based and time-based the self-spacing environment. Section III focuses on the intent-based trajectory prediction applied when using distance-based self-spacing. Use of the TDDA in arrival streams imposes constraints on the initial separation between aircraft; this issue is addressed in Section IV. The fast-time simulation tool used to simulate arrival streams of different aircraft executing the TDDA in both self-spacing scenarios under actual wind conditions is presented in Section V. The performance of the TDDA in a distance- or time-based self-spacing environment is presented in Section VI. Section VII contains the conclusion.

II. THREE-DEGREE DECELERATING APPROACH

A. Description of Procedure

The TDDA is a straight-in approach along a fixed descent path with a -3° path angle as illustrated in Figure 1. [4-7]. The descent path coincides with the Instrument Landing System's (ILS) glide slope, except the aircraft intercepts the descent path at an altitude that lies well above the altitude the aircraft normally intercepts the ILS glide slope and starts with the final 3° descent. The aircraft descends with a constant IAS to a point where the engines are set to idle, this point is referred to as the point of thrust cutback (TCB). Due to the aerodynamic drag the aircraft decelerates, during the deceleration the flaps and gear are extended. For safety reasons most operators require aircraft to be in a stabilized landing configuration before descending below 1000 ft. This is incorporated in the TDDA procedure by demanding the aircraft to be stabilized at the reference altitude, h_{ref} , which is located at 1000 ft. To accomplish this, the flap extension speeds are such that the final approach speed V_{APP} is reached at h_{ref} in a stabilized landing configuration. The flap extension speeds together form the flap schedule of the aircraft. Below h_{ref} the aircraft maintains V_{APP} by reapplying thrust and continues the approach until touchdown.

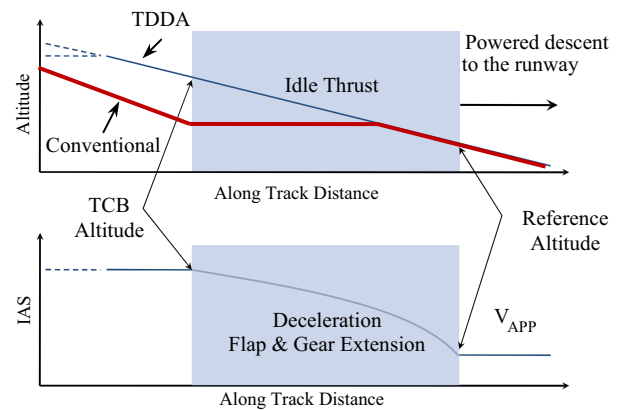


Figure 1. The TDDA Trajectory

The moment of thrust cutback and the flap schedule are the only controls the pilot has to reach V_{APP} (noise goal) at the reference altitude. In addition the pilot has the responsibility to remain safely separated (separation goal) with the preceding aircraft or arrive at the commanded RTA (time goal). The applicable goals depend on the form of self-spacing that is used in the arrival stream.

Research showed that it is difficult for a pilot to determine the correct TCB altitude and a flap schedule [4][5]. Therefore the pilot is supported by a number of optimization and scheduling algorithms fed by wind and trajectory prediction algorithms to meet the noise goal, and separation or time goal [4-6]. Which optimization and scheduling algorithms are active depends on the part of the TDDA the aircraft is in. As soon as the aircraft intercepts the 3° descent path Thrust CutBack (TCB) altitude optimization starts for both the noise and separation/time goal. The algorithm determines the maximum TCB altitude using a binary search method.

When flying below the TCB altitude the Flap Scheduler Algorithm monitors whether the applicable goals will be met. The implemented flap scheduler is based on the scheduler originally described in [5]. If one of the goals is not met and scheduling is possible the flap scheduler updates the schedule. The updated schedule is determined using a binary search algorithm. Optimization of the noise goal is only performed if the time or separation goal is met. Below h_{ref} no scheduling takes place.

B. Two Self-Spacing Scenarios

Because of wake turbulence trailing aircraft must maintain a minimum separation with respect to the preceding aircraft. In the scenario proposed in this research the spacing task is transferred from the air traffic controller to the pilot to carry out CDAs without adversely affecting the runway capacity. Self-spacing can be either distance-based or time-based. Distance-based self-spacing using the relative state of the preceding aircraft might give rise to transient motions in the arrival stream (the slinky effect) resulting in separation violations [8]. Therefore time-based self-spacing was implemented in the TDDA [4]. Time-based spacing concepts proved best for in-trail self-spacing but are hard to implement into the current

spatial working environment of pilots and controllers. Pilots indicated to prefer distance-based over time-based procedures [7].

C. TDDA using Distance-based Self-Spacing

In case of distance-based self-spacing it is the task of the pilot to assure that the separation minimum is never violated. Based on a prediction of the leading aircraft trajectory the own TDDA is planned such that the actual minimum separation lies close to the minimum allowable separation to achieve the highest airport capacity. Figure 2 shows the structure of the TDDA algorithm under distance-based self-spacing. The algorithm computes the separation between the aircraft based on trajectory predictions of the own and the preceding aircraft. The prediction of the own trajectory is also used to determine whether the noise goal will be met. If necessary an optimization of the TCB altitude or flap schedules takes place. TCB altitude optimization takes place when the aircraft is flying above the last computed TCB altitude. When the thrust cutback has taken place and the aircraft has not reached the final approach speed, flap schedule optimization takes place.

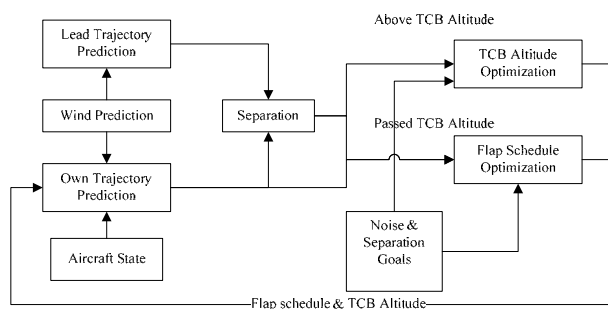


Figure 2. TDDA Algorithm for Distance-Based Self-Spacing

D. TDDA using Distance-based Self-Spacing

Time-based self-spacing makes the on-board leading aircraft trajectory prediction superfluous, instead thereof each aircraft is supplied with an RTA. This does not imply that the separation minima do not have to be obeyed. Determination of RTAs that do not lead to separation violation will be addressed later. The task of the pilot is to arrive at the threshold at the RTA. The resulting TDDA should meet both the noise and time goal. The structure of the TDDA algorithm is identical to the structure under distance-based spacing. The RTA block replaces the lead prediction and separation blocks, see Figure 2.

III. AIRCRAFT INTENT-BASED TRAJECTORY PREDICTION

A. Using Aircraft Intent for Trajectory Prediction

Distance-based self-separation requires a precise trajectory prediction of the preceding aircraft. An aircraft intent based prediction algorithm is proposed here. Aircraft intent is an unambiguous description of how the aircraft has to be operated within a given timeframe. The intent information is the input to a trajectory predictor [10].

Captured in the intent is the outcome of optimization of the trajectory by the TDDA algorithms on-board the leading aircraft. If an aircraft's descent profile is disturbed, for instance by a wind change or delayed pilot action, the TDDA algorithm optimizes the trajectory. The new trajectory is described in aircraft intent that is not in principal the same as the previous intent because of the optimization process. In predictions based on previous states no credit is given to the ongoing optimization process. This can cause unnecessary control actions from the trailing aircraft that can propagate through the arrival stream resulting in the slinky effect.

Ref. [11] shows that a trajectory prediction of the last constant speed segment of the TDDA is sufficient to determine the minimum separation. The change of the separation between two aircraft generally has closing characteristics. Based on the performance characteristics of a number of aircraft and the applicable separation minima it was concluded that the moment of minimum separation will take place when the leading aircraft is flying below the reference altitude.

B. Intent-Based Lead Aircraft Prediction

Prediction of the last constant speed segment is sufficient to determine the minimum separation. This segment can be predicted independent of the other segments using the ETA, V_{APP} , and descent path angle as illustrated in Figure 3.

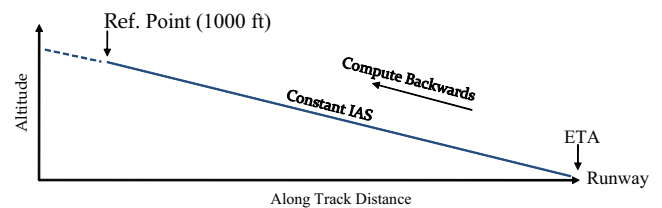


Figure 3. Intent-Based Trajectory Prediction

The speed and flight path angle are kept constant during the last stage of the final approach. No states of the preceding aircraft are used to predict the leading aircraft trajectory, only aircraft intent information is needed. The predictor starts at the runway threshold where the aircraft is at the ETA with speed V_{APP} and computes the trajectory up to the reference altitude. No knowledge about the aerodynamic performance is required, because the airspeed and path angle remain constant. The prevailing wind is the only unknown and is estimated using a wind predictor as described in [11].

IV. TDDAS IN ARRIVAL STREAMS

An arrival stream of aircraft consists of one leading aircraft and a number of trailing aircraft. The aircraft in the stream may vary in type and weight and have different deceleration profiles. The aircraft in the stream can be separated by time- or distance-based self-spacing. The leading aircraft is always supplied with an RTA and optimizes its TCB altitude and flap schedule to arrive at the RTA and meet the noise goal. The trailing aircraft, depending on the self-spacing concept, tries to

meet their separation or time goal and the noise goal. The control space of the TDDA is limited, only if the separation or time goal falls within the control space a TDDA exploiting the noise reduction potential without a capacity loss is possible.

A. TDDA Control Space Impacting Factors

The TDDA control space boundaries are a function of the aircraft type, weight and prevailing wind conditions. The boundaries are set by the TDDA with the shortest and longest duration. To get the shortest duration all flaps are extended at their maximum speed yielding a fast but late deceleration and the lowest possible TCB altitude. The longest duration is achieved by extending flaps at their minimum speed resulting in a slow and early deceleration and the highest possible TCB altitude. Aircraft weight lowers the TCB altitude, shortens the time-to-fly, and reduces the control space. A headwind increases the duration of the TDDA and lowers the TCB altitudes, but also makes the control space smaller. The opposite occurs in case of a tailwind. The variance in TCB altitude and time-to-fly for each type and weight combination reflect the uncertainty in the descent profile of each aircraft ATC has to deal with causing the increase in separation. The impact of the wind conditions justifies the need of a wind predictor.

B. Initial Separation Constraints

To fly a TDDA that does not reduce runway capacity and meets the noise goal, the separation or time goal should fall in the control space. This imposes constraints on the initial separation with respect to the preceding aircraft or entry time. The constraints follow from the control space and leading aircraft trajectory (prediction) and type, and RTA if applicable.

C. Initial Separation - Distance-Based Self-Spacing

The initial separation under distance-based self-spacing is determined as shown in Figure 4. The separation goal implies that the minimum separation should equal the minimum safe separation. The separation goal is visualized by offsetting the lead's trajectory prediction over the required separation away

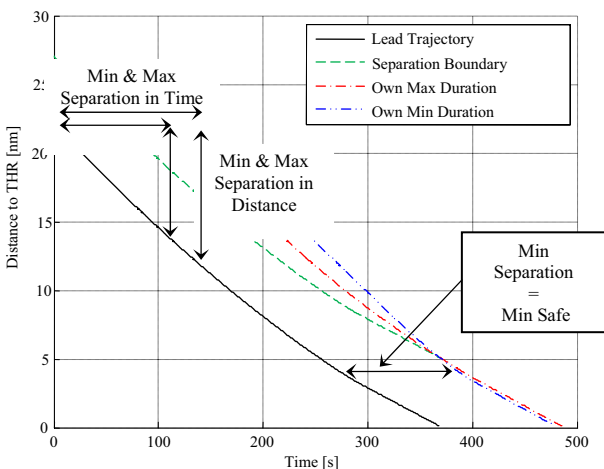


Figure 4. Min and Max Initial Separation for Distance Based Spacing

from the runway (separation boundary). The minimum

allowable distance to the threshold when the lead aircraft is still flying is indicated by the separation boundary. By positioning the control space boundaries such that the minimum separation equals the minimum safe separation the minimum and maximum initial separation expressed in time or distance are determined.

D. Initial Separation - Time-Based Self-Spacing

The constraints in case of time-based spacing are determined in a similar way. The time of arrival of both control space boundaries is set equal to the RTA from where the entry time interval is determined, see Figure 5. For reference the separation boundary is drawn, the minimum separation between both aircraft cannot be violated.

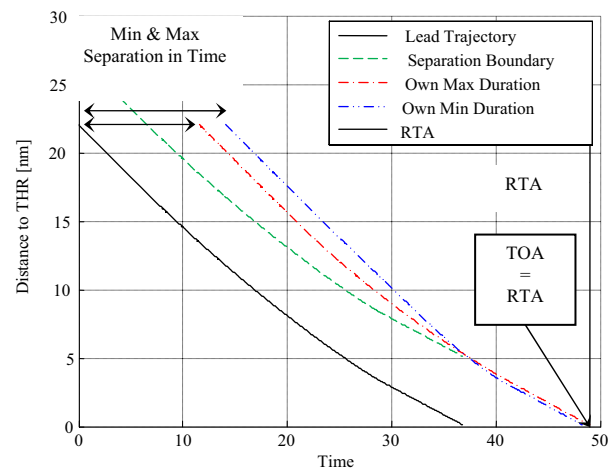


Figure 5. Min and Max Initial Separation for Time Based Spacing

The initial separation constraints are determined based on predictions of the own control space and lead trajectory that have an uncertainty or are subject to changes due to variable wind conditions. An initial separation located in the middle of the interval minimizes the risk that the separation or time goal drops out of the control space resulting in spacing gaps or failure to meet the noise goal. The TCB altitudes of the control space boundaries are the highest and lowest TCB altitude achievable. An initial separation close to the boundary leads to a relatively high or low TCB altitude.

E. Initial Trajectory Optimization

If the initial separation is well determined the control space is reduced by the preceding aircraft as shown in Figure 6. The own aircraft's trajectory should lie as close to the separation boundary as possible. The TCB optimization and flap scheduler search for trajectory with the maximum TCB altitude while still meeting all the goals.

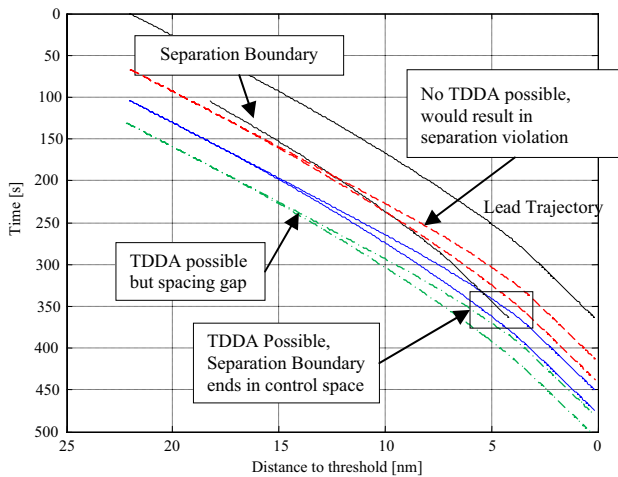


Figure 6. Effect of Leading Aircraft on Control Space

If the separation boundary crosses the lower control space boundary a TDDA that meets the noise goal is impossible without violation of the separation minima. If the separation boundary lies above the control space, execution of the TDDA leads to a spacing gap with an adverse effect on the airport capacity.

V. FAST-TIME TDDA SIMULATION TOOL

The fast-time simulation tool simulates arrival streams of aircraft executing the TDDA in a distance- or time-based self-spacing environment under actual wind conditions. Implemented in the simulator is the TDDA with the optimization and scheduling algorithms depicted in Figure 2. For the aircraft trajectory computation and prediction use is made of point mass models approximating the following aircraft: Boeing 747-400, 777-300, 737-800, 737-400, and the Airbus 321. Randomness in the pilot response time is modeled using the Pilot Response Delay Model described in Ref. [12].

VI. DISTANCE-BASED VS. TIME-BASED SELF-SPACING

Using the simulation tool 5000 arrival streams from eight aircraft in each self-spacing scenario have been generated. The aircraft type and weight were determined randomly. The following aircraft are present in the arrival streams: Boeing 747-400, 777-300, 737-800, 737-400, and the Airbus 321. Per type three different weights were assigned to the aircraft: the Operating Empty Weight (OEW), the Maximum Landing Weight (MLW), and the mean of the OEW and MLW. Weather observations have been used to create 54 time-varying wind profiles. To determine the initial separation interval a 0.2 nm buffer was added to the separation minima to account for uncertainties in the predictions and wind changes. The entry time into the arrival stream was set such that the aircraft were in the middle of their control space computed 10 minutes before the preceding aircraft starts his TDDA. The characteristics of the TDDA are summarized in Table 1.

TABLE I. TDDA CHARACTERISTICS

Aspect	Value
Top of Descent (TOD)	7000 ft

Initial IAS	230 kts
Reference Altitude (h_{ref})	1000 ft
Approach Speed (V_{APP})	$1.3V_{stall} + 10$ kts
Minimum Flap Speed	$1.3V_{stall}$
Maximum Flap Speed	V_{FE}

The performance and feasibility of implementing the TDDA at a high traffic density airport in the two self-spacing scenarios was assessed using the formulated noise, separation, time goal, and the runway capacity.

A. Noise Goal Performance

The noise goal is met if V_{APP} is reached at h_{ref} . In case V_{APP} is reached above h_{ref} , engine thrust needs to be reapplied above h_{ref} resulting in more engine noise. From a safety point of view it is also not desired that the aircraft reaches V_{APP} below h_{ref} . Figure 7 shows the average altitude where V_{APP} was reached, hereafter referred to as $h_{V_{APP}}$, per position in the arrival stream. As expected for the time-based scenario no trend between the position of the aircraft and $h_{V_{APP}}$ could be identified ($R = 0.006$, $p = 0.287$, *Pearson 2-tailed*). On average $h_{V_{APP}}$ lies 20 ft above h_{ref} . When taking into account the 25 ft tolerance used during flap scheduling and TCB optimization it is concluded that on average the noise goal is met. For the distance-based scenario a positive correlation between $h_{V_{APP}}$ and the position in the stream can be identified ($R = 0.145$, $p < 0.001$, *Pearson 2-tailed*). The noise reduction deteriorates towards the end of the arrival stream.

Deterioration of the noise reduction is caused by

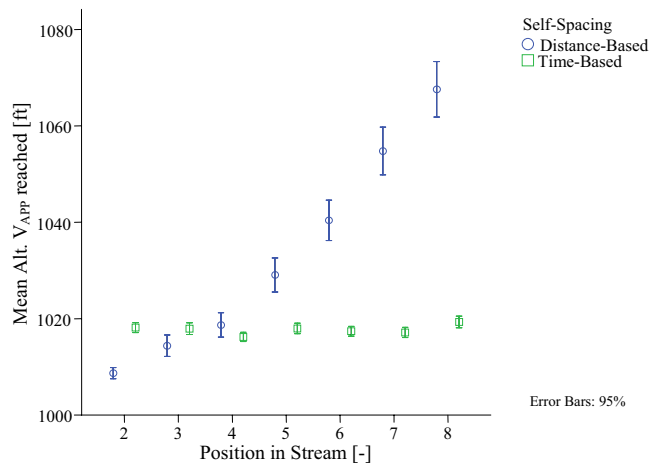


Figure 7. Altitude V_{APP} reached

accumulation of delays (slowdown effect) in the arrival stream. The TDDA is susceptible to time delays. If there is a delay in the arrival stream all trailing aircraft in the stream are affected by this delay. Aircraft further back in the arrival stream are confronted with longer delays than the aircraft in the beginning of the arrival stream, see Figure 8. To remain safely separated, aircraft increase the TCB altitude to the upper bound of their control space. If the aircraft still lose separation flap extension is advanced but this will result in failure to meet the noise goal. In case of a delayed arrival there is a significant correlation between h_{ref} and the magnitude of

the delay ($R = 0.672, p < 0.001, \text{Pearson 2-tailed}$). Arrivals earlier than expected have no effect on the noise goal ($R = 0.057, p < 0.001, \text{Pearson 2-tailed}$).

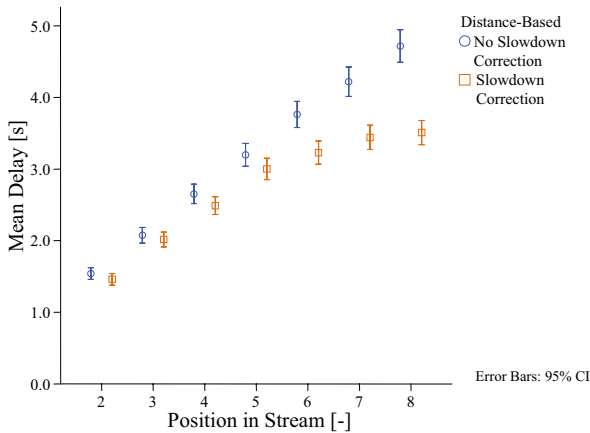


Figure 8. Slowdown effect in distance-based scenario

Deterioration of the noise reduction due to accumulating time delays was suppressed by increasing the initial separation between the aircraft in the end of the arrival stream. In the simulation this has been accomplished by increasing the separation buffer from 0.2 nm to 0.5 nm. In case of a delay the aircraft reduce the spacing to the allowed minimum and still reach V_{APP} at h_{ref} . The increase in time delay flattens and a positive correlation between h_{VAPP} and the position can no longer be identified ($R = 0.040, p < 0.001, \text{Pearson 2-tailed}$).

B. Separation

Under distance-based spacing a separation goal is formulated. The flap and TCB scheduling should be such that the minimum separation equals the minimum safe separation. Although there is no separation goal in the time-based scenario the minimum separation cannot be violated. Separation should be assured by adhering to the RTA. 99% of the aircraft arrive within 6.5 s of the RTA at the threshold. Table 2 lists the mean and median excess separation between aircraft and percentage of aircraft with a loss of separation with respect to the preceding aircraft. In case of time-based separation four times more separation violations occur. A part of the violations in both scenarios would obviously have led to a go-around. Go-arounds do occur during current approaches; London Gatwick reported go-around percentages varying between 0.29% and 0.47% [13]. The percentage achieved in the distance-based scenario falls in this range. To achieve the same result for time-based spacing, the only available measure is addition of separation on top of the minimal separation used for the computation of the RTAs. Based on the separation margin distribution, 0.5 nm additional separation was applied (+0.3 nm). The separation violation percentage dropped to 0.27%.

TABLE II. SEPARATION GOAL PERFORMANCE

Self Spacing	Descriptive		
	Mean	Median	Separation Loss

Time-Based	0.47 nm	0.22 nm	1.42%
Distance-Based	0.37 nm	0.31 nm	0.32%

C. Runway Capacity

In this section the realized runway capacity is evaluated. For the distance-based self-spacing use is made of the arrival stream initially described in Section VI with a correction for the slowdown effect. In the time-based scenario use is made of the same arrival stream as initially described with 0.5 nm additional spacing applied on top of the separation minima to compute the RTAs. The capacity figures are based on 5000 randomly created arrival streams per self-spacing scenario. The runway capacity is a function of the traffic mix, because the traffic mix (on average: 40% heavy, 60% large) is determined randomly, variation in the runway capacity can be expected, see Figure 9. Table 3 lists descriptive statistics.

TABLE III. RUNWAY CAPACITY STATISTICS

Self-Spacing	Descriptive [AC/H]					
	Mean	Median	Std.	Min	Max	Range
Time-Based	35.7	35.3	3.3	26.7	49.7	23.0
Distance-Based	39.2	38.8	3.6	30.9	53.3	22.3

Time-based spacing results in a lower runway capacity than distance-based spacing. An ANOVA indicates that the difference is significant. ($F = 2560.04, p < 0.001$).

To compare the realized runway capacity with the theoretical maximum capacity a ‘packing factor’ is used:

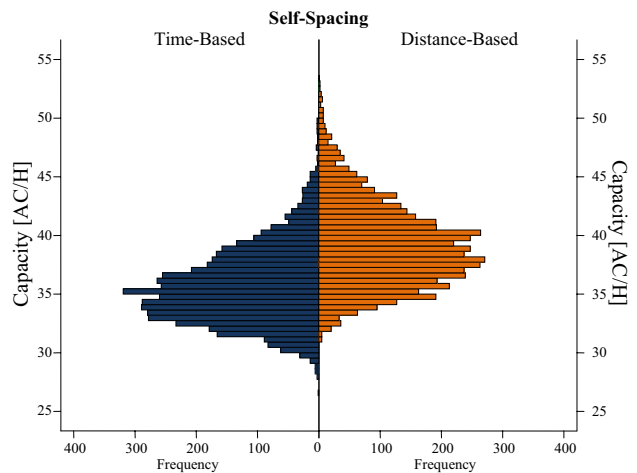


Figure 9. Runway Capacity

$$PF = \frac{\sum_{i=2}^k S_{allowed}}{\sum_{i=2}^k S_{actual}} \quad \forall S_{actual} \geq S_{allowed} \quad (1)$$

where k is the number the aircraft in the arrival stream, S_{actual} is the actual separation minimum between two aircraft, and $S_{allowed}$ the minimum safe separation. Separation violations are not included in the PF calculation. If $PF = 1$ the runway capacity is equal to the theoretical maximum. As expected the PF for distance-based spacing is higher than for time-based spacing, 0.90 and 0.81 respectively. In case of time-based separation it is clear that there is a significant reduction in

capacity. Distance-based self-spacing outperforms time-based spacing in terms capacity by more than three AC/H. Given the changing wind condition and pilot behavior and spacing capabilities of ATC the PF for the conventional approach will always be lower than one.

VIII. CONCLUSION

The aim of this research was to assess the feasibility assessment of implementing the TDDA at a high traffic density airport in a distance- and time-based self-spacing scenario. A fast-time simulation tool was developed and used to quantify the performance differences between distance- and time-based self-spacing in terms of capacity, noise reduction, and loss of separation.

For the time-based scenario no effects of preceding aircraft on trailing aircraft could be identified. However, an increase in separation with a negative effect on the airport capacity to assure safe separation was required. In the distance-based self-spacing scenario a slow-down effect was observed leading to a decrease in the noise reduction towards the end of the arrival stream. The deteriorating noise reduction was solved by altering the initial separation between aircraft in the arrival stream.

After making the aforementioned adjustments, distance-based and time-based self-spacing perform comparable except for the capacity where the distance-based scenario has a three AC/H advantage. Capacity is one of the major factors restraining the use of NAPs and especially CDAs; in that respect distance-based self-spacing is the most promising option.

REFERENCES

- [1] IATA, "Environmental Review 2004," Tech. rep., International Air Transport Association, September 2004.
- [2] ICAO, "Assembly Resolutions in Force," Assembly Resolution A35-5, International Civil Aviation Organization, Doc 9848, October 2004, Appendix C.
- [3] L.J.J. Erkelens, "Research Into New Noise Abatement Procedures for the 21st Century," *AIAA Guidance, Navigation, and Control Conference and Exhibit, Denver (CO), August 14 17*, No. AIAA 2000-4474, 2000, pp. 1-10.
- [4] W.F. de Gaay Fortman, M.M. van Paassen, M. Mulder, A.C. in 't Veld and J.-P.B. Clarke, "Implementing Time-Based Spacing for Decelerating Approaches", *Journal of Aircraft*, 44(1), Jan-Feb 2007, pp. 106-118.
- [5] A.C. in 't Veld, M.M. van Paassen, M. Mulder and J.P. Clarke, "Pilot Support for Self-Separation During Decelerating Approaches," *Proceedings of the International Symposium on Aviation Psychology, 12th, Dayton (OH), April 14-17, 2003*.
- [6] J.L. de Prins, K.F.M. Schippers, M. Mulder, M.M. van Paassen, A.C. in 't Veld and J.-P.B. Clarke, "Enhanced Self-Spacing Algorithm for Three-Degree Decelerating Approaches", *Journal of Guidance, Control & Dynamics*, 30(2), Mar-Apr 2007, pp.576-590.
- [7] J.P. Clarke, *A Systems Analysis Methodology for Developing Single Event Noise Abatement Procedures*, Ph.D. dissertation, Massachusetts Institute of Technology, 1997.
- [8] G.L. Slater, "Dynamics of Self-Spacing in a Stream of In-Trail Aircraft," *Proceedings of the AIAA Guidance, Navigation, Control Conference, Monterey (CA), August 11-14, 2002*, No. AIAA paper 2002-4927, 2002
- [9] J.M.C. de Groot, M. Mulder, M.M. van Paassen, "Distance-Based Self-Spacing in Arrival Streams," *AIAA Guidance, Navigation and Control Conference & Exhibit, San Francisco (CA), August 15-18, , No. AIAA 2005-6275, 2005*.
- [10] M.A. Vilaplana, E. Gallo, F.A. Navarro and S. Swierstra, "Towards a formal language for the common description of aircraft intent," *Digital Avionics Systems Conference, 2005. DASC 2005.*, The 24th Volume 1, Issue , 30 Oct.-3 Nov. 2005 Page(s): 3.C.5 - 3.1-9 Vol. 1
- [11] A.M.P. de Leege, *Three-Degree Decelerating Approaches in Arrival Streams*, MSc. thesis, Faculty of Aerospace Engineering, Delft University of Technology, September 2007.
- [12] N. Ho, and J.P. Clarke, "Mitigating Operational Aircraft Noise Impact by Leveraging on Automation Capability", *1st AIAA Aircraft Technology, Implementatio, and Operational Forum, AIAA*. Los Angeles (CA), No. AIAA 2001-5239, 2001.
- [13] BAA London Gatwick, "Flight Evaluation Report 2001/02," Tech. rep., British Airports Authority, 2002.

Separation Minima Standards:

Research of Current Applicable Minima Laid Down and Foundations

Daniel MOSQUERA BENITEZ

Civil Systems Division
Ingeniería de Sistemas para la Defensa –Isdefe
Madrid, Spain
dmosquera@isdefe.es

Gustavo CUEVAS ANGULO

Civil Systems Division
Ingeniería de Sistemas para la Defensa –Isdefe
Madrid, Spain
tgcuevas@isdefe.es

Abstract— Separation Minima (SM) is the minimum distance a/c need to fly apart from each other at all times to ensure safety. This applies to the three axes: Vertical, Lateral and Longitudinal Separations Minima (See Figure 1. A/c Separation Axes). Many Standards of Separation Minima were defined based on expert judgment and technology available at the time were laid down them, the leap in technology since then makes the SM standards must be updated. However, many of them have not been modified to reflect modern technological capabilities. Due to how SM have been defined (in many cases) makes each region around the world have laid down different values for same operational case or separation rules were laid down with different criteria and context descriptions. As demand is expected to treble by 2020, one of the ATM system challenges is to manage the expected increase in air traffic demand and, reducing SM becomes a potential solution part that would contribute to achieve this challenge, keeping always in mind that a/c Separation Standards reduction increases airspace capacity but can also reduce safety levels, which must be preserved as part of the challenge.

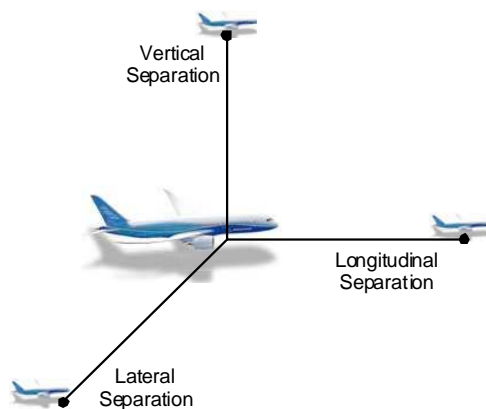


Figure 1. A/c Separation Axes

The best starting point in order to identify which reductions in SM could be realized is undertaking a research of the current separations minima and their foundations. Answering the questions what? and why?, will make it someday possible to answer the question how?. The work that the present paper originated from was focused on extracting information from several international regulations and (ICAO, FAA, British Regulations, Australian Regulations, Canadian Regulations

and Eurocontrol). These regulations/documents include a/c separation minima cases, description of SM values classified by PoF, a/c operation, direction/tracks/routes, conditions and operational context, technology involved, separation axis. In addition, identifying aerodynamic factors, human factor, hazard/risks, equipment precision, surveillance mode, models identification, etc were also investigated. The valuable results of this research are unprecedented in their contents and for the useful way to they are presented.

TABLE I. ACRONYMS

Acronym	Meaning
a/c	Aircraft
ENR	En-Route
FAA	Federal Aviation Administration
ICAO	International Civil Aviation Organization
PoF(s)	Phase of Flight(s)
SM	Separation Minima
SMS	Separation Minima Standard
TMA	Terminal Manoeuvring Area
WV	Wake Vortex
RWY(s)	Runway(s)

I. INTRODUCTION

A. Background

RESET (Reduced Separation Minima) project, funded by the European Commission Directorate General – Transport and Energy inside the 6th Framework Program is a project lead by AENA and formed by 15 partners around Europe with the collaboration of the FAA.

Within this project efforts to research current separation minima status were carried out in Work Package 2 (WP2) and their foundations in WP3.1, both WP were lead by Isdefe with the collaboration of RESET partners.

B. Objectives

The objective of this paper is to describe the current status of SMS laid down (ICAO, Local regulations and Eurocontrol) which is further described by the Separation Minima List. On this list is possible to check different separation minima cases and to identify differences between the different standards. The analysis of the current regulations and separation minima requirements applicable to the different PoFs provides a starting point to make an attempt to reduce separation minima. The current status of separation minima is the product of WP2 and WP3.1 of RESET project. In order to achieve this, several tasks were carried out, hard documentary research and productive results have been obtained. This document describes the objectives pursued, the inputs used, a process description and the methodology followed along the research, the output obtained and its structure. With the information obtained through separation minima research and foundations research, a checklist containing the applicable separation minima for different PoFs was created. In the list, standards will be further categorized and classified, taking into account the PoF and the various factors that are considered in each particular regulation. This will be delivered as an input to studies such as the building a model, prioritization and integration of results, dissemination, and whenever possible, areas where such regulations can be improved, in terms of minima reduction and best practices used, tasks to be carried out inside RESET project.

II. RESEARCH METHODOLOGY

A. SMS

The first activity carried out was getting inputs, searching associated documentation for building input repository; the input documentation repository used to documentary research is listed on TABLE II. PRINCIPAL Documents Analysed for SMS identification hereunder.

TABLE II. PRINCIPAL DOCUMENTS ANALYSED FOR SMS IDENTIFICATION

Source/Autor	Code
ICAO	Docs: 4444, 9476, 9830, 9426, 9574, 9613, 9689, 9643, 7030, 9854, Annex 2, Annex 11.
FAA	Order 7110.65
Civil Aviation Authority	CAP 493
TC Civil Aviation	Standard 821
Civil Aviation Safety Authority	CASR Part 172

A template to be completed was laid out according to ICAO 4444 document. Information from other ICAO documents and regulations from EUROCONTROL and the American (FAA), British (BAA), Canadian and Australian regulatory bodies were incorporated.

Next step was a lay-out for being filled with separation minima data (Excel worksheet template) defined and based on ICAO 4444 document. The separation standards were

classified in a useful way to facilitate the work to be performed by those who consults the Separation Minima List. In order to classify the separation standards according to the methodology used by ICAO document 4444 it was essential that this template, which will be the same one used for all other standards and regulations, was agreed among all different points of view and contributors to increase coherency. The documentary research carried out has consisted of analyzing regulations, extracting from these, values of separation minima laid down, PoF and operational conditions and constrains applicable, as well as merging information. This information was added to the output table named Separation Minima List.

Once the input documentation has been obtained, the methodological road followed starts with the elaboration of a layout of the checklist containing PoFs to classify regulations. The structure of the list is in accordance to ICAO 4444 document. The list was completed with ICAO and non ICAO SMS data, identifying in respective documentation: PoF, operational conditions/constrains/context and their applicable separation minima values.

It was important to decide which local regulations should be included in the list. This study considers British, Canadian and Australian regulations. EUROCONTROL regulation was also considered to investigate the future regulation environment. All separation minima cases analyzed were represented by rows in table. Checklists were later refined to integrate them in a common one. Finally, a draft poster was produced in order to facilitate dissemination work as much as possible.

The current status of SMS is provided by means of separation minima cases contained in the output named Separation Minima List (located on RESET website <http://www.reset.aena.es>), a self-explanatory table that contains information about the standards laid down in regulations. The header columns described in the table, whose fields were completed with information related to each separation minima case, are the following: *PoFs, Operation, Characteristics, Direction/Tracks/Routes, Conditions, Context, Means, Control by, Picture, Separation, Based on, Separation Minima laid down, Reference, Observations.*

The Separation Minima cases were organized according to the above mentioned columns, so the information can be sorted by columns, by means of a filter tool, to find the data needed or identify the differences among different current standards by comparison. Under the qualities of this way to present the results there are the following advantages: it allows the introduction of new separation minima regulations and makes it easy to update the included information.

B. SMS, Foundation Research

Once all SMS were identified, next steps resulted into the identification of the foundations that support the Separation Minima Standard Definition.

The applied methodology follows a five stages approach, each stage with a specific objective as is described hereunder:

- Stage 1. Analyze the current foundation of the SMS.
- Stage 2. Identify how current minima separation have been quantified.
- Stage 3. Complete and classify the list of factors for the selected relevant cases.
- Stage 4. Selection of the SMS that will be analyzed in more detail.
- Stage 5. Group the SMS to be studied by “thematic areas”.

The acknowledge of these five stages, should conclude in answering the following questions:

Where was the definition of the Separation Minima Standard established?, How were the separation minima standard established? How strong are the foundations? and which factors contribute to the Separation Minima definition?, Which from the identified SMS are most relevant to be analyzed in more detail?, Do these SMS have some similar characteristics?, Is there a way to study the standards by groups?, Is all the information needed to analyze each group of Separation Minima Standard available?

The following paragraphs, explains how each step were developed, main conclusions and products.

1) **Stage 1. Analysis of current foundations of SMS.**

The main objective within this stage was to analyze the current foundations of all SMS identified previously.

To acknowledge that, the SMS Current Status list was considered as input and all the main documents were distributed in order to start the “looking & finding process” of all the foundation for each SMS, or (by default) any clue or piece of information that could contribute to the understanding of the SMS definition.

The documents that were analyzed for foundations are listed in the table hereunder:

TABLE III. DOCUMENTS ANALYSED FOR FOUNDATIONS

Doc 4444, Doc 9426, Doc. 7030/4, Doc 9830, Doc 9643, FAA Order 7110.65, Annex 2, Annex 11, British Regulation, Australian Regulation, Canadian Regulation, ATISN 93, CARE-ASAS
--

The results of this process were documented in the SMS Current Status List in order to guarantee the link of each Separation Minima and its Foundations.

2) **Stage 2. Identification of how current SMS have been quantified.**

This stage aims to qualify the stage 1 results, by a complete “auto evaluation” of the research process. To

accomplish this indicator called Foundation Research Assessment (FRA) was created within four possible options: *Success, Few Possibilities, Uncertainty, Unaware*. Each researcher was asked to check each SMS with one of the options, generating this way common criterion to evaluate the foundation strength and availability.

TABLE IV. FOUNDATION RESEARCH ASSESSMENT, OPTIONS

Success	The foundation was found.
Few Possibilities	The foundation research was carried out within a lot of effort and looking in deeply in a lot of information sources but finally the foundation was not found. In some cases just some clues were recorded.
Uncertainty	Within the effort and time allocated it was not possible to carry out a deeper foundation research.
Unaware	Within the effort and time allocated it was not possible to carry out the foundation research.

The results of the Foundation Research Assessment are shown in the table hereunder.

Figure 2. Foundation Research Assessment, Output

Separation Type	(Todas)				Total general
	Phase of Flight				
Foundation Research Assessment	Aerodrome	Arrival (TMA)	Departure (TMA)	En-route	Total general
FEW POSSIBILITIES	80	13	57	156	306
SUCCESS	5	16	14	57	92
UNAWARE	5	1	2	8	16
UNCERTAINTY	93	19	19	85	216
Total general	183	49	90	300	622

3) **Stage 3. Research and Identification of the contributing factors to the SMS.**

To acknowledge this objective after several discussions, brainstorming, etc. some new columns were added to the table. These columns contained the following information:

Identification of Contributing Factors.

Aerodynamic Effects. Aerodynamics factors and/or effects that have influence on the separation case.

Human Factors. Human factors that have influence on the separation case: controllers, pilots, etc.

Hazards/Risk. Identified the separation minima reduction hazards.

Equipment Precision. Precision of system, equipment or device for applying this separation or on which it is based on.

Surveillance. Considerations about surveillance have an influence on the separation application.

Each Separation Minima Standard was analyzed within this factors list in order to identify which of these have impact on the SMS definition and once identified it was recorded in the SM Current Status List.

Product: An excel file with all the SM Standards that perform the three established criteria.

4) **Stage 4. Analysis and selection of the SMS to be studied in more detail.**

The main objective of this stage was to identify the most important and/or relevant SMS where reducing some separation could have an important impact on accomplishment doubling capacity. To acknowledge it criteria for the identification of the Separation Minima cases more relevant to be study in more detail in future steps was created. This criteria was established as it follows.¹:

- Standards associated to operations in Europe will be preferred to others.
- Most commonly used standards are preferable.
- Standards based upon the most advanced technology will be preferred to others.
- All the different PoFs (Airport, TMA/Departures, TMA/Departures, ENR) should be covered.

Starting from these criteria, a Criteria Check Analysis was performed for each separation minima standard. Inside the Separation Minima List a column called Criteria Check Analysis (CRA) was created with two possible options: YES/NO.

After this, and in order to be as much effective as possible, it was decide not only to analyze the SMS that were considered more relevant to be studied in more detail, but also were considered those that have strong foundations and those whose contributing factors have been identified.

Finally the SMS selected to be studied in more detail were those which:

- Have been *marked with a YES in the Criteria Check Analysis* (stage 4 input).
- Have been *marked with a SUCCESS or a FEW POSSIBILITIES in the Foundation Research Assessment* (stage 2 input).
- Have been it corresponding *Contributing Factors identified* (stage 3 input).

At the end of this stage the initial Separation Minima List of 622 standards were filtered and reduced to 157 standards to be studied in future stages.

Product: An excel file with all the SM Standards that match the three established criteria.

5) **Stage 5. Grouping of the SMS by “thematic area”.**

The main objective of this step was to group all the ID’s that correspond to the same or very similar cases, in other words by “thematic area”.

¹ 1st technical meeting minutes, section 4 *Definition of Criteria and selection of Standards for Factors Completion*, page 4.

To acknowledge this objective the final customer of this grouping needed to be identified. It was agreed that this work should focus on the *Modeling Phase* therefore this grouping should address the needs of this phase.

As the *Modeling Phase* needed to document and to compile all the mathematical models, simulation models, collision risk models, formulas or equations, etc. that have been used to define the SM Standard, it was agreed to group the SM Standards selected in step 4 by “thematic area”.

Two “step by step” grouping methodologies were defined for *Aerodrome* and *TMA/ENR*. These methodologies were defined as it follows:

TABLE V. GROUPING METHODOLOGY FOR ENR & TMA

Five Steps were defined in this methodology	
-	First Step: PoF
-	Second Step: Type of Control (Radar/Procedural/ADS)
-	Third Step: Type of Separation (Longitudinal/Lateral/Vertical)
-	Fourth Step: Based on (Time/Distance)
-	Fifth Step: - (RNAV/Navigation aids/WV)

TABLE VI. GROUPING METHODOLOGY FOR AERODROME

Seven Steps were defined in this methodology:	
-	First Step: PoF
-	Second Step: Operation (Land/Take off/Interlaced)
-	Third Step: Rwy Configuration (Same/Parallel/Crossing)
-	Fourth Step: Type of Control (Radar/Procedural)
-	Fifth Step: Type of Separation (Longitudinal)
-	Sixth Step: Based on (Time/Distance)
-	Seventh Step: - (WV/Rwy Separation)

At the end of this stage the FILTERED Separation Minima List of 157 standards were grouped and reduced to 21 groups of SMS.

Product: An excel file with all the SM Standards grouped by thematic area, and two PDF files with the methodology applied when grouping the SM Standards for TMA/En route and Aerodrome phases.

III. SEPARATIONS MINIMA AND FOUNDATIONS CURRENT STATUS

All Separation Minima and Foundations information obtained from this research was integrated in an excel table, which is a Separations Minima Data Base (located on RESET website).

A. POFS (PoF)

During a flight an a/c goes trough different PoFs which have different hazards and risks and therefore different Separation Minima are applied. RESET research, specifically all the foundation assessment was focused on the flight path. The four phases defined in RESET were:

- *Aerodrome (including take offs, lands and taxi).*

- TMA Departures (including climb, same level).
- ENR (including same level, climbing and descent).
- TMA Arrivals (including descent, same level).

In the lines hereunder these each PoF and their associated operations are explained in more detail.

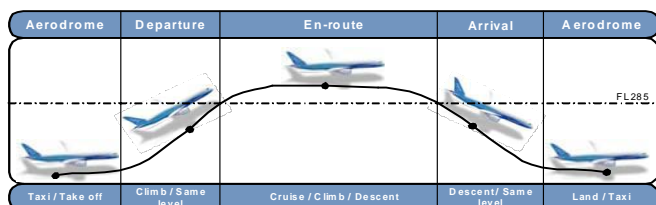


Figure 3. PoFs

1) Aerodrome

Once the push back & start up clearance is given to the pilot and the a/c has started taxiing to the active rwy its interaction with other a/c in both the apron and the taxi ways requires the establishment of Separation Minima to guaranty the a/c's safety. This PoF (within RESET project) is defined from that event to the take off (including initial climb).

The research carried out in this PoF identified in a total of 183 SMS identified. Under this category 3 Operations were analyzed: *Taxi, Take off and Land.*

The information that could be found in the *RST-WP3-ISD-004-Separation Minima List* regarding this PoF is exemplified in the following paragraph:

ID0062

This Standard is a longitudinal separation based on distance in a radar scenario that is defined by ICAO document 9643 in section 2-3 as 3NM between both a/c doing a dependent parallel instrument approaches (mode 1) or a dependent parallel instruments approaches (mode 2) unless more separation is required due to WV influence.

During the research process on this standard, have been identified several contributing factors as aerodynamics effects, human effects, and so on and despite a deep research was carried out the foundation was not found within the time and effort allocated in the project.

2) TMA/Departure & TMA/Arrival

This PoF was split into two, one from the initial climb to FL285 (TMA Departures) and the other one from FL285 to the initial approach (TMA Arrival).

The research carried out within this PoF identified in a total of 139 SMS identified, divided in 90 SMS for TMA/Departure and 49 for TMA/Arrival. Under this category four Operations were analyzed: *Climb, Cruise (same level), Hold and Descent.*

The information that could be found in the *RST-WP3-ISD-004-Separation Minima List* regarding this PoF is exemplified in the following paragraph:

ID0100

This Standard is a longitudinal separation based on distance in a radar scenario that is defined by Canadian Regulations in chapter 3 section 3.0 as 6NM between two a/c climbing (the preceding heavy and the follower light) in the same route due to WV effects.

During the research process on this standard have been identified several contributing factors as equipment precision, surveillance, aerodynamics, etc. Regarding the foundation, despite a deep research was carried out the foundation was not found within the time and effort allocated in the project, but it was identified CARS part 8 standard 8213.1 as possible reference of understanding the SM standard definition.

3) ENR

ENR PoF (within RESET definition) starts once FL285 is passed and cruise altitude is reached, and ends when descending bellow FL285.

The research carried out within this PoF identified in a total of 300 SMS was identified. Under this category four Operations were analyzed: *Climb, Cruise (same level), Hold and Descent.*

The information that could be found in the *RST-WP3-ISD-004-Separation Minima List* regarding this PoF is exemplified in the following paragraph:

ID0542

This standard is a longitudinal separation based on distance in a ADS scenario that is defined by ICAO document 4444 in section 5.4.2 as 30NM between two a/c flying in the same route and using ADS with a maximum of 14 minutes of periodic reporting interval and with an RNP=4.

During the research process on this standard have been identified several contributing factors as relative a/c position and velocities, a/c reaction, environment, WV profile, etc. Regarding the foundation within the time and effort allocated in the project it was possible to find some foundations based on a Collision Risk Model.

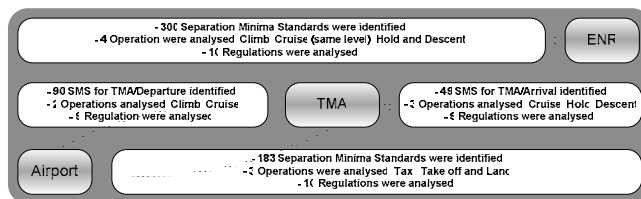


Figure 4. Summary of SMS & Foundation characteristics per PoF.

B. WV Separation: Minima of Minimas

WV separations analysis was carried out separately due to differences in the categorization in each regulation. The only way to compare them was by means of the weight (See Figure 5. WV category broken down in Weight Range).

MTOW Tons	Heavy	Heavy	Heavy	Heavy	Weight Range
200	Heavy	Heavy	Heavy	Heavy	162 < W
189					
182					
175					
168					
162					
154					
147					
140					
136					
133	Medium	Large	Medium	Upper-Medium	136 < W < 162
126					
119					
115					
112					
105					
104					
98					
91					
84					
77	Medium	Large	Medium	Lower-Medium	115,7 < W < 136
70					
63					
56					
49					
42					
40					
35					
28					
21					
18	Light	Small	Light	Small	104 < W < 115,7
17					
14					
10					
7					
5					
2					
1					
0					
0					
ICAO	FAA	UK 4	UK 5		40 < W < 104
CARS					18,6 < W < 40
CASR					17 < W < 18,6
					7 < W < 17
					W < 7

Figure 5. WV category broken down in Weight Range

Looking across different regulations brings out nine weight ranges; this division amongst regulations allows the determination of the smallest minima and a comparison of separation minima applicable among standards (See Figure 6. WV SM laid each regulation down according to weight ranges). As shown, SM between one a/c followed by another varies many times depending on place (regulation) where the a/c operation is being performed. This implies that WV separation would be settled in different way for same a/c. For instance, one A300 a/c followed by one Saab 340 would be separated 6 NM or 5NM depending on if the operations are being carrying out in the USA or in another country, where ICAO regulations are adopted. Apart from, regulations analyzed do not apply to a/c with over 200 tons-MTOW a/c.

Preceding	Following	ICAO	Canadian	Australian	FAA	UK 4	UK 5	Minima of Minimas
162 < W	200 < W	4 NM	4 NM	3 NM	4 NM	4 NM	4 NM	3 NM
162 < W	162 < W < 200	4 NM	4 NM	4 NM	4 NM	4 NM	4 NM	4 NM
162 < W	136 < W < 162	4 NM	4 NM	4 NM	4 NM	4 NM	5 NM	4 NM
162 < W	115,7 < W < 136	5 NM	5 NM	5 NM	4 NM	5 NM	5 NM	4 NM
162 < W	104 < W < 115,7	5 NM	5 NM	5 NM	5 NM	5 NM	5 NM	5 NM
162 < W	40 < W < 104	5 NM	5 NM	5 NM	5 NM	5 NM	5 NM	5 NM
162 < W	18,6 < W < 40	5 NM	5 NM	5 NM	5 NM	5 NM	6 NM	5 NM
162 < W	17 < W < 18,6	5 NM	5 NM	5 NM	6 NM	6 NM	6 NM	5 NM
162 < W	7 < W < 17	5 NM	5 NM	5 NM	6 NM	6 NM	7 NM	5 NM
162 < W	W < 7	5 NM	5 NM	6 NM	6 NM	6 NM	7 NM	6 NM
136 < W < 162	162 < W	4 NM	4 NM	4 NM	4 NM	4 NM		4 NM
136 < W < 162	136 < W < 162	4 NM	4 NM	4 NM	4 NM	4 NM	3 NM	3 NM
136 < W < 162	115,7 < W < 136	5 NM	5 NM	5 NM	4 NM	5 NM	3 NM	3 NM
136 < W < 162	104 < W < 115,7	5 NM	5 NM	5 NM	5 NM	5 NM	3 NM	3 NM
136 < W < 162	40 < W < 104	5 NM	5 NM	5 NM	5 NM	5 NM	4 NM	4 NM
136 < W < 162	18,6 < W < 40	5 NM	5 NM	5 NM	5 NM	5 NM	4 NM	4 NM

Figure 6. WV SM laid each regulation down according to weight ranges

C. TYPE OF SEPARATION

1) Longitudinal Separation

Around 203 SMS were identified. For the Aerodrome PoF, a/c taxiing operation, a 200 meter separation is applied, considering a taxi speed of 30 kt. No influence of aerodynamics is considered, but with separation contributing factors like: Pilot monitoring/situational awareness, Pilot response time, Controller/Pilot communication/coordination, Controller monitoring/situational awareness, the response time of any control function should be less than 0.5 second, Controller display target position error, Accuracy of measured position after processing, Reporting interval, Radar surface. According to Advanced Surveillance Movement Guidance and Control System (ICAO 9830). The contribution factors to separation contributing are: Atmosphere Parameters (temperature, air density, pressure, thermal stratification, Eddy dissipation rate, wind direction), horizontal and vertical positions and closing angles, vertical path separation at crossing point, speed, airplane weight, dimensions and geometry, pilot monitoring/situational awareness, Controller/Pilot communication, Controller monitoring/situational awareness, Controller workload, Surveillance (Update rate, Controller display target position error, accuracy of measured position after processing, reporting interval. The main hazard could be presented are WV Encounter (WVE) and possible crew/passenger injury, loss of control and/or structural damage. For the Departure (TMA) PoF, climbing operation, in ICAO is laid down to be separated 2 min, but if separation needs to be maintained or increased while vertical separation does not exist, then 5 min (while vertical separation does not exist if a departing a/c will be flown through the level of a preceding departing a/c and both a/c propose to follow the same track). In FAA different standards are laid down according to specific conditions: 1 min when preceding a/c turns immediately after take off, 2 min (within 5 min after take off) or 3 min for changing level. Based on distance could be separated 3 NM (within 13NM DME/ATD after take off) or 5 NM (between DME equipped a/c; RNAV equipped a/c using ATD and between DME and ATD a/c provided the DME a/c is either 10,000 feet or below or outside of 10 miles from the DME NAVAID). The main separation contributing factors are: Human (Monitoring/ situation awareness, Pilot response time, Controller response time Controller workload), there is no consideration to on-board equipment and there is no use of any surveillance system except visual means. This separation is based on the probability of mid-air collision due to the leading a/c been caught up by the a/c behind, mainly due to a drift in the speed or own position calculation by any or both a/c. The hazards could be presented principally are: Mid air collision, Lost of longitudinal separation and WV encounter and consequently resulting in a lost of control. According to Australian regulations WV separation varies depending on the following a/c climbing to the higher level or following a/c climbing to the lower level, following a/c climbing to the same level. This regulation considers separation based on distance by means of RNAV, which involves equipment precision factors (Navigation sensor

error, Airborne receiver error, Flight technical error, Navigation reliability, Navigation system integrity, A/c certified for RNP-10 or RNP-4). In Canadian regulations is laid down in following way, 3 minutes until altitude levels are crossed (prior to reaching 15 miles from the departure rwy, the following a/c will climb through the altitude of the leading a/c, and both a/c will follow the same track until vertical separation is established) or 5 min (when the following a/c will climb through the altitude of the leading a/c and both a/c follow the same track until vertical separation is established), and based on distance, 10 NM until altitude levels are crossed (the following a/c will climb through the altitude of the leading a/c, and both a/c use DME and follow the same track to or from the same DME NAVAID immediately after take-off). For the Arrival (TMA) PoF, descending operation are almost the same as climbing operations for changing level, but now for descent. For the ENR PoF, cruising and maintaining same level, there are many different separation cases, the most homogeneous one is control by mach number technique, where all regulations coincide. When the preceding a/c is Mach 0.02, 0.03, 0.04, 0.05 or 0.06 faster than the following a/c, 9, 8, 7, 6 or 5 minutes separation applies respectively. The human factors involved are: Pilot monitoring/situational awareness, Pilot response time, Cockpit Resource Management, Crew workload, Controller/Pilot communication/coordination, Controller monitoring/situational awareness, Controller response time, Controller workload, Controller interaction with displays / automation / decision aids. Here it is important that longitudinal separation minima are based upon quality of meteorological information available. For this PoF there are so many and diverse specific conditions laying down different separation minima cases.

2) Vertical Separation

Around 37 SMS were identified. For the Aerodrome PoF there is not vertical separation. For Departure (TMA), Arrival (TMA) and ENR PoFs, for changing level and cruising maintaining same level operations, the vertical separation minima separation are based on distance and regulations are very standardized and aligned. This is due to this standard is very well documented in ICAO Doc 9536 - Sixth Meeting RGCSP/6 Review of the General Concept of Separation Panel - Volumes 1 and 2, and ICAO 9574 - Appendix A. The value of 300 meters applies for vertical separation below FL 290, between above FL 290 and below FL 410 could applies 300 or 600 meters separation (within designated airspace), and at or above FL 410 the vertical separation is 600 meters. The separation value corresponding to Unlawful Interference is 150 meters (500 feet) according to Attachment B – ICAO Annex 2. Among contributing factors to separation which were considered are aerodynamics factors apart from maneuver response capabilities, vortices will not normally descend more than about 400 – 500 ft they can descend further if there are significant downdraughts, or they may be presented due to an a/c has climbed or descended. Human factors involved for instance are: Controller confidence, Pilot confidence, Consensus of the users, Pilot monitoring/situational awareness, Pilot response time, Cockpit Resource

Management, Crew workload, Controller/Pilot communication/coordination, Controller monitoring/situational awareness, Controller response time, Controller workload, Controller interaction with displays/automation/decision aids, Training/experience. Collision is main risk for this separation, it was calculated considering a Target Level Safety (TLS) value of 2.5×10^{-9} fatal accidents per a/c flight hour. Based on a Collision Risk Model (CRM) the risk of collision modeled is that due to the loss of procedural vertical separation between a/c flying above FL 290 in a given portion of airspace. WV encounter is a potential risk due to occasionally vortices will descend further and be encountered by a/c flying only 1000ft below when Reduced Vertical Separation Minima (RVSM) are in operation, vortices will not normally descend more than about 400 – 500 ft they can descend further if there are significant downdraughts, or they may be present because an a/c has climbed or descended.

3) Lateral Separation

Around 64 SMS were identified. For the Aerodrome PoF, different values of separation minima are applied for landing and taking off operations, considered among them: Independent parallel instrument approaches, Dependent parallel instrument approaches, Simultaneous use of parallel rwys, Segregated operations on parallel rwys, Separation between rwys centre line, Successive departures, Simultaneous departures. For these kind of separations minima equipments such as ILS and/or MLS are necessary on both rwys, suitable surveillance radar available, satisfactory two-way radio communication. For the Departure PoF (climbing operation) according to Australian regulation different values are laid down based on distance for several cases such as: for a/c turns of 16 degrees through 90 degrees, for a/c turns of 91 degrees through 180 degrees, in addition to the 14/17 of a mile protected on the over flown side of the track. For the En Route PoF (cruising and maintaining same level) are laid down separations, some of them based on Angle/distance, the contexts described can be for instance: for turns of 15° or less, for turns from 16° through 90°, for turns from 91° through 180°, oceanic procedures, North Atlantic ICAO region, Caribbean ICAO region, Pacific ICAO region, North America ICAO region-Arctic CTA, by requiring a/c to fly on specified tracks which are separated by a minimum amount appropriate to the navigation aid or method employed, non real time radar monitoring or control of the lateral deviation is exercised; distance between 2 VOR less than 278 km (150 NM), protected on the over flown side of the track.

IV. CONCLUSIONS

The first conclusion is that there are so many separations minima rules, around 622 cases. Among existing SM standards, some differences have been detected between the values applied for the same standards by different regulators. The regulators should agree and study standards in order to improve and standardize SM. For instance, WV separation minima analysis reflects several differences among regulations based on this conclusion the most restricted one

could be extended to the rest of the states or studying if minima of minima complies with safety conditions and applies it, reducing separations. To make sure the studies of what standards are applicable or sensitive to be reduced, it is necessary to carry out in depth studies into the foundations and the models and principles upon which were they are based. The foundations analysis defines the background to these rules and enables to know if it is possible to improve or reduce separations.

Foundations were found mainly in Regulations. For the 600 SMS identified in WP2 a "Foundation Research" was carried out starting from International regulation as ICAO, EUROCONTROL and then looking at local regulation such as: British, Canadian, Australian, FAA.

From the 622 SMS identified.

- For 15% of the cases the foundations were found
- For 49% of the cases the foundation research was carried out within a lot of effort and looking in deep in a lot of information sources but finally the foundation was not found. In some cases just some clues were recorded.
- For 35% of the cases within the effort and time allocated it was not possible to do a deeper foundation research on these SMS.
- For the 1% of the cases within the effort and time allocated it was not possible to do the foundation research on these SMS.

These figures results on the following answer: foundations are not too much strong as they could be, so their definition's improvements could impact directly on the Separation Minima Reduction.

Mainly in the definition of a Separation Minima Standard a group of factors is involved.

The groups of factors identified on the 600 SMS identified were Aerodynamics Effects, Human Factors, Hazards & Risk, Equipment Precision, Surveillance.

Within the effort and time allocated to perform these studies, 600 SMS were too many. Therefore it was discussed which ones were more relevant to be analysed much more deep and it was agreed that: separation minima cases more relevant for further studies must acknowledge the following criteria:

- Standards associated to operations in Europe will be preferred to others.
- Most commonly used standards are preferable.
- Standards based upon the most advanced technology will be preferred to others.
- All the different areas (airport, TMA, en route) should be covered.
- Foundation Research Analysis resulting on SUCCES or FEW POSSIBILITIES.

- At least one Contributing Factors Identified.

Once this analysis was performed it was result on 157 SMS that acknowledge within these criteria, in other words, that will be study in more detail.

Several SMS are quite similar by concept, therefore it has been analysed and them created a Grouping Methodology which main objective was to bring together all the SMS that are similar in order to facilitate their further studies. Finally 21 groups were created for Aerodrome, TMA and ENR SMS.

For these kind of studies "too much information" is not enough, specifically for the modeling definition, therefore the answer should be no. But in terms of "what we have found" the answer is:

- 25 documents were analyzed (ICAO, FAA, AIAA, WakeNet, NLR, Eurocontrol, Cambridge University, etc.).
- 36 references to mathematical models, simulation models, collision risk models, SM foundation, etc. were captured.
- The groups with more quantity of references were the ones related to Wave vortex.

ACKNOWLEDGEMENTS

We would like to recognize the valuable contributions to this research carry out by all RESET partners, specially the ones involved in Reset WP2 and WP3.1.

Note: All documents included on this paper are available on RESET website <http://reset.aena.es>.

DISCLAIMER

Opinions, interpretations, recommendations, and conclusions contained herein are those of the authors and are not necessarily endorsed by the European Commission or other RESET partners.

REFERENCES

- [1] RESET WP2 - D2.1 Separation Minima Current Status
- [2] RESET WP3 - D3.1 Separation Budget Componets
- [3] RST-WP3-ISD-004-Separation Minima List,
- [4] RST-WP3-ISD-005-FILTERED/Separation Minima List,
- [5] RST-WP3-ISD-006-GROUPED/Separation Minima List,
- [6] RST-WP3-ISD-007-Aerodrome,
- [7] RST-WP3-ISD-008-TMA/ENR
- [8] ICAO Document 4444.
- [9] ICAO Document 9426.
- [10] ICAO Document. 7030/4.
- [11] ICAO Document 9830.
- [12] ICAO Document 9643.
- [13] FAA Order 7110.65.

Track 3

Safety & Security

A Review of the Research on Risk and Safety Modelling in Civil Aviation

Fedja Netjasov

Division of Airports and Air Traffic Safety
The Faculty of Transport and Traffic Engineering,
University of Belgrade
Belgrade, Serbia
f.netjasov@sf.bg.ac.yu

Milan Janic

Department of Transport and Infrastructure
OTB Research Institute,
Delft University of Technology
Delft, The Netherlands
janic@otb.tudelft.nl

Abstract— Risk and safety are always considered the most important operational characteristics of contemporary civil aviation. Usually, they refer to the potential occurrence of air traffic accidents which might result in loss of life, damage to infrastructure and third party property damage. Consequently, they have been regarded as externalities in addition to other adverse effects such as noise, air pollution, land-use, water/soil pollution, waste, and congestion. Due to their inherent very high importance, risk and safety have been issues of continuous research ranging from purely technical/technological aspects to strictly institutional. These issues warrant the setting up of adequate regulations on system technology designs and operations. This paper deals with a review of part of the research on risk and safety modeling in civil aviation. In such a context, the basic (generic) concepts and definitions of risk, safety and their evaluation are described. A review of the research is focused on four categories of methods/models for risk and safety assessment: causal for aircraft and air traffic control/management (ATC/ATM) operations, collision risk, human factor error and third-party risk. The review is carried out with respect to their purpose, problems, recommendations and relation to new technologies.

Keywords: *civil aviation, risk and safety, models/methods, new technologies*

I. INTRODUCTION

Nowadays, the air transport system is recognized as one of the fastest growing areas within the transport sector as well as in overall regional and world economies. According to many forecasts this growth will continue at an average rate of 5% in passenger and 6% in freight transport demand over the next two decades. It will primarily be driven by overall economic growth, further globalization of the regional and world's economy, and even further decreasing of airfares thanks to among other factors the growth of the low-cost carrier's market share. The system infrastructure – airports and Air Traffic Control/Management (ATC/ATM) although in many cases acting as temporal “bottlenecks” are expected to be able to support such growth safely, efficiently and effectively.

Physically and operationally, the air transport system is a rather complex system with the main components - airlines, airports and air traffic control services - interacting with each other on different hierarchical levels constituting a very

complicated, highly distributed network of human operators, procedures and technical/technological systems. In particular, risk of accidents and related safety in such a complex system is crucially influenced by interactions between the various components and elements. This implies that providing a satisfactory level of safety (i.e., low risk of accident) is more than making sure that each of the components and elements functions safely [1]. Due to such inherent complexity and severe consequences of accidents, risk and safety have always been considered as issues of the greatest importance for the contemporary air transport system [2]. Consequently, they have been a matter of continuous research from different aspects and perspectives ranging from the purely technical/technological to the strictly institutional. In general, the former have dealt with design of safe aircraft and other system facilities and equipment. The later have implied setting up adequate regulations for system design and operations.

The objective of this paper is to present a review of part of the research dealing with risk and safety in the contemporary civil aviation system.

II. CONCEPTS AND DEFINITIONS OF RISK AND SAFETY

For a long time, risk and safety have been differently and ambiguously interpreted depending on the system and purpose [3]. For technical systems, risk is related to the chance of failure of components or of the entire system causing exposure to hazard and related consequences. In economic business systems, risk is a chance of being exposed to the hazard of losing business opportunities and/or money due to making decisions under uncertain circumstances. In social systems, risk is the chance of being exposed to the hazard of injuries and/or losing of life. Consequently, risk could be considered as combination of the probability (or frequency of occurrence) and the magnitude of consequences (or severity) of a hazardous event [4].

In the air transport system, risk and safety have always been related to air traffic accidents which resulted in the significant loss of life and property (aircraft and the property on the ground). Assuming that making an air trip is an individual choice and that the system deploys some resources to satisfy such choice, four types of risks can be identified in the air transport system [2]: i) real risk to an individual (determined on

the basis of future circumstances after their full development, frequently incorporated in decisions on introduction of new aerospace technologies in any system component); ii) statistical risk of occurrence of an accident (important for companies providing insurance, determined by the available statistical data on the incidents and accidents); iii) predicted risk (important for air transport authorities while introducing changes in technologies and air traffic patterns, determined from methodologies using some relevant historical research); and iv) perceived risk (important for users of the air transport system and determined by the individual's intuition, feeling and perception).

In addition, air traffic accidents may have some features distinguishing them from accidents in other transport modes as follows [2]: i) they may occur at any point in time and space mainly because flights may take place over large areas; ii) the primal target groups exposed to the risk exposure are passengers and crew; in addition, individuals on the ground may be exposed but generally have a lower probability of losing life or property; iii) they are relatively rare events but usually with severe consequences; iv) conditionally, each of them can be classified as an inherently risky although highly unlikely (but still possible) event; and v) risk of an accident is inherently present during the flight.

Risk implies exposure of an individual to the hazard of an air traffic accidental event (collision between aircraft, and/or collision between the aircraft and terrain). This could result in losing life or getting severe injuries both onboard the aircraft and/or on the ground, damaging and/or destroying property (the aircraft and eventually buildings on the ground), and contamination of the environment (water and soil) by burning and/or leaking fuel and oil, and hazardous cargo.

In the above-mentioned context, assessing the risk of occurrence of an air traffic accident with the associated consequences can be used as a measure of the system safety for people, systems and environment.

III. OVERVIEW OF THE METHODS/MODELS FOR ASSESSMENT OF THE RISK AND SAFETY

Many developments in aviation are initiated as a direct result from aircraft accidents. One of them is development of risk and safety methods/models at beginning of 1960's. As a reaction on accidents, first causal methods/models are developed with aim to find out their main causes in order to prevent further accidents. In the same time, collision risk methods/models appeared with proactive role in redesigning the air traffic system in order to safely accommodate increasing traffic demand. Since 1970's, aviation community become more concerned in a human roles in accidents, resulting in development of Human factor errors methods/models. Latter on, during 1990's, airports appear to be a bottleneck of an air traffic system, so the general public become aware of severity of accidents in airports vicinity and their influence on surrounding inhabitants and environment. Increased awareness was resulting in development of Third-party risk methods/models. Causal methods/models for risk and safety assessment of aircraft and ATC/ATM operations, in particular, deals with failures of particular technical systems and

components resulting in the aircraft crash or collision. The failures can be due to many interrelated causes and happen either in the aircraft or at ATC/ATM. Collision risk methods/models are dealing with assessment of the risk of aircraft collision while airborne and/or on the ground due to deterioration of ATC/ATM separation rules. Human factor error methods/models deals with risk and safety assessment of air traffic incidents and accidents due to human error. Third-party risk methods/models consider the risk assessment for people on the ground, who might be affected by the aircraft crash.

The main criterion for selection of particular methods/models has been the authors' judgment about their both theoretical importance and practical contribution (although authors were well aware of existence of many other models and similar previous studies). Also, authors' are focusing on proactive modeling approach, i.e. on methods/models which are attempting to anticipate problems before accidents occur, presenting their purpose and related problems.

A. Causal methods/models for the risk and safety assessment of aircraft and ATC/ATM operations

Causal methods/models of assessment of risk and safety of aircraft and ATM/ATC operations establish the theoretical framework of causes that might lead to aircraft accidents. These methods/models can be qualitative or quantitative. The former provide a diagrammatic or hierarchical description of the factors that might cause accidents. They are useful for improving understanding of causes of accidents and proposing preventive interventions. The later estimate the probability of occurrence of each cause and hence estimate the risk of accident. They might be restricted to pure statistical analysis based on the available data or combine these data with expert judgment on the accident causes. In addition, they can estimate the relative benefits of different interventions aiming at preventing accidents in the future [5], [6]. Some of the methods/models are as follows:

- Fault Tree Analysis (FTA) is method developed by Bell Telephone Laboratories, US in 1961 [3] and has been used for analyzing events or combinations of events that might lead to a hazard or an event with serious consequences. Usually, the analysis has been carried out using a fault tree with several paths representing different combinations of instant-direct and intermediate causes described with logical operators ("and" and "or"). At the top of the tree there is a hazard event or a serious consequence. Then, for a given tree the minimum cut set has been determined, i.e., the minimal set of failures of which if all happen causes the top event to happen too. One fault tree might have several minimal cut sets, and if only one happens, the top event also happens. The probability of occurrence of given minimum cut sets is equivalent to the product of probabilities of occurrence of each event within the set. Consequently, the probability of the occurrence of the top event is equal to the sum of probabilities of particular minimum cut sets. The method has been frequently applied (as the best recommended) to assessment of risk and safety as well as reliability of the aircraft and ATC/ATM computer (hardware) components;

- Common Cause Analysis (CCA) is the method, which can be used for identifying sequences of events leading to an aircraft accident. In particular, the method appears useful to extract common causes of several aircraft accidents. For such a purpose, it “divides” the aircraft into “zones” implying that the system and components in each zone are ultimately independent. Consequently, it is possible to identify the common causes of failures of particular components of such independent systems. The NASA has used this method for a long time (since 1987) although the method itself is probably older than 1975. In addition, it has been recommended for assessment of the risk of failures of aircraft systems and equipment;

- Event Tree Analyses (ETA) method is developed in 1980 and is used for modeling sequences of events arising from a single hazard and consequently describe seriousness of the outcomes from these events. The hierarchy of presenting a hazard, the sequence of events causing failures of the system components, and their state in terms of functioning and failure, represent the core of the method. Consequently, a tree with branches of events and functioning and failing components displays probabilities of failures along particular branches. These in combination with the probability of the hazardous event enable quantification of the probability of the system or component failure. This method has shown it is applicable in combination with FTA (Fault Tree Analysis) for almost all technical systems including the aircraft and ATC/ATM components. Bow-Tie Analysis presents a combination of ETA and FTA. Origins are from 1970’s and 1980s, but since 1999 have been popularized as a structured approach for risk analysis;

- TOPAZ accident risk assessment methodology is a complex method that uses scenario analysis and a Monte Carlo simulation technique for assessment of the risk and safety of ATC/ATM operations modeled as a Petri Nets. It has been developed by NLR (The Netherlands National Aerospace Laboratory) during the 1990’s. The method addresses all types of system safety issues such as technical/technological, organizational, environmental, and human-related and other hazards and their combinations. Risk and safety assessment is performed through few steps enabling identification of safety bottlenecks. The method has been widely applied to risk assessment of ATC/ATM operations [5];

- Bayesian Belief Networks (BBN) is a method based on probability theory, which has been developed to improve understanding of the impacts of different causes on the risk of aircraft accidents (originating from mid of 1980’s, applied in aviation filed at beginning of 2000’s). The method is supposed to capture the wide range of failures of aircraft systems both qualitatively and quantitatively and thus provide rather objective and unambiguous information on the state of system safety relevant for the managerial decisions [7], [8], [9]. The method has been applied as a decision-support tool to calculate effects of specific changes to the aviation system on the overall risk as well as support in developing a proactive policy by providing an insight into the effects of anticipated system changes on risk.

1) Purpose

Increasingly interesting causal methods/models have mainly been used for: i) better understanding of effects of different influencing factors on level of risk; ii) evaluation of overall risk, risk communication, and cost-benefit analysis of new technologies; iii) training of aviation staff and identification of system components that could be improved; and iv) identifying “critical” causes of the aircraft accident as well as measures for reducing risk. For example, in order to decide which measures for risk reduction should be adopted; regulators and safety managers need an understanding of causes of accidents and an ability to evaluate benefits of various interventions. These methods/models can support these decisions [6]. All mentioned methods/models are quantitative except the CCA. Related to risk types given in Section II, it could be mentioned that FTA, ETA and CCA are generally used to determine “statistical risk” of occurrence of an accident or failures, while Bow-Ties, TOPAZ and BBN - “predicted risk” of system changes such as introduction of new technologies, procedures, operations, etc.

2) Problems

The causal methods/models are data driven and highly dependant in their quality on the one hand and the expert judgment about combinations of particular causal factors of the air traffic accidents on the other. Quantification of these methods/models has appeared extremely difficult and time consuming mainly due to the complexity of combinations of causal factors leading to possible accidents. In addition, calculation of probabilities and conditional probabilities in situations where dependencies between particular causal factors have not completely been known further complicates quantification of the methods/models. As well, one important problem has been the cumulative nature of these methods/models, which could make assessment of particular probabilities difficult due to the large number of causal factors and their combinations [8]. Consequently, in some cases it has been rather difficult to express results from these methods/models in a transparent and comprehensible way [6].

B. Collision risk methods/models

One of the principal matters of concern in the daily operation of civil aviation is preventing conflicts between aircraft either while airborne or on the ground, which might escalate to collision. Although aircraft collisions have actually been very rare events contributing to a very small proportion of the total fatalities, they have always caused relatively strong impact mainly due to relatively large number of fatalities per single event and complete destruction of the aircraft involved. In general, separating aircraft using space and time separation standards (minima) has prevented conflicts and collisions. However, due to reduction of this separation in order to increase airspace capacity and thus cope with growing air transport demand, assessment of the risk of conflicts and collisions under such conditions has been investigated using several important methods/models as follows [10], [11]:

- The Reich-Marks model is developed in early 1960’s by Royal Aircraft Establishment, UK [12]. It is based on the assumption that there are random deviations of both aircraft positions and speeds from the expected.

The model was developed to estimate the collision risk for flights over the North Atlantic and consequently to specify appropriate separation rules for the flight trajectories [11]. The model computed the probability of aircraft proximity and the conditional probability of collision given the proximity. Aircraft were represented as three-dimensional boxes, i.e., rectangular parallelepipeds, of given length, width and height reflecting the ATC/ATM minimum separation rules. The collision might occur whenever any two boxes intersected. As well, when one aircraft was represented as the dimensionless point, conflict occurred when the point entered the box. In such a context the collision risk with the vertical, lateral and longitudinal neighbor could be determined independently of each other bearing in mind that the position errors of boxes and points representing the aircraft along their tracks were random variables with zero mean and given standard deviations. Consequently, the prescribed lateral distance between aircraft could be specified with given probability of violation reflecting the acceptable collision risk [10], [13];

- The Machol-Reich model was developed after the ICAO had established the NAT SPG (North Atlantic System Planning Group) in 1966 with the idea of creating the Reich-Marks model as the workable tool as well as increase of airspace capacity. The modified model using actual data for the position error (collected for about 14000 flights) enabled prediction with moderate confidence of each of the vertical, horizontal and longitudinal collision risks. Consequently, the ICAO NAT SPG has adopted the threshold for risk of collision of two aircraft due to the loss of planned separation [10], [14];

- The intersection models belong to the simplest collision risk models. They are based on assumptions that aircraft follow pre-determined crossing trajectories at constant speeds. The probability of a collision at the crossing point is computed using the intensities of traffic flows on each trajectory, aircraft speeds, and the airplane geometry [15], [16], [17];

- The geometric conflict models are similar to the intersection models. In these models (developed in 1990's) the speed of any two aircraft is constant, but their initial three-dimensional positions are random. Based on extrapolating their positions in time, it is possible to geometrically describe the set of initial locations that eventually lead to a conflict. The conflict occurs when two aircraft are closer than the prescribed separation rules. After integrating the probability density of the initial aircraft positions over the conflicting region, the conflict probability can be estimated [18], [19], [20];

- Generalized Reich model was developed by removing restrictive assumptions of Reich model based on the fact that Reich model does not adequately cover some real air traffic situations. The model was based on the hybrid-state Markov processes, aiming to cover a larger variety of air traffic situations. The resulting collision risk equals the probability of collision between two aircraft. Such a generalized collision model was developed during 1990's and has been used as part of the TOPAZ methodology (mentioned in Section II, A) [1], [11], [21], [22], [23], [24].

1) Purpose

The main driving force for developing collision risk methods/models during the 1960's was the need for increasing

airspace capacity over Atlantic through decreasing aircraft separation minima. The methods/models were expected to show if reduction of separation and spacing between the flight tracks would be sufficiently safe, i.e., determine the appropriate spacing between tracks guaranteeing a given level of safety. The collision risk methods/models have gradually been developed from Marks, Reich and Machol to the latest versions used in TOPAZ methodology. The main purpose has always remained to support decision-making processes during system planning and development through evaluation of the risk and safety of proposed changes (either in the existing or new system). Methods/models from this category, according to risk classification from Section II, generally provide an assessment of "predicted risk" and implicitly "real risk to an individual" due to the fact that collisions are usually leading to fatalities.

2) Problems

Despite the collision risk methods/models having been successfully used for a long time (more than 40 years), some problems, which could make their further use even more complex have continued to exist as follows:

- a) Complexity and cost of collecting the enormous amount of data on aircraft three-dimensional positions necessary to define the related statistical distributions [14], [25];

- b) Inherent complexity of the generic collision risk method/model as the result of the modeling approach (closer to the reality). New versions of these methods/models such as those used in TOPAZ are even more complex because they embrace more details when calculating risks, such as possible failure of some technical systems (engine, avionics, etc.) or flight crew awareness or fatigue; and cover complex relationships between elements of the system (flight crew, aircraft, ATC/ATM system, other aircraft, etc.) [26];

- c) Inherent danger of misunderstanding or no understanding from the average user's point of view mainly due to complexity. This requires of the specialists a long and costly familiarization time [27];

- d) The lack of risk-predicting capability with high degree of confidence and bias and uncertainty of the obtained results. Additional time and expertise for calculation of the credible risk intervals are needed [28];

- e) Relying on expert judgment in cases where historical data are not available, or when their collection is very expensive: the experts are used for setting up the value of parameters, value and dispersion of the random variables, and the dependence between variables. In such contexts, there is always the problem of engaging credible experts, especially in cases involving new system concepts;

- f) Complexity in validation particularly of new system concepts. In cases of non-existent systems, the ICAO has recommended comparison with the reference system and evaluating risk against its given threshold value.

C. Human factor error methods/models

Investigation of causes of particular air traffic accidents has identified "human error" as one of the most frequent causes [29]. Human error is considered as an incorrect execution of a particular task, which as an event, triggers a series of consecutive errors in execution of other tasks, finally resulting in serious consequences – an aircraft accident – crash.

Therefore, monitoring and modeling of human errors in the aircraft and ATC/ATM operations aiming at discovering and preventing them have always been high on the research agenda of both academics and practitioners dealing with civil aviation. Consequently, many methods for detection and prevention of “human errors” have been developed; some of them are [5]:

- HAZOP (Hazard and Operability) method (developed in early 1970’s) aims at discovering potential hazards, operability problems, and possible deviations of the actual from the system intended operational conditions (states) including estimating the probability of escalation into a serious event. The method was intended to deal with human errors in complex technical systems such as chemical and nuclear plants having human operator in their control loop. Later on, the UK NATS (National Air Traffic Service) applied the method to different aspects of planning and assessing hazard in operation of the national ATC/ATM, particularly for identification of hazards due to human failures that might develop into risk of air traffic accidents (HAZOP can provide input to FTA and ETA, mentioned in Section III, A);

- HEART (Human Error Assessment and Reduction Techniques) was developed in 1985 for identifying and quantifying errors in an operator’s task. It simultaneously considers particular ergonomic and other environmental factors, which might compromise the required operator’s performance. The impact of a particular (each) factor on the operator’s error while performing particular tasks can be quantified. Then the probability of error in executing a given task (or a series of tasks) can be estimated. The method has been applied by the UK NATS in combination with other methods for identification of the human errors in ATC/ATM;

- TRACER-Lite (Technique for the Retrospective Analysis of Cognitive Errors) was developed in 1999 by NATS, for predicting human errors and deriving error prevention measures in ATC/ATM. The method is retrospective, i.e., it is used for classifying types of errors contributing to the air traffic incidents, which have already happened. The method has a modular structure with three modules: the context; the error discovery; and the error recovery. Hierarchical Task Analysis enabling identification of the “set of critical” tasks, critically influencing safety, usually classifies the human errors;

- HERA (Human Error in ATM) is the retrospective method providing insight into ATC/ATM controllers’ cognitive processes while dealing with air traffic incidents (developed at EUROCONTROL at beginning of 2000’s). The method consists of two parts: a retrospective part for the incident analysis; and a prospective part using the information collected on the assessment of probability of human error in cases of compromised safety. Consequently, the method enables better understanding of the constraints and conditions under which ATC/ATM controllers operate. These conditions are important for understanding ATC/ATM controllers’ non-compliance with existing procedures and skill-related errors;

- HFACS (Human Factor Analysis and Classification System) is method developed at beginning of 2000’s in USA, as a system to categorize latent and immediate causal factors that have been identified in aviation accidents. It is based on analysis of hundreds of aviation accident reports and main

purpose is to provide a framework for accident investigations and to serve as a tool for accident trends assessment. HFACS uses four levels of failure: i) unsafe acts; ii) preconditions for unsafe acts; iii) unsafe supervision and iv) organizational or cultural influences. The method is very promising for analysis of air traffic controller errors and failures in ATC/ATM and is effective for understanding the antecedents of operational errors for air traffic safety analysis.

1) Purpose

The methods/models dealing with human factor errors in civil aviation have been developed to identify and eventually prevent errors (particularly of aircraft crew and ATC/ATM controllers), which could cause aircraft incidents and accidents. In addition, these models have investigated factors from the operational environment, which could cause errors, as well as calculating the probability of making errors in performing given activities. Consequently, it will be expected that they will be applied to both operational and design stages of developing aviation systems. Specific types of methods/models have given insight into the cognitive processes of the ATC/ATM controllers operating in the incidental situations, analyzed these situations, and calculated probability of making errors. In addition, these methods/models have possessed some ability for predicting errors and specifying the error reduction measures. According to risk types in Section II, those methods/models are mostly intended to determine “statistical” and “predicted” risk for given probability of error.

2) Problems

Human factor errors methods/models possess some shortcomings, which might compromise their more efficient and effective application to the ATC/ATM as follows:

- a) Most activities in ATC/ATM and in particular, factors influencing human operator performance and possible errors have usually been considered in isolation, i.e., independently on each other; in many cases the quantitative information has exclusively relied on expert judgment;

- b) Only specialists in “human factors” have been able to use these methods/models efficiently and effectively; i.e., it has been time consuming and almost impossible to apply these methods/models in an operational environment without specialists;

- c) The methods/models have been constrained exclusively to the operational processes and activities in the ATC/ATM.

D. Third-party risk methods/models

Third-party risk implies risk if an individual on the ground to be killed by crashing aircraft. In such a case, the accident is called a “grounding accident” or “grounding crash” and the fatality a “grounding fatality”. Since most air traffic accidents (about 70% according to [29]) happen around airports, the concept and assessment of third-party risk has been mainly focused on areas around airports. In a given context, the basic assumption has been that risk always exists, cannot be reduced to zero and should be predictable, transparent, and controllable, as well as quantifiable and measurable. Modeling of third-party risk has shown promise in resolving these problems including setting up thresholds for acceptable risk around airports [30], [31], [32]. Three cases of assessment of the third-party risk are illustrated as follows:

- USA case - generally implies assessment of the risk an individual is exposed to when at some distance from a given airport during the period of a year. For such a purpose, relevant statistics on fatalities from official sources have been collected and the prospective number of ground fatalities estimated. The estimation has been carried out by multiplying two independent variables – the number of crashes around airports and the number of fatalities per individual crash. The model has shown that the probability of being killed by crashing aircraft has decreased more than proportionally with increasing distance from the airport and increased with increase in the volume of the airport traffic at distances up to about two miles. The model has not considered spatial variability of the risk due to changing residence locations and the aircraft flight paths around the airports, which might be considered as its main disadvantage [32];

- The Netherlands case - this method was developed by the NLR, inspired by the crash of cargo aircraft in the Bijlmer district of Amsterdam in 1992. Method contain the following elements [31], [33]: i) the accident probability model, which calculates the probability of an aircraft accident in the vicinity of an airport depending on the probability of an accident per aircraft movement and the annual volume of airport traffic; ii) the accident location probability model, which calculates the probability of a given location becoming an accident scene depending on its position relative to airport runways and the incoming and outgoing aircraft trajectories; and iii) the accident effect model, which combines output from both previous models to calculate the probability of an accident at each location within the area surrounding a given airport. Individual and societal risks have been used as measures of third-party risk. After calculating the individual risks for the entire area around given airport, the risk contours can be plotted on the horizontal plane [31]. Societal risk applies to the entire area around a given airport and actually exists only when people are actually present in the area [31], [33];

- UK case - has become important after Public Safety Zones (PSZs) were introduced in 1958. The PSZ was defined as an area adjacent to the end of a runway in which development of land had to be restricted if it would likely significantly increase the number of “residing, working or congregating people there” [31]. In the 1997 the method for third-party risk assessments around airports and the proposal of the appropriate risk assessment criteria was developed in a NATS. The method was based on distinguishing aircraft regarding their manufacturer, country of origin, type (large, small, jets, turbo-props), and category (passenger, cargo), modeling of the aircraft crash location and the crash consequences both based on a limited sample, and simplified approach, to draw the risk contours around a given airport. In addition, cost-benefit analysis was applied to set up criteria for acceptable (tolerable) risk [31].

1) Purpose

The third-party methods/models have been mainly used for decision-making and policy purposes related to airport development and operations as follows: a) forecasting risk for an individual to be killed by a crashing airplane in the vicinity of given airports. The information has been used for comparing the risk around airports and that around chemical or nuclear plants; b) zoning around airports using individual risk contours

and societal risk values, i.e. determining areas, which should be considered dangerous for building houses or other vulnerable infrastructure; c) indicating changes in risk contours arising from airport development or changes in using existing infrastructure (changes of runways in use, arrival or departure trajectories, etc). Relative to the classification of risk given in Section II, it could be mentioned that third-party methods/models are used for assessment of “predicted” and “real risk to an individual”.

2) Problems

The third-party methods/models have been permanently improved and updated. The main problems identified during that process have been as follows [33]: a) lack of generality, i.e., the specific method/model has been developed for the specific airport; b) proactive assessment of the risk could not be carried out due to the risk control measures being already in place; c) scarcity of data on real accidents and risk exposure around the airports in the official statistical sources; d) difficulties in setting up threshold values for individual and societal risk; if too high it might compromise the airport operations and development; if too low, it might put individuals at an unacceptable jeopardy.

IV. RECOMMENDATIONS AND RELATION TO NEW TECHNOLOGIES

The methods/models for assessment of risk and safety in civil aviation described in the previous section have been reviewed aiming at identifying, from the engineering perspective, eventual shortcomings which might significantly compromise their usability, as well as points for their eventual improvements. For such a purpose, based on the available literature, a review framework containing the recommendations (requirements) and relation to new technologies for each method/model type, has been designed (the term “new technology” is referring to the new technologies, systems, procedures, concepts, operations, etc). Finally, some commonalities between them are presented in form of prospective research agenda.

A. Recommendations

1) Causal methods/models for risk and safety assessment of aircraft and ATC/ATM

It is desirable that causal methods/models possess some predictive capabilities, i.e., not only predicting the risk level and causal breakdown but also indicating their variations within changing input assumptions. Such capability would enable these methods/models to reflect better the already adopted safety measures as well as eventual benefits of further improvements. In addition, they should be able to assess the safety bottleneck in the existing system, i.e., its most vulnerable component. Due to the very complex and demanding modeling process; modular development could eventually be a compromise solution for these methods/models. This could imply starting with official statistics on air traffic accidents, and later on, allowing integration of particular modules into more complex networks. In addition, these methods/models could be developed specifically for airports, ATC/ATM, and airlines as components of the civil aviation system.

2) *Collision risk methods/models*

Regarding the purpose and existing structure, certain compromise in terms of obtaining some kind of balance between complexity and usability (due to enormous amount of input data and high level of the necessary expertise) might be recommended. Additional recommendations would be development of the method/models for specific purposes such as collision risk assessment in the en-route and terminal airspace or at the airport as well as devotion to their use at local level particularly while assessing the effects of new equipment on the collision risk. Finally, these methods/models should have better predictive capability because their usage will be more and more related to collision risk assessment when new systems, procedures, concepts and operations are introduced.

3) *Human factor error methods/models*

Further development of these methods/models should focus on dealing with human error at all ultimately interrelated levels of ATC/ATM such as operations, maintenance, organization, and management. They should be able to consider mutual dependency between errors from particular interrelated activities as well as dependability of factors causing particular errors. In addition, the methods/models will have to focus more on dealing with existing and new technologies and systems in their both operational and design stages.

4) *Third-party risk methods/models*

Certainly, development of more general methods/models for assessment of third-party risk could be recommended. They should have flexible structure in order to appropriately handle differences and specificities of traffic, layout and surrounding environment at particular airports. In addition, these methods/models should be able to handle proactive managerial, organizational, technical and/or other changes, and to represent their effects on the overall risk and safety around given airport. As well, they should have some predictive capabilities. Last but not least, there is an increasing need for common frameworks for managing third party risk by developing methodologies and tolerability criteria for comparable risk assessment in order to ensure fair competition between airports (in Europe) [34].

B. *Relation to new technologies*

1) *Causal methods/models for risk and safety assessment of aircraft and ATC/ATM*

The causal methods/models could contribute to the proactive development of policies on implementing changes by providing insight into the effects of changes in existing systems on risk and safety [8]. In particular, under conditions when the system changes due to implementation of new technologies, these methods/models could provide feedback about their contribution to lowering risk and consequently increasing the overall system safety.

2) *Collision risk methods/models*

Used for reduction of aircraft separation for more than 40 years, the collision risk methods/models have proved their viability. However, further reductions in aircraft separation by the use of new technologies will be needed as an option for increasing airspace capacity. Therefore, existing modified and new methods/models will have to be able to assess collision risk under such circumstances [10]. Some models such as

TOPAZ are already in place. Use of this method/model is in line with methodology proposed by the ICAO, which points out the necessity for evaluation of risk of new technologies against threshold values and its comparison with the reference system [35]. In cases where there is lack of reference systems or large scale changes in existing systems, expert judgment is recommended. In addition, setting up threshold values for risk while implementing new technologies, which are expected to be of lower risk, is also a matter for further elaboration of existing systems and the development of new collision risk methods/models.

3) *Human factor error methods/models*

Human factor error methods/models with necessary modifications should be applicable to new technologies and systems in ATC/ATM for identifying human errors at all levels of system functioning and they should be able to generate measures for error prevention and/or reduction already at the design stage. For such purposes, they will have to be able to handle careful specification of activities and tasks throughout the system in a way, which will not be highly if not crucially dependent on the highly specialized staff.

4) *Third-party risk methods/models*

Predictive capabilities and flexibility of third-party risk methods/models will be essential to produce new (updated) individual and societal risk estimates based on the expected number of fatalities after introducing new technologies and operational procedures at given airport. On the one hand these are expected to increase airport capacity and on the other they should decrease the accident rate in the vicinity of airports.

C. *Prospective research agenda*

The overview and review of the mentioned methods/models for assessment of risk and safety in civil aviation have uncovered some commonalities between them, which could be, after being summarized, used for generating prospective research agenda. These are as follows:

- Regarding the *purpose* all models have been developed to support decision-making processes during system planning, development and management, through evaluation of the risk and safety of proposed technological, organizational and managerial changes;
- Regarding *problems* that all methods/models have been confronted with: i) Necessity to have a good, statistically significant data bases on air traffic accidents and their causes (the lack of such data has been compensated by the expert judgment inherently containing unreliability and uncertainty); and ii) Complexity in quantification of risk and safety due to dependability of particular air traffic accidents on many interrelated dynamic and stochastic causes;
- Regarding the *recommendations*, all methods/models should have some predictive capabilities, flexibility, and modularity as well as should be generic;
- Regarding *application to new technologies*, all methods/models should be able to investigate their risk and safety under given circumstances. However there might be some limitations in such application due to the inherent limitations of existing models to appropriately handle the risk and safety of new technologies [35].

Mitigating the above-mentioned and other problems in line with recommendations how to improve existing and develop new methods/models for assessment of risk and safety in civil aviation particularly for new still non-existing technologies have been identified as the main research challenge for the prospective research.

V. CONCLUSIONS

The paper has provided a review of some of the methods/models for assessment of risk and safety in civil aviation. The main findings have provided insight into the efforts already carried out in developing these methods/models, their inherent complexity and lack of sufficient flexibility, lack of the available data for calibration and testing, and lack of the sufficient predicting capabilities enabling easier application to the assessment of risk and safety of new technological, procedural and operational concepts. These have aimed at increasing system capacity on the one hand and reducing acceptable risk and safety thresholds on the other. In many cases, the need for developing "specialized" or "dedicated" methods/models for particular parts of the system have been discovered. In addition, difficulties such as the lack of real-life data have been overcome by including expert judgment despite awareness of its uncertainty and biases. The structured need for balance and compromise between methods/models complexity, time and cost of development, and transparency of results have also been pointed out. Prospective research has been considered to further improve existing models in line with recommendations, which have generally implied capability of risk and safety assessment during development and after implementation of new technologies, generality on the one hand and dedication on the other, predictive capabilities, flexibility and easier understood and handled modular system structures.

REFERENCES

- [1] H. Blom, G. Bakker, P. Blanker, J. Daams, M. Everdij, M. Klompstra, "Accident risk assessment for advanced ATM", Proceedings of 2nd USA/Europe Air Traffic Management R&D Seminar, USA, 1998.
- [2] M. Janic, "An Assessment of Risk and Safety in Civil Aviation", Journal of Air Transport Management, Vol. 6, No. 2, pp 43-50, 2000.
- [3] H. Kumamoto, E. Henley, Probabilistic Risk Assessment and Management for Engineers and Scientists, IEEE Press, USA, 1996.
- [4] N. Bahr, System Safety Engineering and Risk Assessment: A Practical Approach, Taylor & Francis, United Kingdom, 1997.
- [5] ATM Safety Techniques and Toolbox, Safety Action Plan – 15, Federal Aviation Administration and EUROCONTROL, USA, 2005.
- [6] J. Spouge, A Demonstration Causal Model for Controlled Flight into Terrain, Det Norske Veritas, United Kingdom, 2004.
- [7] J. Luxhoj, D. Coit, "Modeling Low Probability/High Consequence Events: An Aviation Safety Risk Model", Proceedings of the Reliability & Maintainability Symposium (RAMS), USA, pp. 215 – 220, 2006.
- [8] A. Roelen, R. Wever, A. Hale, L. Goossens, R. Cooke, R. Lapuhaa, M. Simons, P. Valk, "Causal Modelling for Integrated Safety at Airports", Proceedings of ESREL 2003 - European Safety and Reliability Conference, The Netherlands, pp. 1321 – 1327, 2003.
- [9] A. Roelen, R. Wever, R. Cooke, R. Lapuhaa, A. Hale, L. Goossens, "Aviation Causal Model Using Bayesian Belief Nets to Quantify Management Influence", Proceedings of ESREL 2003 - European Safety and Reliability Conference, The Netherlands, pp. 1315 – 1320, 2003.
- [10] R. Machol, "Thirty Years of Modelling Midair Collisions", Interfaces, Vol. 25, No.5, pp. 151-172, 1995.
- [11] J. Shortle, Y. Xie, C. Chen, G. Donohue, "Simulating Collision Probabilities of Landing Airplanes at Non-towered Airports", Simulation, Vol. 80, Issue 1, pp. 21-31, 2004.
- [12] P. Reich, "Analysis of long range air traffic systems: Separation standards – I, II and III", Journal of the Institute of Navigation, No. 19, pp. 88-96, 169-176, 31-338, 1966.
- [13] A Concept Paper for Separation Safety Modelling, Federal Aviation Administration and EUROCONTROL, USA, 1998.
- [14] R. Machol, "An Aircraft Collision Model", Management Science, Vol. 21, No. 10, pp. 1089-1101, 1975.
- [15] W. Siddiquee, "A mathematical model for predicting the number of potential conflict situations at intersecting air routes", Transportation Science, No. 7, pp. 158-167, 1973.
- [16] K. Geisinger, "Airspace Conflict Equations", Transportation Science, No. 19, pp. 139-153, 1985.
- [17] A. Barnett, "Free-flight and en route air safety: A first-order analysis", Operations research, No. 48, pp. 833-845, 2000.
- [18] R. Paielli, H. Erzberger, "Conflict probability estimation for free flight", Journal of Guidance, Control and Dynamics, No. 20, pp. 588-596, 1997.
- [19] R. Paielli, H. Erzberger, "Conflict probability estimation generalized to non-level flight", Air Traffic Control Quarterly, No. 7, pp. 195-222, 1999.
- [20] R. Irvine, "A geometrical approach to conflict probability estimation", Air Traffic Control Quarterly, No. 10, pp. 85-113, 2002.
- [21] G. Bakker, H. Blom, "Air Traffic Collision Risk Modelling", Proceedings of 32nd IEEE Conference on Decision and Control, USA, pp. 1464-1469, 1993.
- [22] H. Blom, G. Bakker, "Conflict probability and In-crossing probability in Air Traffic Management", Proceedings of 41st IEEE Conference on Decision and Control, USA, (preprint), 2002.
- [23] G. Bakker, H. Kramer, H. Blom, "Geometric and Probabilistic Approaches towards Conflict Prediction", Proceedings of 3rd USA/Europe Air Traffic Management R&D Seminar, Italy, 2000.
- [24] H. Blom, G. Bakker, M. Everdij, M. van der Park, "Collision risk Modelling of Air Traffic", Proceedings of European Control Conference, United Kingdom, (preprint), 2003.
- [25] L. Stachtchenko, "An Investigation of Aircraft Collision Risks over the North Atlantic", CORS Journal, Vol. 3, No. 2, pp. 55-71, 1965.
- [26] J. Rouvroye, E. van den Blik, "Comparing safety analysis techniques", Reliability Engineering and System Safety, No. 75, pp. 289-294, 2002.
- [27] Guide to Methods & Tools for Safety Analysis in Air Traffic Management, Global Aviation Information Network, USA, 2003.
- [28] M. Everdij, H. Blom, S. Stroeve, "Structured Assessment Of Bias And Uncertainty in Monte Carlo Simulated Accident Risk", International Conference On Probabilistic Safety Assessment And Management (PSAM 8), USA, (preprint), 2006.
- [29] Statistical Summary of Commercial Jet Airplane Accidents: Worldwide Operations 1959 – 2005, Boeing Commercial Airplanes, USA, 2006.
- [30] B. Ale, "Risk Assessment Practices in the Netherlands", Safety Science, Vol. 40, pp. 105 –126, 2002.
- [31] B. Ale, E. Smith, R. Pitblado, Safety around airport – developments in 1990s and future directions, Det Norske Veritas, United Kingdom, 2000.
- [32] R. Rabouw, K. Thompson, R. Cooke, "The Aviation Risk to Groundlings with Spatial Variability", Proceedings of ESREL 2001 - European Safety and Reliability Conference, Italy, pp. 1219-1226, 2001.
- [33] A. Hale, "Risk Contours and Risk Management Criteria for Safety at Major Airports, with Particular Reference to the Case of Schiphol", Safety Science, Vol. 40, pp. 299–323, 2002.
- [34] Safety in and Around Airports, European Transport Safety Council, European Commission, Belgium, 1999.
- [35] F. Vismari, B. Camargo "Evaluation of the impact of new technologies on aeronautical safety: an approach through modelling, simulation and comparison with legacy systems", Journal of the Brazilian Air Transportation Research Society, No.1, pp 19-30, 2005.

Accident Risk Analysis Benchmarking

Monte Carlo Simulation versus Event Sequences

Henk A.P. Blom, Sybert H. Stroeve, Jelmer J. Scholte
National Aerospace Laboratory NLR
Amsterdam, The Netherlands
blom@nlr.nl; stroeve@nlr.nl; scholte@nlr.nl

Hans H. de Jong
DFS Deutsche Flugsicherung GmbH
Langen, Germany
hans.de.jong@dfs.de

Abstract—Fault and event trees are the dominantly used safety risk models in air traffic. Systemic accident risk assessment by Monte Carlo simulation is a more recent technique, the power of which is less explored. In this paper we compare the two approaches for an accident risk analysis of an active runway crossing operation that is supported by a runway incursion alerting system for the runway controller. For this example we show and explain remarkable differences in results obtained using the two approaches.

Keywords- runway crossing; runway incursion; collision risk; stochastic systems; Monte Carlo simulation; event sequence analysis; alert system

I. INTRODUCTION

In following Hollnagel [11], accident models can be categorized in sequential models, epidemiological models, and systemic models.

Sequential accident models describe an accident as the result of a limited number of sequences of events that occur in a specific order. These models assume that there are well-defined cause-effect links that propagate the effects of chains of events leading to an accident. Examples of sequential accident models are the domino theory, event trees and fault trees. Many methods used in practice are based on the traditional fault/event tree. However, as argued by Hollnagel and several other leading researchers, e.g. [15], [16] and [18], they may not be adequate to account for the complexity of modern socio-technical systems. Sequential accident models are commonly known and often applied in system dependability and safety requirement studies in aviation and air traffic, e.g. [8].

Epidemiological accident models describe an accident in analogy with the spreading of a disease, i.e. as the outcome of a combination of factors, such as performance deviations, environmental conditions, barriers and latent conditions. Like sequential accident models, epidemiological accident models rely on cause-effect propagation in accidents. Epidemiological models provide a broader basis to represent the complexity of accidents than sequential models by better accounting for interactions between relevant factors. Epidemiological models have been used in aviation and air traffic, in methods such as the Human Factors Analysis and Classification System of

Wiegmann and Shappell [21] and Bayesian belief networks, e.g. [1], [10], [13].

Systemic accident models describe the performance of a system as a whole, rather than at the level of cause-effect mechanisms or epidemiological factors. The systemic view considers accidents as emergent phenomena from the variability of a system, for instance due to the concurrent and interacting behaviour of multiple agents (humans, technical systems) in a safety critical and dynamic operation. In such case the interacting multiple agents together form a joint cognitive system (Hollnagel and Woods, [12]). Hollnagel [11] explains that the foundation of systemic models lies in systems, control and chaos theories. Subsequently, he describes the key principles of these theories in terms of an elegant functional resonance accident model. Leveson [15] directly exploits the control theory view for the development of a systemic accident modelling approach, named STAMP. Corker, Pritchett and co-workers [14], [17] have shown that agent based Monte Carlo simulation allows to predict emergent behaviour in advanced air transport developments. Blom et al. [4] exploit stochastic system and control theory to develop a multi-agent Monte Carlo modelling and simulation approach for the evaluation of safety critical air traffic scenarios, named TOPAZ. Stroeve et al. [19] show that this Monte Carlo modelling and simulation approach is of systemic accident type.

To practitioners of accident risk analysis of future air traffic management designs (e.g. in NEXTGEN and SESAR), the development of three rather different types of accident modelling types raises the question whether the developments beyond sequential models yield better safety analysis results or not. In the latter case the novel approaches would be nice to have only. In the former case, the novel approaches even may be of critical design value for the design of future air traffic management. In order to bring more clarity in this question, the aim of this paper is to compare a fault/event tree approach with a Monte Carlo simulation approach for an active runway crossing operation, that is supported for safety reasons by a runway incursion alert system (RIAS) for the runway controller. This active runway crossing example incorporates concurrent and interacting behaviour of pilots and controllers in the dynamics of the operation.

The paper is organized as follows. Section II presents the active runway crossing operation considered. Section III

presents the safety relevant scenarios that have been identified for this operation, and the scenario selected for benchmarking. Section IV presents the risk assessment by the event sequence modelling and analysis. Section V presents the risk assessment through Monte Carlo modelling and simulation. Section VI compares the accident risk results obtained by the two approaches. Finally, Section VII draws conclusions.

II. THE ACTIVE RUNWAY CROSSING OPERATION

The subject of our risk assessment is an active runway crossing operation proposed at Amsterdam Airport Schiphol. In this proposed operation, Runway 18C is used for departures, whereas taxiing aircraft have to pass it on their ways from the gate to departure from Runway 36L, or from landing on Runway 18R to the gate. Figure 1 shows Runway 18C/36C with surrounding taxiways.

During the early development of the infrastructure and operation for simultaneous use of the aforementioned runways, the air navigation service provider opted for crossings over an active Runway 18C/36C, to keep taxi times between airport centre and the far-off runway as low as possible. For this operation, a runway incursion alerting system (RIAS) was foreseen to give stop bar violation alerts and runway incursion alerts. In support of this early development phase, a safety requirements analysis has been performed using a functional hazard analysis (FHA) for, inter alia, the runway incursion alerting system and its usage by the runway controller (RC). In [9], a number of hazardous scenarios have been considered, such as collisions between a departing aircraft and an aircraft making a runway incursion, and between an aircraft sliding off the runway and an aircraft holding at a crossing point. The safety risk analysis of the runway incursion collision scenario delivered, inter alia, the following design requirements under poor visibility condition:

- The probability that the Runway incursion alert system (RIAS) fails to detect a runway incursion is at most 10^{-5} ;
- The probability that the runway controller fails to react appropriately to an alert is at most 5×10^{-5} ;
- The probability that it takes a minute for contingency procedures to become effective when the radio frequency is blocked, is at most 10^{-5} ; and
- The probability that pilots of both aircraft fail to react to stop taxiing and cancel take-off/perform missed approach is at most 10^{-5} .

These FHA based design requirements subsequently formed the basis for the design of a RIAS supported active runway crossing operation of runway 18C/36C. As has been explained in [7], eventually this proposed RIAS supported runway crossing design of active runway 18C/36C has not been selected for implementation at Amsterdam Airport Schiphol. For the purpose of comparing two risk analysis methods, however, this RIAS supported runway crossing design forms a valuable example.

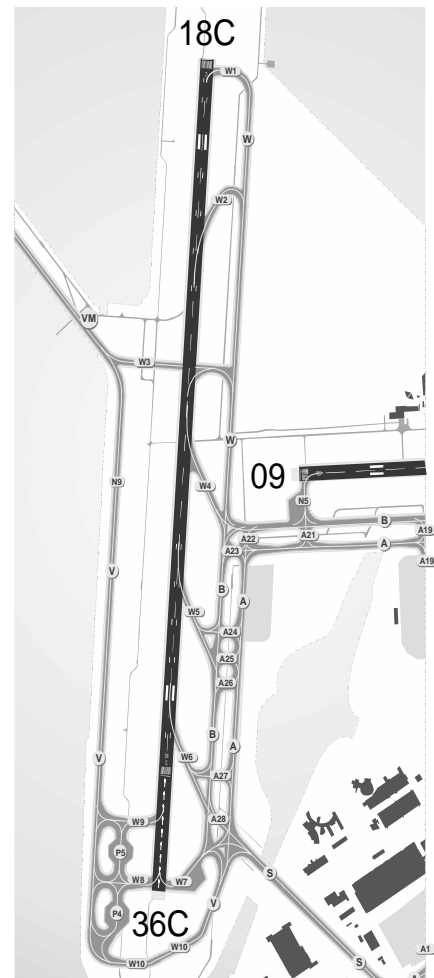


Figure 1. Runway 18C/36C and its surrounding at Amsterdam Airport Schiphol

III. COMMON ACCIDENT RISK ANALYSIS STEPS

In order to validate whether the RIAS supported design of the active runway crossing operation would be sufficiently safe, it has been evaluated following a formal risk assessment process. An overview of the steps taken in this safety risk assessment is given in Figure 2. This cycle has been developed over many years at NLR, and presented in [4]. This cycle is generic in the sense that it is used both in preparation of an event sequence-based assessment as well as a Monte Carlo simulation based assessment.

In step 0, the objective of the assessment is determined, as well as the safety management and regulatory context, the scope and the level of detail of the assessment. The actual safety assessment starts by determining the operation that is assessed (step 1). Next, hazards associated with the operation are identified (step 2), and aggregated into safety relevant scenarios (step 3), for which the potential severities are identified (step 4). The risk quantification is done in the frequency assessment (steps 5). Subsequently, the safety risk associated with each safety relevant scenario is classified (step 6). For each safety relevant scenario with a (possibly) unacceptable safety risk, the main sources (safety bottlenecks)

contributing to safety risks are identified (step 7), which help operational concept developers to learn for which safety issues they should develop improvements in the ATM design. If the ATM design is changed, a new safety risk assessment cycle of the operation must be performed in order to investigate how much the risk posed by previous safety issues has been decreased, and to assess any new safety issues that may have been introduced by the enhancements themselves.

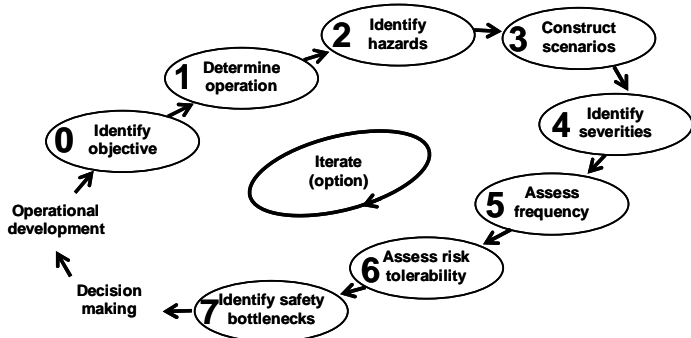


Figure 2. Safety risk assessment cycle.

Steps 0 through 4 are the common accident analysis steps for Monte Carlo simulation and fault/event tree based approaches in step 5. For the active runway crossing operation considered these common steps are described in the remainder of this section.

Step 0: Identify objective

Before starting the actual risk assessment, the objective and scope of the assessment are set. This is done in close co-operation with the decision makers and designers of the advanced operation. Also, the appropriate framework of safety management and safety regulations must be made clear, such that the assessment is performed in line with these.

An important issue in setting this context is the choice of risk tolerability criteria/Target Level of Safety (TLS) with respect to which the assessment is performed and the scope of risks to which these are applied. Depending on the application, such criteria are defined for particular flight condition categories (e.g. flight phases or sub-phases) and for particular severity categories (e.g. accident or serious incident). Typically, within the chosen context, these criteria define which flight condition/severity categories have to be evaluated and which frequency level forms the Target Level of Safety (TLS) threshold per flight condition/severity category.

Step 1: Determine operation

Step 1 serves for the risk assessors to obtain a complete and concise overview of the operation, and to freeze this description during each cycle of analysis. Main input to step 1 is a description of the operational concept from the designers, while its output is a sufficiently complete, structured, consistent and concise description of the operation considered. The operation should be described in generic terms, it should provide any particular operational assumption to be used in the safety assessment, and it has to be verified by the operational concept designers. Typically during this step, holes and

inconsistencies in the concept as developed are also identified and immediately fed back to the design team.

Step 2: Identify hazards

The term hazard is used in the wide sense; i.e. an event or situation with possibly negative effects on safety. Such non-nominal events or situations may evolve into danger, or may hamper the resolution of danger, possibly in combination with other hazards or under certain conditions. The goal of step 2 is to identify as many and as diverse hazards as possible. Hazard identification brainstorming sessions are used as primary means to identify (novel) hazards.

Based on the experience gained in using the hazard identification part of HAZOP in a large number of safety analyses and on scientific studies of brainstorming, NLR has developed a method of hazard identification for air traffic operations by means of pure brainstorming sessions. This method has been reported in [6]. In such a session no analysis is done and solutions are explicitly not considered. An important complementary source is formed by hazards identified in previous studies on related operations. Example hazards are mentioned in the explanation of step 3.

In total, about 100 hazards have been identified. A first ordering of these hazards is made by distinguishing

- root hazards, which describe safety relevant events and conditions that cause the initiation of a conflict in a safety relevant scenario, and
- resolution hazards, which describe events and conditions that influence resolution of the conflict, which is aimed at limiting the severity of consequences.

Step 3: Construct scenarios

When the list of hazards is as complete as reasonably practicable, it is processed to deal with duplicate, overlapping, similar and ambiguously described hazards. Then, per flight condition selected in Step 0, the relevant 'conflict' types which may result from the hazards are identified using a full list of potentially relevant conflict types, such as for instance 'aircraft erroneously crossing and other aircraft in take-off' and 'collision between aircraft sliding off runway and aircraft near crossing'. Although these situations are simply called 'conflicts', it is important to note that not only ordinary conflicts between aircraft are considered; 'conflict types' rather indicate general potentially dangerous operational situations.

Each potentially relevant conflict type is subsequently used as crystallization point upon which all applicable hazards and their combined effects are fitted as elements of additional event sequences. If hazards cannot be appropriately addressed by the crystals developed so far, then additional conflict types need to be defined and corresponding scenarios developed. The output of such a crystallization process is a bundle of event/condition sequences and effects per conflict type/crystallization point, and each resulting crystal is referred to as a safety relevant scenario (see Figure 3). This way of constructing scenarios aims to bring into account all relevant ways in which hazards can play a role.

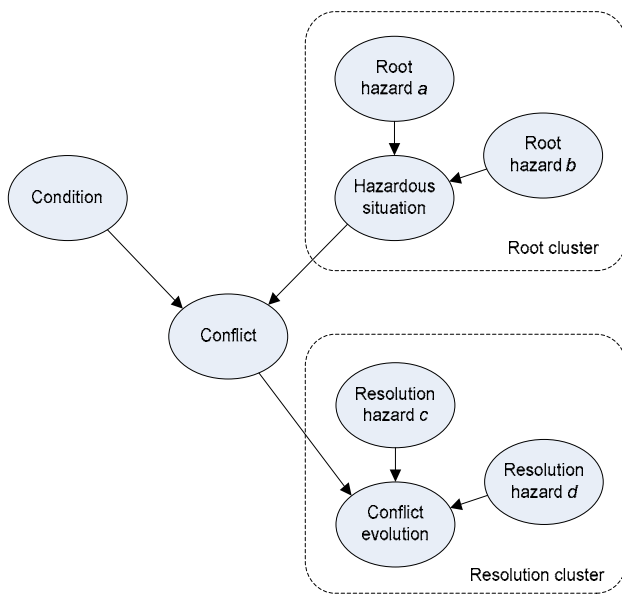


Figure 3. Generic diagram of a safety relevant scenario.

The outcome of this crystallization process are crystals for the following safety relevant scenarios of the active runway crossing operation:

- Scenario I: Aircraft erroneously in take-off and crossing aircraft on runway;
- Scenario II: Aircraft erroneously crossing and other aircraft in take-off;
- Scenario III: Aircraft taking off and runway unexpectedly occupied;
- Scenario IV: Aircraft crossing and runway unexpectedly occupied by aircraft;
- Scenario V: Aircraft crossing and vehicle on runway;
- Scenario VI: Collision between aircraft sliding off runway and aircraft near crossing;
- Scenario VII: Aircraft taking off and vehicle crossing;
- Scenario VIII: Jet-blast from one aircraft to another; and
- Scenario IX: Conflict between aircraft overrunning, or climbing out low, and an aircraft at a nearby taxiway.

Because Scenario II ranks high in expected safety risk, this scenario has been selected for the comparison of the two risk assessment approaches in this paper. Scenario II covers the situation where there is one aircraft that takes off from runway 18C, and has been allowed to do so, and there is one aircraft that crosses the runway while it should not, over the runway crossing position marked by 'W3', somewhat to the north of the middle of the runway (see Figure 1).

In the context of safety relevant scenario II, examples of its elements in figure 3 are:

- Root hazard a: Pilots react on clearance for another aircraft and start crossing;
- Root hazard b: Pilots cross without clearance;
- Hazardous situation: Aircraft crossing runway while it should not;
- Condition: Other aircraft has initiated take-off;

- Conflict: Aircraft is erroneously crossing the runway, while other aircraft is taking off;
- Resolution hazard c: Pilots of crossing aircraft do not frequently look for conflicting traffic;
- Resolution hazard d: Pilots of crossing aircraft are not tuned to frequency of runway controller; and
- Conflict evolution: Possible ways of evolution of the runway incursion conflict, e.g. leading to an accident or an incident of certain severity.

Step 4: Identify severities

For each of the safety relevant scenarios identified in step 3, it is determined which of the severity categories selected in step 0 are applicable to its possible effects. Usually, a range of severities applies to a safety relevant scenario. For all nine safety relevant scenarios, except scenario VIII, all four severities (minor, major, hazardous, accident) have been identified as being applicable [5]. For safety relevant scenario VIII, the severity of accident has been judged not to be applicable.

The sequel of this paper focuses on step 5 (assess frequency) of the safety analysis for the most severe possible effects of scenario II, i.e. accidents. In Section IV this is done using a fault/event tree analysis approach. Next, in Section V, this is done using a Monte Carlo modelling and simulation approach. Finally the results of both approaches are compared in Section VI.

IV. EVENT SEQUENCE BASED RISK ASSESSMENT

For safety relevant scenario II, the accident risk is modelled through a combination of fault and event trees. Two separate fault trees (see Figure 4) have been developed for two specific cases:

- Case p, i.e. pilot of taxiing aircraft starts crossing without contacting the runway controller (e.g. by misunderstanding the ground controller); and
- Case p-not, i.e. pilot of taxiing aircraft has contacted runway controller well, though starts crossing while it should not.

Subsequently, for each of these two fault trees an event tree has been developed. Both event trees make use of the following fixed sequence of twelve branching points:

- There is an aircraft on 18C in take-off (yes/no)
- Early recognition and resolution by pilots (yes/no)
- Early recognition by RC (yes/no)
- Stopbar violation alert and RC becomes aware of it (yes/no)
- Early communication by RC and pilot resolution (yes/no)
- Medium recognition and resolution by pilots (yes/no)
- Medium recognition by RC (yes/no)
- RIAS alert and RC becomes aware of it (yes/no)
- Medium communication by RC and pilot resolution (yes/no)
- Medium recognition by RC (yes/no)
- RIAS alert and RC becomes aware of it (yes/no)

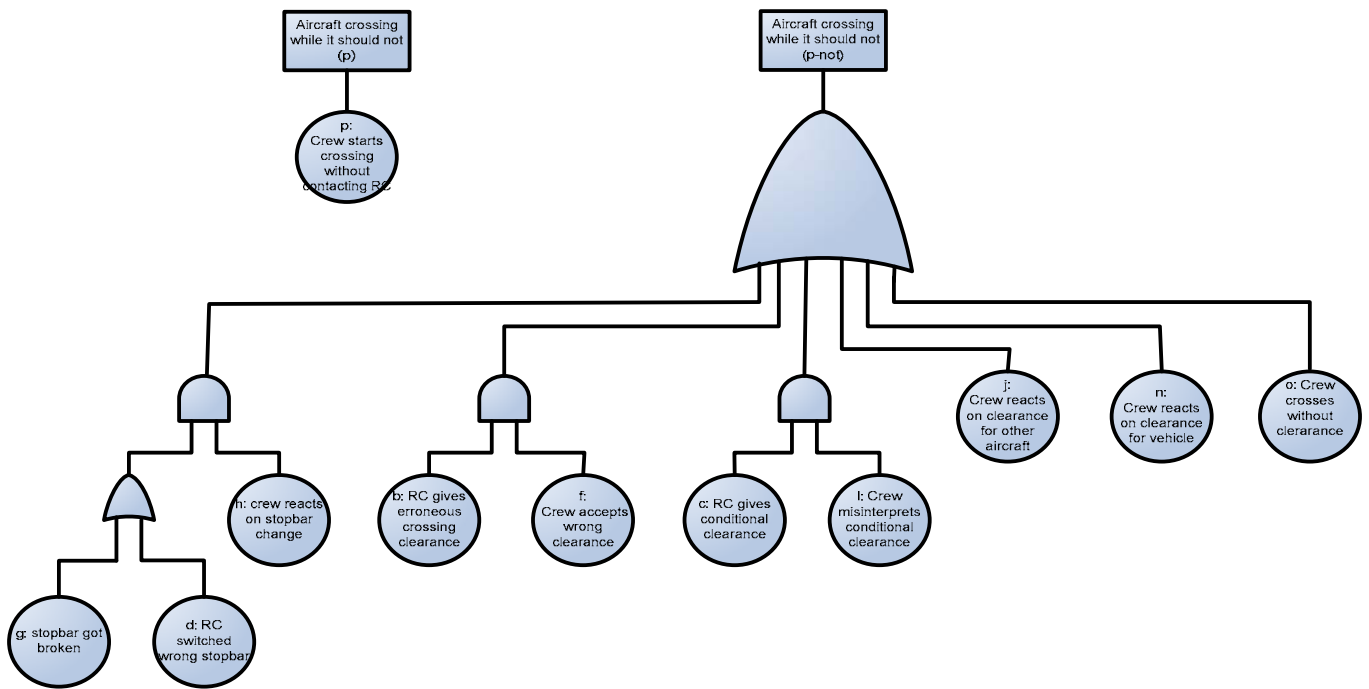


Figure 4: Fault Trees for safety relevant scenario II.

- Medium communication by RC and pilot resolution (yes/no)
- Late recognition and resolution by pilots (yes/no)
- Late recognition by RC and pilot resolution (yes/no)
- Late communication by RC and pilot resolution (yes/no)

The branching points in the event tree differentiate between early, medium, and late recognition of the conflict by the pilots and RC. This approach was chosen as a systematic means to get hold on the large variety in the timing of particular events to happen in combination with the timing of stop bar violation and RIAS alerts and the remaining braking distance.

Of all feasible branching point combinations, there are six ending at accident. All others end at late, medium or early resolution of the conflict. The estimation of the yes/no probability values per branching point has been done on the basis of statistical data and expert based estimates, for each of the two cases (p and p-not) separately [5]. This way, for each branching point both expected, and upper/lower bound probability values have been assessed. Subsequently, the quantified fault and event trees have been used to calculate accident risk in terms of expected value, and upper and lower bound values. The resulting upper bound accident probability values for scenario II are given in Table I.

TABLE I. RESULTS OF THE EVENT SEQUENCE-BASED RISK ASSESSMENT FOR THE TWO CASES CONSIDERED. THE PROBABILITIES ARE UPPER BOUND VALUES PER TAKE-OFF.

Case	Conflict probability	Conditional accident probability given conflict	Accident probability
p-not	$3 \cdot 10^{-4}$	$1.6 \cdot 10^{-5}$	$4.8 \cdot 10^{-9}$
P	$6 \cdot 10^{-5}$	$7.3 \cdot 10^{-5}$	$4.4 \cdot 10^{-9}$
Total	$3.6 \cdot 10^{-4}$	$2.6 \cdot 10^{-5}$	$9.2 \cdot 10^{-9}$

In Figure 5 it is shown which branching points play a role in the accumulation of contributions to the accident probability. For contributions to case p, the risk is dominated by situations in which RC is alerted by stopbar violations. For contributions to case p-not, the risk is dominated by situations in which RC is not yet aware of the conflict after a stopbar violation alert.

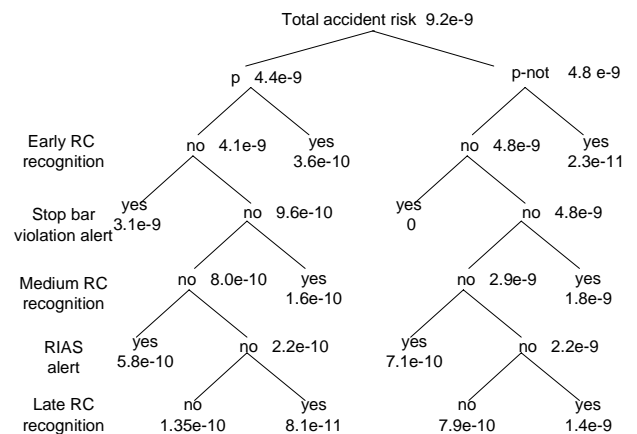


Figure 5. Contributions to the collision risk for events related to early, medium or late RC recognition of the conflict, and the occurrence of stopbar violation or RIAS alert. The probabilities are upper bound values per take-off.

V. MONTE CARLO BASED RISK ASSESSMENT

A. Monte Carlo simulation model

Prior to running Monte Carlo simulations for accident risk analysis, a simulation model is developed that captures the nominal and non-nominal (stochastic and dynamic) behaviour of the aircraft, the relevant technical systems, the relevant

human operators (pilots and runway controller), and the interactions between all these entities. Figure 6 gives an overview of the main entities and their interactions for the runway incursion Monte Carlo simulation model [4], [19], [20]. An arrow from one entity to another entity indicates that the former (directly) influences the latter.

For the development of the models for the human operators, the key aspects are taken into account, such as situation awareness / task performance / task scheduling of a human operator, flight phases / performance modes of aircraft, and availability / status of an alert system. Subsequently, the interactions between human operators and/or technical entities are also modelled, e.g. the effect of task performance of a pilot on the flight phase of an aircraft, or the effect of an alert on the situation awareness of a controller. The resulting human performance models and their interactions have been calibrated and validated through a comparison with an Air-Midas model for surface operations [2], [3]. A complementary validation approach consisting of a systematic bias and uncertainty assessment has been provided in [19].

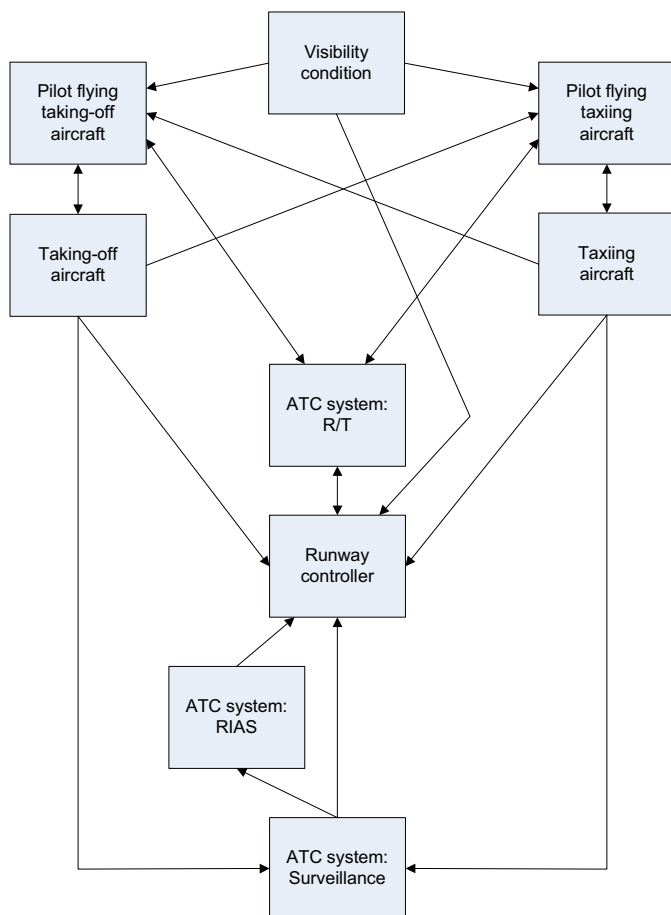


Figure 6. Interactions between the main entities of the runway incursion simulation model: aircraft, pilots flying, runway controller, visibility condition and ATC system (R/T, RIAS and surveillance).

B. Risk assessment results

Table II shows the values assessed for the event probabilities and conditional accident probabilities of the runway crossing operation considered, at a distance of 1000 m from the runway threshold. Two non-nominal situation awareness (SA) conditions for the pilot of the taxiing aircraft are distinguished*):

- The pilot flying of the taxiing aircraft believes to be proceeding on a normal taxiway (without being aware to be heading to a runway crossing), and
- The pilot flying of the taxiing aircraft starts to cross the runway without being aware that crossing is currently not allowed.

Table II shows that the conditional accident probability value is 35 times higher when the PF of the taxiing aircraft believes to be proceeding on a taxiway rather than being crossing a runway.

Through a deeper analysis of the Monte Carlo simulation results, the reasons of this large difference appeared to be as follows. When a pilot aims to cross a runway, then he will stop prior to the runway and expecting a clearance before actually starting the crossing. However, when a PF aims to proceed on a taxiway then there is no reason to stop, and its aircraft maintains taxiing speed. Moreover, being unaware of the runway, the pilot has no reason to frequently scan this runway. As a result of this, at the moment that a pilot detects the runway incursion, then the time period that is left to stop the aircraft is much shorter than the time period that would be available when the aircraft starts from a hold before crossing the runway, first has to build up taxiing speed, and where the pilot frequently scans this runway.

TABLE II. MONTE CARLO SIMULATION BASED RISK ASSESSMENT FOR RUNWAY CROSSING AT 1000M FROM TAKE-OFF STARTING POINT. THE VALUES ARE POINT ESTIMATES PER TAKE-OFF.

SA by PF of Taxiing aircraft	Probability of event	Event conditional accident probability	Accident probability
Cross runway	$2.3 \cdot 10^{-4}$	$4.8 \cdot 10^{-6}$	$1.1 \cdot 10^{-9}$
Proceed taxiway	$3.5 \cdot 10^{-5}$	$1.7 \cdot 10^{-4}$	$6.0 \cdot 10^{-9}$
Total	$2.7 \cdot 10^{-4}$	$2.6 \cdot 10^{-5}$	$7.1 \cdot 10^{-9}$

Figure 7 provides a view on various contributions to the collision risk, such as the influence of situation awareness conditions of the pilot of the taxiing aircraft (Cross runway/Proceed runway), and the functioning of ATC alert and communication systems (Up/Down). These results show that accident risk is dominated by situations where a pilot flying of a taxiing aircraft is not aware of the nearby runway due to erroneous situation awareness, whereas neither failure of alert systems for ATC, nor failure of communication systems contribute noticeably to collision risk per departure.

*) The identification of these non-nominal SA conditions is in fact a direct result of the systematic modelling of multi-agent SA within our Monte Carlo simulation model [20].

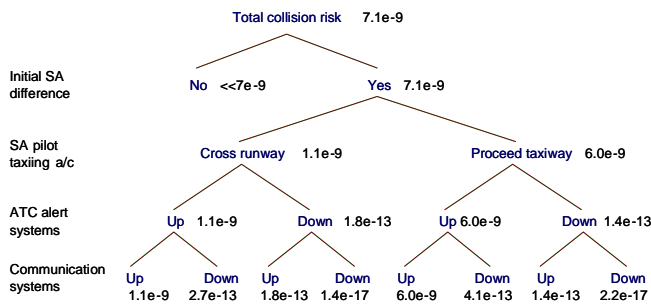


Figure 7. Contributions to the collision risk for various combinations of events related to pilot situation awareness and functioning of ATC alert and communication system, for crossing distance of 1000 m, under good visibility.

C. Conditional risk increase under hypothetical assumptions

Using the Monte Carlo simulation model, it also is possible to attain insight in the conditional risk increase when the monitoring capability of any of the involved human operators or ATC alert system is assumed to be out of the loop. Table III shows the conditional collision risks obtained for the hypothetical situation where an aircraft taxis towards a runway crossing while the pilot believes to taxi on a normal taxiway. The conditional collision risks in Table III refer to hypothetical cases where traffic conflict monitoring by specific human operators, or ATC alerting system are assumed to be within the model ('yes') or out of it ('no'). A risk increase factor is determined by comparing the conditional collision risk with the situation in which none of the human operators are taken out of the loop.

Table III shows that out-of-the-loop placing of monitoring function by PF of taxiing aircraft ranks highest on risk increase. Second ranks risk increase due to out-of-the-loop placing of monitoring by PF of taking-off aircraft. Third ranks risk increase due to out-of-the-loop placing of monitoring by RC, and fourth ranks risk increase due to out-of-the-loop placing of ATC alert systems.

TABLE III. CONDITIONAL RISK INCREASE FACTORS ACHIEVED IN THE SIMULATION MODEL DUE TO OUT OF THE LOOP ASSUMPTIONS OF MONITORING BY HUMAN OPERATORS OR ATC ALERT SYSTEM, WHEN THE PF OF A TAXIING AIRCRAFT INTENDS TO PROCEED ON A NORMAL TAXIWAY UNDER GOOD VISIBILITY (CROSSING IS AT 1000 M FROM RUNWAY THRESHOLD)

Hypothetical Case	Monitoring in the loop of				Conditional risk	Increase Factor
	PF taxi	PF take-off	RC	ATC alert systems		
0	Yes	yes	yes	yes	$1.7 \cdot 10^{-4}$	1
1	no	yes	yes	yes	$9.4 \cdot 10^{-3}$	55
2	Yes	no	yes	yes	$4.0 \cdot 10^{-4}$	2.4
3	Yes	yes	no	yes	$2.3 \cdot 10^{-4}$	1.4
4	Yes	yes	yes	no	$2.2 \cdot 10^{-4}$	1.3

VI. COMPARISON OF RESULTS

For scenario II, the upper bound of estimated accident risk assessed by the event sequence approach equals $9.2 \cdot 10^{-9}$ per

take-off [5]. This means that the upper bound value estimated by the event sequence approach is almost equal to the Monte Carlo simulation based point estimate value of $7.1 \cdot 10^{-9}$ per take-off.

Comparison of the values in Table I with the values in Table II, shows that the conditional collision risk total/mean values also are about the same for the scenario based upper bound estimates and the Monte Carlo simulation based point estimates. Through a bias and uncertainty assessment conducted in [19], it has been shown that the Monte Carlo simulation based upper bound value of the conditional risk is almost a factor five larger than the point estimated value. This means that the event sequence approach in Section IV leads to a significantly lower accident risk level per take-off than the Monte Carlo simulation approach in Section V.

In order to better understand the reason of this difference, Table IV compares conditional risk increase factors under MC simulation and event sequence approaches for the hypothetical cases of Table III plus two combinations, i.e. case 1&2 and case 1&2&3 respectively. The Monte Carlo simulation-based results are given for the case that the pilots of the taxiing aircraft believe to be proceeding on a taxiway and the event sequence-based results are given for event p 'Pilot starts crossing without contacting RC (e.g. by misunderstanding the ground controller)'. These two conditions have in common that they both imply that pilots of the taxiing aircraft are not aware of the need to be in contact with the runway controller, which makes their comparison particularly relevant.

TABLE IV. COMPARISON OF CONDITIONAL RISK INCREASE FACTORS DUE TO ASSUMING (COMBINATIONS) OF HUMAN OPERATOR MONITORING OUT-OF-THE-LOOP. IN THE MC SIMULATION RESULTS THE PILOT OF THE TAXIING AIRCRAFT (ERRONEOUSLY) BELIEVES TO TAXI ON A NORMAL TAXIWAY. THE EVENT SEQUENCE BASED RESULTS ARE UPPER BOUND ESTIMATES THAT APPLY TO EVENT P.

case	Monitoring in-loop-of			MC simulations		Event sequences	
	PF taxi	PF take-off	RC	Conditional risk	Factor	Conditional risk	Factor
0	yes	yes	yes	$1.7 \cdot 10^{-4}$	1	$7.3 \cdot 10^{-5}$	1
1	no	yes	yes	$9.4 \cdot 10^{-3}$	55.3	**)	**)
2	yes	no	yes	$4.0 \cdot 10^{-4}$	2.35	**)	**)
3	yes	yes	no	$2.3 \cdot 10^{-4}$	1.35	$1.2 \cdot 10^{-3}$	17
1&2	no	no	yes	*)	*)	$1.4 \cdot 10^{-2}$	200
1&2&3	no	no	no	$8.9 \cdot 10^{-2}$	530	$2.5 \cdot 10^{-1}$	3400

*) In the MC simulation approach, not monitoring by pilots of both aircraft was not evaluated

***) In the event sequence approach, distinction between pilots of different aircraft was not modelled

The event sequence results in Table IV show that the product of the risk increase factor for the case without monitoring of the controller (case 3) and the risk increase factor for the case without monitoring of the pilots of both aircraft (case 1&2), is equal to the risk increase factor for the case where pilots and controller do not monitor (case 1&2&3). This reflects that in the event sequence approach the risk reducing contributions of pilots and controller are considered to be independent. In contrast, for the Monte Carlo simulation results, it follows from Table IV that the product of the risk

increase factors due to not monitoring of individual human operators (case 1, case 2 and case 3) is considerably smaller than the risk increase factor for the case when both pilots and controller are not monitoring (case 1&2&3). This reflects that in the Monte Carlo simulation approach, the risk reducing contributions of pilots and controller are interdependent, such that the risk increase due to not monitoring by one of the actors is moderated by the performance of the others. In particular, Table IV shows that the increase in the conditional collision risk by excluding monitoring by RC is much higher in the event sequence approach (factor 17.1) than it is in the Monte Carlo simulation approach (factor 1.35). It is the simulation based approach that makes clear that although RC identifies a good share of conflicts, its contribution to avoiding a collision is much smaller than the event sequence based approach predicted. Deeper analysis yields that a significant proportion of the instructions issued by the runway controller arrive late, hence these instructions either concern conflicts that are already solved by the pilots, or even arrive too late for any of the pilots involved to successfully avoid a collision.

VII. CONCLUSION

In safety risk analysis literature there is an ongoing debate regarding the advantage of systemic modelling over sequential modelling for complex safety critical socio-technical systems, such as future air traffic management. In order to contribute in a concrete way to this debate, we performed a benchmark of both approaches for the same demanding application, i.e. accident risk analysis of active runway crossing operation. The sequential modelling approach considered is fault/event tree modelling. The systemic modelling approach considered is Monte Carlo modelling and simulation.

Within the dynamics of an active runway crossing operation, the concurrent and interacting behaviours of pilots and controllers make that accident risk analysis on the basis of an event sequence based approach may be more demanding than what can be managed in a controlled way. The introduction of a differentiation between early, medium and late responses did not prevent a significant underestimation of (conditional) collision risk contributions of up to an order in magnitude. More specifically, for the runway incursion scenario considered, the event sequence based approach provided results which would imply that, under good visibility conditions, RIAS support to RC could serve as an effective means of reducing accident risk by an order in magnitude. The Monte Carlo simulation approach, however, showed that the effective contribution of RIAS support is almost zero for the simple reason that the pilots often will receive a RIAS triggered instruction from RC at a moment that one of the pilots already has recognized and started to resolve the conflict. Nevertheless, in such case, RC may perceive him/herself to have played a key role in resolving the conflict well.

The key difficulty is that coping with time and dynamic dependencies within event sequence modelling, is complicated by the interactions and concurrencies that play a key role in runway incursion. The power of a Monte Carlo simulation approach is that it can handle concurrency and interactions well.

VIII. REFERENCES

- [1] Ale BJM, Bellamy LJ, Cooke RM, Goossens LHJ, Hale AR, Roelen ALC, Smith E. Towards a causal model for air transport safety - an ongoing research project. *Safety Science* Vol. 44 (2006):657-673
- [2] Blom HAP, Corker KM, Stroeve SH, Van der Park MNJ. Study on the integration of Air-MIDAS and TOPAZ: Phase 3 Final Report. National Aerospace Laboratory NLR, report CR-2003-584, October 2003
- [3] Blom HAP, Corker KM, Stroeve SH. Study on the integration of human performance and accident risk assessment models: Air-MIDAS & TOPAZ. 6th USA/Europe ATM R&D Seminar, 2005, Baltimore, USA
- [4] Blom HAP, Stroeve SH, De Jong HH. Safety risk assessment by Monte Carlo simulation of complex safety critical operations. In: Redmill F, Anderson T (eds.), *Developments in risk-based approaches to safety*, Springer-Verlag, London, 2006, pp. 47-67
- [5] De Jong HH, Tump RS, Blom HAP, Van Doorn BA, Karwal AK, Bloem EA. Qualitative safety assessment of a RIAS based operation at Schiphol Airport including a quantitative model: Crossing of departures on 01L/19R under good visibility conditions. National Aerospace Laboratory NLR, Memorandum LL-2001-017, May 2001
- [6] De Jong HH. Guidelines for the identification of hazards; How to make unimaginable hazards imaginable? NLR Contract report 2004-094 for EUROCONTROL, March 2004 (included in EUROCONTROL, 2006: FHA, Ch. 3, GM B.2)
- [7] De Jong HH, Blom HAP, Stroeve SH. Unimaginable hazards and emergent behavior in air traffic operations. *Proceedings ESREL 2007*, 25-27 June 2007, Stavanger, Norway
- [8] Eurocontrol. A-SMGCS Levels 1 and 2 Preliminary Safety Case. Edition 2.0, 2006.
- [9] Gleave DP, Thompson KS. Amsterdam Airport Schiphol Risk Assessment of Runway 18/36 Taxiway Options, Report No: 72/98/R026R, July 1998, Issue 1, Roke Manor Research Limited, 1998
- [10] Greenberg R, Cook SC, Harris D. A civil aviation safety assessment model using a Bayesian belief network (BNN). *The Aeronautical Journal*, November 2005, pp. 557-568
- [11] Hollnagel E. *Barriers and accident prevention*. Ashgate, Hampshire, UK, 2004
- [12] Hollnagel E, Woods DD. *Joint cognitive systems: Foundations of cognitive systems engineering*. CRC Press, Boca Raton (FL), USA, 2005.
- [13] Kardes E, Luxhoj JT. A hierarchical probabilistic approach for risk assessments of an aviation safety product folio. *Air Traffic Control Quarterly*, Volume 13 (2005), pp. 279-308.
- [14] Lee SM, Pritchett AR, Corker KM. Evaluating transformations of the air transportation system through agent-based modeling and simulation. 7th USA/Europe ATM R&D Seminar, Barcelona, Spain, 2007.
- [15] Leveson N. A new accident model for engineering safer systems. *Safety Science*, Volume 42 (2004), pp. 237-270.
- [16] Pariès J. Complexity, emergence, resilience... In: Hollnagel E, Woods DD, Leveson N (eds.). *Resilience engineering: Concepts and precepts*. Ashgate, Aldershot, England, 2006.
- [17] Shah AP, Pritchett AR, Feigh KM, Kalaver SA, Jadhav A, Corker KM, Holl DM, Bea RC. Analyzing air traffic management systems using agent-based modelling and simulation. 6th USA/Europe ATM R&D Seminar, Baltimore, USA, 2005.
- [18] Sträter O. *Cognition and safety: An integrated approach to system design and assessment*. Ashgate, Hampshire, England, 2005.
- [19] Stroeve SH, Blom HAP, Bakker GJ. Safety risk impact analysis of an ATC runway incursion alert system. *Eurocontrol Safety R&D Seminar*, Barcelona, Spain, 25-27 October 2006
- [20] Stroeve SH, Blom HAP, Van der Park MNJ. Multi-agent situation awareness error evolution in accident risk modelling. *Proceedings of the 5th USA/Europe ATM R&D Seminar*, Budapest, Hungary, 2003
- [21] Wiegmann DA, Shappell SA. Human error analysis of aviation accidents: Application of the Human Factors Analysis and Classification System (HFACS). *Aviation, Space, and Environmental Medicine*, Volume 72 (2001), pp. 1006-1016.

Analyzing Relationships Between Aircraft Accidents and Incidents

A data mining approach

Zohreh Nazeri
The MITRE Corporation*
7515 Colshire Drive
McLean, Virginia, USA
nazeri@mitre.org

George Donohue, Lance Sherry
George Mason University
4400 University Drive
Fairfax, Virginia, USA
{gdonohue, lsherry}@gmu.edu

Abstract—Air transportation systems are designed to ensure that aircraft accidents are rare events. To minimize these accidents, factors causing or contributing to accidents must be understood and prevented. Previous research has studied accident data to determine these factors. The low rate of accidents however, makes it difficult to discover repeating patterns of these factors. In this research we employed a data mining technique to conduct a holistic analysis of aircraft *incident* data in relation to the *accident* data. The analysis identifies relationships between the accident and incident data and finds patterns of causal and contributory factors which are significantly associated with aircraft accidents.

Keywords- aviation safety; aircraft accidents; aircraft incidents; data mining; contrast-set mining

I. INTRODUCTION

Levels of safety are typically measured by the number of accidents and incidents and their rates. An aircraft accident is defined as an occurrence associated with the operation of an aircraft in which people suffer death or injury, and/or in which the aircraft receives substantial damage. An aircraft incident is an occurrence which is not an accident but is a safety hazard and with addition of one or more factors could have resulted in injury or fatality, and/or substantial damage to the aircraft [1]. Throughout the history of air transportation, along with the continuous growth in air travel, remarkable improvements have been made in lowering of accident rates. Nevertheless, further improvements are needed.

Figure 1 indicates annual rates of accidents that meet the selection criteria used in this study (as explained in section II.A). The accident data is obtained from the National Transportation Safety Bureau (NTSB) database.

* The author's affiliation with The MITRE Corporation is provided for identification purposes only, and is not intended to convey or imply MITRE's concurrence with, or support for, the positions, opinions or viewpoints expressed by the author.

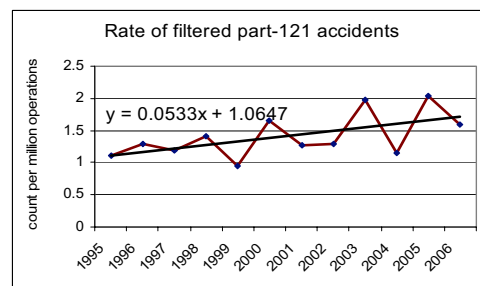


Figure 1. Rates of part-121 accidents that meet this study's filtering criteria

To further improve the flight safety, traditionally researchers focus on studying accident data to determine causal factors leading to accidents [2, 3, and 4]. The low rate of accidents however, makes it difficult to discover patterns of these factors. Other approaches study larger sets of data available on incidents but don't analyze relationships between these data and the accident data [5, 6, and 7].

Two influential theories in the safety studies are Swiss Cheese model [8] and Heinrich Pyramid [9]. The Swiss cheese model, introduced by James Reason in 1990's, views a hazardous situation culminating in an accident equivalent to passing through successive slices of Swiss cheese. Each "slice" is a system or process designed to prevent harm, however, the slices have holes representing errors in the system or process. Each error may occur frequently without harmful results, but when combined (i.e., holes are lined up), accident occurs. This model is widely used for investigating human factors (See Figure 2).

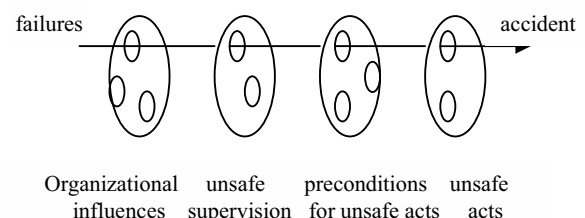


Figure 2. Swiss Cheese model

The Heinrich Pyramid, introduced by H. W. Heinrich in 1930's, represents the accidents as low-frequency and high-risk safety hazards at the top of a pyramid. Moving down on the pyramid, the next layers consist of incidents and unsafe acts which are less hazardous but are more frequent. (See [Figure 3](#).) An adaptation of the Heinrich pyramid in the aviation safety domain suggests that for every major accident there are 3-5 non-fatal accidents, 10-15 incidents, and hundreds of unreported events [10].

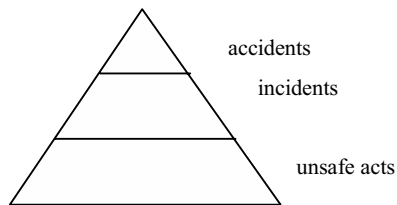


Figure 3. Heinrich Pyramid

In this study, similar to Heinrich pyramid, we look for the relationships between accidents and incidents. However, unlike the Heinrich pyramid which considers a quantitative relationship between accidents and incidents, our study looks for correlations between the underlying factors of accidents and incidents. The study considers the factors individually as well as combinations of the factors similar to the line-up of the holes in the Swiss Cheese model.

We analyzed the incidents in contrast to accidents and identified factors which are present in both classes of events but are significantly associated with accidents. We then studied the identified patterns of factors in the context of incidents.

II. DATA

The data used in this study consists of accidents and incidents pertaining to commercial flights (part-121) from 1995 through 2004, which provides a large enough sample size for the analysis. The accident data is obtained from the National Transportation Safety Board (NTSB) database. The incident data is obtained from four national databases: Federal Aviation Administration Accident and Incident Database System (FAA/AIDS), National Aeronautics and Space Administration Aviation Safety Reporting System (NASA/ASRS), FAA Operational Errors and Deviations (FAA/OED), and FAA System Difficulty Reports (FAA/SDRS).

Each report of accident or incident in these databases consists of structured fields plus a narrative explaining the event. Causal and contributory factors are identified either directly by the person who submits the report, or indirectly by a domain expert who reviews the report. These factors are in the structured fields. Our analysis used these factors.

A. Data Selection

Since the purpose of the analysis was identifying patterns of accident factors related to the routine operation of the flight,

accidents and incidents due to the following causes were filtered out from the data:

- Medical and alcohol related events, such as pilot being sick or drunk
- Terrorism and security related events, such as bomb threats
- Passenger and cabin-crew related problems, such as passengers being injured due to hot coffee spilling on them
- Bird/animal striking the aircraft
- Events during the phases of operation when the aircraft is not operating (parked, standing, preflight)

Also, reports pertaining to the Alaska region were filtered out since flight environment and procedures in this region are different from other regions in the United States and require a separate study.

After applying the filters, there were 184 accidents, and the following sets of incidents in the data for analysis: 2,188 reports in the FAA/AIDS dataset, 29,922 reports in the NASA/ASRS dataset, 10,493 reports in the FAA/OED dataset, and 85,687 reports in the FAA/SDRS dataset.

B. Data Constraints

All accidents in the United States involving civil aircraft are investigated by the NTSB, an independent organization, and reported in the NTSB database. Accident data, therefore, can be assumed complete and free of bias. These assumptions cannot be made about the incident data. Incidents are under-reported and are subject to self-reporting bias. Voluntary reports represent a fraction of incidents [11] and recent audits indicate reporting of the incidents mandated by the FAA are under-reported [12 and 13].

Our study analyzed the underlying factors of accidents and incidents. The historical data on incidents is large enough to represent these factors qualitatively. Also, we consider all factors that are present in the events, primary or contributory. This minimizes the impact of the bias in reporting a factor as contributory versus primary.

III. METHODOLOGY

We first developed a common taxonomy across the accident and incident databases to identify common fields (factors) between the two classes of events. We then transformed each report into a vector consisting of the common fields populated with their corresponding values for each report. Next, we applied the STUCCO [14] algorithm to the accident and incident vectors and identified patterns of factors which are significantly associated with accidents or with incidents. The findings were ranked using Factor Support Ratio, a measure introduced in this study as described below in this section under 'Ranking'. Results of the analyses conducted on multiple databases were compared for cross-database validation. The results are discussed in the next section.

A. Common Taxonomy

The process of deriving the common taxonomy is data-driven. After reviewing each individual database structure and unique values for each field, we developed a hierarchy of factors and sub-factors common across the databases. Eight high-level categories of factors were identified in the data, each containing corresponding sub-factors. These factors and examples of their sub-factors are shown in Figure 4. The 'Other' category contains all sub-factors which were not big enough to have a separate category for themselves.

A normalization of the values was needed so that all databases use the same word/phrase to refer to the same factor/condition. For example, to refer to the action where pilot has to execute a maneuver to avoid a vehicle or object on the runway, 'ground encounter' is used in one database and 'object avoidance' in another.

The reports were converted to vectors consisting of fields that indicate presence or absence of each of the common factors and sub-factors in the accident or incident. These vectors were then used in the analysis.

Factor	Sub-Factor examples
Aircraft	Engine, Flight control system, Landing gear
Airport	Snow not removed from runway, Poor Lighting, Confusing marking
Air Traffic Control	Communication with pilot, Complying with procedures
Company	Procedures, Management, Training
Maintenance	Compliance, Inspection
Pilot	Visual lookout, Altitude deviation, Decision/Judgment
Weather	Wind, Thunderstorm, Ice
Other	Factors not in the other categories, e.g., FAA oversight, Visibility

Figure 4. Cross-database common taxonomy

B. Contrast-Set Mining

Since the objective of the study was to identify factors and factor combinations that are precursor to accidents, we needed an analysis technique that could take advantage of both sets of data (accidents and incidents) and determine which factors are more likely to lead to accidents. We applied the STUCCO algorithm [14] to analyze accident and incident vectors by contrasting them. The algorithm finds conjunctions of attribute-value pairs that are significantly different across multiple groups. In the case of our data, there are two groups: accidents and incidents. Attribute-values are binary values indicating presence or absence of the factors in the event.

The factors and their children (combinations of factors) are examined for their frequency (support) in each group. For each factor-set, *deviation* is calculated as absolute value of the difference between accident support and incident support. In the first step, factor-sets for which deviation is more than a

minimum threshold proceed to the next step to be tested further. We used a minimum of 1% threshold for the deviation.

In the next step, Chi Square test is performed to test statistical significance of the distribution of factor-set over the two groups. The contingency table shown in Figure 5 is used for this test. A p-value of 0.05 is used as the threshold. Factor-sets with a p-value of more than 0.05 are rejected. A p-value of less than 0.05 is the equivalent of being in the 95% confidence interval and is accepted.

	accidents	incidents
factor-set true	accidents containing the factor-set	incidents containing the factor-set
factor-set false	accidents not containing the factor-set	incidents not containing the factor-set

Figure 5. Contingency table used for Chi Square significance test

C. Ranking

Once significant factor-sets were identified by the algorithm, we ranked them based on the *Factor Support Ratio* measure. As shown in equation (1), we calculate the Factor Support Ratio for each factor-set as the ratio of the factor-set's support in accident dataset over its support in the incident dataset.

$$\text{Support Ratio} = \frac{\text{Support}_{\text{accident}}}{\text{Support}_{\text{incident}}} \quad (1)$$

The Support Ratio is the probability of a factor-set being involved in an accident divided by its probability of being involved in an incident. The information conveyed by this measure about the factor-set is different than that of the *deviation* (the difference between the factor-set's accident and incident supports) that is used in the algorithm. To understand the Support Ratio better, consider factor-sets A and B and their corresponding measures in Table 1.

TABLE I. SUPPORT RATIO VS. DEVIATION

factor-set	accident supp	incident support	Dev	Support Ratio
A	60%	50%	10%	1.2
B	11%	1%	10%	11
C	60%	10%	50%	6

Both factor-sets A and B have a deviation of 10% between their accident support and incident support. However, in the case of factor-set B, the support in accidents is 11 times more than in incidents. This can be interpreted as: occurrence of factor-set B in an accident is 11 times more likely than its occurrence in an incident. This is a more distinctive distribution than that of factor-set A which has a Support Ratio of 1.2. We can use this measure to compare factor-sets A and

B, and say factor-set A is more likely to be involved in accidents than factor-set B.

To understand the significance of Support Ratio for ranking the factors consider factor-sets A and C in [Table 1](#). Both factor-sets appear in 60% of accidents. But the fact that factor-set C appears in 10% of incidents raises its Support Ratio (compared to factor-set A). We interpret this as: factor-set C is qualitatively a more significant accident factor than factor-set A. When factor-set C occurs it is more likely (by a factor of 6) to be involved in an accident than in an incident. But when factor-set A occurs, the likelihood of having an accident versus having an incident is smaller (1.2).

Accident support of a factor (frequency of the factor in accidents) by itself shows how many times the factor has been involved in accidents but does not show how frequently the factor has occurred (in accidents and incidents). Similarly, incident support of a factor only indicates how many times the factor has occurred in incidents without an indication of the factor's role in accidents. In some cases, a factor seen frequently in incidents might rarely be involved in accidents. This means the factor is not a significant accident factor. One explanation could be that the factor could be stopped from leading to accidents once it occurred.

IV. RESULTS

We performed separate analyses on four pairs of datasets, each pair consisted of accident reports and their corresponding incident reports in one of the four incident databases. The results of the analyses were compared at the end. Below are major findings of the study that were consistent across the multiple analyses of incident/accident database pairs.

A. Combination of factors

Factors are more likely to yield to accidents (rather than incidents) when they are combined together. Ranking of the results by the Factor Support Ratio showed that likelihood of a factor being involved in an accident rises as more factors co-occur with the factor. For example, the Support Ratio for combination of *pilot+airport* factors was 7.2 compared to the Support Ratio of 3.9 for the *pilot* factors, signifying that *pilot* factors combined with *airport* factors are 1.8 times more likely to result in accidents than the *pilot* factors alone.

B. Company factors

Company factors are referred to factors such as mistakes by the company (or airline) personnel, inadequate or non-existing procedures by the company for performing a task, and lack of management by the company management. The analyses identified these factors as significant accident factors. Ranking of the results by their Support Ratios identified *company* factors as the highest ranked category of accident factors among the eight categories of factors in the data.

C. ATC factors

The analyses identified *Air Traffic Control (ATC)* factors as the next highest ranked category of accident factors following the *company* factors. Among the ATC factors, *ATC*

communications are identified as the most significant sub-factors associated with accidents. *ATC communications* refer to factors such as controllers issuing traffic advisories, controllers providing weather information to the pilot, and controllers checking for correct readback of instructions by the pilot.

D. Pilot factors

Pilot factors are more frequent than other factors in accidents but they are also more frequent in incidents and therefore their Support Ratio is lower and ranks them after the *company* and *ATC* factors. Among the *pilot* factors, *visual lookout* is identified as the most significant pilot sub-factor.

E. Aircraft factors

Aircraft factors are referred to mechanical problems with the aircraft or its components and systems. Examples are problems with landing gears, flight control systems, and wings. Without presence of other factors, *aircraft* factors are identified as incident factors, meaning that they are more likely to cause incidents than accidents when occurring alone. But when *aircraft* factors are combined with other factors, such as severe weather or pilot errors, the combination becomes an accident factor.

V. CONCLUSION

We further studied the results in the context of the historical databases where data was available. The data over a ten-year-period (1995-2004) showed that *pilot* and *aircraft* factors are decreasing and *Air Traffic Control (ATC)* factors are increasing (see figures 6, 7, and 8). The operational error reports in the OED database show that *ATC* factors are influenced by a variety of conditions, referred to as *complexity factors*. The data available to us included eleven of these complexity conditions: *airspace design, emergency event, controller experience, flow control, number of aircraft, runway conditions, runway configuration, terrain, special event, weather, and other*. The ten-year historical data showed top-most frequent complexity conditions influencing the *ATC* factors are *number of aircraft, airspace design, runway configuration, and controller experience*.

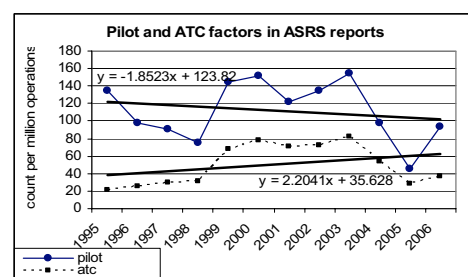


Figure 6. Incident rates containing *pilot* and *ATC* factors, in the ASRS database

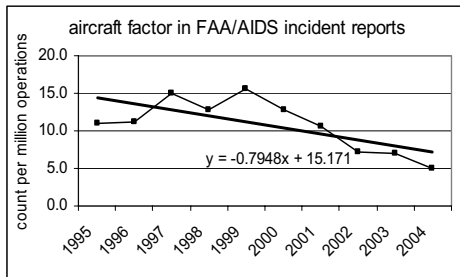


Figure 7. Rate of the incidents containing *aircraft* factors, in the AIDS database

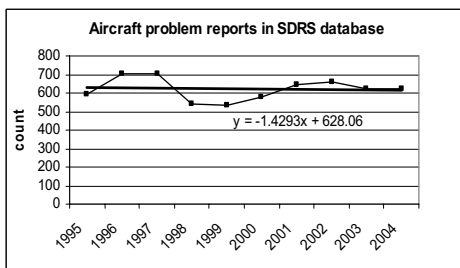


Figure 8. Rate of incidents containing *aircraft* factors, in the SDRS database

The *number of aircraft* complexity condition can be expected to rise even further considering the continuous growth of air transportation projected by the FAA [15]. The projected growth will impact *Runway configuration* and *airspace design* complexity conditions indirectly. With an increased volume, airports will have to use runway configurations that accommodate more departures and arrivals. Difficulties with the airspace design, such as limited space for complying with altitude changes when moving the aircraft from one airspace to the other, will be aggravated when there are more aircraft in the airspace. The *controller experience* will also be impacted by the projected growth, since more controllers will be needed to handle the increased operations. In addition, the number of controllers retiring in the past few years has exceeded the projections [16] and a large number of existing controllers are expected to retire within the next few years [17]. FAA plans to hire over 1,000 controllers per year for the next 10 years. The current experienced controllers will be replaced with a new generation.

Considering the accident factors identified in this study, the projected growth in air transportation and the consequent aggravation of the conditions affecting the accident factors, accident rates can be expected to increase beyond their current levels unless changes are made to current conditions.

VI. FUTURE WORK

The study conducted here is a starting point for further research on the relationships between accidents and incidents and identification of more detailed accident factors. As a continuation of this study, the methodology applied here can be applied to other safety databases that were not available to this

study, such as the Aviation Safety Action Program (ASAP) and Flight Operations Quality Assurance (FOQA) databases maintained by airlines. These databases offer safety data which could yield to discovery of more detailed accident factors. In addition, upon availability of more detailed data, the approach in this research can be taken one step further to study patterns of factors within each of the identified categories.

This study covered accidents and incidents pertaining to commercial flights within the United States. A similar study could be conducted on the General Aviation (GA). Depending on the availability of the data, the studies could be extended to regions in other countries as well.

ACKNOWLEDGMENT

Zohreh Nazeri thanks Christina Hunt, manager, at the FAA Aviation Safety Information Analysis and Sharing (ASIAS), for providing much of the data needed for the study, and Linda Connell, director, at the Aviation Safety Reporting System (ASRS) program at NASA, for helping with downloading the ASRS data from their on-line query interface.

REFERENCES

- [1] Federal Aviation Administration, Air Traffic Organization, *Aircraft Accident and Incident Notification, Investigation, and Reporting*. Order 8020.16 http://www.faa.gov/airports_airtraffic/air_traffic/publications/at_orders/media/AAI.pdf
- [2] P. Ladkin, *ATM Related Accidents*. Eurocontrol. From website: http://www.eurocontrol.int/corporate/public/standrd_page/cb_safety.html, 2006.
- [3] K. Dimukes, *The Limits of Expertise: The Misunderstanding Role of Pilot Error in Airline Accidents*. ASPA/ICAO regional seminar, March 2005.
- [4] G. Van Es, *Review of Air Traffic Management-Related Accidents Worldwide: 1980-2001*. Fifteenth Annual Aviation Safety Seminar (EASS). Geneva, Switzerland, March 2003.
- [5] A. Majumdar, M. D. Dupuy, and W. O. Ochieng, *A framework for the Development of Safety Indicators for New Zealand Airspace: the Categorical Analysis of Factors Affecting Loss of Separation Incidents*. Transportation Research Board (TRB) annual conference, 2006.
- [6] M. Hansen and Y. Zhang, *Safety Efficiency: Link between Operational Performance and Operation Errors in the national Airspace System*. Transportation Research Record, journal of Transportation Research Board, no. 1888, p 15. 2004.
- [7] National Aeronautics and Space Administration, *Air Traffic Management System*. From website: <http://quest.arc.nasa.gov/aero/virtual/demo/ATM/tutorial/tutorial1.html>. 2007.
- [8] J. Reason, *Human Error*. Cambridge, U.K.: Cambridge University Press. 1990.
- [9] H. W. Heinrich, *Industrial Accident Prevention*. New York. McGraw-Hill. 1931.
- [10] C. A. Hart, *Stuck on a Plateau: A Common Problem*. Accident Precursor Analysis and Management: Reducing Technological Risk through Diligence, National Academies Press. 2004.
- [11] B. Strauss and M. Granger, *Everyday Threats to Aircraft Safety*. Issues in Science and Technology 19(2); 82-86, 2002.
- [12] Federal Aviation Administration, *Audit of Control over the Reporting of Operational Errors*. Report number AV-2004-085, September 20, 2004.
- [13] U.S. Office of Special Council. *Report of Disclosure Referred for Investigation*. OSC Files DI-06-1499 and DI-07-2156, July 2007.
- [14] S. D. Bay and M. J. Pazzani, *Detecting Change in Categorical Data: Mining Contrast Sets*. Fifth ACM SIGKDD International Conference on

Knowledge Discovery and Data Mining. New York: The Association for Computing Machinery.

- [15] Federal Aviation Administration, *FAA Aerospace Forecasts FY2006-2017*. From http://www.faa.gov/data_statistics/aviation/aerospace_forecasts/2006-2017/media/FAA%20Aerospace%20Forecast.pdf
- [16] Federal Aviation Administration. A Plan for the Future, 2007-2016. The Federal Aviation Administration's 10-year strategy for the air traffic control workforce. March 2007.
- [17] United States General Accounting Office (GAO), Air Traffic Control, FAA Needs to Better Prepare for Impending Wave of Controller Attrition. GAO-02-591, June, 2002.

ATC Complexity as Workload and Safety Driver

Jelena Djokic, Bernd Lorenz

EUROCONTROL CRDS

Budapest, Hungary

jelena.djokic@eurocontrol.int,

bernd.lorenz@eurocontrol.int

Hartmut Fricke

Institute of Logistics and Aviation

Technical University of Dresden

Dresden, Germany

fricke@ifl.tu-dresden.de

Abstract— This paper describes an investigation into ATC complexity as a contributory factor in changes of safety level. ATC complexity, together with equipment interface and procedural demands comprise the task demands on the controller; subsequent controller activities are mediated by performance shaping factors to create workload. In order to establish a link between ATC complexity, a controller's subjective workload and safety, complexity factors are identified and subsequently related to both workload and safety indicators. The studied data comes from a real-time simulation using controller-pilot data-link communication (CPDLC) technology, recently completed at EUROCONTROL CRDS in Budapest.

Keywords - ATC complexity; task demands; controller's activity; workload; safety

I. INTRODUCTION

The EUROCONTROL Statistics and Forecast Service (STATFOR) predicts that in 2025 the number of commercial flights in Europe will be between 1.7 and 2.1 times the traffic in 2005 [1]. This is an average growth of 2.7%-3.7% per year. The most pressing problem facing the European Air Traffic Management (ATM), therefore, will be to provide sufficient capacity to meet this increased air traffic demand, while at the same time the safety level of air travel has to be maintained or even improved. Airspace capacity that lags behind air traffic demand inevitably leads to flight delays, which in turn means an economic loss to airlines.

In the current air traffic control (ATC) environment the key limiting factor to increasing sector capacity is the workload of the air traffic controller. Therefore proposed solutions for increasing airspace capacity aim at reducing controller workload - which includes: the delegation of separation tasks from ground to the aircraft (e.g. the free-flight concept [2]), a re-sectorisation of the airspace, and the introduction of new controller support tools in order to reduce the amount of work, or at least the difficulty of the controller tasks. As the work of air traffic controllers is foremost cognitive in nature a considerable amount of research has been undertaken to understand the complex task demands that drive the workload of a controller (see [3] for a recent review). The term "workload" denotes a subjective quality reflecting the individual controller's perception of the task demand imposed on him/her by the current air traffic situation. Thus,

many studies implicitly assume that controller workload varies as a function of both directly measurable air traffic factors (number of aircraft in the sector, speed variability, proximity of aircraft, etc.) and controller's activity mediated by factors such as the controller's abilities, age, fatigue, level of experience, etc. [4].

As ATC is a safety-critical working environment any changes implying an adverse impact on controller workload have a direct bearing on flight safety (Fig.1.adopted after [4]).

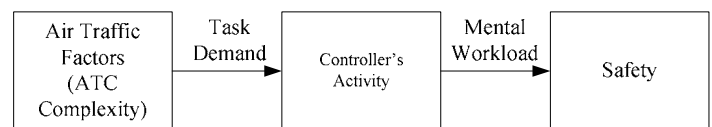


Figure 1. Simplified scheme of the relationship between ATC complexity and safety

The present paper examines the relationship between task demands as defined by a set of ATC complexity factors, controller's actions, subjective workload, and safety. For a safety criterion, a metric was used that was recently developed as part of the EUROCONTROL INTEGRA project [5]. This metric referred to is Propensity and is defined as a proxy for the likelihood of a safety significant event occurring during normal operations. Thus, the present study attempts to predict subjective workload and Propensity as criteria on a moment-to-moment basis using a linear combination of ATC complexity factors and controller's activity measures as predictors.

The paper is organized as follows: In the following section we will give a brief overview of research on ATC complexity and the derivation of task demand metrics from which a selection for the purpose of the present study was made. A more detailed description of controller's activity and workload measures used in the study follows. Then, the INTEGRA safety metric is explained. Next is described the real-time simulation experiment which provided the data base for the calculation/collection of the predictor and criterion metrics. What follows are a description of the approach for statistical analysis and the presentation of the results. Finally, these results will be discussed and conclusions drawn.

II. ATC COMPLEXITY

A straightforward determinant of controller workload is simply the number of aircraft for which the controller is responsible in a specified time and sector. This measure is referred to as the sector load. Predicting sector load and avoiding sector overload is the basic tool upon which current traffic flow management is built. However, the level of difficulty experienced by the controllers depends on additional factors beyond the number of aircraft present in a sector [6]. To be able to capture more accurately ATC complexity, it is necessary to take into consideration also flight characteristics of each individual aircraft as well as interactions between pairs of aircraft. Important flight characteristics of aircraft relate to instantaneous changes of the state of the aircraft, e.g. changes in altitude, heading or speed. Interactions between aircraft are considered not only in terms of potential conflicts but also include the pattern of how aircraft converge and the degree of what in [7] has been referred to as the disorder among aircraft, i.e. the variability in headings and speeds of aircraft. Despite the fact that ATC complexity has been the subject of a significant number of studies (see [3] for a recent review), and many complexity factors have been proposed, up to now a comprehensive and generally accepted set of measures has not been defined yet. For the purpose of the present study, a list of complexity factors was selected that has been consistently found to be important and for which detailed calculation formula have been reported. The factors were partially elicited from work described in [8 - 12].

The selected overall set of 24 complexity factors is presented in Table I. It is out of the scope of this paper to describe all 24 factors in detail. For a more thorough review of the listed factors readers are referred to the indicated source literature.

III. CONTROLLER ACTIVITY – LINK BETWEEN TASK DEMANDS AND CONTROLLER’S WORKLOAD

Even though task demand factors can capture one aspect of the ATC situation, it should be kept in mind that ATC is a dynamic environment and that controllers actively interact with the traffic, and therefore have an important influence on ATC complexity and hence the level of safety.

Several researchers agree that workload is a result of such a complex interaction between the task demand and the way the controller actively manages the situation (e.g. [3], [4], [13], [14]). Moreover controllers, by performing certain activities, regulate the evolution of the task demands with the aim of keeping workload at an acceptable level. Nevertheless, not all controller tasks are observable. As defined by [15], there are four controller tasks while managing the ATC situation: monitoring, evaluating, planning and implementing the formulated plan. Furthermore, out of these four tasks only one is observable, and that is the implementation process. It means that by

taking only objectively measurable (sub)tasks into consideration, it is possible to capture only one aspect of comprehensive controller activity involved. However, as this aspect of the controller’s activity is directly connected with changes made by the controller on the ATC situation, we considered it sufficient for our study.

Thus, in the current study, controller’s input (data entries) and radio communication were used as the representatives of performed controller’s activities (the study is based only on the Executive controller data entries, and not Planning controller, and therefore no phone communication is not considered here).

TABLE I. COMPLEXITY FACTORS

	Complexity Factors	Used in
1	number of aircraft	[8], [10], [12]
2	number of climbing aircraft	[8], [10], [12]
3	number of descending aircraft	[8], [10], [12]
4	number of aircraft with heading change greater than 15°	[10], [9]
5	number of aircraft with the speed change greater than 10 knots	[10], [9]
6	number of aircraft with lateral distance between 0-25nm and vertical separation less than 2000ft above 29000ft	[10], [9]
7	horizontal proximity measure 1 (C5)	[8], [10], [12]
8	vertical proximity measure 1 (C6)	[8], [10], [12]
9	horizontal proximity measure 2 (C7)	[8], [10]
10	vertical proximity measure 2 (C8)	[8], [10], [12]
11	horizontal proximity measure 3 (C9)	[8], [10]
12	vertical proximity measure 3 (C10)	[8], [10]
13	time-to-go to conflict measure (C13)	[8], [10]
14	variance of ground speed (C14)	[8], [10], [12]
15	ratio of standard deviation of speed to average speed (C15)	[8], [10], [12]
16	density indicator (mean)	[7], [11], [12]
17	variability in headings (track_disorder)	[7], [11], [12]
18	variability in speed (speed_disorder)	[7], [11], [12]
19	divergence between pairs of aircraft	[7], [11], [12]
20	convergence between pairs of aircraft	[7], [11], [12]
21-22	sensitivity indicator (a/c converging-mean; a/c diverging-mean)	[7], [11], [12]
23-24	insensitivity indicator (a/c converging-mean; a/c diverging-mean)	[7], [11], [12]

IV. THE INTEGRA CONCEPT OF SAFETY METRICS

Within the INTEGRA concept [16] the term Propensity expresses the likelihood of a safety significant event occurring during the operation of the ATM system. It is defined through the probability function of an aircraft about its calculated position. Therefore, the interactions between aircraft are presented through the interactions between these probability functions. Propensity is calculated for each pair of aircraft that are within defined cut-off criteria for both vertical and horizontal separation.

If the distance between two aircraft is decreasing, the

interaction between their probability functions will be higher, and therefore, the value of propensity metric is increasing. Additionally, the INTEGRA authors introduced into propensity calculation a so-called safety weighting function for the pair of aircraft, determined by the distance between the two aircraft. The purpose of the safety weighting function is to describe the safety significance of proximity between aircraft. This weighting function also takes the use of advanced tools, the density of air traffic, higher than normal information processing loads and severe weather into account.

The propensity defined for an aircraft pair takes values between 0 and 1: the propensity tends to 1 when there is a major reduction in safety margins and tends to 0 when safety margins are assured. More detailed information on the propensity metric can be found in the [16].

V. THE REAL-TIME SIMULATION EXPERIMENT

1) Simulation

In order to obtain relevant values, data were recorded during a two-week LINK2000+ Small Scale Real Time Simulation 2 experiment (LINK 2000+ SSRTS2). The aim of this simulation was to develop and validate new principles of task delegation between the planning and executive controller with the aim to best accommodate the Controller-Pilot Data-Link Communication (CPDLC) capability in an en-route environment [17, 18]. The simulation involved three different sectors of the Central European Air Traffic Services (CEATS) airspace. The data used for the present study are data obtained for the two busiest sectors simulated.

2) ATC Complexity measures

The flight plans and flown trajectories were used as input data for the complexity factors calculations. Customized software was developed to calculate these values for each 2 minute time steps.

3) Controller activity measures

All inputs made by the executive controller recorded during the simulation were extracted. These inputs refer to assignments of vertical rate, exit flight levels/planned entry levels, cleared flight levels, headings, speed instructions, and direct clearances. These were summed across each 2-minute time step and across input, resulting in only one measure named Actions_SUM. Furthermore, cumulative duration of radio calls (= frequency occupancy time per 2-minute time step) was calculated as well as the average duration of single calls. Altogether, we used three measures of the controller's activity – Actions_SUM, Frequency Occupancy Time and Average Radio Duration obtained for every 2-minute time steps.

4) Workload measures

For the same time steps, controllers were providing workload ratings. To collect workload measures during the simulation the Instantaneous Self Assessment (ISA) technique as operator-subjective metric was applied, where the air traffic

controller gives subjective ratings of workload. This tool was developed by the UK NATS and offers 5 points rating scale. On every time step controller can opt a level of workload ranging from very low to very high.

5) Safety measure

As noted before, the propensity metric was used as the safety criterion. The propensity value is calculated using software developed within the frame of the EUROCONTROL INTEGRA project. The values are obtained for each pair of aircraft within a sector unit during the 2-minute time steps. Bearing in mind that there was more than one pair of aircraft in the sector, it was necessary to extract one value per time step. In the study we opted for the maximum value.

Descriptive statistics of the extracted measures are given in table II.

TABLE II. DESCRIPTIVE STATISTICS OF DEPENDENT MEASURES

Measures	Min.	Max.	Mean	Std. Deviation
Actions_SUM (count)	0.000	13.000	3.320	2.372
Frequency Occupancy Time (s)	0.000	50.400	23.991	8.779
Average Radio Duration (s)	0.000	7.200	3.245	0.674
ISA (rating 1 to 5)	1.000	5.000	2.870	0.736
propensity_max (value 0 to 1)	0.010	0.976	0.599	0.147

6) Participants and data extraction procedure

The LINK2000+ SSRTS2 experiment involved a total of 18 controllers out of which 6 controllers worked on the two sectors considered here. The data used for the statistical analysis were derived from these 6 participants only. Each controller completed an overall of 8 exercises of 1 hour and 20 minutes, from which 1- hour recordings were extracted for analysis. Scores were derived for every 2 minutes, resulting in 30 measurements per exercise. These data were obtained for each indicator (ATC complexity measures, workload measures and safety measures). The overall dataset comprised 6 (controllers) x 8 (exercises) x 30 (time segments) = 1440 measurements for each indicator. Prior inspection of the data set revealed that during the whole simulation (all 8 exercises) one of the six controllers always rated workload as 'fair', hence there were no variations in workload measure. The data of this participant was discarded from the analysis, with 1200 measurements remaining. In 58 time segments (4.8%) data was missing. Therefore, subsequently reported results are based on measurements obtained in 1142 time segments.

VI. STATISTICAL EVALUATION AND RESULTS

A. Principal Component Analysis

In a first analysis step, a Principal Component Analysis (PCA) on all 24 complexity metrics was computed in order

to derive a reduced number of uncorrelated predictor variables for the subsequent computation of regression models.

Principal components having an eigenvalue > 1 were extracted and subsequently rotated using the VARIMAX method. This analysis resulted in the extraction of 8 principal components that accounted for 67.26 % of the total variance in the metrics. The table III. displays these components sorted by the sizes of their eigenvalues and along with the percentage of variance they account for. By inspection of the pattern of loadings given in the 8 component x 24 metrics matrix, the following components meanings could be derived. Note, that the loading of a given metric on a given component is equivalent to the correlation between that metric and that component. Therefore, the metric with the highest loading by and large guides the interpretation of the component.

TABLE III. RESULTS OF THE PRINCIPAL COMPONENT ANALYSIS

Components	Eigenvalue	% of Variance	Cum. % of Variance
Comp.1	4.935	20.563	20.563
Comp.2	3.442	14.343	34.906
Comp.3	1.775	7.395	42.301
Comp.4	1.450	6.042	48.343
Comp.5	1.344	5.602	53.944
Comp.6	1.129	4.706	58.650
Comp.7	1.036	4.316	62.966
Comp.8	1.030	4.291	67.257

Comp.1 – ground speed variance and divergence/convergence: strongly related to the variance of the ground speed (0.884) and the ratio of the standard deviation to the mean ground speed (0.845). Also, the strong correlation with divergence and convergence factors (0.787 and 0.785 respectively) was recognised, which is in compliance with speed significance, as divergence/convergence factors actually measure how fast aircraft are moving toward/from each other.

Comp. 2 – aircraft count: this component has the strongest correlation with the number of the aircraft in the sector (0.816)

Comp. 3 – horizontal proximity: this component can be considered as addition to the previous one, as it shows high correlation with the horizontal distance between aircraft taking into consideration the aircraft count - horizontal proximity measure (C5) : 0.894 and density_mean: 0.815 . Together these two components (*Comp. 2* and *Comp. 3*) can be representatives of so-called *sector density*.

Comp. 4 – aircraft vertical transitioning: highly correlated to the number of descending aircraft (0.785) as well as speed

change related to this vertical evolution (0.732)

Comp.5 – conflict sensitivity: this component is loaded highly by both sensitivity indicators ($Sd_{+}(i)$:0.772 and $Sd_{-}(i)$: 0.751). Sensitivity is related to the gradient of the relative distance between aircraft. This indicator measures the change in terms of relative distance in response to changes in speed and heading of the involved aircraft. If sensitivity is high only small changes in heading and speed imply a high impact on relative distance. This is the case, e.g. when two aircraft are heading towards each other. The sensitivity indicators are designed to set a weight on potential conflicts that are difficult to solve (see 12). Note that a situation with high sensitivity is easier to resolve for the controller than one with a low sensitivity [7].

Comp.6 – insensitivity: This component is strongly related to the insensitivity indicators both for convergence and divergence of the aircraft ($insen_c$: 0.723 and $insen_d$: 0.686). It is not simply an analogue with the opposite direction to the previous component. High insensitivity is given for a pair of aircraft when the degree of convergence is high while sensitivity for convergence is low.

Comp.7 – vertical separation: high correlation with the measure of the vertical separation of aircraft in close horizontal proximity (C10) defines this component (0.849)

Comp.8 – horizontal separation: analogously to the previous component, this component is defined based on the correlation with the measure of horizontal separation of the aircraft in close vertical proximity C9 (0.908).

The PCA yielded 8 component scores for each two-minute interval which were used as predictors in the subsequent multiple regression analyses.

B. Multiple Regression Analyses

Two sets of multiple regression models were computed. The first set was performed to assess the effectiveness in predicting ISA workload ratings on the basis of ATC complexity and controller activity metrics. The second set of multiple regression analysis assesses the effectiveness of predicting propensity using ATC complexity, controller activity and ISA workload ratings as predictors.

1) ISA regression models

Instead of using the stepwise linear regression involving all predictors, we first compared two alternative multiple regression equations to fit the data. For the first multiple regression equation all 8 component scores were forced into the model regardless of their single significance. In the second equation the 3 activity metrics entered the equation. This was done in order to assess the contribution of ATC complexity components in relation to the controller activity metrics. Table IV contains the global statistics of these two equations. It can be seen that the first equation containing only complexity components yielded a multiple R of 0.36

corresponding to R^2 of 0.13 or in other words corresponding to 13% of variance of the ISA workload ratings. Adding the controller activity measures in the second equation contributed to a significant increase in the multiple R although it increased the percentage of variance explained by only 3% to a total of 16%. Therefore, it can be concluded that both sources of information, ATC complexity and controller activity, have a unique contribution to the prediction of ISA workload ratings.

TABLE IV. COMPARISON OF ALTERNATIVE MULTIPLE REGRESSION MODELS FOR PREDICTION OF ISA

Regression equation containing	mult. R	R ²	R ² change	F change	df	Sig. F change
complexity components	0.36	0.13	0.13	21.20	8, 1131	0.000
complexity components and controller's activity measures	0.40	0.16	0.03	13.77	3, 1128	0.000

A final multiple regression model was computed using traditional stepwise linear regression approach in order to identify those predictors that are responsible for the significant contribution to workload prediction. This model we refer to as the optimised model as all insignificant variables have been removed. The parameter statistics of this model is given in Table V. The model consists of 8 parameters (Comp.1 – Comp.6, Frequency Occupancy Time and Average Radio Duration)

TABLE V. PARAMETER STATISTICS OF THE OPTIMISED MODEL FOR THE PREDICTION OF ISA WORKLOAD RATINGS

	B	Std. Error	Beta	t	Sig.
Comp.1	0.052	0.020	0.071	2.596	0.010
Comp.2	0.102	0.021	0.139	4.921	0.000
Comp.3	0.101	0.020	0.137	5.009	0.000
Comp.4	0.092	0.020	0.124	4.561	0.000
Comp.5	-0.147	0.021	-0.200	-7.166	0.000
Comp.6	0.074	0.020	0.100	3.676	0.000
Frequency Occupancy Time	0.012	0.003	0.142	4.627	0.000
Average Radio Duration	-0.182	0.033	-0.164	-5.488	0.000

The stepwise regression analysis revealed that the first 6 out of 8 complexity components remained in the prediction model. The components that showed the strongest correlation with ISA ratings are Comp. 3 and Comp. 5 which consider horizontal proximity and the conflict sensitivity. The higher horizontal proximity, i.e. the closer the aircraft in the horizontal plane, the higher was controller workload. When sensitivity of the conflict increased, the workload ratings of the controller decreased, which is consistent with [7].

Frequency Occupancy Time and Average Radio Duration representing the communication load also remained in the model. When Frequency Occupancy Time, i.e. overall frequency occupancy time is increasing, the workload rating is also higher. On the other hand, the increment of the average duration shows the decrease in controller's workload.

2) Propensity regression models

Analogously to the ISA regression models, we compared alternative multiple regression equations in order to evaluate how the three sources of indicators (ATC complexity, controller activity and ISA workload) contributed to the prediction of propensity. Four propensity regression equations were considered: the first equation contains only the scores of the ATC complexity components, two "intermediate" equations in addition contain the three activity measures or the ISA rating, respectively. The fourth equation contains all input variables. The global statistics of these equations are shown in table VI. The equation that contains only ATC complexity components yields a multiple R of 0.53, i.e. accounts for 28% of variance of the propensity metric. When comparing the two "intermediate" equations, one can see that adding the 3 activity measures to equation 1 improved the prediction only by 1% (see R^2 change in equation 2 in table VI), while adding ISA improved the prediction by 2%. However, there is no gain in predictive power when the measures of controller activity are added to equation 3 as indicated by the statistics obtained for equation 4, the full model.

TABLE VI. COMPARISON OF ALTERNATIVE MULTIPLE REGRESSION MODELS FOR PREDICTION OF PROPENSITY

Regression equation containing	mult. R	R ²	R ² change	F change	df	Sig. F change
1.complexity components	0.53	0.28	0.28	55.99	8, 1133	0.000
2. complexity components and controller's activity measures	0.54	0.29	0.01	3.63	3, 1130	0.013
3. complexity components and ISA ratings	0.55	0.30	0.02	26.05	1, 1130	0.000
4. complexity components, ISA and controller's activity measures	0.55	0.30	0.00	1.83	1, 1127	0.139

Finally, stepwise regression analysis was performed for the identification of an optimized model. The parameter statistics this model are shown in table VII. The model consists of 5 parameters (Comp.2 – Comp. 6 and ISA ratings).

TABLE VII. PARAMETER STATISTICS OF THE OPTIMISED MODEL FOR THE PREDICTION OF PROPENSITY METRIC

	B	Std. Error	Beta	t	Sig.
Comp. 2	0.027	0.004	0.184	7.268	0.000

Comp. 3	0.041	0.004	0.280	11.104	0.000
Comp. 4	0.014	0.004	0.095	3.795	0.000
Comp. 5	-0.049	0.004	-0.332	-12.924	0.000
ISA ratings	0.028	0.005	0.138	5.192	0.000

The Comp.1, which stands for ground speed variance and divergence/convergence of the pair of aircraft, did not remain in the model. Even though through Principal Components Analysis it resulted in highest eigenvalue and accounted for 20.563% of the total variance in the metric, it did not contribute to the prediction of propensity metric. The components that showed the strongest correlation with propensity metric are Comp. 3 and Comp. 5, i.e. horizontal proximity and the conflict sensitivity. Also, of great significance are correlations of propensity metric with number of aircraft in the sector (Comp.2) and vertical movements of the aircraft in the sector (Comp.4). As it could be anticipated through the previous comparison of alternative regression models, the activity measures did not remain in the model. Only ISA ratings contributed to the prediction of propensity.

VII. DISCUSSION AND CONCLUSIONS

The focus of the paper was the investigation of the relationship between ATC complexity, controller's activity measures, subjective workload and safety measures. Based on the previous work in the field an initial set of 24 complexity factors was defined. In order to reduce this set, a Principal Component Analysis (PCA) was performed, which resulted in 8 components. [12] also performed a PCA using a set of 27 complexity indicators as input variables. There is a big overlap between their and our set of input variables (see complexity factors that is used also in [12] in table I). However, their data was extracted in one-minute time steps from real traffic recorded in a total of 103 sectors across one day of traffic. Their PCA revealed 6 components (using the same extraction criterion of eigenvalue > 1 as in the present paper) that accounted for 76% of the total variance. Aircraft count had the highest loading on the first component accounting for 46.7% of the variance in their PCA which corresponds to the second component in the PCA of the present study. Comparing the loadings of the complexity indicators on the remaining five components suggests that the first component of our PCA corresponds to a mix of their second, fourth and fifth component. Finally, their sixth component is more or less equivalent to our seventh component. Their third component in [12] was correlated with a metric representing the degree of incoming sector flows, which was not considered here. Therefore, a good agreement between the PCA results obtained in our simulation study and their real-traffic study can be concluded.

The scores for the eight components were calculated and further entered in different multiple regression models in order to reveal their correlation with ISA workload measures,

controller activity indicators and INTEGRA propensity as a safety measure.

First of all, it was found that subjective controller workload as measured by the ISA ratings depends on additional factors rather than only on aircraft count. This is in agreement with a couple of other studies (e.g. [7], [9], [12]).

The results suggested that subjective workload hinges on other aspects of the ATC complexity as well as on the communication load of the controller. Both the total frequency occupancy time and average radio duration significantly correlate with ISA workload ratings. Moreover, as it was hypothesised in [19], the present study suggested that the average time for an individual communication is negatively related to workload. In other words, the amount of time that a controller spends on a single communication should decline as the situation gets busier.

Furthermore, the results of the propensity regression model provided some insight into the construct validity of this metric in the sense that we first can pinpoint what aspects of the sector situation influence the degree of propensity, and second, in what way propensity is related to subjective workload and measures of controller activity. Four complexity components were found to be correlated with propensity: aircraft count, horizontal proximity, vertical transitioning, and conflict sensitivity. It is assumed that it is due to the calculations of the propensity metrics which relies on the calculated position of aircraft rather than their speed, the latter being reflected in component 1 of the PCA.

It is still an issue for further validation to demonstrate that propensity is a valid safety metric. This should be taken into account in future work, which should consider also other the actual occurrence of safety critical events into account.

REFERENCES:

- [1] EUROCONTROL, "EUROCONTROL long-term forecast: IFR flight movements 2006-2025 (DAP/DIA/STATFOR Doc216)", EUROCONTROL Statistics and Forecast Service (STATFOR), 2006.
- [2] RTCA, "Report of the RTCA board of directors' select committee on free flight", Washington DC, 1995.
- [3] B. Hilburn, "Cognitive complexity in air traffic control - a literature review (EEC Note No. 04/04)", EUROCONTROL, 2004.
- [4] S. Loft, P. Sanderson, A. Neal and M. Mooij, "Modeling and predicting mental workload in en route air traffic control: critical review and broader implications", in Human Factors: The Journal of the Human Factors and Ergonomics Society, 2007, vol. 49, pp. 376-399.
- [5] EUROCONTROL, "INTEGRA metrics & methodologies - detailed specification of safety metrics; inputs-processing-outputs", EUROCONTROL, 2000.
- [6] B. Sridhar, K.S. Sheth and S. Grabbe, "Airspace complexity and its application in air traffic management", In Proceedings of the 2nd USA/Europe Air Traffic Management 1998 R&D Seminar, Orlando, Florida, 1998.
- [7] D. Delahaye and S. Puechmorel, "Air traffic complexity: towards intrinsic metrics", In Proceedings of the 3rd USA/Europe Air Traffic Management 2000 R&D Seminar, Napoli, Italy, 2000.
- [8] G.B. Chatterji and B. Sridhar, "Measures for air traffic controller workload prediction", In Proceedings of the First AIAA Aircraft

- Technology, Integration, and Operations Forum, Los Angeles, CA, 2001.
- [9] I.V. Laudeman, S.G. Shelden, R. Branstrom and C.L. Brasil, "Dynamic density: an air traffic management metric (NASA-TM-1998-112226), San Jose State University Foundation, San Jose, CA, 1998.
- [10] P. Kopardekar and S. Magyarits, "Measurement and prediction of dynamic density", In Proceedings of the 5th USA/Europe Air Traffic Management 2003 R&D Seminar, Budapest, Hungary, 2003.
- [11] F. Chatton, "Etudes de nouvelles métriques de complexité de la circulation aérienne", Master's thesis, Ecole Nationale de l'Aviation Civile (ENAC), 2001.
- [12] D. Gianazza and K. Guittet, "Evaluation of air traffic complexity metrics using neural networks and sector status", In Proceedings of the 2nd International Conference on Research in Air Transportation (ICRAT), Belgrade, Serbia, 2006.
- [13] Majumdar, W. Ochieng, G. McAuley, J. M. Lenzi, and C. Lepadatu, "The factors affecting airspace capacity in Europe: a cross sectional time-series analysis using simulated controller workload data.", in Journal of Navigation, 2004, vol. 57, pp. 385-405.
- [14] W. S. Pawlak, C. R. Brinton, K. Crouch and K. M. Lancaster, "A framework for the evaluation of air traffic control complexity", In Proceedings of the AIAAGuidance, Navigation and Control Conference, San Diego, CA, 1996.
- [15] J.M. Histon and R.J. Hansman, "The impact of structure on cognitive complexity in air traffic control (Report No. ICAT-2002-4)", MIT International Center for Air Transportation, Department of Aeronautics & Astronautics, Massachusetts Institute of Technology, Cambridge, MA 02139 USA, 2002.
- [16] EUROCONTROL, "INTEGRA metrics & methodologies - safety metrics; technical definitions", EUROCONTROL, 2000.
- [17] EUROCONTROL, "LINK2000+ small scale real-time simulation No.2 (LINK2000+ SSRTS2), final report (CRDS Note No.26)", EUROCONTROL, 2007.
- [18] R. Schuen-Medwed, B. Lorenz and S. Oze, "Empowerment of the planning controller in CPDLC environment", In Proceedings of the Human Factors and Ergonomics Society, Europe Chapter Annual Meeting., Braunschweig, Germany, 2007
- [19] C. Manning, S. Mills, C. Fox, E. Pfeleiderer and H. Mogilka, "The relationship between air traffic control communication events and measures of controller taskload and workload", In Proceedings of the 4th USA/Europe Air Traffic Management 2001 R&D Seminar, Santa Fe, 2001

Assessment of local aircraft crash risk

Application of a cluster analysis as a statistical method for detecting similar airports

Christoph Thiel

Institute of Logistics and Aviation
Technische Universität Dresden
Dresden, Germany
thiel.christoph@googlemail.com

Hartmut Fricke

Institute of Logistics and Aviation
Technische Universität Dresden
Dresden, Germany

The assessment of local aircraft crash risks in the vicinity of airports is of primary importance in numerous safety studies relating to the determination of Third Party Risk due to aircraft accidents. This paper presents an approach of determining local aircraft crash rates by means of a cluster analysis. This statistical method detects similarities between airports in consideration of safety relevant parameters.

Safety, aircraft crash risk, accident ratio, External Risk, similarity analysis, cluster analysis

I. BACKGROUND

A. Introduction

In the context of planning approval procedures at many major European airports, safety relevant issues today gain increasing importance over e.g. noise or environmental issues. When analyzing the safety in air traffic, one essential model for airport related safety studies is the determination of External or Third Party Risk, that is the risk of death due to an aircraft accident for people who do not participate in the air transport system, generally people living or working in the vicinity of an airport.

Despite the lack of regulations relating to External Risk in most European countries (with the exception of The Netherlands and Great Britain), determining External Risk becomes a central instrument for evaluating risks due to aircraft accidents for people living around airports.

The External Risk model consists of three sub-models: the accident ratio (AR), the accident location (AL) and lastly the accident consequence (AC) sub-model. The scope of this paper is the accident ratio sub-model, for which a statistical method of determining a local accident ratio for a specific airport by means of cluster analysis is presented.

In order to further clarify the topic the following section gives an introduction into the External Risk model itself.

B. The External Risk Model

The External Risk (ER) expresses the statistical potential of a human being receiving fatal injuries as the result of a severe aircraft accident or its potential consequences in form of secondary effects on the ground (damages to an industrial plant, for example). This potential is important around airports,

because the operational accident risk for aircraft is highest during takeoff and landing and so ER calculations generally refer to an airport. The term *external* refers to the fact that the risk is calculated for those people who are not formally participating in the air transport system during a given time period. Typically, this is the population residing in the area around that airport, or people who work there (employees). More precisely, these are people located at least temporarily within a selected investigation area around the airport.

The External Risk effectively consists of two types of measurable risk figures, the local or Individual Risk and the group or Societal Risk:

- The Individual Risk is the probability of an imaginary person being killed in a particular location within the investigation area as a result of an aircraft accident during a period of one year. It is therefore not important to know whether a person is actually present or not.
- When calculating risk, it might additionally be important to consider the population that is actually present and the distribution of this population around the airport rather than an imaginary person. Calculations are made in order to determine the size of the risk of one or more simultaneous casualties within this population. This probability of a disaster of a certain size is known as the Societal Risk.

The External Risk model furthermore consists of three sub-models:

- an accident probability model, providing a local Accident Ratio (AR) for fatal aircraft accidents according to the definitions given in ICAO Annex 13 [1],
- an accident location model, providing an Accident Location (AL) distribution probability function referred to the Air Traffic Route System and linked to a runway and/or threshold, and
- an accident consequence model, providing the local Accident Consequence (AC) Area with regard to local terrain and industrial site details.

With these three sub-models and all required airport related parameters (e.g. movements per departure/ arrival route, traffic mix, industrial sites) Individual Risk will be calculated and shown in individual risk contours around the airport area. Following Figure 1 demonstrate such individual risk contours as an example at an imaginary airport:

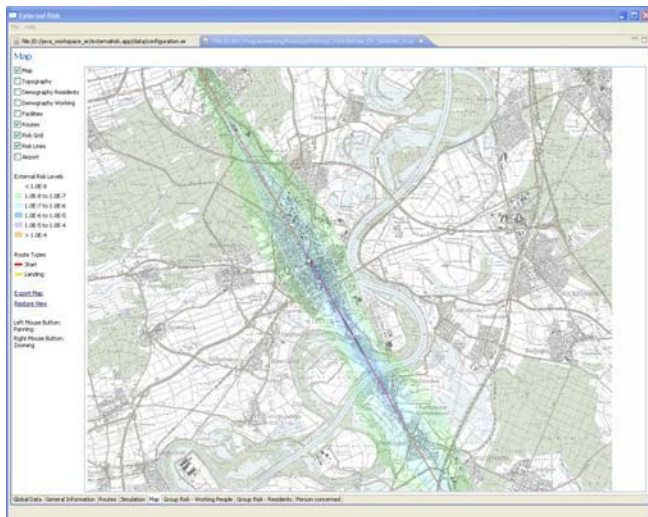


Figure 1. Individual Risk

With information about population density (residents/ employees) within the investigation area (usually an area of 40 km x 40 km centred in the airport reference point) societal risk will be calculated as so called F/n- curves.

As an essential element of External Risk calculation, the accident ratio sub-model and current approaches for calculation are described more precisely in the following chapter II. Furthermore in chapter III is presented an alternative approach of assessing a local accident ratio at a certain airport by means of statistical analysis.

II. CURRENT METHODS OF LOCAL AR DETERMINATION

A. Definition of local Accident Ratio

Generally, aircraft crash risk may be assessed by using theoretical models which would use the measured probabilities of all possible causal factors to predict the probability of a crash. However, such a theoretical approach is very problematic, since aircraft accidents are usually the result of a combination of many separate causal factors with unknown probabilities and complex interrelationships. An alternative method is to use empirical data on accidents and on aircraft movements to calculate aircraft crash risks. This non-causal approach assumes that the historical accident ratio will continue into the future, which, if there are future safety improvements, as they are currently aimed by the SESAR Consortium [2], may lead to an overestimation of accident ratio in future years.

The accident ratio itself is defined as the probability of an aircraft accident per movement and is calculated by dividing the number of accidents in a certain time period by the number of relevant movements within this period:

$$AR = \frac{\text{Number Of Accidents}}{\text{Number Of Movements}} \quad (1)$$

Because crash probabilities differ considerably between airports, a selection of data from the available worldwide accident data is required to make the accident ratios suitable for a specific airport.

Many sources provide global AR values, calculated by a simple division of worldwide accidents by worldwide movement data. For example the "Statistical Summary of Commercial Jet Airplane Accidents" annually published by Boeing [3] provides a good overview about the global AR per year in worldwide commercial operations, as well as a trend in global AR.

But this global AR may not be valid for a specific airport; in fact, a local AR differs from the global AR, depending on the airport, because there are numerous factors which influence the safety within a certain investigation area (certain airport) and lead to a local accident crash probability.

Calculating a local accident crash probability suitable for a specific airport may be done in many different ways. The current essential (more or less standardized) methods, which are used by several national organizations to determine local aircraft crash risks (e.g. for calculating External Risk), are presented in the following sections.

B. NATS-Method

The National Air Traffic Services (NATS) of the United Kingdom uses an AR Model specific for generic aircraft groups within their model of Public Safety Zone (PSZ) calculation [4]. On an empiric basis, crash rates per classified aircraft group will be calculated.

The NATS defines a classification of aircrafts according to their type of engine (jets, turboprops and piston-engine), the region of their manufacturer (eastern, western) and their date of first delivery. The essential breakdown of aircraft classification is therefore:

- Class I western airliner jets (e.g. Boeing B707, Comet)
- Class II – IV western airliner jets (Boeing B727, B747, Airbus A310, etc.)
- Other jets (Eastern Jets / Executive Jets)
- Turboprops (western airliner Turboprops before 1970/ after 1970, unclassified Turboprops)
- Piston-engine

By means of historical aircraft accident data and statistics, NATS calculates an AR for every group of aircraft and with knowledge about the share of every aircraft group in the traffic mix at a certain airport, the overall AR for this certain airport can be calculated.

For most of the major, worldwide airports the current traffic-rate of modern western airliner jets (Class II – IV) is more than 95% and the share of e.g. eastern jets or older turboprops is marginal. So, according to the NATS-method this

approach would calculate nearly the same overall AR for every major airport.

This approach is therefore not very useful for calculating an AR suitable for a specific airport, because a causal differentiation primarily for major airports is not given. Furthermore a differentiated analysis of e.g. the airports Air Traffic Control (ATC) infrastructure or operational performance which may influence a local AR is not taken into consideration.

C. DOE- Standard

The U.S. Department of Energy (DOE) Standard DOE-STD-3014-2006 [5] describes a method of risk analysis for hazardous facilities due to aircraft crashes. The chosen approach is very similar to the method of External Risk calculation, as here also the probability of an aircraft crash independent from the crash location is one part of the risk calculation process.

For near-airport facilities, the aircraft crash rate is empirically calculated for accidents during take off or landing at a specific airport (close to a hazardous facility) differentiated by type of traffic (general aviation, commercial aviation and military). Furthermore, there are some defined sub-categories: for general aviation by type and number of engines and for commercial operations the two sub-categories: air carrier and air taxi.

This classification by type of traffic implies the same problem as the classification by aircraft type as favoured by NATS: a differentiation for many major airports is not given, because one type of traffic is omnipresent (here commercial air carrier) and therefore this method would also calculate nearly the same overall AR for most major airports.

D. NLR- Method

In 1993 the National Aerospace Laboratory of the Netherlands (NLR) published the first documentation about External Risk Calculation [6]. This document gives, amongst others, a detailed description how to determine a local accident ratio. Despite of numerous model updates [7] this accident ratio model remains more or less unchanged until today.

The NLR does not evaluate local AR by means of classification by aircraft or type of traffic only; in fact it uses a method of detecting a certain set of airports similar to the airport under investigation by means of expert justice. The similarity analysis may differ depending on the airport under investigation, but it takes following essential criteria into consideration:

- ATC Infrastructure (precision/ non-precision approaches, terminal approach radar, etc.)
- Size of airport (e.g. number of runways)
- Operational performance (annual movements)
- Local geographical peculiarities (e.g. mountainous area)

So, the NLR- approach does not include the disadvantages of the NATS or DOE method, as it uses more than one parameter for the AR assessment. However this expert based analysis is a very subjective and irreproducible method, as there are no clearly defined selection criteria (this differs depending on the airport) and the selection process itself is not traceable.

E. Conclusion

Each of the presented methods for assessing a local aircraft crash risk includes certain disadvantages: the NATS and DOE methods do not apply a real similarity analysis, as they only take one parameter into consideration and the NLR method by means of expert justice is irreproducible. To avoid these disadvantages, an alternative method of detecting similar airports is presented here. This method takes into consideration the operational performance of the airports as well as reproducibility, since it is calculated by means of statistical analysis.

III. SIMILARITY BETWEEN AIRPORTS, A STATISTICAL APPROACH

A. Introduction

1) AR determination process overview

A selection of a certain number of airports empirically similar to the airport under investigation and the total count of accidents at these similar airports has to be quantified to determine a local accident ratio.

So, the first step of determining a local AR is the selection of a certain number of similar airports. Here, similarity is understood as similar operational performance of the airports in terms of yearly passenger traffic, quantity of handled cargo and air traffic movements. By means of correlation analyses, it could be shown, that the chosen parameters (flight operations, cargo, passengers) provide the highest correlation between traffic load and traffic safety: a high correlation between the available Air Traffic Control Infrastructure, type of traffic at the airport (IFR/VFR) and the size of airport, for example, could be identified.

Once a certain set of similar airports is defined, all relevant accidents at these airports have to be investigated. The local AR for the airport than is calculated as the division of the sum of all accidents by sum of all movements. The whole process of determining a local AR at a certain airport is presented in following figure 2:

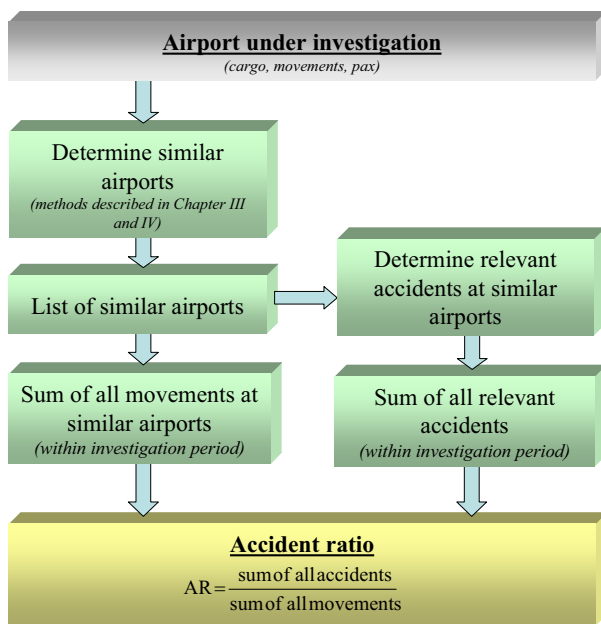


Figure 2. AR determination process

In order to estimate crash rates from historical data, it is necessary to have complete data on airport related crashes and the corresponding numbers of movements. The following sections give an overview how to investigate this information.

2) Investigation of traffic data

The first step to determine a local AR, suitable for the airport under investigation, is the selection of a certain number of similar airports. As described above, operational performance of the airports in terms of yearly passenger traffic, quantity of handled cargo and air traffic movements gives a good indication for similarity analysis. Therefore these three parameters for a huge set of airports must be investigated.

The database of the Airports Council International (ACI) [8] is a very comprehensive source of worldwide airport operations data. It provides operational data for more than 1500 airports worldwide from 1991 until today and is updated monthly. As mentioned above, there are three variables:

- Number of air traffic movements per year
- Number of passengers per year and
- Amount of handled cargo per year.

These three variables are used to filter similar airports based on the mean value per variable from 1991 until today (17 years) for each airport. This period was selected to provide an empirically stable reference data set while taking into consideration the poor data quality of the 80ies decade. Furthermore for validation of the analysis results, the data used should be applicable to current aviation which implies that only recent data should be taken into account.

The similarity selection process itself will be performed by means of statistical methods and is described later in section B of this chapter.

3) Investigation of relevant accidents

In order to determine relevant accidents at the identified similar airports there are various international accident databases available. A very comprehensive and easy accessible database for worldwide aircraft accidents is the Aviation Safety Net (ASN) [9]. It covers nearly every aircraft accident having occurred during the last 50 years and provides a huge amount of additional information (e.g. investigation agency, link to full accident report). Another comprehensive and free online accessible database is the NTSB- Database [10] which today includes more than 65,000 accidents having occurred since 1982.

Accidents taken into consideration should be at least consistent with following selection criteria:

- Occurrence with at least one fatal injured person on ground or on board the accident aircraft (according to definition of fatal accident in ICAO Annex 13),
- Occurrence during take off or landing phase and within a certain area around the airport (as mentioned above, usually 40 km * 40 km),
- Occurrence within the investigation period (as mentioned above, usually about 15 years)
- Occurrences not involving sabotage, hijacking or military action

Depending on the airport under investigation, additional selection criteria may be defined, e.g. no occurrences with aircraft below 5.7 to MTOM, if the airport under investigation does not include this kind of traffic or its traffic count is marginal.

B. Clusteranalysis in AR- Determination

1) Introduction

Cluster analysis is a multivariate procedure for detecting natural groupings (clusters) in data. Cluster analysis classification is based upon the placing of objects into more or less homogeneous groups, in a manner such that the relationship between groups is revealed. This means, the formed clusters should be internally homogenous (members are similar to each other) and externally heterogeneous (members are different to members of other clusters). Figure 3 below shows such a clustered dataset by means of a schematic diagram (note, that the given figure does not represent the true spread of airports operational performance and grouping, as it is only used for purpose of clarification):

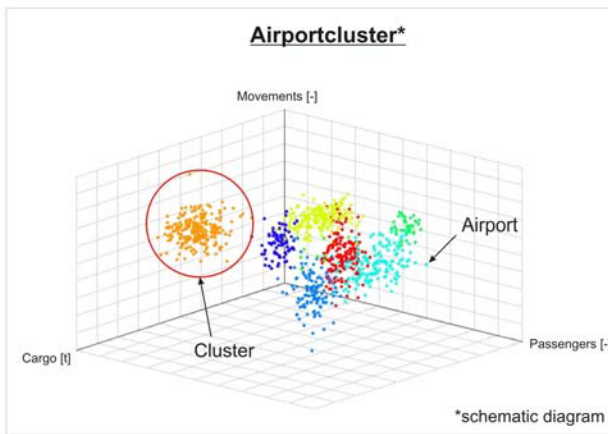


Figure 3. schematic cluster diagram

The selection of airports similar to a certain airport is derived here by a statistical similarity analysis – the cluster analysis. There are some more statistical methods for such similarity analyses, but irrespective of the method every similarity analysis underlies the same diametrical effect: The higher the number of similar elements (airports) is, the lower is the similarity to the reference element (airport under investigation). This means, the number of similar airports has to be high enough to achieve a statistically stable data basis for AR calculation, as well as the similarity to the airport under investigation has to be given for every considered airport. Otherwise, a set of e.g. 5 to 10 similar airports may induct a high similarity, but it's not enough to ensure a statistically stable data base. Therefore, finding a balanced proportion between the number of airports and similarity measure is of primary importance.

2) Standardization of variables

Standardization of variables has to be executed to enable the comparison of variables to minimize the bias in weighting which may result from differing measurement scales and ranges.

The variables used for clustering - the yearly passenger traffic, quantity of handled cargo and air traffic movements, may show very high differences, e.g. depending on the airport the number of yearly passengers is about 100-times higher than the number of yearly movements (generally speaking, average number of passengers per movement). For comparative purposes and in order not to overestimate the passenger variable, every variable has to be normalized before starting the cluster algorithm. The best method of normalization for the chosen approach is the so called *Z-Transformation*, which results in a mean value of zero and a variance of one for each variable:

$$z_i = \frac{x_i - x_m}{S_i} \quad (2)$$

With x_m as mean value and S_i as standard deviation:

$$x_m = \frac{1}{N_x} \sum_{i=1}^{N_x} x_i \quad (3)$$

$$S_i = \sqrt{\frac{1}{N_x - 1} \sum_{i=1}^{N_x} (x_i - x_m)^2} \quad (4)$$

3) Clustering with Ward's linkage

Generally, hierarchical cluster analyses are comprised of agglomerative and divisive methods. The divisive methods start with all of the observations in one cluster and then proceed to split (partition) them into smaller clusters. The agglomerative methods initially consider each observation as a separate cluster and then proceeds to combine them until all observations belong to one cluster. The here applied cluster algorithm is of agglomerative nature.

The two key steps within cluster analysis are in the first place the measurement of distances between objects and secondly to group the objects based upon the resultant distances (linkages). The distances are a measure of similarity between objects and may be measured in a variety of ways, e.g. as Euclidean distance. The criteria used to link (group) the variables may also be undertaken in a variety of manners, as a result significant variation in results may be seen. Linkages are based upon how the association between groups is measured. Four of the better-known algorithms for hierarchical clustering are average linkage, complete linkage, single linkage and Ward's linkage. Ward's is a popular default linkage which produces compact groups of well distributed size. The final algorithm according to Ward's procedure is applied here: it uses an analysis of variance approach to evaluate the distances between clusters. This method is regarded as very efficient for the existing approach, as it forms clusters of approximately the same size.

Ward's linkage is a specific weighting method applied in the hierarchical cluster analysis. The linkage function specifying the distance between two clusters is computed as the increase in the error sum of squares (ESS) after fusing two clusters into a single one. Ward's Method seeks to choose the successive clustering steps so as to minimize the increase in ESS at each step.

The ESS of a set x of N_x values is the sum of squares of the deviation from the mean value x_m . For a set x the ESS is therefore described by the following expression:

$$ESS(x) = \sum_{i=1}^{N_x} (x_i - x_m)^2 \quad (5)$$

The distance or linkage function for the distance between the two Clusters X and Y is described as:

$$D(X, Y) = ESS(XY) - [ESS(X) + ESS(Y)] \quad (6)$$

with XY being the combined cluster resulting from fusion clusters X and Y. So, the distance function describes the increase in ESS by fusing Cluster X and Cluster Y into one combined cluster XY. This distance function has to be calculated for every combination of clusters within the dataset and put into the so called distance matrix. The minimum increase in ESS between two certain clusters marks the next

agglomeration step. The two clusters with minimum increase in ESS will be combined into one cluster.

The essential process of agglomerative clustering with Ward’s method is presented in following figure 4:

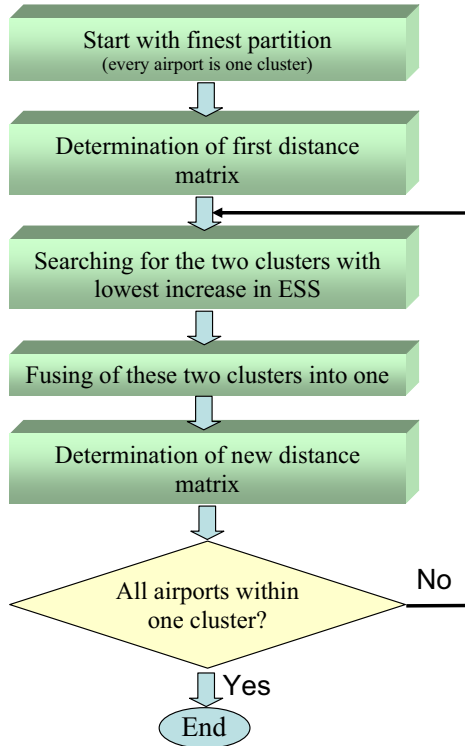


Figure 4. Process of agglomerative clustering

The results of the applied cluster analysis may be presented as a so called dendrogram. This presents the fusion of clusters per cluster step as shown in the following Figure 5, at an imaginary example:

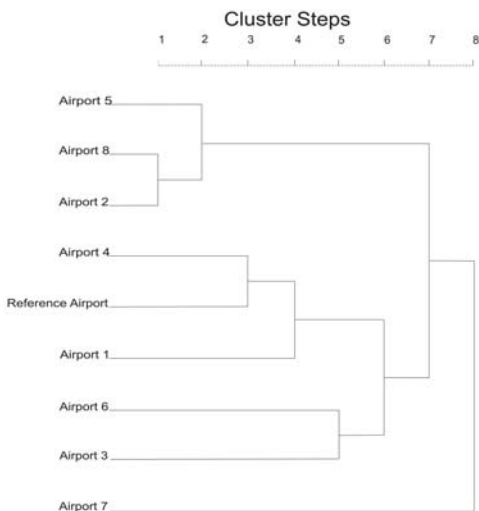


Figure 5. Dendrogram – an imaginary example

As seen, within the first step of cluster analysis, *airport 8* and *airport 2* will be combined to one single cluster and step

three indicates the first clustering of the *reference airport* (combined with *airport 4*). Within step 6 the master cluster contains 5 airports and within step 8 all airports belonging to one cluster, which marks the end of the cluster agglomeration.

This imaginary example also shows the difficulty of setting the best point of interrupting the cluster algorithm. Theoretically every cluster step reaching the minimum count of similar airports may be chosen as final cluster step.

4) *Interrupting the Analysis*

Once the similarity threshold through a minimum reachable data variance is reached, that cluster will be set to the master cluster, which contains the reference airport. All other airports belonging to the master cluster form the set of data that will be used in calculating the AR.

Detecting this similarity threshold (finding the best point of interrupting the cluster agglomeration) may be unfrugal. As described earlier there must be a balanced proportion between the number of similarity airports and similarity measure. The higher the number of included airports in the master cluster is, the lower is the similarity to the reference airport. A statistical method for finding this point may be the so called *F-test*, which provides an indication about the homogeneity of a certain group of airports. Therefore for each step of the cluster agglomeration the F-value for each variable of the reference cluster has to be calculated as the quotient of variance of reference cluster and variance of the entire data set:

$$F = \frac{V(J, G)}{V(J)} \tag{7}$$

With:

$V(J, G)$ = variance of variable J in group G

$V(J)$ = variance of variable J in the entirety

With variance defined as:

$$V(x) = \frac{1}{N_x - 1} \sum_{i=1}^{N_x} (x_i - x_m)^2 \tag{8}$$

As a result, it has to be mentioned that the smaller the F-value is, the higher is the homogeneity of the reference cluster. A cluster has to be considered as homogeneous if the F-value for each variable is not bigger than one.

Generally, every cluster step can be chosen as final master cluster, whose F-value is lower than one and whose reaches a minimum sample size (from experience more than at least 20 airports). Practically, this cluster step should be used which marks the similarity threshold: where the F-value for each variable is still slightly below one and within the next cluster step one of the variables would exceed an F-value of one.

C. *Summary*

Once this similarity threshold is defined, all airports belonging to the master cluster within the derived cluster step can be set as similar airports. Afterwards all relevant accidents at these airports in the given time period have to be investigated as well as the total count of movements at all

airports in the given time period to determine a local accident ratio, suitable for the airport under investigation.

Note, that the given mathematical descriptions in this paper do not represent the full Ward’s cluster algorithm, as the determination process of such multivariate clustering process is much more complex. The full mathematical process is documented in a comprehensible way in various statistical volumes (e.g. [11], [12]). Presenting the full cluster algorithm, would go beyond the scope of this paper, nonetheless it could be shown that the application of a cluster analysis is a good instrument for detecting similar airports.

The following chapter IV gives an overview of the results of a hierarchical cluster analysis conducted according to the method described above, in order to identify prevailing trends.

IV. RESULTS

Based on the described method of airport clustering a cluster analysis was performed by using traffic data from the ACI- airports data base [8]. From this database airports with more than 30,000 movements per year (mean value from 1991 to 2006) and with traffic data from at least six out of 17 years per variable were selected. Within these selection criteria a dataset of 398 worldwide airports was used as input value for the cluster analysis.

The cluster analysis was performed without a reference airport, as the results should not be used for determination of a single AR for a specific airport, but should be used for identifying prevailing trends and correlations between traffic load and AR.

As a result, several clusters that reach a minimum sample size of at least 20 airports and with an F- value lower than one could be found. For each of the 398 airports within these clusters all relevant accidents were examined by using various international accident databases (e.g. [9], [10]). Following Table I present four clusters as an example:

TABLE I. SAMPLE OF AIRPORT CLUSTER

Sample of airport cluster				
Cluster- No.	1	7	9	11
No of airports	54	111	35	21
movements ^a	132,252	102,266	215,236	430,603
cargo ^a	156,436	27,016	331,266	338,934
passengers ^a	8,684,462	2,600,680	17,040,752	24,767,204
No of accidents (last 17 years)	23	95	8	9
AR [per movement]	1.89E-07	4.92E-07	6.25E-08	5.85E-08

a. mean value per airport and per year (last 17 years)

Based on these “characteristic” clusters, a correlation between traffic load and accident ratio was to be verified. Therefore, a large amount of random airports with typical passenger/ movement ratios and cargo/ movements ratios were produced and based on the determined clusters a local AR was calculated for each of these airports.

Finally, it could be found that the higher the number of movements (respectively passengers and cargo volume) is, the lower is the number of accidents per movement. Generally speaking, a decreasing AR by increasing traffic volume could be detected.

Following Figure 6 gives an example of this general trend, for imaginary airports with a passenger/ movement ratio of 60 (60 passengers per movement) and a cargo/ movement ratio of 1.25 (1.25 t cargo per movement):

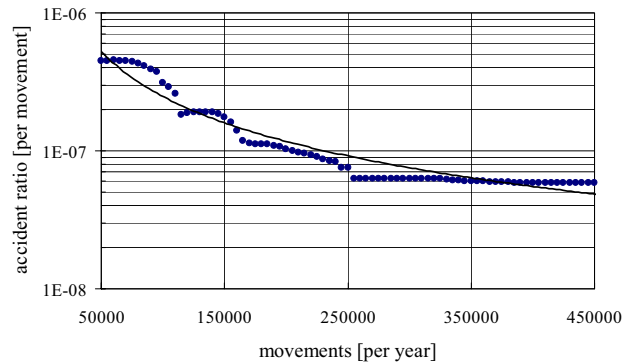


Figure 6. Decreasing accident ratio with increasing traffic volume

As seen in Figure 6 the AR values ranges between $5 \cdot 10^{-7}$ per movement and $5 \cdot 10^{-8}$ per movement, with a significant decreasing trend.

This decreasing trend may be due to the increasing navigation infrastructure (ILS Equipment, RNP-RNAV procedures, etc.) and increasing professionalism of all stakeholders with increasing traffic volume. On the other hand the results lead to the assumption that the complexity of the airport layout (e.g. numbers of runways) does not have a negative influence on the local accident ratio. However, more research is needed to identify all underlying causes for this decreasing trend.

V. CONCLUSION

The aim of the applied hierarchical cluster analysis is to achieve a reproducible, statistically based algorithm of detecting airports which are empirically similar to a certain airport, in order to assess a local accident ratio for this airport.

The cluster analysis is a well known statistical grouping method and applied for the given problem it gives a very good indication which set of airports can be assumed as similar to a specific airport under investigation. In terms of a legalization of External Risk calculations, a fully expert based analysis of similar airports (as favoured by the NLR), which is irreproducible and subjective method, may not lead to a legalization process, as it is not a specific and standardized calculation method. The applied cluster analysis avoids this disadvantage by means of a statistical and reproducible algorithm.

Finally a cluster analysis was performed, that uses traffic and accident data of nearly 400 worldwide airports. It could be

found, that the higher the number of traffic volume at a specific airport the lower is the number of accidents per movement.

ACKNOWLEDGMENT

The author thanks Mr. Norbert Gronak and Mr. Daniel Fiedler from the Gesellschaft für Luftverkehrsforschung (Organization of Air Traffic Research) for supporting him with their expertise in External Risk calculations, and for granting access to their traffic and accident databases. A special thanks goes to Ms. Angela Mantay for her enormous support with the set-up of the cluster algorithm and with the time consuming accident researches.

REFERENCES

- [1] ICAO, "Aircraft accident and incident investigation (Annex 13)", 9th Edition, July 2001
- [2] SESAR Consortium, "D2 - The performance target", SESAR Definition Phase – Deliverable 2, 22.12.2006
- [3] Boeing Commercial Airplane Group, "Statistical summary of commercial jet airplane accident, worldwide operations 1959 – 2006", Seattle (USA), Juni 2007
- [4] A. W. Evans, P. B. Foot, S. M. Mason, I. G. Parker and K. Slater, "Third party risk near airports and public safety zone policy", National Air Traffic Services Ltd (NATS), R&D Report 9636, London, June 1997
- [5] U.S. Department of Energy (DOE), "DOE Standard, Accident analysis for aircraft crash into hazardous facilities", DOE-STD-3014-2006, Washington DC, May 2006
- [6] M. A. Piers, M. P. Loog et al, "The development of a method for the analysis of societal and individual risk due to aircraft accidents in the vicinity of airports" , CR 93372 L, National Aerospace Laboratory (NLR), Netherlands, November 1993
- [7] A. J. Pikaar, M. A. Piers and B. Ale, "External risk around airports. A model update", NLR-TP-2000-400, presented at the 5th international conference on probabilistic safety assessment and management, Osaka, Japan, 2000
- [8] Airports Council International (ACI), Top 500 - Annual Airport Traffic Statistics, Internet: www.airports.org
- [9] Aviation Safety Network (ASN), Aircraft accident database, on permanent update, Internet: <http://www.aviation-safety.net>
- [10] National Transportation Safety Board (NTSB), US Aircraft accident database, on permanent update, Internet: <http://www.ntsb.gov/aviation>
- [11] W. Härdle and L. Simar, "Applied Multivariate Statistical Analysis", Springer Verlag, New York, 2003
- [12] C. Romesburg, "Cluster analysis for researchers", LuLu Press, North Carolina, 2004

Hybrid System Framework for the Safety Modelling of the In Trail Procedure

Marco Colageo, Antonio Di Francesco
MSc Final-year students
College of Engineering, Center of Excellence DEWS
University of L'Aquila (Italy)

Abstract—The purpose of this paper is to provide a framework based on hybrid systems theory for safety modelling in air traffic management applications. This framework can be used to represent complex multi-agent applications in which a wide set of possible abnormal scenarios has been considered. In the aviation context possible catastrophic events can take place due to an error of a single agent involved in the procedure. It will be shown how the hybrid system framework allows a description and detection of these errors and their effects on the evolution of the procedure. At first it is proposed a description of the *ASEP-In Trail Procedure* which has been chosen to illustrate the methodology. Then, a general view about hybrid systems is proposed in order to explain the mathematical environment. Once basic concepts have been introduced, the hybrid model of the *ASEP-ITP* is explained and the concept of critical observability is introduced. Finally, an hybrid observer is proposed in order to detect unsafe situations associated with the hybrid system evolution.

Index Terms—Safety Modelling, Hybrid Systems, Critical Observability, Air Traffic Management, In Trail Procedure.

I. INTRODUCTION

THE volume of air traffic in the oceanic airspace is quickly increasing inducing the necessity of an improved efficiency in the management of the air traffic flows along these routes. The new procedures that are developed to satisfy this necessity have to increase capacity without affecting safety. This has to be proved using advanced methods. The complexity of the safety analysis of new procedures comes from the specific structure of the environment on which they are applied. The main aspect to be considered is that operations are the result of interactions between many entities of various types and at multiple locations. Furthermore the air traffic management systems are characterized by a mixed environment with human-controlled and computer-controlled subsystems the behaviors of which evolve following completely different logics that cannot be represented using the same class of mathematical models. This complexity can easily be modelled by means of agents in the context of hybrid system theory. Each decision taken by a single agent, either human operator or computer aid, influences the actions of all other agents involved. An hazardous decision induced by a wrong situational awareness can then be reflected into a catastrophic event. When modelling this kind of multi-agent systems all the decision making processes of each agent and their interactions have to be taken into account in order to

identify non-nominal situations and act accordingly to prevent them to evolve into accidents.

Up to now the methodologies used for safety analysis can be classified in three main categories which reflect the temporal evolution of the complexity of airborne scenarios. As proposed in [5], [6] these categories are Sequential Modelling, Epidemiological Modelling and Systemic Accident Modelling: the *Sequential modelling* represents the accident as the outcome of a series of individual steps that occur in a given and (in principle) predictable order, using hierarchies such as the event tree or networks (Critical Path models or Petri networks); the *Epidemiological modelling* describes accidents as the outcome of a combination of manifest and latent factors that happen to exist together in space and time; the *Systemic accident modelling* considers accidents as something that must be expected. Systemic models have their roots in control theory and emphasize the need to base accident analysis on an understanding of the functional characteristics of the system rather than on assumptions or hypotheses about internal mechanisms or cause-effect chains. Systemic models deliberately try to avoid a description of an accident as a sequential or ordered relation among individual events or even as a concatenation of latent conditions.

In this paper we propose to apply a new methodology for safety modelling that has been developed in [3], [7], [9], [10]. This methodology is based on hybrid systems theory that provides a powerful framework to develop multi-agents models. Using this methodology it is possible to link the changes of the physical systems behaviour with the actions made by each agent. These actions can be right decisions taken by human operators, like pilots and controllers, but also decisions due to situational awareness errors. In this context each decision can represent an instantaneous change inside the continuous dynamics of an agent. Using hybrid model it is possible to describe the behaviour of single agent by means of discrete states. Different continuous dynamics that are associated with each discrete state and represent different aspects of the behaviour of the agent; the decisions taken by the agent and by the other agents involved generate the switches between the different discrete states. In this way, a complex behaviour of an agent can be suitably represented with simplified dynamics whose descriptive power is enhanced using the event-driven discrete systems, without making use of a more complex mathematical model. Once all the agents have been modelled, the behaviour of the whole system can

be analyzed by following the evolution of each agent and, at the same time, their interactions. In this way non-nominal and abnormal situations can be identified and subsequently inserted in the model as an additional state.

The paper is organized as follows. In Section 2 a description of the *In Trail Procedure (ITP)* application which has been chosen to illustrate the hybrid system framework is proposed. In Section 3, the hybrid model of the airborne procedure is presented explaining how it describes the procedure's steps. In Section 4, the hybrid observability problem is introduced and a hybrid observer is proposed. Section 5 provides some concluding remarks.

II. DESCRIPTION OF THE IN TRAIL PROCEDURE

The In Trail Procedure (ITP) is part of the *Airborne Separation Assistance Systems (ASAS)* area. ASAS embraces the goal of improving flight management by introducing a stronger interaction between pilots and controllers. The In Trail Procedure (ITP) here considered is envisioned as an *Airborne Separation (ASEP) Application* which is one of the four ASAS application categories. ASEP applications involve the transfer of responsibilities for the separation from the controller to the flight crew during the execution of the procedure. This can happen when the flight crew does have the most appropriate surveillance equipments (i.e. ADS-B and ASAS equipment) and is therefore able to monitor separation and act if necessary.

The ASEP-ITP [1], [2] described hereafter is a procedure that aims at improving flight efficiency along oceanic routes where procedural control is performed. The procedure provides a safe and practical method for air traffic controller to approve, and flight crew to conduct, climb and descent through different flight levels with less stringent applicability conditions than today's operations.

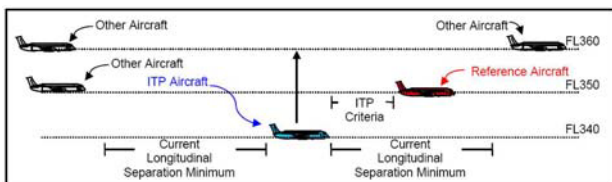


Fig. 1. Example of ITP geometry

A. ITP Criteria

The ASEP-ITP allows climb or descent through only one flight level for a maximum of 2000 feet in RVSM airspace (and 4000 feet in non-RVSM) and the ITP speed/distance criteria are designed so that under nominal conditions the proposed 5NM separation minimum is preserved throughout the ITP manoeuvre. The proposed ITP speed/distance criteria are the following:

- initiation ITP distance of no less than 10 NM and positive ground speed differential of no more that 20 kts, or
- ITP distance of no less than 15 NM and positive ground speed differential of no more that 30 kts.

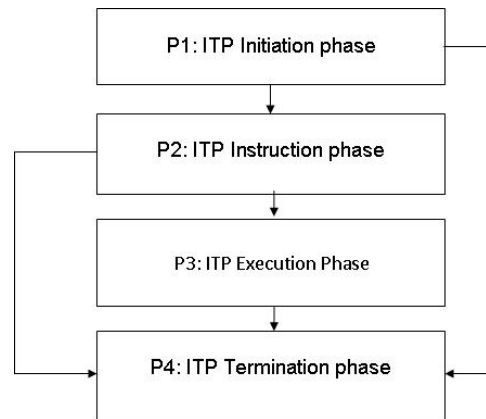


Fig. 2. ASEP-ITP phases diagram

The ITP encompasses a set of six vertical geometries: leading climb (as shown in Figure 1), leading descent, following climb, following descent, combined leading-following climb and combined leading-following descent. These geometries are designed on the basis of the relative position of the ITP aircraft and one or two reference aircraft.

The initiation criteria (*ITP speed/distance criteria*) that are necessary to start an ITP procedure are designed so that the estimated distance between the airliner which performs the climb or descent (*ITP aircraft*) and one or two ADS-B equipped aircraft (*reference aircraft*) in the surrounding area should get no closer than the *ITP separation minimum* of 5 NM until vertical separation is again achieved. These ITP speed/distance criteria are based on combinations of relative speed and relative distance values between the ITP aircraft and the reference aircraft are necessary conditions which have to be verified by the flight crew before requesting an ITP manoeuvre to the air traffic controller (ATC).

The ITP aircraft must maintain a minimum 300 ft/min of climb or descent and constant cruise Mach number throughout the ITP manoeuvre. The reference aircraft must be non-maneuvring and it is not expected to manoeuvre during the ITP. Given these conditions, it can be shown that a 4000 ft flight level change would result in a reduction in the initial distance of 4.5 NM assuming a positive ground speed differential of 20 kts. To ensure that the ITP separation minimum of 5NM will be guaranteed during the flight level change under these conditions, the initial distance between the aircraft must exceed 9.5 NM. So using 10 NM of initial distance the separation minimum is guaranteed. In the same way it could be proved that with positive ground speed differential of more than 20 but less than 30 kts, an initial distance of 15 NM ensures that ITP separation minimum is respected.

A compact view of the ASEP-ITP phases is illustrated in Figure 2, and is now described.

B. ITP Initiation phase

The decision to request an ITP rather than a standard flight level change will typically be based on a number of

factors outside the scope of the ITP application, such as crew preference and judgment, the magnitude of the desired flight level change, and any other information available to the crew about the flight's progress and proximate traffic situation.

Once the flight crew has decided to consider requesting an ITP, the flight crew proceeds through the following steps to formulate and initiate the request:

- 1) Identification of ITP flight levels
 - The crew identifies a requested flight level, which is a flight level above (for a climb) or below (for a descent) one flight level and that is no more than 4000 ft from the initial flight level.
- 2) Checking ITP aircraft Performance by the crew:
 - The ITP aircraft is capable of performing a rate of climb or descent of at least 300 fpm at the assigned Mach number to the requested flight level.
 - The ITP aircraft is not expected to manoeuvre except for a climb or descent or a change of course to remain on their clearance.
- 3) Identification of reference aircraft The crew selects as reference aircraft up to two potentially blocking aircraft which meet the following criteria:
 - The ITP aircraft has the same direction with potentially blocking aircraft.
 - Qualified ADS-B data are available from potentially blocking aircraft.
 - The ITP speed/distance criteria are met with potentially blocking aircraft.
- 4) ITP Request
 - If the ITP criteria are met, the ITP aircraft crew requests the ITP, using the required ITP phraseology which provides the controller with the requested ITP flight level change geometry (i.e., leading or following), the ITP distance and the flight ID of reference aircraft.

C. ITP Instruction Phase

- 1) Issue of ITP Clearance by controller ATC determines if standard separation will be met with all aircraft at the requested flight level and at all flight levels between the ITP aircraft's initial flight level and requested flight level. If so, a standard (non-ITP) flight level change clearance can be issued. *If not*,
 - Determine that the ITP request message format is correct and that the flight crew has correctly identified the reference aircraft at the intervening flight level.
 - Determine that standard separation will be met with other aircraft (i.e., all but the reference aircraft) at the requested flight level and at all flight levels between the ITP aircraft's initial Flight Level and requested flight level.
 - Determine that the ITP aircraft is not a reference aircraft in another ITP clearance;
 - Determine that the ITP aircraft and the reference aircraft are on the same track.

- Determine that the reference aircraft are non-maneuvring and not expected to manoeuvre during the ITP. A change of course (only) to remain on the same identical Track as the ITP aircraft would not be considered a manoeuvre. The controller will not issue an ITP clearance if a reference aircraft is in the process of a manoeuvre or expected to manoeuvre.
- Determine that the positive mach differential is no greater than 0.03 Mach.

Based on the ITP aircraft's request and the controller's determination of the previous six conditions, the controller would issue the ITP clearance.

2) ITP Crew Re-Assessment

- After the ITP clearance is issued, the flight crew of the ITP aircraft must again determine that the ITP criteria continue to be met with respect to the reference aircraft immediately before initiating the climb or descent. If the ITP criteria are no longer met, the crew refuses the clearance and remains at the initial flight level.

D. ITP Execution Phase

1) ITP Aircraft Crew Tasks during the ITP Manoeuvre

- As after a standard climb or descent clearance, the crew must initiate the ITP without delay after receipt of the clearance. Note that the crew re-assessment should not cause an undue delay in the initiation of this manoeuvre.
- The crew must maintain the original cruise Mach number during the climb or descent.
- The ITP aircraft must maintain a minimum 300 fpm climb or descent rate, or the minimum rate required by regulation, whichever greater, throughout the ITP manoeuvre.
- The ITP aircraft crew shall monitor the ITP distance to the reference aircraft during the climb or descent. The crew monitors the ASAS equipment indicating the range of the blocking aircraft. If the separation minimum is predicted to be violated a temporary speed change is allowed.
- The ITP flight crew reports the establishment at the new flight level.
- If the ITP cannot be successfully completed as cleared once the climb or descent has been initiated, an abnormal termination occurs. ATC must be notified immediately when this condition occurs.

2) Controller Tasks during the ITP Manoeuvre

- The controller will not issue any manoeuvre clearance to the reference aircraft until the ITP Aircraft reports establishment at the new flight level or the ITP is abnormally terminated.

E. ITP Termination Phase

- 1) The ITP is completed when the ITP flight crew reports established at the new flight level.

- 2) If the ITP aircraft cannot successfully complete the ITP once the climb or descent has been initiated, an abnormal termination occurs.

III. HYBRID MODEL OF THE ITP PROCEDURE

In this section the hybrid model of the ASEP-ITP is proposed.

A. Preliminaries on Hybrid System Theory

The following description provides a general view of the hybrid systems (i.e [3], [4]). Thus, only the basic definitions are presented in order to facilitate the understanding of the ITP hybrid model proposed.

Definition 1. (Non Deterministic Hybrid System [3])

A hybrid system is a tuple $H = (Q \times X, Q_0 \times X_0, U, Y, \varepsilon, E, \Psi, \eta, Inv, G, R)$ such that:

- $Q = \{q_1, q_2, \dots, q_N\}$ is a set of **discrete states**.
- $X \in \mathbb{R}^n$ is a set of **continuous states**.
- $Q_0 \subseteq Q$ is a set of **initial discrete states**.
- $X_0 \subseteq X$ is a set of **initial continuous states**.
- $U \subseteq \mathbb{R}^m$ is a set of **continuous control input**.
- $Y \subseteq \mathbb{R}^p$ is the set of **continuous observable output**.
- $\{\varepsilon_q\}_{q \in Q}$ associates to each discrete state $q \in Q$ the continuous time-invariant dynamics $\varepsilon_q : \dot{x} = F_q(x)$ with output $y = g_q(x)$.
- $E \subseteq Q \times Q$ is a **collection of edges**, where each edge $e \in E$ is a ordered pair of discrete states, the first component of which is known as source and is denoted by $s(e)$, while the second is the target and is denoted by $t(e)$.
- Ψ is the finite set of discrete output symbols $\varepsilon, \psi_1, \psi_2, \dots, \psi_r$ where ε is the empty string that corresponds to unobservable output.
- $\eta : E \rightarrow \Psi$ is the output function, that associates to each edge a discrete output symbol.
- $\{Inv_q\}_{q \in Q}$ associates to each discrete state $q \in Q$ an invariant set $Inv_q \subseteq X$.
- $\{G_e\}_{e \in E}$ associates to each edge $e \in E$ a guard set $G_e \subseteq Inv_{s(e)}$.
- $\{R_e\}_{e \in E}$ associates to each edge $e \in E$ a reset map $R_e : Inv_{s(e)} \rightarrow 2^{Inv_{t(e)}}$, from $Inv_{s(e)} \subseteq X$ to the power set (i.e. the set of all the subsets) of $Inv_{t(e)}$.

The system so defined can be compactly described using the graph depicted in the Figure 3. It should be noticed that this representation contains all the mathematical attributes introduced in the definition 1. The evolution of an Hybrid System can be synthesized in this way: supposed $(q_1, x_0) \in Init$ the initial hybrid system state, the continuous state x evolves according to the continuous dynamic \dot{x} with $x(0) = x_{1,0}$, as long as $x \in Inv_{q_1}$, whereas the discrete state q remains constant $q(t) = q_1$. If at some point, state reaches guard G_{e1} then the discrete transition from q_1 to q_2 is enable. In this situation, when the continuous state leaves the Inv_{q_1} the discrete transition is forced, and the state x changes value according to the reset map R_{e1} . Next the process is repeated starting from $(q_2, x_{2,0})$.

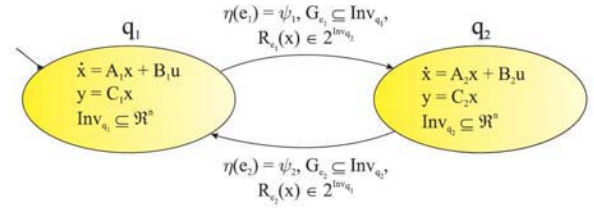


Fig. 3. Non-deterministic Hybrid System

A particular class of non-deterministic hybrid systems is represented by the *rectangular automata*. This subclass is introduced here and is the one that will be used in the hybrid model of ASEP-ITP. Considered the space \mathbb{R}^n with variables x_1, \dots, x_n , a rectangular set B of dimension n is the product of n intervals $B_i \subseteq \mathbb{R}$ of the real line, where each B_i is a bounded or unbounded interval.

Definition 2. (Rectangular Automaton [4]): A rectangular automaton is a hybrid system, as defined in Definition 1, that also satisfies the following constraints:

- For every discrete state $q \in Q$, the set of initial continuous states $X_0 \subseteq X$ and the invariant set $Inv_q \subseteq X$ are rectangular sets.
- For every discrete state $q \in Q$, there is a rectangular set B^q such that the continuous time invariant dynamics $\varepsilon_q : \dot{x} = F_q(x) \in B^q$ for all $x \in \mathbb{R}^n$.
- For every edge $e \in E$, the set $Guard_e$ is a rectangular set, and there is a rectangular set B^e and a subset $J^e \subseteq \{1, \dots, n\}$ such that for all $x \in \mathbb{R}^n$ the reset map is $R_e = \{(x_1, \dots, x_n) \in \mathbb{R}^n \mid \text{for all } 1 \leq i \leq n, \text{ if } i \in J^e \text{ then } x_i \in B_i^e \text{ else } x_i = x_i\}$.

Therefore, in a rectangular automaton, the derivative of each variable stays between two fixed bounds, which can be different in different discrete states. Then in each discrete state $q \in Q$ the continuous dynamics can be defined as $\dot{x}_i \in B_i^q \subseteq B^q$ for all $1 \leq i \leq n$. With each discrete jump across an edge e , the value of the variable x_i either does not change if $i \notin J^e$, or resets non-deterministically to a new value within some fixed constant interval $B_i^e \subseteq B^e$ if $i \in J^e$.

The hybrid model proposed below is slightly different from the one of the Definition 2. This model also embeds a set Σ of discrete input signals, and each edge $e \in E$ is associated to a symbol $\sigma \in \Sigma$ that triggers the discrete transition between the states linked by e . These inputs can be considered as discrete disturbance or control inputs which model the communication among the agents.

B. Assumptions

The ASEP-ITP can be decomposed in various subsystems representing the agents involved in the procedure, each with hybrid dynamics modelling its specific operations. It should be remarked that to exploit the descriptive power of hybrid system each agent must be considered by itself and subsequently the effects of their actions on the dynamics of other agents can be considered merging the models so obtained.

The agents considered are:

- Air crew flying of ITP aircraft
- Oceanic controller

The approach used for selecting the agents does not provide the modelling of the reference aircraft as an agent. The main reason is that the flight crew of the reference aircraft does not have the awareness of existence of an ITP manoeuvre in which it is involved. In fact, there is no communication between the controller or the flight crew of the ITP aircraft and the flight crew of the reference aircraft. Furthermore any hazardous actions of the reference aircraft can be considered inside the hybrid dynamics of other agents.

The model proposed considers the simplest case of ASEP-ITP execution where the ITP aircraft requests a climb through one flight level, with only one leading reference aircraft involved and without other blocking aircraft. Furthermore, no wind is assumed. The continuous dynamics used in this approach are intentionally simplified. In fact due to the configuration of the traffic flows in the oceanic airspace (i.e. organized parallel tracks system) it is possible to focus on longitudinal and vertical dynamics without considering the lateral dynamics. Moreover, for safety analysis of this ITP, using a more complicated model that considers a complete dynamic of the aircraft would not be relevant.

C. Pilot flying of ITP aircraft Agent

Before explaining the model, the following variables are introduced:

- 1) z_i initial flight level of the aircraft
- 2) z_f requested flight level of the ITP aircraft
- 3) $v_{x,min}$ minimal ground speed of the ITP aircraft
- 4) $v_{x,max}$ maximal ground speed of the ITP aircraft
- 5) $v_{z,max}$ maximal vertical speed of the ITP aircraft
- 6) x_r longitudinal position of the reference aircraft
- 7) v_{rx} the ground speed of the reference aircraft
- 8) M_i assigned Mach number for the ITP aircraft
- 9) a speed of sound, assumed as a constant value

Furthermore the following interesting areas of the airspace can be identified:

- 1) A safe region in which the ITP aircraft performing the ITP manoeuvre respects the ITP minimum distance separation. The safe zone is defined as $\Omega_S = \{(x, z) : x \in [-\infty, x_r - 5], z \in (z_i, z_f)\}$.
- 2) Thus, an unsafe zone can be defined as follows: $\Omega_U = \{(x, z) : x \in [x_r - 5, +\infty], z \in (z_i, z_f)\}$.

The agent \mathcal{H}_p *Pilot Flying of ITP Aircraft* can be described using a model based on Definition 2. The following are the objects of the system:

- $Q = \{q_1, q_2, q_3, q_4, q_5, q_6, q_7, q_8, q_9, q_{10}, q_{11}, q_{12}\}$ is the set of discrete states each associated with a node inside the graph depicted in Figure 4;
- $X = \{(x, z) : x \in \mathbb{R}_0^+, z \in \mathbb{R}^+\}$, with $\mathbb{R}_0^+ = \mathbb{R}^+ \cup \{0\}$ is the set of the continuous state values where x represents the longitudinal position of the aircraft expressed in nautical miles, z the altitude of the aircraft expressed in hundred of feet (i.e. flight level inside the International Standard Atmosphere).

- The initial discrete state is q_1 ;
- The continuous dynamics are the followings:
 - $F_{q_1}(x, z) = \{\dot{x} = M_i a, \dot{z} = 0\}$
 - $F_{q_7}(x, z) = \{\dot{x} \in [v_{x,min}, v_{rx} + 30], \dot{z} \in [300, v_{z,max}]\}$
 - $F_{q_8}(x, z) = \{\dot{x} \in [v_{x,min}, v_{x,max}], \dot{z} \in [300, v_{z,max}]\}$
 - $F_{q_9}(x, z) = \{\dot{x} \in [v_{x,min}, v_{x,max}], \dot{z} \in [0, 300]\}$
 - $F_{q_{10}}(x, z) = \{\dot{x} = M a, \dot{z} = 0, M \neq M_i\}$
 - $F_{q_i}(x, z) = F_{q_1}(x, z)$ for $i = 2, 3, 4, 5, 6, 11, 12$
- $\Sigma = \{\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5, \sigma_6, \sigma_7, \sigma_8, \sigma_9\}$ is the set of discrete inputs, where σ_1 means decision to make an ITP, σ_2 represents the reassessment failed, σ_3 represents the ITP criteria are not verified, σ_4 means the ITP criteria verified, σ_5 represents the clearance denied, σ_6 means the clearance issued, σ_7 means detection of an abnormal event, $\sigma_8 = \varepsilon$, σ_9 represents a situational awareness error;
- $\Psi = \{\psi_1, \psi_2, \psi_3, \psi_4, \psi_5, \psi_6\} \cup \{\varepsilon\}$ is the set of discrete outputs, where ψ_1 means the clearance rejected by the crew, ψ_2 represents the clearance request, ψ_3 represents the clearance accepted by the crew, ψ_4 means the abnormal termination communication by the crew to the controller, ψ_5 means the report established at the new flight level, ψ_6 represents the confirmation by the crew of the reception of the denied clearance;
- $E \subseteq Q \times Q$ is the set of transitions given by the graph depicted in Figure 4. A label $\sigma \in \Sigma$ is associated to each edge as shown in Figure 4;
- $\eta : E \rightarrow \Psi$ the discrete output function defined by the graph depicted in Figure 4;
- The domains of the discrete states are the following:
 - $Inv_{q_1} = \{(x, z) : x \in \mathbb{R}_0^+, z = z_i\}$
 - $Inv_{q_7} = \{(x, z) \in \Omega_S\}$
 - $Inv_{q_8} = \{(x, z) \in \Omega_S \cup \Omega_U\}$
 - $Inv_{q_9} = Inv_{q_8}$
 - $Inv_{q_{10}} = \{(x, z) : x \in \mathbb{R}_0^+, z = z_f\}$
 - $Inv_{q_{12}} = Inv_{q_{10}}$
 - $Inv_{q_i} = Inv_{q_1}$ for $i = 2, 3, 4, 5, 6, 11$
- The guards are the empty set for all the discrete transitions excepted for:
 - $G(q_7, q_{12}) = \{x \in \Omega_S, z = z_f\}$
 - $G(q_8, q_{12}) = G(q_9, q_{12}) = \{(x, y) : x \in \mathbb{R}^+\}$
- The reset function is always the identity function excepted for:
 - $R(q_7, q_{11}) = \{x_{q_{11}} = x_{q_7}, z_{q_{11}} = z_i\}$
 - $R(q_8, q_{11}) = R(q_9, q_{11}) = R(q_7, q_{11})$

The direct graph of this hybrid model is shown in the Figure 4. The evolution of an ITP could be followed on the graph in this way. Initially the aircraft is in the *cruise* (i.e. discrete

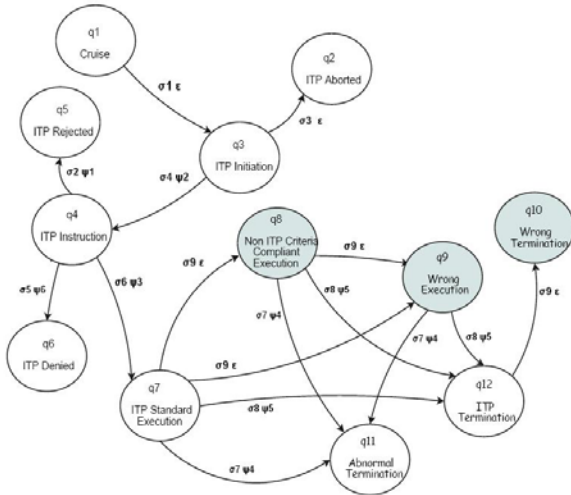


Fig. 4. Direct graph of pilot flying of ITP aircraft agent. The shadowed locations are the critical states

state q_1) phase. When the flight crew performs an ASEP-ITP manoeuvre, the discrete transition to the *ITP Initiation* (q_3) takes place. Then, the flight crew has to verify if the ITP speed/distance criteria are met. If the criteria are not satisfied, the flight crew aborts the ITP initiation phase and there exists the discrete transition to the *ITP Aborted* (q_2) state. If the criteria are verified, the flight crew requests the clearance to the ATC and the discrete transition to *ITP instruction* (q_4) takes place. In this phase, if the clearance is denied by the ATC the ITP is not executed and the discrete switch to *ITP denied* (q_6) takes place. When the clearance is issued, the flight crew has to recheck the ITP speed/distance criteria in order to evaluate if the criteria are still met. If the criteria are not met, the flight crew rejects the clearance and this is represented by the discrete switch to the *ITP rejected* (q_5) state. If the ITP criteria are still met, the flight crew accepts the clearance, communicates it to the ATC and the discrete state *ITP Instruction* (q_4) changes to *Standard ITP execution* (q_7) state. It can happen that during the first or the second verification of ITP speed/distance criteria the flight crew makes an error due to a wrong situational awareness. This scenario is modelled using an unobservable transition from the *Standard ITP execution* (q_7) to *Non ITP criteria compliant execution* (q_8); this transition is not detectable because the flight crew does not know that an error occurred. From both these discrete states, it is possible to jump to the *Wrong execution* (q_9) state, which models the situation where, again due to a situational awareness error, the flight crew is performing the manoeuvre without compliance with the performance criteria (i.e. vertical speed more than 300 ft/min and Mach number constant). Starting from the discrete states *Standard ITP execution* (q_7), *Non ITP criteria compliant execution* (q_8) and *Wrong execution* (q_9), the manoeuvre is terminated in two different ways. In the first case the flight crew detects an abnormal event and the manoeuvre is terminated in an abnormal mode. The flight crew communicates to the ATC the abnormal termination and the discrete transition takes place to *Abnormal Termination*

(q_{11}). In the second case, the ITP terminates in the correct way; the flight crew communicates to the ATC established in the requested flight level and the discrete state changes to *ITP termination* (q_{12}) state. From this discrete state a situational awareness error can bring to an unsafe situation. In fact, if the flight crew has changed the Mach number during the manoeuvre for safety reasons and it does not revert to the assigned Mach number when the requested flight level is reached, an unobservable transition to *Wrong termination* (q_{10}) takes place.

D. Oceanic controller Agent

The hybrid model of the oceanic controller agent does not include continuous dynamics and all the discrete transitions take place because of the occurrence of a discrete input. Thus, this hybrid model can be considered as a *discrete event system*. The objects of the model are the followings:

- $Q = \{q_1, q_2, q_3, q_4\}$ is the set of discrete states which are associated with the corresponding vertices of the graph shown in Figure 5.
- The initial discrete state is q_1 .
- $\Sigma = \{\sigma_1, \sigma_2, \sigma_3, \sigma_4\}$ is the set of discrete inputs, where σ_1 represents the request of an ITP, σ_2 means the abnormal termination communication, σ_3 means a situational awareness error and σ_4 represents the communication of ITP terminated.
- $\Psi = \{\psi_1, \psi_2, \psi_3, \psi_4\} \cup \{\epsilon\}$ is the set of discrete outputs where ψ_1 means the clearance issued, ψ_2 represents the ITP request denied, ψ_3 represents the abnormal termination confirmation, ψ_4 means the confirmation of a standard ITP termination.
- $E \subseteq Q \times Q$ is the set of transitions given by the graph depicted in Figure 5. A label $\sigma \in \Sigma$ is associated to each edge as shown in Figure 5.
- $\eta : E \rightarrow \Psi$ the discrete output function defined by the graph depicted in Figure 5.

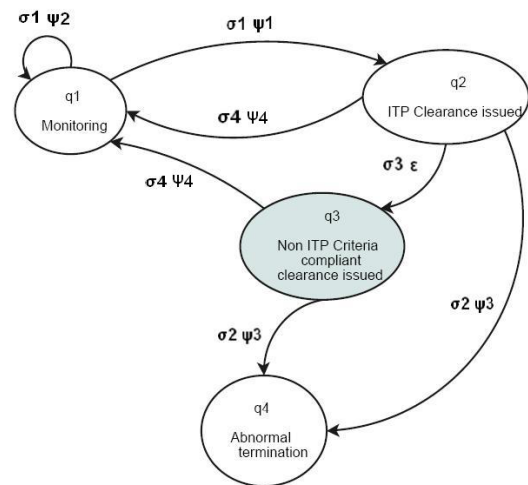


Fig. 5. Discrete graph of the Oceanic Controller agent

At the beginning of the ITP application, the discrete state is *Monitoring* (i.e. discrete state q_1), which models the usual monitoring of the controller. When the ITP request from the flight crew is received, the controller can decide to issue or deny the clearance on the basis of the verification of some criteria. If the request is not accepted, the controller communicates the deny to the flight crew and the discrete state does not change; if the clearance is issued, the discrete transition to the *ITP clearance issued* (q_2) takes place. During the verification of the criteria, the controller can make an error due to a wrong situational awareness. In this case, an unobservable transition from *ITP clearance issued* (q_2) to *Non ITP criteria compliant clearance issued* (q_3) takes place. From both *ITP clearance issued* (q_2) to *Non ITP criteria compliant clearance issued* (q_3) it is possible to jump to *Abnormal Termination* (q_4), when the flight crew communicates the occurrence of an abnormal event; otherwise, if the controller receives the confirmation of a standard ITP termination by the flight crew, the discrete state is reverted to the initial *Monitoring* (q_1) state.

IV. HYBRID OBSERVER OF THE ITP AGENTS

The hybrid model presented in the previous section describes in detail the procedure and identifies safe and unsafe scenarios. For safety analysis it is important to detect, instantaneously or with an acceptable delay, the discrete states of the hybrid model associated with hazardous situations. This issue represents a typical discrete observability problem of hybrid systems. The idea is to design a finite state machine, known as an "observer", which is able to discriminate the current discrete state using only the observable output generated by the transitions.

In the literature, several definitions of observability for hybrid systems have been proposed. As defined in [7], [11], an hybrid system is K -current-state observable if any discrete location of the hybrid system can be identified by the use of the discrete outputs, after a finite number $K > 0$ of discrete transitions. It should be noticed that this notion cannot allow for the immediate detection of critical states (i.e. $K = 0$). An alternative definition is presented in [8] and requires that all the states of the system, both safe and unsafe, have to be immediately observable. For safety analysis it is sufficient to consider the observability only of the set of the critical states instead of the whole discrete state space. This approach is considered in [9] where *critical observability* is proposed.

The next section presents the hybrid observer designed for the Pilot Flying of ITP Aircraft Agent. This observer checks for the critical observability of the agent, assuming $Q_c = \{q_8, q_9, q_{10}\}$ as set of critical states. The same approach can be used to design the observer of the other agents involved in the ITP procedure.

A. Hybrid Observer of Pilot flying of ITP aircraft Agent

The algorithm presented in [9] provides a method to design the observer \mathcal{O}_p of the hybrid system \mathcal{H}_p starting from the direct graph associated to the system. In this way, the observer

obtained is a finite state machine $\mathcal{O}_p = (\hat{Q}, \hat{q}_0, \hat{Q}_m, \hat{\Psi}, \hat{E}, \hat{\eta})$ defined as follows:

- $\hat{Q} = \{\{q_1, q_2, q_3\}, \{q_4\}, \{q_5\}, \{q_6\}, \{q_7, q_8, q_9\}, \{q_{11}\}, \{q_{10}, q_{12}\}\}$, where q_i are the discrete states of \mathcal{H}_p .
- The initial state is $\hat{q}_0 = \{q_1, q_2, q_3\}$.
- The set of final or accepting states is $\hat{Q}_m = \{\{q_7, q_8, q_9\}, \{q_{12}, q_{10}\}\}$
- The set of discrete inputs $\hat{\Psi} = \{\psi_1, \psi_2, \psi_3, \psi_4, \psi_5, \psi_6\}$, where each ψ_i represents a discrete output of the hybrid system \mathcal{H}_p .
- The set of transitions $\hat{E} \subseteq \hat{Q} \times \hat{Q}$ given by the graph in Figure 6.
- The discrete output function $\hat{\eta}$ defined as the identity for all the edges.

Roughly speaking, assuming ψ_i a discrete output of \mathcal{H}_p and \hat{q}_i the current state of \mathcal{O}_p , each state of the observer is designed by grouping together the discrete states q_i which can be reached from all the states $q_i \in \hat{q}_i$ by a transition labeled with ψ_i , and all the discrete states q_i which can be reached from the first ones by an unobservable transition. The discrete outputs of the hybrid model now are used to trigger the transitions of the observer (i.e. they are considered discrete inputs of \mathcal{O}_p). For this reasons, the discrete states of the observer are defined as sets of q_i . The graph of the observer \mathcal{O}_p is depicted in Figure 6: the initial state groups the initial states of \mathcal{H}_p (i.e. q_1), and the states that can be reached from q_1 through transitions with unobservable output (i.e. q_2, q_3). As the first observable output (i.e. ψ_2) is available, the associated transition of \mathcal{O}_p (i.e. from $\{q_1, q_2, q_3\}$ to $\{q_4\}$) takes place. Then, each time that a new observable output is generated, a new transition of the observer is triggered according to the graph.

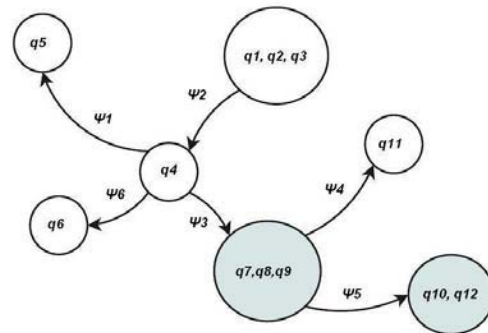


Fig. 6. The observer \mathcal{O}_p

It should be noticed that the observer \mathcal{O}_p cannot be used to identify immediately the critical discrete states. In fact, there exists two discrete states of the observer where both safe and unsafe states q_i coexist. However, critical observability can be recovered by generating a set of extra output signals which can be used to distinguish when the system reaches a critical state. These signals can be generated with a non-zero time δ from the

continuous inputs, outputs and dynamics. The generating time of the extra outputs creates a delayed detection of the critical states. This kind of observability is known as δ -observability (i.e. [3]). The observer \tilde{O}_p defined after the generation of the extra-output is able to discriminate the critical states with a delay δ which has to be acceptable. \tilde{O}_p is shown in Figure 7: within this graph, the new transitions triggered by the extra output signal (i.e. from $\{q_7, q_8, q_9\}$ to $\{q_8, q_9\}$ and $\{q_{10}, q_{12}\}$ to $\{q_{10}\}$) allows to discriminate the unsafe states.

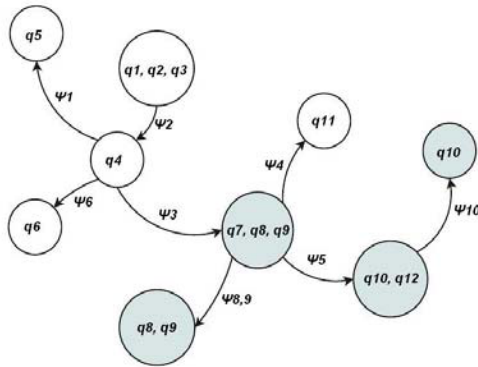


Fig. 7. The observer \tilde{O}_p with delay δ . New outputs are $\psi_{8,9}$ and ψ_{10}

V. CONCLUSION

In this paper the hybrid system framework for safety modelling in air traffic management applications has been discussed. The need to develop new sophisticated modelling methodologies originates from new challenges in safety and from the increasing of inherent complexity in the airborne procedures. A specific procedure, the ASEP-ITP, has been investigated to show how this framework can be used to represent a complex multi-agent application in which a wide set of possible abnormal scenarios may happen. In the aviation context, possible catastrophic events can take place due to an error of a single agent involved in the procedure. It has been shown how the hybrid system framework allows the description and the detection of these errors and the understanding of their effects on the evolution of the procedure. The observers which have been proposed here will allow to perform a formal safety analysis, which investigates unforeseen circumstances originated by the interaction of the hybrid agents.

ACKNOWLEDGMENTS

This work was partially supported by European Commission under STREP project n.TREN/07/FP6AE/S07.71574/037180 IFLY. It was carried out within the Erasmus Student Placement program, involving the College of Engineering, Center of Excellence DEWS, of University of L'Aquila (Italy) and DSNA-DTI-R&D of Toulouse (France). The authors are grateful to Pascal Lezaud, Thierry Miquel (DSNA-DTI-R&D), Maria Domenica Di Benedetto and Alessandro D'Innocenzo (Center of Excellence DEWS) for their support and suggestions.

REFERENCES

- [1] "D6.1b Qualitative Risk Assessment for ASEP-ITP", ASSTAR Projects, 01 February 2007 v.1.0
- [2] "In-Trail Procedure in Procedural Airspace (ATSA-ITP) Application Description", Package I Requirements Focus Group, 21 June 2007 v8.0
- [3] A. D'Innocenzo, PhD november 2006, "Observability and Temporal Properties of Hybrid Systems: Analysis and Verification"
- [4] R. Alur, T.A. Henzinger, G. Laferriere and G.J. Pappas, "Discrete Abstractions of Hybrid Systems", Proceedings of the IEEE, vol. 88, NO. 7, July 2000
- [5] A. Dijkstra, "Resilience Engineering and Safety Management Systems in Aviation", KLM Royal Dutch Airlines / TU Delft
- [6] E. Hollnagel, O. Goteman, "The Functional Resonance Accident Model"
- [7] E. De Santis, M.D. Di Benedetto, S. Di Gennaro, G. Pola, "Hybrid observer design methodology", Public deliverable D7.2, project IST-2001-32460 HYBRIDGE, August 19, 2003, <http://www.nlr.nl/public/hosted-sites/hybridge>
- [8] M. Oishi, I. Hwang, C. Tomlin, "Immediate Observability of Discrete Event Systems with Application to User-Interface Design", Proceedings of the 42nd IEEE conference on decision and control, Maui, Hawaii, USA, pag. 2665-2672, 2003.
- [9] M.D. Di Benedetto, S. Di Gennaro, A. D'Innocenzo, "Error Detection Within a Specific Time Horizon", public deliverable D7.4, project IST-2001-32460 HYBRIDGE, January 26, 2005, <http://www.nlr.nl/public/hosted-sites/hybridge>
- [10] M. D. Di Benedetto, S. Di Gennaro, A. D'Innocenzo, "Situation Awareness Error Detection", Public Deliverable D7.3, Project IST200132460 HYBRIDGE, August 18, 2004, <http://www.nlr.nl/public/hostedsites/hybridge>.
- [11] A. Balluchi, L. Benvenuti, M. D. Di Benedetto, A. L. Sangiovanni-Vincentelli, "Design of Observers for Hybrid Systems", In Claire J. Tomlin and Mark R. Greenstreet, Editors, Hybrid Systems: Computation and Control, Vol. 2289 of Lecture Notes in Computer Science, pp. 7689, Springer-Verlag, Berlin Heidelberg New York, 2002.

Proactive, Reactive, and Interactive Risk Assessment and Management of URET Implementation in Air Route Traffic Control Centers

Grace Uang*
Parsons Corporation
San Francisco, CA
gracie.uang@gmail.com

Jasenka Rakas
University of California at Berkeley
NEXTOR
Berkeley, CA 94720, USA
jrakas@berkeley.edu

Tatjana Bolic
Venice International University
Isola San Servolo
Venice, Italy
Tatjana.bolic@univiu.org

Abstract

The current trend within air traffic management (ATM), as a part of the Next Generation Air Transportation System (NextGen), is to increase the airspace system capacity to operate in diminishing capacity conditions while improving standards of safety. An extensive body of research exists regarding introducing automation into air traffic control in order to create more flexible and cost-efficient operations. The User Request Evaluation Tool (URET) is a strategic support tool designed to assist controllers with timely detection of conflicts; it offers tools for checking the conflict resolution clearances. This study develops general proactive, reactive, and interactive approaches for the risk assessment and management of the system in order to achieve quality (safety and serviceability) and reliability; it also presents a case study of URET implementation in Air Route Traffic Control Centers in the past ten years. First, the reactive approach, used in URET deployment, is developed, followed by developing the complementary and necessary proactive and interactive approaches. Safety Management and Assessment (SMAS) evaluation is performed for the reactive approach. Findings show that many factors led to cases of URET usage deviating from that provided for in the original design, and for using URET less often than it was originally intended.

Keywords: Air Traffic Control, en route, URET, performance shaping factors, proactive, reactive, interactive, risk assessment and management, SMAS

I. INTRODUCTION

Human error contributes to air traffic management (ATM) incidents in the order of 90% or more [1], compared to the human error contributions in nuclear power (70-90%) or medical (98%) industries. However, ATM is unique because of its highly dynamic and time-critical environment, which is also very cognitive in the nature of its tasks. It is important to recognize that “human error is not, however, the only causal factor involved: incidents or accidents generally happen when several causes—possibly including human error—are combined.” In addition, many of these factors are “latent,” existing dormant within the system long before a major incident occurs. According to [2], some of the frequent latent

contributors to incidents are: complex system design, poor man-machine interfaces, inappropriate work organization, awkward work procedures or policies, reduced or altered communication between human operators.

The ATM system is very reliable considering that every day air traffic controllers handle high numbers of aircraft movements without any incidents. Many redundant components in the system, as well as the structured communications between its players “have generally allowed a graceful recovery from failures, without accidents.” Such errors, due to the system reliability and humans in the loop are “not likely to be damaging to system performance if they can be caught and corrected by error-tolerant systems” [5].

Many studies analyzed the means of meeting growing air traffic demand, as well as introducing automation and surveillance tools into the existing system. However, automation can be a “mixed blessing” and can actually heighten the importance and impact of human error.

Due to pressure on ATC to handle an increasing number of aircraft, the “tolerable error margin both for the pilot and the controller is shrinking as more traffic is packed into already crowded airspace” [6]. The current ATM struggle is to increase the airspace system capacity to operate in lower margin conditions while continuing to improve the current standards of safety.

As a response to the NextGen additional capacity and safety needs, this study addresses the following goals and objectives:

- addresses performance shaping factors relevant to air traffic controllers and provides a research and development foundation for the “next generation” ATC system
- discusses automation issues and introduces the specific automation tool, User-Request Evaluation Tool (URET), its intended use, and its actual use
- develops a reactive approach (Safety Management Assessment System—Braille Chart) for the “failure” of URET to be used in accordance with its designed usage

* The author was a graduate student at the University of California at Berkeley, Dept. of Civil and Environmental Engineering, at the time this study was conducted.

- develops a proactive and interactive approach for applicability of URET deployment in Air Route Traffic Control Centers (ARTCCs)

II. ATC SYSTEM OPERATIONS

Because the presented study focuses on Air Route Traffic Control Centers (ARTCCs), a general description of air traffic controller sector teams is presented below.

A sector controller team ranges from one to three persons, depending on the traffic-induced workload. A two-person controller team most commonly consists of:

1. Radar Controller (R-side) - fully certified controller responsible for maintaining prescribed separation between aircraft under his/her control using radar-displayed information. He issues altitude, heading or airspeed change clearances to accomplish this duty and handles all communication with pilots.
2. Radar Associate Controller (D-side) - either fully certified or developmental controller assists the R-side controller. Duties include flight strip management, coordination with other controllers, and identifying potential conflicts between aircraft that are not yet under the active control to the R-side controller.

According to Bea [7], quality can be defined as the ability to satisfy the requirements of serviceability (use for purpose for conditions), safety (acceptability of risks), compatibility (acceptability of impacts), and durability (freedom from unanticipated degradation). Because the goal of ATC is to facilitate the "safe, orderly and expeditious flow of air traffic" [6], to satisfy goals of both safety and efficiency can, however, be contradictory. Indeed, in the ATC, safety always supersedes efficiency. Given the economic pressures from airline management and the increasing number of airplanes trying to utilize the airspace, it is easy to see that this could become a starting point for compromises in system integrity that could ultimately lead to failures. Airspace capacity maximization needs to take into account the fact that controller's workload increases with the number of aircraft under control, and therefore we should not insist on reducing aircraft separation to the levels where the systems failure recovery is difficult to achieve. Means to increase capacity, which bring about increase in the controller workload, and sometimes need for separation reduction, have been turning to automation for help.

III. AUTOMATION IN ATC

Traditionally, automation is implemented in an attempt to reduce the operator's workload during peak periods of task load. However, this may not always be the actual outcome. For example, automation can reduce mental workload when workload is already low, or increase mental workload when workload is already high, better known as "clumsy automation" [10].

Automation technology in ATC has been advancing over time. Even so, humans are still essential to the system to monitor the system using automation tools, act as controllers, and need to be able to keep the system operating even when

automation fails, which is a major concern in automation implementation. Several studies indicate that operators may require more time to intervene under automation than under manual control because they first need to regain awareness of the state of the system. When operators are actively involved in creating the state of the system (as opposed to passively monitoring automation), they develop a more complete situational awareness of the system state [10], which is of paramount importance in ATC, and the state-of-the-art in many areas of ATC.

User Request Evaluation Tool is one of the automation tools introduced. It was intended to be a strategic support tool for the D-side controller of an en route sector team. With URET, the "D-side controller should be able to help the R-side controller to resolve potential conflicts of aircraft that are not yet under the sector's active control, to check if the clearances the R-side controller is issuing are conflict free, and to better perform other D-side duties" [11]. URET notifies the controller of the potential conflicts by continuously checking current flight plan trajectories for strategic conflicts up to 20 min into the future. It was expected that URET would enable the controller team to handle more aircraft due to the decreased workload resulting from the automation as well as the availability of more accurate information. In addition, URET was expected to help controllers provide more direct routings and better flight profiles to airlines. Currently, URET is installed and in use in all Air Route Traffic Control Centers (ARTCCs). However, URET's actual usage differs from the designed one. For example, a very small percentage of controllers use it to check for the potential conflicts of aircraft not under the active control. On the other hand, they all use the electronic flight strip replacement functionality and a majority of controllers find it very helpful, above all because it reduces their manual workload.

URET deployment and integration into the system took ten years. Many factors contributed to such a long period of integration: ATC system safety, organizational factors (FAA and controller union), automation tool itself, etc. Still, the tool is not being used in a designed way, which can be considered as an implementation failure in terms of serviceability, and reliability as viewed from the engineered systems' reliability point of view. The following text describes the performance shaping factors further used to develop approaches to achieving reliability in the URET implementation management engineered system.

IV. PERFORMANCE SHAPING FACTORS

Performance Shaping Factors (PSF) are "influences that can result in an increase in the mean rates of human errors" and are "useful in helping develop quantification of the potential effects of changes in organization, hardware, procedures, and environments on the base rates of human errors" [7]. PSF can be divided into the following categories: impairments, training, workload, organizational, communications, societal, and environmental.

Impairment of a subject can come from one of four major causes: fatigue, well-being, medical, and drugs. Fatigue is the most studied in the human performance field. Work-rest schedules and shift work address these fatigue issues.

Training can be subdivided into routine task performance, unfamiliar events, and emergency response. The effectiveness of training can greatly impact a controller's performance.

The workload PSF "entails the effects of demands imposed upon the subject by assigned and unassigned tasks." An individual's demand can be subdivided into the following categories: occupation (direct and indirect), regulatory (laws, codes governing the subject's work and/or personal life), societal, and personal [7].

The organizational PSF is a very powerful one, with culture (in particular, risk acceptance) being a dominant PSF. Teamwork both among controllers and between controllers and pilots is critically important for safe and efficient air traffic control [7].

The communications PSF can be subdivided into oral, written, and nonverbal. The effectiveness of communications also depends on shared assumptions, a shared mental model [9], or shared situation awareness between speaker and listener, which is critically important for a safe and efficient ATC

Societal and external environments are both PSFs that are not addressed in this study as they are not as significant or relevant to the ARTCCs as the other PSFs.

With the given background on performance shaping factors for air traffic controllers, the consideration of automation in the future air traffic control system is logical. The next section proceeds with the discussion about the research dedicated to learning about the benefits and costs of automation into the system.

V. METHODOLOGY

This paper focuses on the quality attributes of safety (acceptability of risks), serviceability and reliability. Reliability is defined as the "probability that a given level of quality will be achieved during the design, construction, and operating life-cycle phases of an engineered system [7]. The authors view the system as consisting of seven components (see Table 1).

There are three fundamental approaches to achieving reliability in engineered systems: proactive, reactive, and interactive. These approaches are inter-related, interdependent, interactive, and complimentary. The proactive approach includes measures employed before accidents and incidents; the reactive approach includes measures employed after accidents and incidents, and the interactive approach includes measures employed during the evolution of accidents and incidents. Each of the three approaches has its own strengths and weaknesses. The objective is to "define a combination that can be most effective and efficient in maintaining the desirable and acceptable quality and reliability of systems" [7].

First, we introduce and define the following important terms:

- System = the implementation of URET in ARTCCs sectors (1995-2005) and the organizations involved in its implementation
- Failure = system usage different from the designed one

In the first step, an overview of the existing methodology was conducted based on [11] that involves interviews of Subject Matter Experts (SMEs) (air traffic controllers) at the Oakland center (which did not have URET at the time), and the Indianapolis, Jacksonville, and Washington centers (with URET). The interview results provide useful information about:

1. The specifics of controller team operations, in both URET and non-URET environments
2. Specific uses of URET
3. Center culture

In the second step, data for an in-depth reactive approach are gathered, collecting 3 grades (low bound, most probable, and high bound) for each factor in each of the 7 system components, and an explanation of why the particular grades are chosen.

Further, based on the system (URET implementation in ARTCCs), a reactive approach is developed in order to understand why the system usage differed from its original usage design. This is necessary in order to further develop a proactive and interactive approach (plan and monitor) for a successful implementation of URET.

VI. REACTIVE APPROACH

The reactive approach is based on analysis of the failure of a system, where the system usage differs from the designed one.

Development of the approach relies on gathering all available information on the failure and the life-cycle characteristics of the system. The information should address the following three categories:

- *Initiating events and factors* that may have triggered the accident sequence
- *Propagating events and factors* that may have allowed the accident sequence to escalate and result in the accident, and
- *Contributing events and factors* that may have encouraged the initiating and propagating events.

Rather than considering a particular "accident" or "incident" that many reactive approaches are based on, the considered "incident" in this study is defined as follows: (i) URET is not achieving its original purpose, (ii) URET is not used as much as expected, and (iii) if used, it is used mostly for unintended purposes.

Information gathered in the three mentioned categories should address the seven system components, as well as the life-cycle characteristics and history of the system including design, construction, operation, and maintenance.

Furthermore, information is used in the Safety Management Assessment System (SMAS) protocol. SMAS is intended to be used as a proactive measure to help identify potentially important problems or flaws in systems, and to help determine how these potential problems or flaws might best be remedied. The focus of SMAS is on those human and particularly organizational factors (HOF) that influence the safety of the system. SMAS can be used through all three approaches (proactive, reactive and interactive).

The SMAS evaluation incorporates a qualitative assessment of the factors listed in Table 1 in each of the seven key system components: 1) operators, 2) organizations, 3)

procedures, 4) equipment, 5) structure, 6) environments, and 7) interfaces. The evaluation of component factors relies upon experienced and trained assessors to assign a threefold grade: the lower and upper bounds of the most probable estimate, allowing assessors to capture their uncertainty when making an evaluation. Grading scale is composed of discrete values from one to seven where: 1 is “outstanding in all standards and requirements”, 4 is average, 7 is very poor, and the intermediate grades are used to express evaluations between these anchor points. A mean grade, standard deviation of the grade (St. Dev.) and coefficient of variation (COV) of the grade are determined for each attribute.

TABLE I. EVALUATION CATEGORIES AND FACTORS [7]

Operating Teams Process auditing Safety culture Risk perception Emergency preparedness Command and control Communications	Organizational Process auditing Safety culture Risk perception Emergency preparedness Command & controls Training Communications resources	Procedures Operating Maintenance QA/QC Contractor selection Pre-start up review Emergency response Management of change Validations
Equipment Design guidelines and specification Materials Demand systems Power systems Configurations Control systems	Structure Design guidelines & specs Materials Loadings Structure configuration Computer programs Research, development and testing background	Environmental External (Weather) Internal Social external – (Regulatory, Society) Social (internal) (within organization and operating team)
Operators & other Organizations & other	Interfaces Procedures & other Environmental & other	Equipment & other Structure and other

A “Braille” (Pareto) chart is then developed, summarizing the mean grades developed by the assessment team for each of the factors. The ‘high’ grades, those above 4, indicate components and their factors that are mitigation candidates. The coefficients of variation associated with each factor indicate the range of uncertainty associated with the ratings. The assessors are able to back track and identify the factors and attributes that result in a particular grade along with the recorded comments that provide the rationale for the grading. This provides a strong interpretative and evaluative component in identifying the best actions to improve a grade.

Although the SMAS factors gradings are performed for each of the seven components, we are presenting an explanation for the grading of the first system component only (Operating Teams), whose results are shown in Table II.

A. SMAS Evaluation of Operating Teams Component

The following six factors were evaluated for the first system component, Operating Teams:

- **Process Auditing** – The controllers do not have an actual, formalized auditing process of URET. In the exploratory interviews, controllers emphasized they only perform the processes for which they were adequately trained and consequently, they were not fully utilizing URET.
- **Safety Culture** – Safety is their top priority (above efficiency and strategy).

- **Risk Perception** – Air traffic controllers do not want to come into any risk-related situation; they want to avoid such situations at all costs.
- **Emergency Preparedness** – The most probable grade of 2.5 depends on how long the controllers have been using URET, and how they control traffic. For example, those that extensively used paper strips felt that handling paper strips helped them be more aware of situation, and hence, felt more prepared to respond quickly to emergencies. Switching to the electronic flight strip replacement made them feel that their mental picture was inadequate. Furthermore, they felt that procedures for use are insufficient to prepare them for emergencies.
- **Command & Control** – The most probable grade of 4 is based on the fact that the tool is used differently from its designed purpose. The controllers are not using URET for its intended use (conflict probe), but more for its electronic flight strip functions that replaced paper strips.
- **Communications** – The implementation of URET should have improved controller team communications. Some centers’ communications did indeed improve, but primarily on an individual basis, rather than on a system-wide basis. Some centers’ communications actually worsened. Interestingly, before the use of URET, most controllers worked in teams of 2; with URET, there were more occurrences where controllers were working alone.

TABLE II. OPERATING TEAMS GRADING

Operating Teams	Lower Bound	Most Prob.	High Bound	Mean	St. Dev.	COV
Process Auditing	2	3	5	3.3	0.6	0.2
Safety Culture	1	2	4	2.3	0.6	0.3
Risk Perception	1	1	2	1.3	0.2	0.2
Emergency Preparedness	1	2.5	4	2.5	0.6	0.2
Command & Control	3	4	6	4.3	0.6	0.1
Training	4	5	7	5.3	0.6	0.1
Communications	2	4	6	4.0	0.8	0.2
GRADE (Category Mean)				3.3		

Braille chart is developed based on the grading of all the factors listed in Table I, including the remaining six system components (Organizational, Procedures, Equipment, Structure, Environment and Interfaces). The chart summarizes the mean grades developed for each of the factors, as depicted in Figure 1.

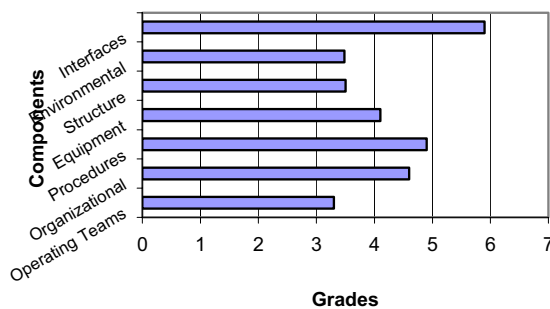


Figure 1 Braille Chart for URET Evaluation Categories and Factors

Based on the previously presented grading scale, this Braille Chart indicates that the majority of our system components do not meet the average standards (grade of 4). Four out of the seven components—Organizational, Procedures, Equipment, and Interfaces—are below average (between grades 4 and 7), whereas the other three components—Structure, Operating Teams, and Environmental—were barely above average (between grades 1 and 4). The operating teams (controllers) are not ranked poorly because the responsibility cannot be placed on them for using URET less often than was intended, as well as for not using URET for its intended purpose. The causes of the current style of URET implementation are rooted in the organizational factors, upstream from the operators themselves (i.e. lack of funding and attention for training and procedures). The worst ranked components were the Interfaces between Procedural + Operating, Procedural + Organizational, and Hardware + Procedural. These grades support the argument for integration of automation into the existing system, where separate system components (e.g. procedures, operating, organizational, hardware) are developed separately and independently.

B. Event Category Findings

The initiating events—that may have triggered the outcome of URET not being implemented as originally planned—are predominantly at the organizational level.

URET's adoption decision was made at the organizational level, but individual users were then left to "adopt, re-invent or reject the innovation during its implementation" [12]. The FAA decided to deploy URET to all ARTCCs in the United States in 2002, after more than 20 years since the beginning of URET development. Its continued development is a result of its robustness through the changing objectives of ATC management. The final deployment decision was based on automation performance metrics and cost-benefit assessment rather than on the analysis of controller performance when using URET.

The compounding/propagating factors—those that may have allowed the event sequence to escalate and result in the accident (e.g. the unintended usage and lack of usage of URET)—include the performance shaping factors of workload and stress. It was expected that by fully using URET, the controller team will not only be able to handle more aircraft (because of reduced manual workload from automation and availability of more accurate information), but will also be

able to provide more direct routings and better flight profiles to airlines. Since URET is designed as a decision support tool for the D-side controller to support the sector team strategic planning function, "full use of the tool is virtually impossible if only one controller works the sector" [12]. URET's electronic flight strip replacement function caused an unintended consequence - the reduction of teams from two-controller into one-controller teams. However, URET was designed for a two-controller team. In situations of increased traffic, if only one controller is working, and does not have the other controller's support and communication, the workload and stress do not allow controller to apply strategic approach (what URET is designed for), but make him/her to work tactically. Moreover, in such situations URET becomes a low priority—the controller does not even have the time to use it, when presumably he would need it the most - an unintended outcome.

The contributing factors—those that may have encouraged the initiating and propagating events—are insufficient training, lack of procedures, cultural differences, staffing, and structure of the center airspace.

Training was one of the most significant contributing factors that led to the system usage that was different from the designed one. According to [11], "there was a lack of requirement for new controllers completing any on-the-job training to have URET proficiency." Currently, training is done primarily on the job with more experienced controllers acting as teachers. Furthermore, this form of training is not efficient because most of these controllers are "more proficient in the pre-URET ATC, and less in the 'proper' URET use" [11].

For four of the new centers, URET training lasted only for four days, with the majority of the time spent on "buttonology"—explanations of how each function works and how to use the computer-human interface. The training did not address how URET would be integrated into existing practices, nor the creation of new procedures incorporating it. The training was not extensive enough to provide the controllers with full knowledge of URET and its use. Hence, they lack the ability to use the full range of URET functionality [11].

Although URET is installed in all ARTCCs, the FAA Order on Air Traffic Control, which contains the ATC procedures, covers Decision Support Tools (e.g. URET and Ocean-21) in a single, very brief chapter (Chapter 13). Most of the URET procedures assume a pre-URET environment. For example, flight strips are rapidly becoming obsolete, and yet, the chapter that covers the use of flight strips is quite long and elaborate. Although URET is intended to change the way controllers do their jobs, these changes are still not addressed in any training or work procedures. According to exploratory interviews with the SMEs and controllers, it was found that no appropriate ARTCC procedures exist for using URET [11].

VII. PROACTIVE APPROACH

The proactive approach is intended to study the physical aspects of systems and procedural-human aspects, identify potential improvements and critical flaws, and identify ways to improve the quality of the systems and procedures. Proactive approaches include system design measures (e.g. design for robustness), and life-cycle ergonomic design.

According to [7], a large variety of sample cases were studied in detail in which “errors made during and in the design of the system lead to the failure (lower than desired quality) of the system.” Organizations that were involved in the system designs are the dominant cause of system design related failures. Many of these errors are caused by the lack of the culture to promote quality in the design process. To improve and assure sufficient quality in system design, the priorities where one should devote attention and resources to are organizations, individuals, and procedures [7].

A. Design Organizations

High-reliability organizations (HRO) can reduce the probabilities of errors by enacting the following [13]: command by exception, redundancy, procedures and rules, training, appropriate rewards and punishment, and the ability of management to ‘see the big picture.’

Command by exception is essentially migrating decision making responsibility to the persons with the most expertise in making the decision when unfamiliar situations arise. In systems like ATC, and its automation, the decision making (when automation introduction is in question) should be made by a group of experts that is comprised from controllers, tool designers, human factors experts—an interdisciplinary group.

Redundancy involves people, procedures, and hardware. Currently, URET is on a different power source than the other tools, providing a form of redundancy in case one power source goes out. However, in some ARTCCs where URET unintentionally reduced the two-controller team to a one-controller team, redundancy in terms of controller decision-making was reduced. Thus, it is recommended that a minimum of two-controller teams at all times becomes a standard, even when workload is low.

Another important aspect of HRO are procedures that are “correct, accurate, well organized, well documented, and are not excessively complex” [7]. In addition, HRO could develop constant and high-quality training programs. As discovered in exploratory interviews, training and procedures were two of the major problems contributing to the failure of intended and sufficient URET use. The current ATC training is performed on the job by experienced controllers who are more proficient in pre-URET than post-URET environments. More attention could be devoted to developing high-quality training.

According to [11] it was suggested that in order for the developers to become a HRO, they need to focus on integration, rather than leaving controllers to “adopt it and adapt it on their own”.

B. Design Teams and Quality Assurance

According to [7], the design teams are the “first line of defense to prevent and/or detect and correct system design engineering malfunctions.” This category includes personnel selection and training, as well as the formation of cohesive teams and teamwork.

It is suggested that these design teams become more cross-functional, where the design teams incorporate researchers and programmers, as well as the operators themselves. If the end-users are incorporated early on in the design process, this could lead to less re-work downstream and would serve the purpose better.

Quality assurance (QA) includes the activity conducted prior to an operation to ‘assure’ that the desired quality is developed. QA is the proactive element of QA/QC, where Quality Control (QC) is the interactive element (discussed in the following section). QA methods take into account the quality attributes of the system (serviceability, safety, durability, and compatibility), and focus both on error prevention as well as error detection and correction.

C. Design Procedures

In the next generation system design procedures and guidelines, it is now clear that human and organizational factors (HOF) should be taken into even greater and more serious consideration.

It is suggested that the design teams take into account the lessons learnt in the past and incorporate them into their design procedures. Not only could these organizations focus on developing and incorporating QA/QC into their procedures, but they could also continue to work on proactively incorporating HOF in order to continually improve the design. It is further suggested that the QA/QC is incorporated into both the procedures of how to use URET, and a way for the controllers and auditors to continually and easily check that URET is being used as intended. Another suggestion would be to institutionalize QA/QC into everyday work, rather than sporadically through large-scale studies. For example, it could become part of the controller’s daily or weekly routine during their short breaks to evaluate their performance relative to URET.

D. Ergonomic Design

Ergonomics is the “science and practice of designing systems to fit people” [12]. Micro-ergonomics addresses the man-machine interface design at the local work station level, whereas macro-ergonomics “addresses the design of the work system as a whole” [15]. In micro-ergonomics, engineered systems are designed in order to decrease the likelihoods and consequences of failures associated with human-system interfaces, and sometimes, increase the likelihoods of detection and correction of failures. By understanding the impacts and the significance of micro-ergonomics, there will be a better grasp and implementation of proactive risk assessment and management.

Since many of the ARTCC sectors have reduced the teams to a one-controller team, different aspects of URET design could be considered to allow more flexibility for each individual ARTCC. URET has been demonstrated to be unsuitable for use by a one-controller team. URET has a separate trackball for information input, meaning that in situations of only one-controller working, he will need “to use two trackballs and two keyboards if he wants to use URET functions other than flight strip replacement one,” [11] which makes URET more complex to use. It is further suggested to re-design URET to accommodate both situations of one-controller and two-controller teams.

VIII. INTERACTIVE APPROACH

Since the aviation industry is highly dynamic and time critical, this interactive approach is crucial to the local operators to deal with threats to quality and reliability themselves. In other

words, the proactive and reactive approaches are not sufficient. Reference 3 has developed the proactive and reactive approaches, but there is a third and crucial element missing: the interactive (real-time) approach. According to [7], this interactive approach needs to be recognized and further developed.

This interactive approach is based on the argument that in essence, the aspects that influence or determine system failures in the future are unpredictable and unknowable. Reference 7 provides an interactive theory that is based on organizations and teams “interacting” with the system to return it to a safe state, hence turning an accident or failure into an incident or “near miss.” The goal is to increase the proportion of successful interventions as events unfold by developing the operators’ cognitive skills so that they can manage an unimaginable event before them.

Interactive management has two fundamental approaches with the objective to achieve quality in systems: 1) to improve the management of the causes to reduce the incidence of HOE, and 2) to improve the management of the consequences to reduce the effects of HOE. There are three time frames in which one can focus HOE management activities: 1) to prevent errors before the activity, 2) to detect and correct errors during the activity, and 3) to reduce the consequences of the errors after the error is committed [7].

In managing a crisis, training can reduce the amount of cognitive processing required to determine what should be done. In addition, observations themselves are crucial in crisis management. They provide clues about whether implementation is producing the desired results. A crucial aspect of operator training managing a crisis is a true, detailed understanding of URET in relation to ATC. It is not enough that the operator knows how to “push buttons” and use the Computer-Human interface. The operator needs to know each of URET’s 4 functions deeply, how they function in *relation* to one another and to the system as a whole.

There are two fundamental approaches to improving crisis performance [9]: 1) providing people support, and 2) providing system support. Providing people support includes selecting the right personnel and in the case of ATC, it is crucial that these operators are audited, trained, and re-trained in order to fully implement all the functions and intentions of URET. The training should include crisis management strategies. System support is the second approach to improve crisis performance, which includes factors such as maintenance of equipment and procedures so that they can be relied upon when a crisis occurs.

In ATC, human factors have traditionally been considered only during the design of the new tool (proactive), where the designed is expected to address all potential problems and consequences. However, once the tool has been implemented, the design is assumed to be sound, and the human factors specialists move to research the next innovation—there is not enough follow-up and observation on the implementation process [11]. It is suggested that the usage of URET is continuously being monitored and observed throughout. More importantly, the auditors would need to know and understand what exactly they are measuring because the end result could be misleading.

In summary, it is suggested that the developers could learn more from field evaluations, and implement an interactive, real-time approach to monitoring the

implementation of URET, in order to achieve a quicker evolution of air traffic management capabilities. In order to achieve this successful interactive approach, the selection of well-qualified operators and perpetual training and re-training is necessary to produce highly cognitive operators who can apply crisis management strategies while successfully navigating the crisis management loop. There needs to be more interactive, real-time monitoring on how controllers adopt/adapt URET and how its usage in turn impacts the overall ATC system.

IX. SUMMARY OF THE STUDY RESULTS

A Braille Chart is produced as a result of the analysis of seven factors as shown in the previous section in Table I. Based on the grading scale with 1 being outstanding, exceeding all standards and requirements, and 7 being very poor, and not meeting any standards or requirements, a Braille Chart indicates that the majority of the system components do not meet the average standards (grade of 4). Four out of the seven components—Organizational, Procedures, Equipment, and Interfaces—are below average (between grades 4 and 7), whereas the other three components—Structure, Operating Teams, and Environmental—were barely above average (between grades 1 and 4). The operating teams (controllers) are not ranked poorly because the responsibility cannot be placed on them for using URET less often than was intended, as well as for not using URET for its intended purpose. The causes of the current style of URET implementation are rooted in the organizational factors, upstream from the operators themselves (i.e. lack of funding and attention for training and procedures). The worst ranked components were the Interfaces between Procedural + Operating, Procedural + Organizational, and Hardware + Procedural. These grades support the argument for the integration of automation into the existing system, where currently the separate system components (e.g. procedures, operating, organizational, hardware) are developed almost separately and independently.

For the purpose of developing the Factors Grading Braille chart, in calculating the means for each system component, it is assumed that each of the factors in the components had distributed weights. For example, in the Operating Teams component, the Training factor most likely contributed the most to the overall component grade. However, the purpose of this study was not to go into such detail, but rather, to bring to the surface the factors of concern that the proactive measures should address.

In the next step, the study addressed proactive measures to implement in terms of 1) design organizations (high reliability organizations), 2) design teams, 3) quality assurance, 4) design procedures, and 5) ergonomics. These proactive measures are developed from lessons learned from the reactive approach, and are intended to identify potential improvements, critical flaws, and ways to improve the quality of the systems and procedures. However, it is important to acknowledge that proactive measures cannot predict everything, especially in such complex and dynamic systems such as ATC, where it is critical that safety never be compromised. Thus, it is crucial that, in this stage of intense research, development, and transition, a strong, robust interactive approach be developed to ensure the smooth and successful implementation of automation tools (e.g. URET)

into the system. For the current state, several interactive measures are proposed to constantly monitor the impact of URET on the controllers and the system. Also, it was pointed out that it is important for the organizations to know what they are measuring. For example, [7] has also mentioned that even if one is correctly measuring something, this may not mean anything since he or she may be measuring the wrong thing.

X. VALIDATION

It is important to determine the validity and reliability of the study's analytical methods and processes:

A. Validity

This paper has both external and internal validity. External validity is the extent to which the method (approach) is generalizable—the degree the results of its application to a sample population can be attributed to the larger population—or transferable—the degree the method's results in one application can be applied to another similar application. Internal validity “is the basic minimum without which the method is uninterpretable,” addressing the rigor with which the method was conducted—in terms of the design, the care taken to conduct measurements, and decisions concerning what was/wasn't measured [17]. The design of the method and care with which it was conducted was very carefully performed. We graded nearly every component and factor in the ATC-URET system, and provided justification and explanation for each grade.

The method of interviewing and developing three risk assessment and management approaches is very appropriate for its intended purpose. Reference 3 has developed extensive studies in the proactive and reactive approach. Reference 7 has extended these studies further, spending much time in developing the theories and applications of the missing element—the interactive approach. By applying this study on the implementation of URET in ARTCC sectors, this work also validates the work performed in [7] in how necessary and crucial it is to ensure that URET is monitored and interactively studied to improve its integration into the current system.

The generalizability of this method is both intended to assess the current state, as well as being applicable to the future state of URET, since it is constantly being studied and improved, in addition to other innovations. Also, the method and three approaches developed in this study have been validated by [7], where proactive, reactive, and interactive approaches were efficiently applied to the specific field of geotechnical engineering. The authors recommend that this way of approaching problems can be applied to any industry, so long as there are human and organizational factors involved (e.g. hospitals, nuclear plants, NASA)—finding an industry upon which to apply these concepts should not prove too difficult.

XI. CONCLUSIONS AND FUTURE WORK

This study has developed reactive and proactive approaches to the integration of URET in en route air traffic control centers. Following these approaches, we then developed an interactive approach for ARTCCs.

There is a great deal more research needed to proactively develop, in particular, improved training and procedures as

well as effective interactive measures. In order to develop a strong proactive approach, both the reactive and interactive approaches need to continuously feed information and lessons into one another—they are each complementary and necessary. By interactively studying the impacts of URET (and other future automation tools to be adopted and implemented) on the performance of air traffic controllers, as well as its very adoption and adaptation, usage, and integration, the gap between our current system and the next generation ATC will finally begin to diminish.

When automation introduction is in question, decision making should be made by a group of experts that is comprised from controllers, tool designers, human factors experts—all in an interdisciplinary group. By having these interdisciplinary groups apply these three approaches to studying automation introduction and implementation, they get closer to the ultimate goal of achieving quality (i.e. safety and serviceability).

We believe that the proposed approach to studying URET automation tool can be applied to other emerging automation, communication or surveillance tools in the near future. The lessons learnt from URET can be incorporated in applying sound proactive measures in other new implementations (e.g. other tools such as data link or ADS-B), so that the organizations will not have to start with learning from a reactive standpoint, and can move directly to a proactive and interactive system. We believe that the work completed in this paper is an excellent starting place for bringing proactive, reactive, and interactive quality management tools to this dynamic, complex field of introducing automation into the already stressed national airspace system.

REFERENCES

- [1] Isaac, A., Shorrock, S.T., Kirwan, B. (2002). “Human error in Europe air traffic management: the HERA Project.” *Reliability Engineering and System Safety*, Vol. 75, 257-272.
- [2] Javaux, D. (2002). “Human error, safety, and systems development in aviation.” *Reliability Engineering and System Safety*, Vol. 75, 115-119.
- [3] Maurino, D.E., Reason, J., Johnston, N., Lee, R.B. (1995). *Beyond aviation human factors*, Ashgate Publishing Company, Vermont.
- [4] Endsley, M.R., and Rodgers, M.D. (1998). “Distribution of attention, situation awareness, and workload in a passive air traffic control task: Implications for operational errors and automation.” *Air Traffic Control Quarterly*, 6(1), 21-44.
- [5] Wickens, C.D., Mavor, A.S., McGee, J.P. (1997). *Flight to the future: human factors in air traffic control*, National Academy Press, Washington, D.C.
- [6] Young, W.C., Tsai, M.T., Chuang, L.M. (2000). “Air traffic control system management.” *IEEE*, 494-498.
- [7] Bea, R.G. (2007). “Human and organizational factors: quality and reliability of engineered systems.” *CE 290A Course Reader*, Vol.1, 60-65.
- [8] Harwood, K. (1993). “Defining human-centered system issues for verifying and validating air traffic control systems.” CTA Incorporated, Moffett Field, CA.
- [9] Rakas, J., and Yang, S. (2007). “Analysis of multiple open transactions and controller-pilot miscommunications.” 7th USA/Europe Air Traffic Management R&D Seminar, Barcelona, 1-11.
- [10] Mertzger, U., and Parasuraman, R. (2005). “Automation in future air traffic management: effects of decision aid reliability on controller performance and mental workload.” *Human Factors and Ergonomics Society*, 47, 1.

- [11] Bolic, T.S. (2006). "Automation adoption and adaptation in air traffic control," thesis, presented to University of California at Berkeley, CA, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.
- [12] Rogers, E. (2003). *Diffusion of innovations*. The Free Press, New York.
- [13] Roberts, K.H. (1989). "New challenges in organizational research: high reliability organizations." *Industrial Crisis Quarterly*, Vol. 3.
- [14] Moore, W.H., and Miller, G. (1997). "Human factor engineering applications to the design of offshore platforms." *Proceedings of the optimizing offshore safety conference*, IBC Asia, Ltd., Kuala Lumpur, Malaysia.
- [15] Brown, O., Jr., and Hendrick, H.W. (1996). "Human factors in organizational design and management." *Proceedings of the Fifth International Symposium on Human Factors in Organizational Design and Management*, North-Holland Publishers, Amsterdam, The Netherlands.
- [16] Bellamy, L.J. (1994). "Gaining a better understanding of human behavior in emergencies." *Proceedings of the Conference on Mastering the Human Factors in Shipping*, IIR Ltd, London.
- [17] Campbell, D.T., and Staneley, J.C. (1963). *Experimental and Quasi-Experimental Design for Research*, Houghton Mifflin, Co., Boston.

Track 4

Decision Support Tools

Decision Support Tool for Predicting Aircraft Arrival Rates, Ground Delay Programs, and Airport Delays from Weather Forecasts

David A. Smith
Center for Air Transportation
Systems Research
George Mason University
Fairfax, VA 22030
Telephone: (703) 975-1386
Email: dsmit4@gmu.edu

Dr. Lance Sherry
Center for Air Transportation
Systems Research
George Mason University
Fairfax, VA 22030
Telephone: (703) 993-1711
Email: lsherry@gmu.edu

Abstract—The principle “bottlenecks” of the air traffic control system are the major commercial airports. Atlanta, Detroit, St. Louis, Minneapolis, Newark, Philadelphia, and LaGuardia all expect to be at least 98% capacity by 2012. Due to their cost and the environmental and noise issues associated with construction, it is unlikely that any new airports will be built in the near future. Therefore to make the National Airspace System run more efficiently, techniques to more effectively use the limited airport capacity must be developed.

Air Traffic Management has always been a tactical exercise, with decisions being made to counter near term problems. Since decisions are made quickly, limited time is available to plan out alternate options that may better alleviate arrival flow problems at airports. Extra time means nothing when there is no way to anticipate future operations, therefore predictive tools are required to provide advance notice of future air traffic delays. This research describes how to use Support Vector Machines (SVM) to predict future airport capacity. The Terminal Aerodrome Forecast (TAF) is used as an independent variable within the SVM to predict Aircraft Arrival Rates (AAR) which depict airport capacity. Within a decision support tool, the AAR can be derived to determine Ground Delay Program (GDP) program rate and duration and passenger delay. The introduction of this decision support tool will expand the amount of time available to make decisions and move resources to implement plans.

I. PROBLEM STATEMENT

Air traffic congestion has become a widespread phenomenon in the United States. The principle bottlenecks of the air traffic control system are the major commercial airports, of which at least a dozen currently operate near or above their point of saturation under even moderately adverse weather conditions [1]. The Macroscopic Capacity Model (MCM) analyzed 16 airports within a 1000 nmi. triangle from Boston, Massachusetts, to Minneapolis, Minnesota, to Tallahassee, Florida. Based on this analysis, the MCM showed that in 1997 these airports were operating at 74% of maximum capacity. The model further went on to predict that these airports will be at 89% capacity by 2012 [2].

The congestion problem is made worse because most airline

schedules are optimized without any consideration for unexpected irregularities. When irregularities occur, the primary goal of the airlines is to get back to the original schedule as soon as possible, while minimizing flight cancellations and delays [4]. When trying to get back on schedule, sometimes it is the complexity of the situation, coupled with time pressure, which results in results in quick decisions that may be less than optimal [5]. Therefore, it would be advantageous to develop techniques to lessen the complexity of the situation and increase the time available.

One way to increase the time available is to create a tool that can predict the impact of weather on future inbound flight operations. Weather reports such as the TAF, Aviation Routine Weather Report (METAR), and the Collaborative Convective Forecast Product (CCFP) all provide raw weather forecast information. None of these forecasts though inform National Airspace System (NAS) stakeholders what the effect of that weather will be on flight operations. This research intends to fill this void by developing a process from which a forecast can be entered to produce estimate of the delay and capacity of the airport within the forecast area. Capacity estimates, in the form of AARs are produced for four time periods of the operational day. Ground Delay Program estimates of duration and program AARs along with expected delays can be derived from the predicted AARs. Now the forecast will not only provide the winds and ceiling, but also the AARs, GDPs, and expected delay.

II. BACKGROUND

For efficient operation of the NAS, there is a need for the weather forecasting services and TFM products to estimate the reduction in capacity due to adverse weather. Weather forecast products are uncertain and the uncertainty increases with lead-time. Useful applications of weather forecasts requires either refinement, consultation, and application of the weather forecast to estimate air traffic capacity or decision support tools that take forecasts and make predictions based on past

forecasts and those forecasts connections to NAS capacity [6]. This paper describes a methodology used to create one such decision support tool known as the Weather Delay Prediction Tool. With this tool, the user enters the TAF for a given day and airport and the tool provides AAR predictions which can be derived to estimate delay and GDP time and duration.

Initially, this research focused on the CCFP as the weather forecast. The CCFP is a thunderstorm forecast for the entire United States and Canada and the research focused on its use as a predictive tool. After conversations with traffic management personnel and airline management, it was concluded that they rarely used the CCFP for any weather planning and relied on the TAF instead. The TAF has a good collection of available archived forecasts, so it was a good fit for the research objectives. To measure delays, a tool to predict GDPs was first considered. Over the course of the research it was determined that measuring delays may be more appropriate and then derive GDPs from the results. However, after presenting the work to air traffic management experts at the National Airspace System Performance Workshop, it was determined that it was better to use the AAR, since that was a common used factor to measure degraded airport capacity due to irregular operations. Also, GDPs and delays can be derived easily if the AAR is known.

III. METHOD

The general procedure used to determine a connection between weather forecast and airport capacity was:

- Collect data from the various available data sources,
- using assorted tools, format the data into a usable layout,
- use a classification tool to connect the two sets, and
- test the data to ensure there is a correlation.

A. Data Collection

FAA officials, airlines, air traffic controllers and others say Philadelphia plays a major role in delays up and down the coast thanks to poor airport design, bad weather, heavy traffic and close proximity to New York. Through September 2007, 68% of departures were on time in Philadelphia, better only than New York's JFK International, Chicago's O'Hare International and Liberty International in Newark, N.J. Fewer than two-thirds of arrivals were on time in Philadelphia during that period. The FAA has deemed Philadelphia a "pacing" airport that, because it sits in the middle of the busy East Coast air corridor, causes delays nationwide. Because of these facts, Philadelphia was chosen as the airport to evaluate the weather prediction tool [7]. The data used in this paper came from three areas:

- The TAF data was collected from a website provided by the National Climatic Data Center (NCDC).
- The Aircraft Arrival Rate data was collected from the Aviation System Performance Metrics (ASPM) database based maintained by the FAA.
- The delay data was found on the Bureau of Transportation Statistics website for summary statistics for destination airports.

1) *Terminal Aerodrome Forecast:* The TAF is an operational forecast consisting of the expected meteorological conditions significant to a given airport or terminal. TAFs always include a forecast of surface wind speed and direction, visibility, and clouds. Weather type, obstructions to vision, and low level wind shear are included as needed. The National Weather Service (NWS) produces over 570 TAFs. A TAF is a report established for the 5 statute mile radius around an airport. In the U.S., TAFs are produced four times a day starting at approximately 30 minutes before each main synoptic hour (00Z, 06Z, 12Z, and 18Z). All the forecasts produced starting one hour before the main synoptic hour up to four hours past the main synoptic hour are considered to be for the same cycle [8]. NWS is responsible for providing terminal forecasts to commercial and general aviation pilots for the protection of life and property and in response to requirements levied by International Civil Aviation Organization (ICAO) via the FAA in order to promote the safety and efficiency of the NAS.

2) *Aircraft Arrival Rates:* A Strategic Plan of Operations for managing flows during severe weather events in the NAS takes into account reduced AARs due to weather constraints. If the predicted capacity (number of aircraft that the airport can safely land in a given time period) falls short of scheduled demand (number of aircraft that wish to land at an airport in a given time period), traffic flow managers may implement a GDP [9]. GDPs are implemented by the Air Traffic Control System Command Center (ATCSCC) after consultation with regional Federal Aviation Administration (FAA) centers and with airline operations centers. A GDP applies to a particular airport, has specified start and stop times, and sets an allowable arrival rate.

Originally this research we focused on predicting GDPs by using the SVM. However, after discussions with air traffic managers, it was decided that it was more appropriate to predict AARs. AARs offer several advantages. First, each airport tends to revert to a finite set of AAR rates when airport capacity had to be reduced due to weather. This allowed grouping the possible outcomes into only a few distinct bins. Then a value was chosen between each bin and tested whether the day was \geq to the in between value or $<$ the between value. Finally, a predictor function was developed for each of these values and from the results we were able to predict the future AAR.

The second advantage of the AAR was that GDPs could be predicted based on the conclusions of the predictor function. GDPs occur when the AAR is below the rate for a normal operations when the weather is favorable. AAR predictions are made for four times during the day based on the demand level of the airport. This generated a graph found in Figure 1. For this airport, the greatest demand hours were at 0700, 1100, 1500, and 2000 local time. Table I shows the demand hour and the assumed coverage hours for the airport. This airports normal AAR was 44, so Figure 1 predicts a GDP from 1300 to 2400.

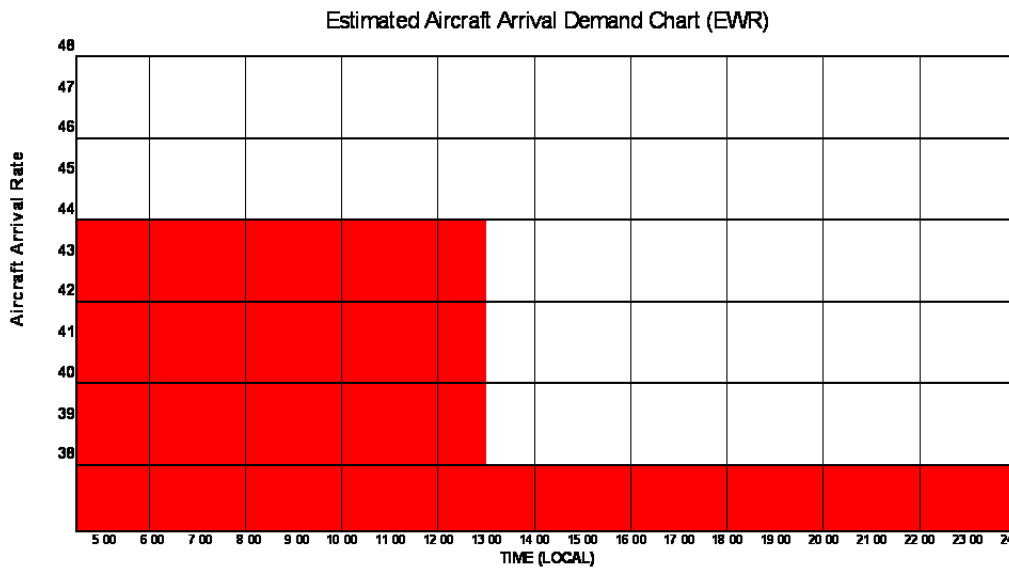


Fig. 1. Aircraft Arrival Demand Chart

Demand Hour	Assumed Time Block
0700	0500-0900
1100	0900-1300
1500	1300-1730
2000	1730-2400

TABLE I
DEMAND HOUR AND ASSUMED TIME BLOCK

B. Support Vector Machines

The Support Vector Machine (SVM) is a supervised learning method that generates input-output mapping functions from a set of labeled training data. In our case we are using the mapping function as a classification function. In addition to its solid mathematical foundation in statistical learning theory, SVMs have demonstrated highly competitive performance in numerous real-world applications, such as bioinformatics, text mining, face recognition, and image processing [10]. SVMs are based on the concept of decision planes that define decision boundaries. A decision plane is one that separates between a set of objects having different class memberships. A schematic example is shown in the Figure 2. In this example, the objects belong either to class square or circle. The separating line defines a boundary on the right side of which all objects are squares and to the left of which all objects are circles.

Figure 2 is a classic example of a linear classifier, i.e., a classifier that separates a set of objects into their respective groups (square and circle in this case) with a line. Most classification tasks, however, are not that simple, and often more complex structures are needed in order to make an optimal separation, i.e., correctly classify new objects (test cases) on the basis of the examples that are available (training cases). This situation is depicted in Figure 3. Compared to Figure 2, it is clear that a full separation of the square and circle objects would require a curve (which is more complex than a line). Classification

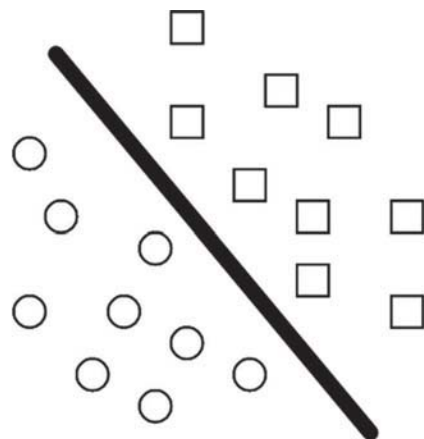


Fig. 2. Separating line defines a boundary

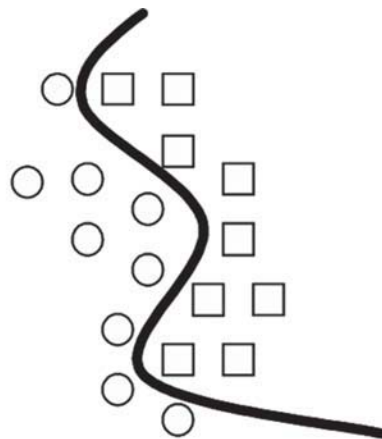


Fig. 3. Full Separation requires a curve.

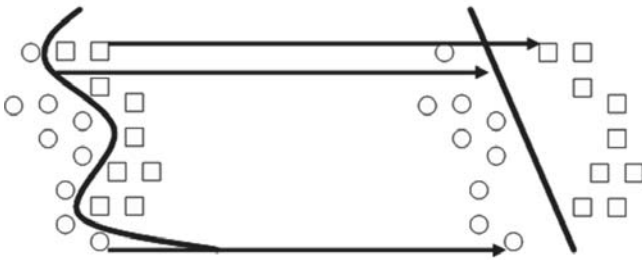


Fig. 4. Objects are mapped using a set of mathematical functions.

tasks based on drawing separating lines to distinguish between objects of different class memberships are known as hyper-plane classifiers. Support Vector Machines are particularly suited to handle such tasks.

Figure 4 shows the basic idea behind Support Vector Machines. Here we see the original objects (left side of the schematic) mapped, i.e., rearranged, using a set of mathematical functions, known as kernels. The process of rearranging the objects is known as mapping (transformation). Note that in this new setting, the mapped objects (right side of the schematic) is linearly separable and, thus, instead of constructing the complex curve (left schematic), all one has to do is to find an optimal line that can separate the square and the circle objects. Support Vector Machine (SVM) is a method that performs classification tasks by constructing hyper-planes in a multidimensional space that separates cases of different class labels. SVM supports both regression and classification tasks and can handle multiple continuous and categorical variables. For categorical variables a dummy variable is created with case values as either -1 or 1. For this type of SVM, training involves the minimization of the error function:

$$\begin{aligned} & \frac{1}{2}w^T w + C \sum_{i=1}^N \xi_i \\ \text{s.t. } & y_i (w^T x + b) \geq 1 - \xi_i \\ & \xi_i \geq 0, i = 1 \dots N \end{aligned}$$

Where C is the capacity constant, x is the vector of coefficients, b a constant, y the dummy variable, and ξ_i are parameters for handling non-separable data (inputs). The index i labels the N training cases. The kernel w is used to transform data from the input (independent) to the feature space. It should be noted that the larger the C , the more the error is penalized. Thus, C should be chosen with care to avoid over fitting [11].

C. Predictor Function

After collecting the TAF data as the independent variable matrix and the ASPM AAR as the dependent variable vector the SVM was applied to determine a function to predict future AAR's. The quadratic program introduced earlier was coded into AMPL. AMPL is a comprehensive and powerful algebraic modeling language for linear and nonlinear optimization problems, in discrete or continuous variables. After coding, the

program was submitted and the associated data to the NEOS Server for Optimization.

1) *Creating the TAF Vector:* The x in the quadratic program represented the 57 character long vector from the TAF weather data collected from 2002 through 2006. To create the vector, TAF data was collected from a website provided by the National Climatic Data Center (NCDC). These files tend to be long, up to 100 pages of text data, because all reports received are placed in these files as they are received and they are updated approximately every five minutes as data becomes available. Also, forecasts may be duplicated within the files and multiple forecasts received from a station may appear in a file [12]. To transform the raw TAF data into usable vector form, data was pasted into an Excel Spreadsheet. Then the text to column function was used to put each part of the data into a separate cell. After the data was transformed into a linear format, it was then parsed down to include only the 0600 Zulu TAF reports. It was assumed that planning would take place early in the morning and the 0600 Zulu TAF, which equates to 0100 EST, was the first of the day.

2) *Support Vector Machine Method:* The first step in the process was to find the common AARs for each airport in the study. Using the ASPM database, AARs were collected for each of the four peak hours for the 1826 days in the dataset. Airports tend to have a set of common AARs that they use, so there are a consistent set of values to perform the classification algorithm.

In the quadratic program, y represents a binary variable that indicates whether or not an AAR was set at a certain numerical rate for a given airport. Values equal to -1 indicate that day was greater than or equal to the numerical rate while values equal to 1 indicate that day was less than the numerical rate. One advantage of the SVM method is the way it deals with data outliers. For most methods, statistical techniques are used to eliminate values that are considered abnormalities. The SVM has an error function in the objective function, where the C variable is set to a value that increases or decreases the number of incorrect classifications within the data. A high C allows fewer outliers, while a smaller C allows more. For our analysis C was set at 1000 after experimenting with other values. This helped to determine a ξ vector, which was only used to relax the function, so a solution was possible. The ξ vector was not used in the final prediction function.

For the independent variables, the five years worth of data included 1826 days so this created an 1826×57 data matrix for the independent variable. The AMPL code was run on the NEOS Server and found a solution vector w and variable b for each airport. After determining the w vector and the y variable the current TAF forecast could be used to develop an x vector using that data and then use Equation 1 to develop a prediction value.

$$w^T x_i + b \quad (1)$$

If the prediction value was greater than 0, then the algorithm predicts that less than an AAR will occur on that day.

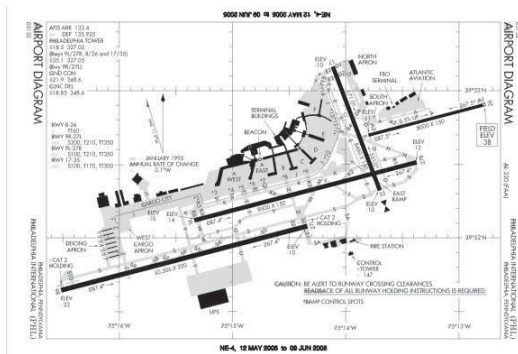


Fig. 5. Philadelphia Airport Map [13]

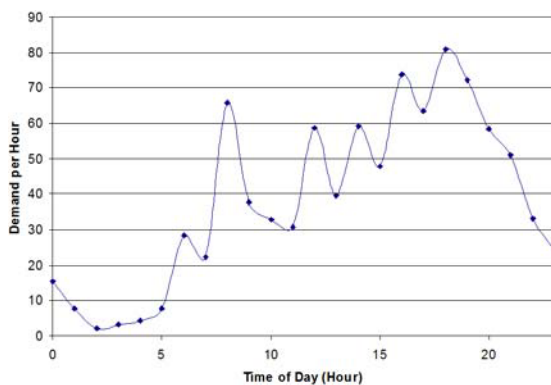


Fig. 6. PHL Hourly Demand

Conversely, if the value is less than zero then the algorithm predicts greater than an AAR value on that day.

IV. RESULTS

Figure 5 shows that the primary runways at Philadelphia are 9L/R and 27 L/R. Poor runway arrangement limits the number of planes that can take off from the airport at once, especially during bad weather. Although a small runway was added in 1999, most of the layout dates back to the 1970s or earlier [7].

The SVM only classifies the data for a given AAR value. In order to create a useful tool, several SVM operations had to be done for one airport. The first step in the process was to find the average demand rates for each hour during the day. These peaks are highlighted in Figure 6.

Figure 6 shows six peaks, but to reduce the dimensionality of the problem we chose only 0800, 1200, 1600, and 1800. For those time periods the most common AAR was 52, indicating normal operations, which occurred 60% of the time. During times of irregular operations the AARs are reduced to 48 for 20% of the time and 36 for 9% of the time. Because these three AARs constitute 89% of the possible AARs, these were set as the only possible solutions that the model will predict. Two SVMs were solved for each time period. The first SVM will classify whether or not the day had an AAR less than 52 or greater than or equal to 52. If the SVM classifies a given

Time	Divider	Sensitivity	Specificity	PPV	NPV	Correct
0800	48	0.38	1.00	1.00	0.84	0.86
	52	0.61	0.90	0.79	0.78	0.78
1200	48	0.35	0.96	0.64	0.88	0.86
	52	0.50	0.91	0.79	0.74	0.75
1600	48	0.31	0.98	0.75	0.89	0.88
	52	0.49	0.91	0.74	0.76	0.75
1800	48	0.32	0.98	0.75	0.89	0.88
	52	0.48	0.90	0.72	0.75	0.75
Combined		0.46	0.95	0.77	0.82	0.81

TABLE II
PHILADELPHIA TRAINING DATA

Time	Divider	Sensitivity	Specificity	PPV	NPV	Correct
800	48	0.40	1.00	1.00	0.83	0.85
	52	0.49	0.94	0.81	0.79	0.80
1200	48	0.44	0.99	0.85	0.92	0.91
	52	0.40	0.92	0.70	0.77	0.76
1600	48	0.36	0.98	0.77	0.89	0.88
	52	0.39	0.92	0.66	0.80	0.78
1800	48	0.28	0.98	0.73	0.88	0.87
	52	0.35	0.92	0.61	0.79	0.76
Combined		0.40	0.96	0.75	0.84	0.83

TABLE III
PHILADELPHIA TESTING DATA

day and time period as greater than or equal to 52, then the tool will show a AAR of 52. This AAR would also indicate no GDP during this period. If the SVM predicts less than 52, then we would develop an SVM to test to see if the given day and time period is less than 48 or greater than or equal to 48. Again, if the SVM indicates greater than 48, then the tool sets the AAR to 48 and indicates a GDP. If the SVM indicates less than 48, then we set the AAR is set to 36 and also a GDP is predicted during this period. The duration of a predicted GDP is based on what time periods have GDP AARs.

A. Philadelphia Results

To evaluate how the SVM worked for Philadelphia, two methods were applied. The first method observed the success rate of the SVM prediction functions for the two test points for each time period. Data was also separated between training data, which was the data from January 2002 through December 2006, and testing data which is data from January 2007 through June 2007. The results for the training data are found in Table II and the results for the testing data are found in Table III.

Table II and Table III indicate that the SVM algorithm was correct 81% of the time for the training data and 83% for the testing data. To create a meaningful tool containing these algorithms a set of rules was established to estimate the AAR. The tool only considers three possible AARs, one associated with normal operations, one associated with a slight reduction in capacity, and one associated with a large reduction of capacity.

The first rule tested whether or not the point, that represents a day, was below 48. If it was below 48, then the AAR was determined based on a weighted average of the observed AARs

Time	Actual	Accuracy	Predicted AAR		
			36	48	52
0800	36	0.718	158	41	21
	48	0.409	102	135	93
	52	0.778	96	187	993
1200	36	0.640	105	32	27
	48	0.464	87	141	76
	52	0.736	109	249	1000
1600	36	0.746	85	21	8
	48	0.396	89	126	103
	52	0.736	94	238	1026
1800	36	0.754	89	17	12
	48	0.389	88	126	110
	52	0.754	98	242	1044

TABLE IV
TOOL RESULTS FOR PHILADELPHIA TRAINING DATA

Time	Delay	Predicted AAR		
		36	48	52
0800	Low	0	0	0
	Mean	9	0	0
	High	19	8	1
1200	Low	0	0	0
	Mean	15	0	0
	High	37	9	3
1600	Low	1	0	0
	Mean	25	6	0
	High	48	19	9
1800	Low	13	0	0
	Mean	54	19	7
	High	96	42	22

TABLE VI
PHILADELPHIA DELAY PREDICTIONS

Time	Actual	Accuracy	Predicted AAR		
			36	48	52
0800	36	0.760	19	3	3
	48	0.250	5	3	4
	52	0.793	15	15	115
1200	36	0.923	12	0	1
	48	0.300	6	6	8
	52	0.772	8	26	115
1600	36	0.769	10	2	1
	48	0.294	7	5	5
	52	0.829	12	14	126
1800	36	0.727	8	1	2
	48	0.278	7	5	6
	52	0.810	15	14	124

TABLE V
TOOL RESULTS FOR PHILADELPHIA TESTING DATA

below the tested rate, which for all four time periods was 36. If the SVM indicated the point was equal to or greater than 48 or less than 52, then we assumed the AAR was 48. All other results were assumed to be 52. Table IV and Table V show the tool performance for the training and testing data.

Table IV and Table V show that the accuracy is better at the extreme points than the points in the middle. This shows that the SVM method is better at finding extreme points on the edge instead of points inside.

B. Delay Prediction

Within the airline industry and air traffic management the AAR determines the airport capacity and is used to highlight the severity of a GDP, therefore it is the preferred prediction variable. Most flying consumers do not understand what AARs are and prefer to know what are the potential delays. The Weather Channel uses a Red, Amber, or Green rating system to highlight the airport impact. Although, the website does not explain the rating system, one would assume that Red impact means the most delays and Green impact means little to no delays. Amber is somewhere in the middle. To make the Weather Delay Prediction Tool applicable to the traveler, a delay prediction needed to be added.

1) *Delay Prediction Method:* Since the SVM model only predicts three potential AAR outcomes, then the average delays during the time of those AARs would provide not

only a mean value, but also a range. Using the AAR data from ASPM and the delay data from Bureau of Transportation Statistics website, the average delay was calculated for each corresponding AAR. To provide a range of values, the standard deviation was calculated and added and subtracted to the mean value to provide a range.

2) *Delay Results:* Delay values are negative if the average arrival is early. Therefore, many of the average values are negative. If this was true then the value was changed to zero. Delay values and ranges are rounded to the nearest minute in Table VI. Table VI indicates the delay mean and the high and low range for each time period and predicted AAR. For instance, for the 1800 time period, if the model predicts an AAR of 36, then the delay mean is 54 minutes with a range as high as 96 and as low as 13 minutes. Now the flying consumer has information that they can use to plan their travel day.

C. Strategies for the Weather Delay Tool

Hub and spoke networks have become the most popular type of airline scheduling. In this type of scheduling, several points of departure feed into a single hub airport from which connecting flights carry passengers to their final destination. The advantage of the cross-connections is the multiplier effect as to the number of city pair that can be served. However, airports that are designated as the “hub” are subjected to increased congestion that are exasperated by irregular operations [4].

Meyer et al. (1998) [14] introduces the reliever and floating hub concept in a 1998 paper. Since most airline schedules are made without regard to unexpected daily changes due to severe weather conditions, there is very little slack time which means that any delay early in the day is likely to affect the schedule for the rest of the day unless the airline can take effective steps to correct the problem. Most carriers have developed procedures to follow in the event of unexpected disruptions in operations. However, most of these procedures are implemented manually, with little or no reliance on automated decision support systems [4].

The reliever hub is a strategy to reassign and optimize airport and airline schedules when experiencing a disruptive disturbance at a major hub airport and still maintain reasonable service. Figure 7 shows an example of a hub and spoke system.

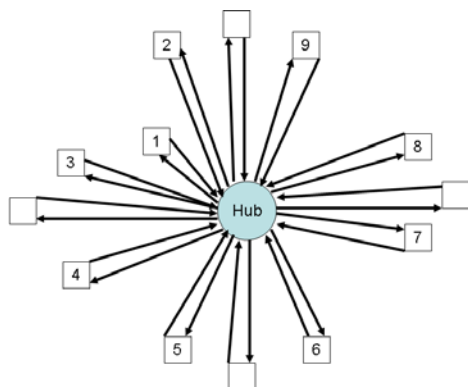


Fig. 7. Hub and Spoke System

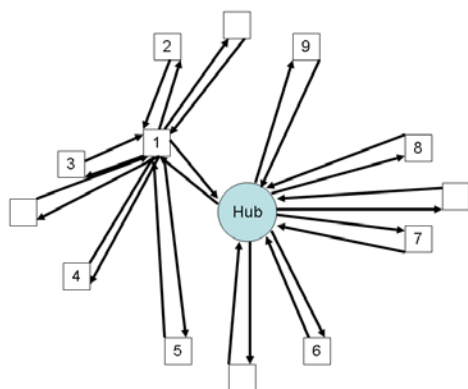


Fig. 8. Reliever Hub Option

The strategy is to temporarily use a nearby airport to act as a connecting hub, which can reduce the delays caused by a capacity reduction at the major hub. Figure 8 shows city 1 as a reliever hub. All cities to the west of the hub are sent to the reliever hub while all cities to the east continue to go to the main hub which reduces the demand on the main hub and decreases delays within the system. Service from city 1 to the hub would probably have to use a larger plane or more flights to insure passengers that need to get to the eastern cities or to the hub city arrive at their final destination.

Determining the best location for a reliever hub is a complex task. Even if the alternate hub is chosen ahead of time, the airline still must reconfigure schedule and passenger itinerary to minimize the total delay. Whether this reconfiguration is done by hand or is automated, it still requires time and employee manpower. It also requires a decision to use this manpower. Manpower has a cost associated with it, so the airline manager has to reasonably sure that the labor cost will help reduce any future loss due to extended delays. Because the TAF Delay Prediction Tool predict future delays it provides time and justification. Since the tool has been trained by historical data, it makes an AAR prediction based on what happened with similar TAFs in the past. Now the manager has justification to begin planning for the reliever hub.

Because the manager can enter the TAF a day in advance, there is now time to implement the plan by rescheduling flights

and even informing passengers of any changes. Larger or smaller aircraft can be swapped to account for the change in passenger. This time also allows ground crews at the reliever hub more time to ready themselves for unusually high activity. The reliever hub may not have the permanent infrastructure to support an increase in passengers, therefore temporary solutions may have to be implemented. Since there may not be enough gates, planes may have to be serviced on the apron. This may require the use of buses and bus drivers to drive passengers from the planes to the terminal. Temporary shelter may have to be set up to shield passengers from the heat or cold.

V. CONCLUSION AND FUTURE WORK

A. Weather Delay Prediction

The paper shows the possibilities of a Weather Delay Prediction Tool and what it can do to help NAS stakeholders. The algorithm is capable of classifying weather forecasts into three sets, where each set represents a specific AAR. Typically the highest AAR represents the airport during normal operations, while the two lower values represent reduced capacity due to weather or other congestion issues. Analysis showed that the SVM was more effective at predicting the normal AAR and the lower reduced capacity AAR. Therefore, for the weather delay prediction tool, it is appropriate to set a red, amber, or green scale. If the tool indicates green operations, then it is likely that the capacity at the airport will be at the maximum AAR and delays will be minimal. If the tool indicates red operations, then it is likely that the capacity at the airport will be significantly reduced and delays may be excessive. The Amber response indicates that the prediction is more uncertain, however, planners should prepared to have reduced operations at that airport. This appears to be the same rating system employed by the Weather Channel website, but it is also used by the military to rate progress of projects, describe the suitability of terrain for armored vehicles, or any other situation that requires a general rating. Since the tool provides only a general assessment of airport capacity through AARs, then a general prediction of delays is also included based on AAR and time period. This tool helps the flying public know how long they can expect to be delayed due to weather.

GDPs will be predicted based on the tool prediction for each time period. The time periods for each airport were determined based on the demand peaks during the operational day. The two important pieces that came from a GDP are the programmed AAR and the duration. Programmed AAR is predicted based on the tool's prediction for each time period. The length of the GDP is determined based on which time period are below the normal rate. For instance, the peak time periods at Philadelphia were 0800, 1200, 1600, and 1800. Therefore, if the AAR prediction at 1600 and 1800 were below normal and the AARs at 0800 and 1200, then we assume that the GDP begins half way between 1200 and 1600 at 1400 and lasts until the the end of the operational day at 2400.

B. SVM Disadvantages

A disadvantage of the SVM is that it does not show if any factor has more influence on the outcome than another. For each individual prediction equation developed, there were same factors that were weighted higher than others. The prediction equation is not an intuitive answer. However, across all of the prediction equations, there was not a value that consistently had more influence than another. By the nature of the algorithm, recursive partitioning searches for the value that best divides the data, so if determining which factors have the most influence on the final solution, the recursive partitioning method is more appropriate.

It is difficult to determine the effect of some of the data sets on the SVM. For instance, construction and airport upgrades at an airport can create inconsistent data. Analysts can attempt to normalize the data to try to maintain a consistent data set. However there was no way to determine how this affected the actual results. Results from airports that required normalization may not be as accurate as airports that had a consistent data set. Also, the SVM can not predict rare occurrences. For instance if a AAR rarely happened and the SVM tried to separate it from the rest of the data, a prediction vector of all zeros was the output with a b value of either 1 or -1.

C. Future Work

1) *Proper Data Set Size*: One of the issues with the SVM is what is the right amount of training data needed to produce a prediction equation that produces the most accurate predictions without overfitting the data. The process in this paper was to develop a prediction equation and then compare it to the training data and then the testing data. The analysis used 57 factors in the prediction function. The number was based on the factors found in the TAF for four time periods. If more factors were added to the data, then the performance of the predictor function applied to the training data improved. Unfortunately this did not improve the performance of the testing data which indicates that the data was overfitted. This situation is similar to using a 8th order polynomial to predict a line with 8 variables. It performs well with the 8 variables, but has little prediction value for any new independent variable. An optimization algorithm could be developed to determine the proper mix of training data and factors. Additional data factors could be added by adding time periods and the size of the training data could be reduced, although not increased since TAF data only goes back as far as January 2002.

2) *Factors Other than Weather*: This research focused on weather and using a forecast product to predict reduced airport capacity. Within the paper, it was discussed that other factor such as schedule congestion and runway construction also affect airport capacity. Further study should determine if there are any factors, besides weather factors, that can be added to the set of independent variables to produce a better predictive model. Techniques such as linear regression, have methods available to add or remove variables to the equation. At present though, there is no such standard process for SVMs other than adding the variable and testing the results to see if there is

improvement. Developing this technique in itself would entail extensive analysis.

ACKNOWLEDGMENT

We would like to thank the Center for Air Transportation System Research (CATSR) at George Mason University for providing support to this study. We would also like to thank Dr. George Donohue from George Mason University and Dr. David Rodenhuis from the University of Victoria, British Columbia. Earlier conversations with both were the genesis of this work and their advice on air traffic management and weather forecasting were vital. We would also like to thank Ved Sud from the FAA, Mike Brennan from Metron Aviation, and Mike Wambsgans from Flight Explorer for their technical contributions. Danyi Wang and Bengi Manley from the CATSR lab provided valuable information on delays and ground delay programs and provided input to early study results. I would also like to thank Dr. Rajesh Ganesan, Dr. Andrew Loerch, and Dr. Aimee Flannery for the time and effort they set aside to help us complete this research.

REFERENCES

- [1] M. Terrab and S. Paulose, "Dynamic strategic and tactical air traffic flow control," Rensselaer Polytechnic Institute, Tech. Rep., August 1992.
- [2] G. Donohue and W. Laska, *United States and European Airport Capacity Assessment Using the GMU Macroscopic Capacity Model*. Lexington, MA: American Institute of Aeronautics and Astronautics, 2001, vol. 193, ch. 5, pp. 61–73.
- [3] Y. Ageeva, "Approaches to incorporating robustness into airline scheduling," Masters Thesis, Massachusetts Institute of Technology, August 2000.
- [4] F. Durso, T. Truitt, C. Hackworth, D. Ohrt, J. Hamic, and C. Manning, "Factors characterizing en route operational errors: Do they tell us anything about situation awareness?" in *Proceedings of the International Conference on Experimental Analysis and Measurement of Situation Awareness*, D. Garland and M. Endsley, Eds. Daytona Beach, FL: Embry-Riddle Aeronautical University Press, June 1996, pp. 189–196.
- [5] D. Rodenhuis, "Hub forecast prototype test," in *Paper J3.9, Proc. Aviation, Range, and Aerospace Meteor (ARAM), American Meteor. Soc.*, June 2006.
- [6] P. Walters, "Delays at Philly airport caused by poor design, bad weather," *Aviation*, p. 20, December 2007.
- [7] *National Air Traffic Training Program, Air Traffic Guide, Aviation Routine, Weather Report (METAR), Aerodrome Forecast (TAF)*, Aviation Weather Center, Washington, D.C., May 2007.
- [8] B. Hoffman, J. Krozel, and R. Jakobavitis, "Potential benefits of fix-based ground delay programs to address weather constraints," Metron Aviation, Inc. Herndon, VA 20170, Tech. Rep., August 2004.
- [9] V. Kecman, "Studies in fuzziness and soft computing," in *Support Vector Machines: Theory and Applications*, L. Wang, Ed. Berlin: Springer, January 2005, ch. Support Vector Machines - An Introduction, pp. 1–47.
- [10] T. Hill and P. Lewicki, *STATISTICS Methods and Applications*. StatSoft, 2006.
- [11] "HDSS access system, national climatic data center," <http://Hurricane.ncdc.noaa.gov>, June 2007.
- [12] *U.S. Terminal Procedure Publications*, Federal Aviation Administration, Washington, D.C., August 2007.
- [13] E. Meyer, C. Rice, P. Jaillet, and M. McNerney, "Evaluating the feasibility of reliever and floating hub concepts when a primary airline hub experiences excessive delays," University of Texas at Austin, Tech. Rep., 1998.

Pilot Support for Flying Curved Decelerating Approaches in Realistic Wind Conditions

Alexander in t Veld, Rob Groenouwe, Max Mulder and René van Paassen
Faculty of Aerospace Engineering
Delft University of Technology
Delft, The Netherlands
a.c.intveld@tudelft.nl

Abstract—The most promising aircraft noise abatement approach procedures are those that combine flying longer at high altitude with continuous descents in a near-idle thrust setting. Although very effective at mitigating noise impact on the populated areas that surround airports, these procedures reduce runway capacity with respect to standard ILS approaches. Large uncertainties in descent trajectories force air traffic controllers to apply large separations in order to ensure safe operation. In this paper, a solution is presented that addresses the problems of variability in deceleration profiles and wind uncertainty. Spacing is done by providing pilots with a required time of arrival. A support system then helps the pilot in meeting this time goal. A wind prediction algorithm has been developed that creates a wind profile estimate along the intended three dimensional approach track, using filtered wind data observations broadcast by nearby aircraft. By combining accurate wind estimates with a flap scheduling algorithm, accurate track and speed guidance is available on-board. An interface has been designed that aids the pilot both in flying a controlled continuous descent approach and in meeting the time target set by air traffic control. To test the combined support system, a piloted simulator experiment was set up. Performance in terms of time goals was found to be consistent under all tested conditions and significantly better in comparison with the non-supported condition. Also, workload is significantly lower with the display optimization present. Providing the pilot with continuously updated time performance information based on actual meteorological circumstances was shown to be an important requirement for the implementation of CDAs in a time based spacing environment.

Index Terms—continuous decent approach, wind prediction, trajectory prediction, pilot guidance

I. INTRODUCTION

All over the western world, and especially in Europe, aircraft noise is one of the major limiting factors of airport capacity growth[1]. A lot of attention is paid to mitigating airport nuisance from the surrounding communities. Many different approaches are taken, for instance in the fields of aircraft engine technology and airport infrastructure planning[2]. The approach under study here will focus on noise abatement approach procedures, and how these can attribute to lowering aircraft noise impact on the ground. In this field there is still significant room for improvement, since the procedures currently in place hardly make use of advances in guidance, navigation and surveillance technology.

Over the years, several noise abatement approach procedures have been developed[3], [4]. Variants include general procedures like the Low-Power Low-Drag Approach

or the Continuous Descent Approach, and airport-specific measures[5], [6], [7], [8]. One characteristic aspect of many of these proposed procedures is that part of the approach is carried out with more or less idle thrust settings. Also, aircraft should avoid flying level segments at low altitude, which are typical for the current standard procedures, including ILS approaches. These new procedures have been shown to reduce aircraft noise, but at a cost: the different continuous descent profiles of various aircraft cause air traffic controllers to apply large initial separations to ensure safe operation. As a result, runway capacity is reduced[6], [9].

In this paper a solution to this problem is proposed by introducing a pilot support system that enables time-based separation during the approach. In other words, pilots are given a Required Time of Arrival (RTA), rather than radar vectors. It then becomes the pilot's task to comply, within bounds, with this time goal, whereas final responsibility for safe separation remains with ATC. The resulting system enables continuous descents, while still guaranteeing safe separation. This research focuses in particular on curved approach procedures under realistic wind conditions. Wind has a major influence on the accuracy of time based separation, and prediction of the wind profile encountered during the approach could be of great importance. A tool capable of accurately predicting wind conditions along a three-dimensional approach track was developed. Together with an algorithm that calculates the optimum settings for parameters such as the altitude where thrust is reduced to flight idle, this forms a support system that helps the pilot in flying idle thrust, continuous descent approaches while meeting arrival times commanded by ATC. This should allow ATC to sequence and space aircraft in the TMA more tightly, eliminating the capacity reduction currently associated with many noise abatement procedures.

This paper discusses (i) the characteristics of the particular Continuous Descent Approach procedure investigated in this paper, (ii) the algorithms that form the pilot support system and (iii) the results of a piloted simulator experiment that was set up to test the behavior of the system under realistic circumstances.

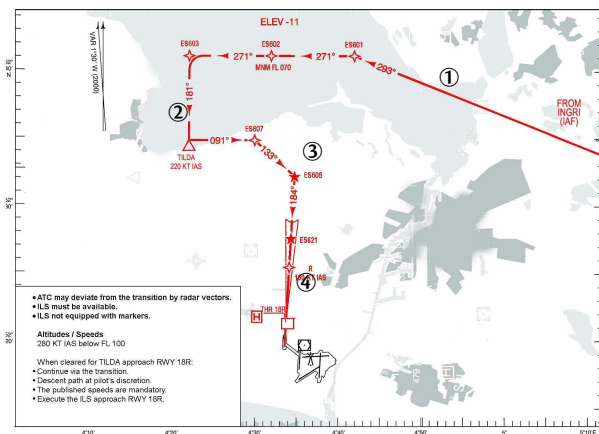


Fig. 1. The approach route, top view. The starting points of the phases of the CDA procedure are indicated. ① Level flight at FL70, ② Constant IAS descent along 3° glide path, ③ Idle thrust descent, ④ Constant final approach speed along ILS.

II. CONTINUOUS DESCENT APPROACHES

A. Description of the procedure

Based on practical experience with Noise Abatement Procedures at Amsterdam Airport Schiphol in the Netherlands and previous research[9], [6], a procedure resembling a standard nighttime transition[10] was chosen as the scenario for this research. These transitions typically involve a number of turns to avoid flying directly over the most densely populated areas. Obviously, these turns will cause great variation of the headwind and crosswind components when flying these transitions.

As can be seen from Fig. 1, the procedure consists of four parts. In phase ①, the aircraft is flying level at a relatively high altitude, but within TMA boundaries, for instance FL70. Nominal airspeed in this part is 220 kts IAS. At about 22 nm out, the aircraft intercepts a 3° glide path, but maintains its nominal indicated airspeed (phase ②). At a predetermined altitude, thrust is reduced to flight idle, marking the beginning of phase ③. When the aircraft reaches its final approach speed, thrust is reapplied and this speed is maintained (phase ④) until touchdown on the runway. For safety reasons, the aircraft should reach this approach speed no later than when it reaches 1000 ft altitude, approximately 3 nm from the runway threshold. This point will be later in this paper referred to as the reference window and is located at RNAV waypoint R ('Romeo'). At this point, the aircraft should be fully configured for landing, with full flaps extended and landing gear down. From here, the remainder of the approach is identical to a standard ILS approach procedure.

This type of approach procedure requires certain technologies to be available in aircraft throughout the arrival stream. For example, following a 3° glide path, while not yet aligned with the runway centerline, would require VNAV-path or Microwave Landing System (MLS) capabilities on board of the aircraft. Although not yet widely implemented, these

technologies are already available today.

B. Time based separation

Variations in the characteristics of this approach trajectory would make it difficult for an air traffic controller to space incoming traffic. Three factors are identified as having an important influence on the CDA's characteristics[11]:

- Different aircraft types with their own characteristic idle-thrust deceleration profiles,
- Varying wind conditions,
- Uncertainties in pilot behavior.

As all of these factors need to be accounted for in spacing, uncertainty adds up and controllers apply large initial separations as a safety buffer. Transferring all or part of the spacing task to the cockpit could greatly reduce the problem of the different deceleration profiles, since in general the flight crew will have access to more aircraft-specific and situation-specific information on own aircraft characteristics than an air traffic controller on the ground[12]. One way to go about this is the concept of time-based separation, providing each pilot in the chain with a Required Time of Arrival (RTA) at touchdown and other waypoints along the approach trajectory. These RTAs allow the controllers to increase runway landing capacity, by providing the aircraft under their control with optimized arrival times. It then becomes the pilot's task to navigate his aircraft to the runway, respecting the time constraints as demanded by ATC. The focus of this research is to investigate whether implementation of a system of time-based separation is feasible under actual operating conditions (curved trajectories, varying winds), without putting too much workload on the flight crew.

III. SUPPORT SYSTEM DESIGN

To help the flight crew in meeting the goals stated above, a support system has been developed. Its main aim is to provide the pilot with continuous information on the aircraft's status with respect to the time goal and the execution of the continuous descent approach. The support system consists of four modules, shown as encircled blocks in Figure 2. In this section the first three modules (wind prediction (①), track prediction (②) and an optimization module (③)) are explained. The last module (module ④) translates some of the parameters in the system into information for the pilot, which is then presented on the cockpit displays, along with time performance indication. This is described in Section V.

A. Wind profile prediction

As listed in Section II-B, the wind encountered during the approach is expected to have a large influence on the accuracy with which the entire procedure is flown. It is, therefore, very important to have a tool that accurately predicts the wind profile ahead of the own aircraft. An algorithm was developed, capable of predicting horizontal wind speed and direction along any path in a three dimensional space. This wind profile prediction algorithm is based on previous research [11], which assumes measurements of wind data are available, for instance

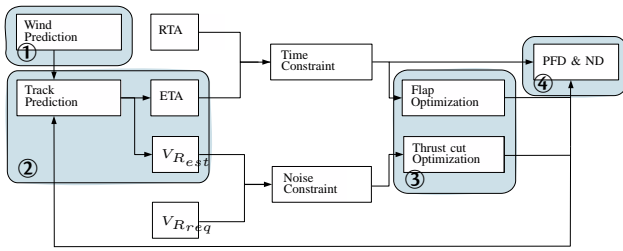


Fig. 2. Schematic of the support system algorithm. The four modules are indicated ① - ④.

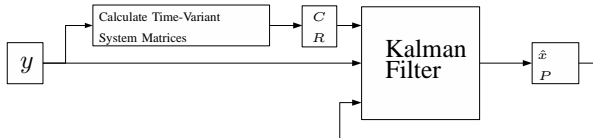


Fig. 3. Schematic of the filtering process in the wind prediction algorithm.

through ADS-B soundings from other aircraft in the vicinity. These measurements are then filtered using a Kalman filtering technique to produce an estimate of the wind profile the aircraft will encounter. The Kalman filter is well-suited to deal with integrating noisy measurements in the prediction[13]. In addition, it is easily implemented in a recursive algorithm, thereby reducing the need to store large quantities of wind data on board. This model was modified to be usable in three dimensions, making wind profile prediction along any curved approach trajectory, from any position and altitude, possible. In addition, functionality to use every incoming observation was incorporated, where in earlier work only data around certain fixed altitude intervals was used. This allows optimal use of the information at hand, and is expected to make the prediction more reliable, especially in situations where data density is low. A description of the way the wind prediction algorithm works is given below.

First, an initial wind speed estimate vector \hat{x} is set up. This vector contains wind speeds for a number of altitudes. In this study, the altitude interval is set to 500 ft. This interval is arbitrary, since the accuracy of the prediction only depends on the amount and accuracy of the available wind measurements. The set up of this initial wind profile guess is arbitrary. It may consist of unfiltered measurements of ADS-B soundings, or a standard profile uploaded from an Air Traffic Services unit. In this case a standard logarithmic wind profile is constructed as follows:

$$\hat{x} = V_0 \left(\frac{h}{h_0} \right)^\kappa \tag{1}$$

with κ the Von Karman constant, equal to 0.4[14]. This equation bases the wind speed \hat{x} on the free stream wind velocity V_0 at a corresponding altitude h_0 . h is the altitude at which we want to determine the wind speed.

Whenever new data comes in, this profile is updated. An incoming observation y_k is split into North and East

components, to be able to estimate these separately. Next, a weight matrix C is set up that determines the influence this observation will have on the state estimation of the wind speed. The weights in the matrix are determined based on the altitude difference between the states of interest and the measurement. With this C -matrix, the current estimate for the altitude of the observation can be calculated, and its value can be compared to the measured value to determine the innovation e . This way, an incoming measurement only influences the states in its altitude vicinity:

$$\hat{y}(k|k-1) = C(k)\hat{x}(k|k-1) \tag{2}$$

$$e(k) = y(k) - \hat{y}(k|k-1) \tag{3}$$

This innovation is multiplied with the Kalman gain to obtain a new state estimate. The Kalman gain is based on the relative magnitudes of the uncertainties in the current estimate and the new measurement, represented by the prediction error covariance matrix P and the measurement noise covariance matrix R , respectively. Since data from aircraft on the same approach track rather than elsewhere in the TMA is of more use for this prediction, the measurement noise covariance matrix R has been made dependent of the distance between the point of the measurement and the own track d :

$$R(k) = f(R_0, d(k)) \tag{4}$$

Here, R_0 represents the uncertainty in an incoming observation, mainly caused by measurement error. The accuracy of wind measurements in ADS-B soundings is approximately 2 kts[15]. The prediction error covariance matrix P is given by:

$$P(k|k-1) = AP(k-1|k-1)A^T + Q \tag{5}$$

In this equation, the matrix A represents the system dynamics. However, since the filter is used only as a noise filtering mechanism, no system dynamics are present and A reduces to the Identity matrix. Q is the (constant) process noise covariance matrix, which is determined empirically. At the same time as when the initial wind profile is set up, the P matrix is assigned a large value. This represents the large uncertainty in the accuracy of the profile at this point, and ensures that in the beginning, incoming observations have a large influence on the wind profile estimate. With this projection, the covariance of the innovation step $e(k)$ can be represented by the matrix S :

$$S(k) = C(k)P(k|k-1)C(k)^T + R(k) \tag{6}$$

The covariance matrices together determine the Kalman gain K :

$$K(k) = AP(k|k-1)C(k)^T S(k)^{-1} \tag{7}$$

A high uncertainty in the current estimate (high P) and much confidence in the accuracy of the measurement (low S)

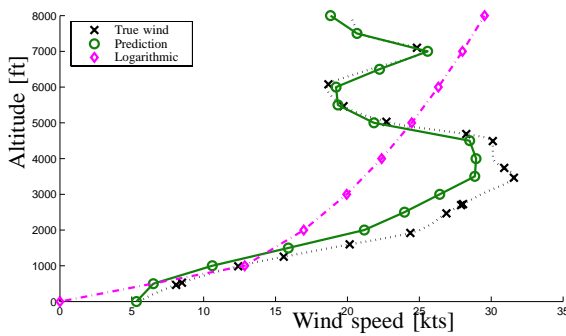


Fig. 4. Wind profile prediction performance. A typical wind profile and its estimate (in 500 ft intervals). For reference, a best-fit logarithmic profile is also shown.

yield a high value of the Kalman gain, which in turn assigns a large weight to the incoming measurement in updating the state estimate through the innovation. By the same rationale, much confidence in the current estimate and little in the accuracy of an incoming measurement yields a small value of the Kalman gain and consequently little influence of the observation on the wind speed estimate:

$$\hat{x}(k|k) = A\hat{x}(k|k-1) + K(k)e(k) \quad (8)$$

In the last step, the prediction error covariance matrix P is updated, according to:

$$P(k|k) = AP(k|k-1)A^T + Q - K(k)C(k)P(k|k-1)A^T \quad (9)$$

This iteration is repeated with a constant frequency of 1 Hz. It can be seen that when no new observation data are available, there will be no innovation, and the loop will be reduced to updating the prediction error covariance P . In this way, the uncertainty about an estimate increases when time goes by without new incoming measurements. A schematic of this loop is shown in Fig. 3.

Interpolation between the updated elements of the state vector yields the wind profile the aircraft is expected to encounter during its approach flight. The results of one such wind profile prediction are shown in Fig. 4. Here, a random wind profile (crosses) is shown, together with its best estimation (circles) based on the available observations. A best fit logarithmic profile (diamonds) is shown for reference. It becomes immediately clear that the Kalman filtering method is much more capable of capturing the random variations in realistic wind profiles. For the wind speed profiles used in the piloted simulator experiment (see Section V), the average root mean squared (RMS) of the wind speed prediction error for the Kalman filter based predictor is 2 kts, corresponding to the accuracy of the wind speed measurements available through ADS-B soundings. Prediction accuracy is much lower for a logarithmic predictor, with RMS values averaging 6 kts, occasionally running as high as 9 kts.

B. Track and time prediction

In the proposed scenario, the track to fly is fixed and determined by RNAV waypoints, as shown in Fig. 1. During the flight, an algorithm estimates the speed and time profiles along this track, taking into account the actual condition parameters such as predicted wind speed profile along this track, aircraft weight, flap setting, etc. The track prediction, consisting of speeds and times calculated for every point on the track, is repeated every second.

Between the constant speed segments of initial airspeed and final approach speed, deceleration takes place by selecting a flight idle thrust setting. This deceleration profile is influenced by varying wind conditions. Flap extension towards landing configuration also takes place in this phase, resulting in very non linear aircraft behavior. To predict the aircraft motion in this phase, a simple three degrees-of-freedom aerodynamic model of the B747-200 was used. The model is a point mass model that only looks at the forces along the flight-path and perpendicular to it. The resulting accelerations along the flight trajectory are integrated over time to yield speed, distance and time profiles. Since the model is two dimensional, it is fed with only the along-track component of the predicted wind speed.

C. Optimization of support system performance

To introduce greater flexibility and robustness in the aforementioned prediction modules, an algorithm was added to optimize two CDA parameters, flap speeds and thrust cut altitude. This third module is based on a flap scheduling algorithm, used in previous research in various forms [16], [11], [17]. It uses the same aircraft model mentioned in Section III-B, to calculate the effects of changed flap settings on the speed and time profiles in the trajectory ahead.

The flap schedule algorithm works in two modes: in HOLD mode and CAPTURE mode. In CAPTURE mode the algorithm calculates the thrust cut altitude, the altitude at which the thrust should be set to idle so that V_{APP} can be reached at h_R using the nominal flap schedule. This information is communicated to the pilot through a cue on the PFD, see Section V.

Once the thrust has been set to idle the module switches to HOLD mode. In HOLD mode the algorithm determines a flap schedule such that the aircraft reaches V_{APP} at h_R . This deviation from the nominal schedule can be used to cope with errors caused by for instance an inaccurate wind prediction, unexpected behavior from preceding aircraft, etc. In this mode, the algorithm predicts the aircraft's trajectory based on the current flap schedule. This yields an ETA, which is then compared to the RTA commanded by ATC:

$$\Delta T = ETA - RTA \quad (10)$$

Based on this difference, the flap scheduler does a rough tuning of the flaps to either their upper or their lower bounds, depending on the sign of this ΔT . In case the aircraft is predicted to arrive early, deceleration needs to be faster than the current (nominal) flap schedule will provide. The flaps

will thus be set to their upper bounds, and the resulting new trajectory is calculated. This process is repeated for the consecutive flaps, until the target is overshoot (ΔT changes sign). From here, the flap scheduler fine tunes the previous flap speed so that the aircraft arrives exactly at its RTA.

The combination of thrust cut altitude and flap selection speeds ultimately determines the CDA performance. Changes in the one parameter necessarily cause changes in the other, if the *safety goal* is not to be violated. For example, if thrust reduction is delayed (executed at a lower altitude), the aircraft will reach h_R with a speed higher than the approach speed. This can be prevented by selecting flaps at speeds higher than according to the nominal schedule, in order to increase the deceleration rate. The upper and lower bounds of the flap speeds hence define the boundaries of the control space the pilot has during the approach. Within this control space, the pilot can maneuver the aircraft to anticipate or delay his arrival time. The *time goal* requires that the aircraft touches down within a small time window around the RTA. With the aircraft on final approach, the majority of the work needed to reach this goal is already done. The flap scheduler algorithm helps the pilot to fine-tune his exact arrival time. Off-line simulations have shown that for a straight-in continuous descent approach from 7000 ft, 250 kts IAS, this control space is limited to 8-30 seconds, depending on wind conditions[11]. For this reason, it is important that the flight crew is able to steer their aircraft to within these bounds, before they start the descent.

IV. PILOT INTERFACE

Conventional Display

In the base-line condition, the pilot interface consists of a conventional Primary Flight Display (PFD), Navigation Display (ND) and a display showing the Mode Control Panel (MCP). The required information for the time based CDA procedure is printed on two cue cards. This cueing system is loosely based on one developed at the Massachusetts Institute of Technology[18]. This system places gates at strategic locations on the track. These gates correspond to information on thrust setting, aircraft configuration and time. One cue card is designed to focus primarily on the *safety goal*, by providing the pilot with continuous descent parameter information in the final phase of the approach. The card shows a profile view of the track to fly, comparable to conventional approach charts. For four wind speeds (0, 15, 30 and 50 kts) and three wind directions (headwind and crosswind from either side on Final), the thrust cut altitude and speeds for Flaps 5, Flaps 10 and Flaps 20 are given in a table. The pilot has to interpolate between these parameters to match them with the actual situation. The printed wind speeds assume a logarithmic wind profile with the reference wind speed measured at 7000 ft altitude.

The second cue card focusses on the *time goal*, by providing the pilot with time gates at certain waypoints. The card shows a top view of the track to fly, comparable to Figure 1. The time slots at these gates are based on the aircraft following the nominal speed (IAS) profile. Taking the wind conditions

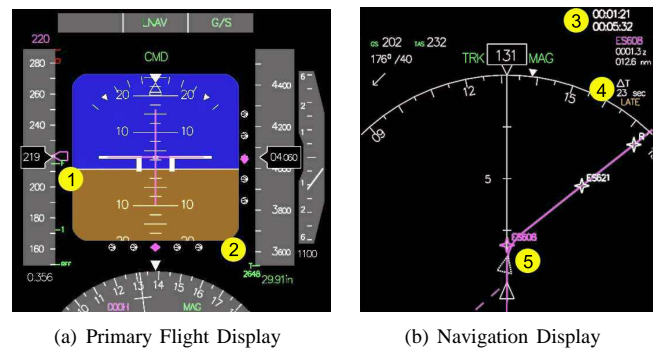


Fig. 5. Display Modifications. (a) PFD, with ① the Flap Cue and ② the Thrust cue. (b) ND, with ③ Elapsed time and RTA, ④ Time performance and EARLY/LATE indicator and ⑤ Ghost symbol.

mentioned above (4 wind speeds, two directions) into account then yields a series of time-over-waypoint datasets. These datasets are displayed in a table.

Augmented Display

The information produced by the prediction and optimization algorithms described in Section III must be presented to the pilot in a logical and intuitive way. Display modifications and augmentations must also be designed to minimize clutter on the current lay-out of displays. The modifications explained below are shown in Figure 5.

One cue was added to help meet the *safety goal*. To indicate the altitude where thrust should be reduced to flight idle in order to meet the approach speed V_{APP} at h_R , a letter 'T' is added on the altitude tape of the Primary Flight Display. This is shown as item ② in Figure 5(a).

To help meeting the *time goal*, a series of display augmentations was introduced. To minimize ΔT in the final phase of the arrival, a letter 'F' is presented on the speed tape of the PFD at the optimal speed for the next flap selection. This is shown as item ① in Figure 5(a). On the Navigation Display, several augmentations are present. The elapsed time since the start of the approach and the Required Time of Arrival are shown as item ③ in Figure 5(b). At ④ an indication of the current situation with respect to the RTA is given as ΔT in seconds, combined with an amber EARLY/LATE indication in case this deviation is greater than 10 seconds. This time indication is also shown in a ghost symbol, through item ⑤. This ghost is an image of the own aircraft flying the intended approach track, keeping a position where the aircraft should be if ΔT were zero. The ghost symbol is a dashed variant of the white (solid line) own aircraft symbol.

V. PILOTED EXPERIMENT

To test the effectiveness of time-based spacing and identify the operational constraints of implementation, a piloted experiment was conducted. Seven professional airline pilots tested the system under various operating conditions in a fixed base flight simulator.

TABLE I
PILOT EXPERIENCE.

	Age	Aircraft types	Flying hours
Pilot A	68	DC3, CV640, DC8, B747-3/400, C550	13200
Pilot B	31	F100	1200
Pilot C	31	B747-400	300
Pilot D	23	PA28, DA42	190
Pilot E	50	military, B737, BA146, DC10, A320	12500
Pilot F	33	B757, B767	6000
Pilot G	26	B747-400	1650

Independent variables

First, two displays were defined. One, corresponding to the baseline condition, was a conventional display layout consisting of a Navigation Display (ND) and Primary Flight Display (PFD). The second display, for the augmented condition, consists of an ND and a PFD extended with information derived from the flap scheduler and optimization algorithms, as described in Section IV.

Second, four different wind conditions are defined that together represent a realistic set of wind conditions that could be encountered during any approach. A typical wind profile is shown in Figure 4. These profiles are all taken from data sets of actual wind measurements, but scaled and rotated to correspond to the wind speeds of interest. These wind conditions, listed in Table II, comprise two headwind conditions on Final, and two crosswind conditions. For each wind direction, two wind speeds at the starting altitude of 7000 ft are defined. The choice for wind speed values is such that the lower value could be encountered under normal, regularly occurring circumstances. The higher wind speed occurs in more rare situations. The wind speeds for crosswind approaches are lower, in correspondence to crosswind and headwind limits for landing.

Experiment design

The experiment design matrix is factorial, combining the four wind conditions with both displays. This yields eight experiment runs per pilot. Seven professional airline pilots (over 4500 flying hours on average, see Table I) flew a set of these eight runs, yielding 56 experiment runs in total. Each set was preceded by four to six practice runs, to familiarize the pilots with the procedure and wind conditions.

Apparatus

The experiment was conducted in the fixed base research simulator at the Control & Simulation Division. The pilots were seated on the co-pilot side, controlling the aircraft with a side stick. The Primary Flight Display, Navigation Display and the Mode Control Panel were shown on two 18" screens. An outside visual was shown of a landscape with a fictitious airport with two parallel runways.

Aircraft

The aircraft model used is a non-linear six-degrees-of-freedom model of the Boeing 747-200. The flight is executed with the autopilot in LNAV mode, leaving only manual pitch

TABLE II
INITIAL CONDITIONS

	Wind speed [kts]	Wind dir. [°]	Airspeed [KIAS]	Distance to go [nm]	RTA [min:sec]
A	26	90	270	47.7	11:23
B	26	180	270	46.3	11:46
C	44	180	250	31.6	09:39
D	12	90	250	30.8	07:59

control and throttle control to the pilot. The reason for maintaining partial manual control throughout the flight was to introduce a basic level of workload during the approach. With autopilot engaged and no radio traffic or ATC present, an experiment run would comprise a lot of idle time between autopilot inputs. For guidance along a three-dimensional glide path, a Microwave Landing System (MLS)-type vertical guidance was available. This allows the pilot to fly a continuously descending path, irrespective of his position with respect to the runway. To improve lateral stability, a yaw damper was added.

Scenario

Initial conditions, such as position along the track and airspeed, are varied per wind condition, to limit the influence of learning effects on the way the track is flown. The required arrival times are tuned to each initial condition, based on a relative deviation from the nominal RTA for that condition, so that only the effects of the wind conditions influence the pilot's performance. The initial conditions are listed in Table II. In the level flight segment following each initial condition, the pilot can position the aircraft as good as possible for meeting the RTA. This is done by choosing a higher airspeed (in case a pilot is late), or a lower airspeed (in case a pilot is early) than the 220 kts chosen for the nominal speed profile. The pilot then has to maintain this airspeed until his ΔT is reduced to zero, after which he can return to the nominal 220 kts until the point of thrust cut.

Procedure

The pilot's task is to fly a Continuous Descent Approach, while meeting both *safety* and *time goals*. The aircraft starts in one of the initial conditions listed in Table II at an altitude of 7000 ft, with Flaps 1° extended and with autopilot and autothrottle engaged. The pilot is given an RTA, which is entered through an interface window in the Mode Control Panel. After this, the autopilot is switched to LNAV mode, and the autothrottle is disengaged.

Dependent measures

The dependent variables consist of both objective and subjective measures. Objective measures include operational performance and pilot control activity. Operational performance was judged by measuring the accuracy with which the targets were reached. For the *safety goal*, the deviations from the final approach speed V_{APP} at the reference window and in the remainder of the approach were measured. For the *time goal*, the deviation from the RTA was measured.

TABLE III
DEPENDENT MEASURES

	Measure	Description
	ΔV_{APP}	Deviation from V_{APP} at 'R'
<i>Safety goal</i>	ΔV_{final}	RMS of the deviation from V_{APP} at 'R'
	Flap setting	Aircraft should be fully configured at 'R'
<i>Time goal</i>	ΔT	Deviation from the RTA at 'R'
<i>Workload</i>		Number of throttle setting changes NASA Task Load Index (TLX) score

Pilot control activity was measured by counting the number of thrust changes during a run. The subjective measures were taken from a questionnaire, aimed at giving an insight into pilot acceptance of the system, and a NASA Task Load Index (TLX)[19] sheet to assess pilot workload for each run. The dependent measures are listed in Table III.

Hypotheses

The first hypothesis is that for the augmented display time performance will improve with respect to the baseline condition. The reasoning for this is that with the help of the support system, the pilot has continuous information about his *time goal* performance at hand, which will allow for smoother and more accurate transitions between the different phases in the flight. When time information is only provided at discrete points (the gates), own performance estimation is clearly more difficult.

Secondly, CDA performance (reaching the approach speed of 150 kts at 'R', preferably no sooner but definitely not later) is expected to increase. The idea is that this performance mainly depends on the moment of thrust cut. The altitude where this is done depends heavily on the wind speed and direction on Final, so a more accurate prediction of this wind profile will increase the *safety goal* performance.

Finally, it is hypothesized that workload will be higher for the baseline condition, since in this case the pilot will have to interpolate continuously between the data on his cue cards to retrieve the appropriate parameters. Moreover, the wind used to set up the cue card data resembles standard logarithmic profiles. The discrepancy between this profile and the actual wind will require extra corrective pilot action, hence increasing the workload.

VI. RESULTS AND DISCUSSION

A. Operational Performance

Two types of performance measures are selected, each related to either the *safety goal* (mandatory performance targets) or the *time goal* (optimization). It appeared throughout the experiment that the variation in wind *speed* does not have a significant influence on performance ($F_{2,6} = 0.244$, $p = 0.784$ for the *safety goal*, $F_{2,6} = 1.930$, $p = 0.156$ for the *time goal*). For that reason, the four wind conditions are reduced to two clusters, defined by the wind direction.

Safety goal: To investigate how well the aircraft is established for landing at the reference window, three performance parameters are defined. The first one is the deviation from the target approach speed of 150 kts IAS, when passing the reference window (waypoint 'R', 3.14 nm from touchdown, 1000 ft altitude). The means and 95% confidence intervals for this score per wind condition (headwind or crosswind) are plotted in Figure 6(a). In this figure the deviation for the optimized configuration is lower in both headwind and crosswind, but this effect is obscured by the large spread in the data. An analysis of variance (ANOVA) shows that this spread is indeed too large to see any differences; the effect of the display configuration on the speed deviation is not significant ($F_{1,6} = 0.023$, $p = 0.883$).

What can be seen from the error bar plots, is that crosswind has a negative effect on safety performance. Although most of the time hidden by the relatively large variances, this effect is significant for the deviation from V_{APP} at 'R', when the optimized display configuration is used ($F_{1,6} = 9.160$, $p = 0.023$). This can be explained by the fact that the display optimization only uses the headwind component of the estimated wind speed to predict its time and speed profile. In a headwind condition, this works out well, but in a crosswind on Final, an error is introduced: the algorithm assumes a near-zero headwind component, while in reality the aircraft needs to compensate for the crosswind in order to stay on its ground track. As this is usually done by 'crabbing' the aircraft, the actual ground speed will be lower than the predicted ground speed. As a consequence, the airplane starts lagging behind the original predicted time profile, which in turn causes pilots to delay flap selection in order to maintain a higher airspeed. In many cases, the *safety goal* suffers from this decision, with the average approach speed going up from little over 151 kts to 156 kts. For the baseline condition, the data on the cue cards is based on trial runs, instead of predictor data. So, since the accuracy of the wind information on the cards (based on logarithmic profiles) is always the same, the deviation from V_{APP} is not influenced by the direction of the wind.

The fact that speed performance goes down when the track prediction is less accurate (in the crosswind condition) suggests that pilots closely followed the cues presented on the displays. At the same time, the large spread of the speed performance in the baseline condition indicates big differences in the flying strategies, adopted by each pilot. This is confirmed by the pilots' answers to a questionnaire, which showed that they used the CDA-parameter cue card mostly to determine the thrust cut altitude, but thereafter relied more on their pilot experience to determine flap selection. In the optimized display configuration, the need to look away from the instruments to check the cue cards is eliminated, and pilots use the displayed instructions.

Time goal: The main check on the time performance is the deviation from the RTA at the reference window. The means and 95% Confidence Intervals for this parameter are shown in Figure 6(b). Clearly, the average time performance is better with the aid of display optimization. An ANOVA

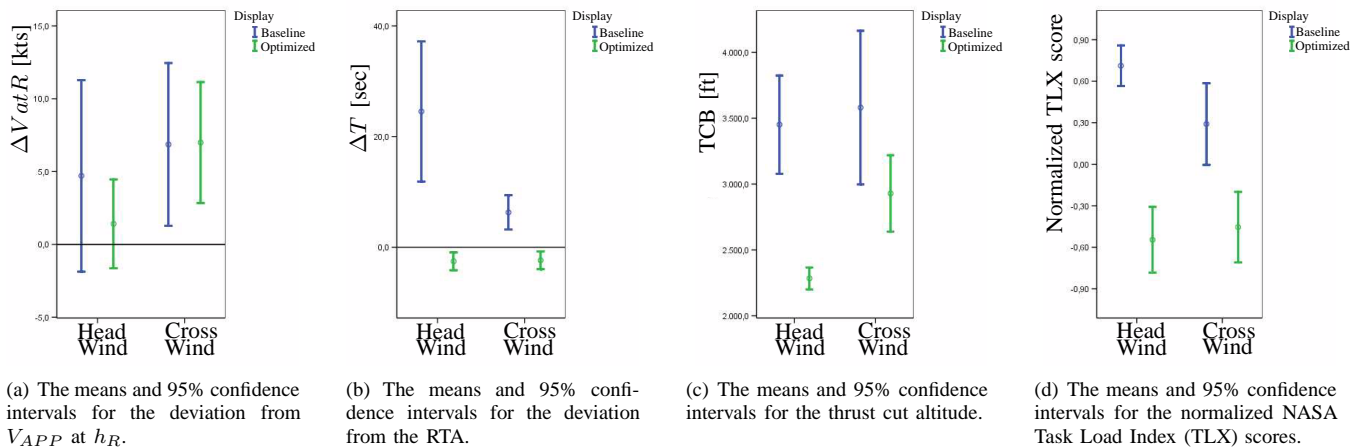


Fig. 6. CDA performance scores.

shows that this influence is indeed significant ($F_{1,6} = 7.368$, $p = 0.033$). Average time performance is within 4 seconds of the RTA. This is definitely accurate enough to guarantee safe separation. This result would allow air traffic controllers to space incoming traffic more tightly, increasing runway throughput capacity. The addition of the proposed automation to existing cockpit displays enables pilots to fly a time based CDA.

The explanation for this improvement is twofold. First, by providing the pilot with continuously updated information on his ETA status, he is able to very accurately adjust the aircraft's speed profile to minimize ΔT . In the baseline, the number of intermediate time gates is limited to four (for conditions C and D in Table II) or five (for conditions A and B). Second, the more accurate wind prediction that is available in the enhanced display condition yields a better estimate of the optimal altitude for thrust reduction. This can be seen in Figure 6(c), showing the altitude where pilots cut back on thrust to decelerate the aircraft to 150 kts. Regardless of wind direction, this altitude is significantly lower in the optimized display condition than in the baseline condition ($F_{1,6} = 3.560$, $p = 0.098$). With the used wind profile class in mind (see Figure 4), it becomes clear that the logarithmic profile prediction is not able to deal with the increase in wind speed around 3500 ft. The aircraft encounters more (head)wind than expected, so its thrust cut altitude should be lowered. The track prediction and optimization routines take this effect into account.

Although the automation provided a major improvement in time performance, speed performance (deviation from V_{APP}) was still in the same range as in the baseline situation. This might be due to the fact that the automation mainly focussed on achieving the time goal. To strike a better balance between the two performance criteria, and to provide a more logical lay out of the presented clues, it is recommended to let the cues on the Primary Flight Display focus on CDA performance (arrive stabilized at the reference window), while the cues on the Navigation Display (ghost and ΔT) focus on

the time goal. Another improvement would be to integrate the cues more tightly with current procedures. For example, in many aircraft the landing gear should be lowered between two fixed consecutive flap settings. To incorporate such information in the pilot support system would ease implementation and increase acceptability by airlines and flight crew.

B. Pilot workload

The results for the subjective workload measurements are represented by the normalized NASA TLX rating scores in Figure 6(d). From this it becomes clear that the workload experienced by the pilots is lower for the optimized display condition, in all wind conditions. This effect is highly significant ($F_{1,6} = 50.390$, $p < 0.001$), an observation that is confirmed by the questionnaire answers with all pilots indicating a higher workload in the baseline condition. In contrast, the workload in this type of noise abatement procedure with the optimization present was considered comparable to that of a conventional ILS approach.

Furthermore, it can be seen that for the baseline condition, the workload score also depends on the wind direction. The score is significantly lower in crosswind conditions on Final ($F_{1,6} = 12.797$, $p = 0.012$). Several factors could influence this phenomenon. First, in a headwind condition, the full force of the sharp changes in wind speed along the altitude profile is felt. In a crosswind, only one component of the wind is of influence, so the absolute change in wind speed is smaller. Another factor might be that the crosswind condition on Final means a headwind condition between waypoints TILDA and EH608 (see Figure 1). In many cases, this is the phase where pilots reach the on-time schedule ($\Delta T = 0$ and reduce speed from 250 or 270 kts to the nominal speed of 220 kts. This is not always an easy task, since a B747 in this situation (Flaps 1° , gear up, 3° glide path, no speed breaks) has a very low deceleration rate. A headwind in this situation reduces the kinematic flight path angle, which makes it easier for the pilot to decelerate the aircraft. This increases the chances of a stabilized approach and hence reduces pilot workload.

Flying a CDA already puts more demand on the flight crew than a regular arrival, where the basic control task is following ATC vectors. Although the tasks of flying and navigation are normally shared between the Pilot Flying and the Pilot Not Flying, the persons interviewed indicated that they would find the sharp increase in workload as experienced in the base-line scenario unacceptable.

VII. CONCLUSIONS

This paper investigated the feasibility of introducing time based spacing in realistic, three dimensional continuous descent approaches under actual wind conditions. For reasons of safety, it is important that (1), the continuous descent ends in a stabilized approach configuration and speed some distance before the runway and (2), pilots adhere to the required arrival times demanded by ATC, in order to maintain safe separation throughout the approach.

The development of an FMS based prediction and optimization system combined with a pilot support interface enables the flight crew to reach these two goals. A wind prediction algorithm that makes use of weather information broadcast by other aircraft in the TMA makes accurate wind profile prediction along the approach trajectory possible. Wind speed prediction error along a 30-45 nm approach trajectory is 2 kts. An accurate knowledge of the wind ahead makes sure an optimal thrust cut altitude and flap speed schedule can be selected. The piloted experiments show a strong improvement in time performance when continuously updated information on this goal is present on the display. Pilots are able to reach their RTAs within an average margin of 4 seconds, regardless of wind conditions. The average speed performance (being stabilized at a reference window) is unaffected by display optimization, although the presence of automation reduces the spread in this performance criterion. The addition of time constraints without extra aids would result in an unacceptable increase in workload, due to the continuous calculation and interpolation the pilots have to perform. Workload is significantly lower with the display optimization present, and at the same level as in current ILS approach procedures.

VIII. RECOMMENDATIONS

The piloted experiment shows the feasibility of time-based continuous descent approach procedures under realistic wind conditions, along a fixed trajectory. The pilot's control space to reach an RTA could be greatly enlarged by adopting a more flexible approach route. Shortening or extending the length of the track to fly, similar to the 'tromboning' technique currently used by ATC, should reduce the need for speed changes, thereby improving predictability of the speed profile during the approach.

REFERENCES

- [1] ICAO, "Outlook for air transport to the year 2015," International Civil Aviation Organization, 2005.
- [2] —, "Assembly resolution a33-7: Consolidated statement of continuing icao policies and practices related to environmental protection," International Civil Aviation Organization, 2001.
- [3] E. H. Bolz and L. M. Reuss, "System study of advanced operational procedures for noise reduction," NASA (Task Order (TO) 56), Tech. Rep., 2001.
- [4] SOURDINE, "Sourdine ii, d9-1, final report," Tech. Rep., 2006.
- [5] A. D. Kershaw, D. P. Rhodes, and N. A. Smith, "The influence of ac in approach noise abatement," *3rd USA/Europe Air Traffic Management R&D Seminar Napoli*, 2000.
- [6] L. J. Erkelens, "Research into new noise abatement procedures for the 21st century," *AIAA GNC Conference and Exhibit, 14-17 August 2000, Denver (CO)*, 2001.
- [7] J. P. B. Clarke, N. T. Ho, L. Ren, J. A. Brown, K. R. Elmer, K. O. Tong, and J. K. Wat, "Continuous descent approach: Design and flight test for louisville international airport," *Journal of Aircraft*, vol. 41, no. 5, 2004.
- [8] F. Wubben and J. Busink, "Night time restrictions at amsterdam-schiphol," to 70 Aviation & Environment, under Contract of the Ministry of Transport, Public Works and Water Affairs, Tech. Rep., 2004.
- [9] L. Ren, J.-P. Clarke, and N. T. Ho, "Achieving low approach noise without sacrificing capacity," *22nd Digital Avionics Conference, 12-16 October 2003, Indianapolis (IN)*, 2003.
- [10] Anon., "Schiphol rnav night instrument approach chart, sugol, river and artp transitions to rwy 18r," Air Traffic Control the Netherlands, Tech. Rep., 2006.
- [11] W. F. de Gaay Fortman, M. M. van Paassen, M. Mulder, and A. C. in 't Veld, "Design of a pilot support system to implement time-based spacing for decelerating approaches," *AIAA GNC Conference and Exhibit, 15-18 August 2005, San Francisco (CA)*, 2005.
- [12] M. F. Koeslag, "Advanced continuous descent approaches," Master's thesis, Faculty of Aerospace Engineering, Delft University of Technology, 1999.
- [13] P. S. Maybeck, *Stochastic Models, Estimation and Control*. Academic Press, 1979, vol. 1.
- [14] P. Frenzen and C. A. Vogel, "On the magnitude and apparent range of variation of the von karman constant in the atmospheric surface layer," *Boundary-Layer Meteorology*, vol. 72, 1995.
- [15] D. Painting, "Amdar reference manual," World Meteorological Organization - AMDAR Panel, Tech. Rep., 2001.
- [16] K. F. M. Schippers, J. L. de Prins, M. M. van Paassen, M. Mulder, A. C. in 't Veld, and J.-P. Clarke, "Investigation of a three-degree decelerated approach of a twin engine jet aircraft under actual flight conditions," *AIAA GNC Conference and Exhibit, 15-18 August 2005, San Francisco (CA)*, 2004.
- [17] A. C. in 't Veld, M. M. van Paassen, M. Mulder, and J.-P. Clarke, "Pilot support for separation assurance during decelerating approaches," *AIAA GNC Conference and Exhibit, 16-18 August 2004, Providence (RI)*, 2004.
- [18] N. T. Ho, "Design of aircraft noise abatement approach procedures for near-term implementation," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, 2005.
- [19] S. G. Hart and L. E. Staveland, "Development of nasa-tlx (task load index): Results of empirical and theoretical research," *Human Mental Workload*, pp. 239-250, 1988.

Developing a decision-support-tool for an air taxi service

A research proposal to develop a decision support tool to analyze an air taxi service on strategic and operational level

Peter A. Sengers BSc.

Dep. of Technology, Policy and Management, section for
Transport and Logistic Organization
Delft University of Technology
Delft, the Netherlands
p.a.sengers@student.tudelft.nl

Stefaan S.A.Ghijis MSc.

Dep. of Aerospace Engineering, Aerospace Management
and Operations
Delft University of Technology
Delft, the Netherlands
s.s.a.ghijis@tudelft.nl

Abstract—this paper is a research proposal to develop a tool to analyze logistic concepts of the air taxi service of Aeolus Aviation in different scenarios. Based on this analysis recommendations can be done for a suitable logistic concept for Aeolus. Based on background analysis of the air taxi service three objectives are formulated; analyzing the air taxi service on strategic level, developing a decision support tool to analyze logistic concepts and finally developing a suitable logistic concept for Aeolus Aviation. Based on these objective research questions are formulated and research methods to answer these questions are given.

Keywords-component; Air taxi, complexity, Research proposal, decision support tool, simulation

I. INTRODUCTION

Business travelers are more and more concerned of efficient flying. When time and money can be saved when traveling, value is added to their business processes. Recent events in the aviation have not made it easy to gain efficiency. For example the security measures after 9/11 have increased airport congestion. But also the congestion on the road decreases the efficiency of traveling. These developments point to an increase demand for efficient and low cost means of ad hoc point to point transportation.

Aeolus Aviation will provide efficient traveling solutions in Western Europe, adding value to customer processes. Where competitors focus on luxury, Aeolus has a more no-nonsense approach: cost and time efficiency. In this effort Aeolus will continuously attempt to improve efficiency and accurately match capacity with demand. The main goals are to develop an air taxi service, which has the lowest price and travel time and the highest profit. Because many factors are still unknown it is hard to develop an optimal logistic concept of the air taxi service. A logistic concept of an air taxi service is a certain service concept providing transport between selected airports influenced by a certain schedule for the crew, a certain price

structure, a certain amount of aircrafts with maintenance schedule and a certain location of a maintenance base. To find a suitable logistic concept for the air taxi service of Aeolus a decision support tool can be built to analyze the logistic concepts. First the air taxi service of Aeolus Aviation will be strategically analyzed. What are the potential customers? Where are those customers? And what are the competitors? Secondly on operational level a tool will be developed to analyze logistic concepts of the air taxi service. When this tool is developed logistic concepts will be developed and tested to come up with a suitable concept for the air taxi service.

In the next chapter the background of the purposed air taxi service of Aeolus Aviation will be analyzed. In the last chapter the research project will be discussed. Based on the background analysis the research objectives will be formulated. Next the main research question and the respective definitions will be described. Based on the research objectives and the main research question a research approach is developed. Finally the research methods and planning will be discussed.

II. BACKGROUND ANALYSIS

This paper is a project proposal of the logistic concepts of the air taxi service of Aeolus Aviation. To get a better view of the background of an air taxi service in paragraph A the air taxi concept in general and the advantages and disadvantages will be discussed. In paragraph B the complexity of operating an air taxi service will be elaborated to get insight into the factors influencing the air taxi service. In paragraph C Aeolus Aviation and their plans will be discussed.

A. The air taxi service concept

To get a better idea of an air taxi service, in this paragraph the air taxi service concept will be discussed. First the concept of the air taxi service will be described; next the strengths,

weaknesses, opportunities and threats will be analyzed by means of a SWOT analysis.

The demand for air travel is increasing [1]. This larger flow of passengers combined with more strict security measures results in congestions at the major airports, resulting in irritations of business travelers[2]. But on the road an increase of congestion is remarked as well and transport by rail only reaches the bigger cities. A solution for these developments is an air taxi service: on demand point to point business aviation for short distances. This air taxi service results in shorter waiting times at airports, the possibility of choosing the departure time, pick up and deliver at an airport closer to the destination. To achieve this, a large amount of smaller, 10-200, aircrafts will be used to fulfill the demand. Such aircrafts are: the new very light jets, piston and propeller aircrafts. These aircrafts can normally carry 4 to 6 persons, as a shared flight or for a group, can be operated just by one pilot, depending on the type of aircraft, and can land on smaller airports than bigger planes. They are cheap in comparison with current similar jets and have lower operating costs, due to the lower fuel consumption [3]. With significant lower costs than the current unscheduled air service operators, these new air taxi operators promise to provide low fare on-demand personal transportation. The concept of an air taxi is born.

To get a better view of the positive and negative sides of an air taxi service in table 1 an SWOT analysis is shown. In a SWOT analysis the strengths, weaknesses, opportunities and threats are analyzed. The strengths of air taxi services are the efficient and lower door-to-door travel time, its reliability and its personalized way of transport. The efficient and lower door-to-door travel time is the result of the possibility to use a wider set of airports. When using a wider set of airports it is possible to use uncongested, smaller airports closer to the customer's door. The difference in travel time will even become larger when the capacity crisis at general airports will generate more delays. An air taxi service is more reliable because the amount of delays will be lower.[4] The personalized way of transport by an air taxi is a strength as well. Customers can choose their own trip and are not depended of other customers. Weaknesses are the complexity, which can cause inefficiency for an air taxi service company [5] and the price of a ticket, which is still higher than the price of scheduled flights. The complexity of an air taxi service will be discussed in the next paragraph. Another weakness is the dependency on the weather and daylight. For example, according to JAR OPS 1 a single engine propeller aircraft cannot fly when the visual range is low, for example in bad weather or at night. JAR OPS 1 is the Joint Aviation Requirement for the operation of commercial air transport (aero planes). Any commercial airline within the European Union flying jet or propeller aircraft has to comply with this standard. The opportunities of an air taxi service are the better affordability for customer in comparison to current comparable services. The better affordability, in comparison to the past, is the result of the lower acquisition and operating costs, lower airport landing fees, but also the result of operational efficiencies. Other opportunities for the success of an air taxi service arise when the congestion on the road and at the airport increases and the demand for efficient flying increases as well. Another opportunity is the decreasing of the strictness of

regulations in Europe. This will result in lower costs and more possibilities. The threats are congestion in the air. When more airlines will provide in air taxi flight, more aircrafts will be in the air, which can cause delays. Another threat is the possible safety problems. When more aircrafts are in the sky, the sky will become busier and less safe. Another threat is the entrance in the market of large scheduled airlines. Differentiation of those airlines will be a threat because those airlines will have a large budget. Finally ageing of the present pilots and more upcoming airlines can result in a scarcity of pilots in the future.

TABLE I. SWOT ANALYSIS OF AIR TAXI SERVICE

Strengths	Opportunities
Efficient and lower travel time	Lower price for customer
Reliable for customer	Increased airport congestion
Personalized transport	Increased road congestion
	Decreasing strictness regulations
Weaknesses	Threats
Complexity for Aeolus	Congestion in the air
Higher price than scheduled service	Safety problems
Depending on weather/daylight	Entrance in market of air taxi services by large scheduled airlines
	Forecasted scarcity of pilots

B. Complexity of an air taxi service

As mentioned in the previous paragraph the air taxi service business is complex, which can cause inefficiencies for the air taxi airlines. In this paragraph the complexity will be elaborated to get a better view of the aspects influencing the air taxi service. This complexity can be divided into three levels: strategic, tactic and operation. Figure 2.2 [6] is based on the flight schedule development process of scheduled airlines of Bootsma and is adapted to the air taxi service of Aeolus. In Figure 2.2 left the general process of planning is shown for each level. As shown on strategic level the market will be analyzed to end up in a plan. In the case of a scheduled flight on tactical level the flight schedule will be developed. On operational level that schedule will be executed. For an air taxi service the tactical level can be combined with operational level because the activities on tactical level of a scheduled service happens during the operational level of an air taxi service, because there is no schedule. In figure 2.2 on the right the process of an air taxi service is presented. On strategic level the service concept, existing out of service between different airports, and fleet planning will be developed and on operational level this service concept of the air taxi will be

executed. First the complexity on strategic level will be described, second the complexity on operational level will be discussed.

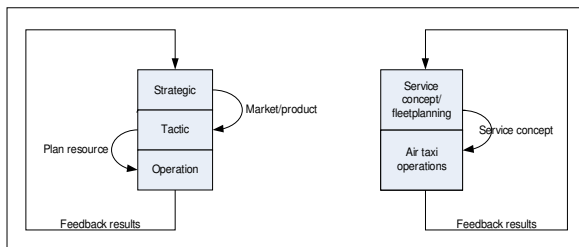


Figure 1. Flight schedule development of scheduled airline

1) Complexity on strategic level

In this sub-paragraph the complexity of an air taxi service on strategic level will be discussed, based on the strategic activities of an air taxi. First the complexity of the demand will be described, next the complexity of the airports will be discussed and finally the complexity of revenue management of an air taxi service airline will be discussed.

a) Complexity of demand

The first factor influencing the complexity is forecasting the demand of on demand air taxi flights. Analyzing the demand for an air taxi service is different than in the case of a scheduled service. First of all the demand for scheduled flight is larger than for air taxi services. Second for scheduled services the fulfillment of travel requests are driven by the supply in terms of the amount of seats on given airline network mostly defines months ago. The demand for air taxi service is driven by individual demand. The passenger can determine the origin, destination and earliest departure time and latest arrival time without being constrained by flight schedules.[7] The individual demand of an airport is hard to determine, because it is very insecure. To make a strategic choice regarding the environment of the air taxi service, the market for air taxi services is based on [8]:

- Type of potential clients:
- Origin and destinations of potential clients
- Competitors

b) Complexity of airports

The complexity of the airports is determined by the constraints of the operated airports. Constraints of airports are:

- Closing times
- Accessibility
- Runway length
- Fueling possibilities

- Landing fees
- Other regulations

Partners are playing a part as well. When developing a door to door service, cooperation with other transport services is needed. Therefore more partners need to be considered when developing an air taxi service.

c) Complexity of service concepts

A service concept of an air taxi service consists of a set of airports with the constraints belonging to those airports and a transport service transporting the passengers by aircraft. The service concept of an air taxi service is complex because an air taxi aircraft flies not often the same route twice consecutively and generates less there and back traffic than scheduled flights: it is a more unbalanced network. It will be unlikely that the arrival airport of a revenue flight corresponds to the departure airport of other revenue flights. Therefore non revenue flights, also called repositioning flights, must exist to connect the destination airport of one revenue flight to the departure airport of the next revenue flight operated by the same aircraft. Because repositioning flights are only generating costs and no revenue and so have influence on the ticket price these repositioning flights needs to be minimized [9]. Several network structures can be developed. Examples of structures are:

- Shuttles between airports with a high stable demand
- Operating a large amount of airports in Western Europe
- Operating a limited amount of airports in Western Europe with a high demand

d) Complexity of ticket pricing

When determine the price of a ticket for an air taxi flight several techniques are possible:[10]

- Ticket price based on fixed costs per mile
- Ticket price based on forecasted demand (revenue management scheduled service)
- Ticket price based on true costs per action plus profit margin (revenue management air taxi service)

Revenue management techniques are techniques to optimize the revenue earned from a fixed, perishable resource. The challenge is to sell the right resources to the right customer at the right time[11]. The demand for scheduled flight is known earlier and can be used easier for pricing of tickets. Revenue management for air taxi flights can be compared with the ticket pricing technique based on true costs per action plus profit margin depends more on the demand at other airports and than for scheduled flights. When implementing revenue management for air taxi service it is important to know the demand on other days and at other airports to minimize repositioning flights. For example when is known that tomorrow the demand at airport b will be large and today it isn't and the customer wants to go from airport a to b, Aeolus can increase the price, because the chance of a repositioning flight is big, or the customer can be advised to

a) Market of Aeolus Aviation

As been mentioned the market in which an air taxi service will operate depends on the potential customers, where these clients are and what the competition is. Aeolus will focus on customers more sensitive to higher costs and until now they have not thought of the advantages of business charters. But because of the low price and travel time Aeolus will be remarked. Aeolus wants to attract customers from scheduled flights where time-efficiency is lacking, due to passengers handling and mass transport. And Aeolus wants to attract customers from road transport, where time-efficiency is lacking, due to congestions of the roads. The specific potential customers and their locations and their influence are unknown, so are the competitors. In the near future an extensive market research will be done.

b) Type of aircrafts

To achieve the objectives Aeolus Aviation uses a different type of aircraft than most air taxi services in Europe. Most of the air taxi services are using jets, but Aeolus Aviation is going to use propeller aircrafts. These propeller aircrafts can be especially used for flights of 300-1000 km. In figure 2.3 can be seen that the demand for air taxi flights with a turboprop starts to increase for flights of 100 km and approximately 50 percent of the business flights are flight of less than 500 km. After that the demand is decreasing, but still present till flights of 1000 km. Thus for a propeller aircraft a large amount flight can be executed on a range of 200 to 1000 km. IFR movements are movements according to the instrument flight rules.

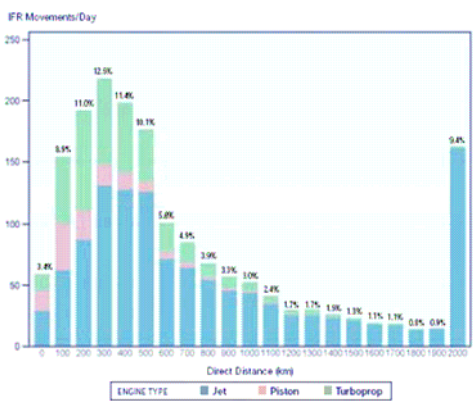


Figure 3. : IFR Movements/day[15]

Especially for these distances the turboprop aircrafts have advantages. The advantages are that propeller aircrafts:

- have cheaper types to purchase
- are economic
- have a shorter runway length

The disadvantages are that propeller aircrafts:

- less comfortable
- slower[16]

- have stricter regulations

But the lower price for flights will compete with these disadvantages, because the customer's top priority is the lowest price[17].

The amount of aircrafts is not yet determined. It will be likely that the amount of aircrafts will increase throughout the years, so that needs to be considered in the trade off of the different logistic concepts.

c) Pricing of tickets

Another difference between Aeolus and other air taxi services is the fact that passengers pay for the aircraft and not for a seat. Seat sharing will therefore not be possible. This will influence the pricing of tickets of the air taxi service of Aeolus Aviation. Which pricing technique Aeolus will use is not yet determined.

Other customer service specifications of Aeolus Aviation are[18]:

- Door to door air service
- Within 6 hours airplane is ready to take-off everywhere in Western Europe
- Flexible route planning
- Capacity for 6 passengers
- Hotel arrangements when necessary

2) Aeolus Aviation on operational level

Operations are needed to fulfill the strategy of Aeolus. Off course providing transport for customers is the main operation of an air taxi service. But the transport of these customers is influenced by many factors. In paragraph 2.2.2 the most important factors influencing the air taxi service on operational level are described. These are:

- Crew scheduling
- Aircraft characteristics
- Maintenance scheduling
- Location of the maintenance base

All these aspects, together with the service concepts between airports, can be combined into logistic concepts of the air taxi service of Aeolus Aviation.

Thus a logistic concept of an air taxi service is a certain service concept providing transport between selected airports influenced by a certain schedule for the crew, a certain price structure, a certain amount of aircrafts with maintenance schedule and a certain location of a maintenance base. Several logistic concepts can be developed, so it is not easy to develop an optimal logistic concept of the air taxi service network of Aeolus Aviation with the lowest costs and the highest revenues. A tool needs to be developed to analyze logistic concepts to eventually come up with a suitable logistic concept.

D. Conclusions

Aeolus Aviation is starting up an air taxi service. Operating an air taxi service is complex on strategic and on operational level. The complexity on strategic level is caused by insecurity of the demand, constraints of airports, minimization of the repositioning flights and pricing techniques. The complexity on operational level is caused by the scheduling of the crew and maintenance and the location of a platform. So many factors are influencing the logistic concept of an air taxi service network.

The main goals of Aeolus Aviation are to develop an air taxi service, which has the lowest price and travel time for flights and the highest profit. Because many factors are still unknown, it is hard to develop an optimal logistic concept with the lowest costs and highest revenues. A tool is needed to model logistic concepts and to analyze the factors influencing the air taxi service of Aeolus Aviation

III. RESEARCH PROJECT

As mentioned in the previous chapter many factors influencing the air taxi service of Aeolus Aviation are still unknown and it is hard to find an optimal logistic concept. Therefore in this research the discussed issues will be further investigated to eventually find a feasible solution for a logistic concept of the air taxi service of Aeolus Aviation.

In this chapter the research project will be discussed. First the main objectives of the research will be mentioned. Next the main research questions and definitions are formulated and finally the research approach with sub questions will be discussed.

A. Research objectives

The analyses of the previous chapter result in three objectives:

- *Analyzing the air taxi service of Aeolus Aviation on strategic level*
- *Developing a tool to analyze logistic concepts of the air taxi service*
- *Developing a suitable logistic concept for the air taxi service of Aeolus Aviation*

B. Main research question and definitions

In order to reach the objectives, mentioned in the previous paragraph, the following main research question need to be answered:

How to analyze logistic concepts of the air taxi service of Aeolus Aviation on strategic and operational level using a decision support tool and what logistic concept would be suitable for Aeolus Aviation?

The strategic level of the air taxi service of Aeolus Aviation can be analyzed by answering the following sub questions.

- *What are the requirements, assumptions and criteria influencing the logistic concept of the air taxi service network of Aeolus Aviation?*
- *What are the potential clients for air taxi flights of Aeolus Aviation and what are the most important regions in Europe to operate from to be able to reach them?*
- *What are the most important competitors in the network and what is their influence on the service concepts of Aeolus Aviation?*
- *Which airports in those regions are most suitable to operate from, regarding Accessibility, regulations, costs, future developments of those airports and cooperation with other parties?*

To be able to develop a tool to analyze logistic concepts of the air taxi service the following sub questions need to be answered.

- *What are the requirements for the tool to model the logistic concepts?*
- *How can the network of airports be modeled?*
- *How can the demand for air taxi flights be modeled?*
- *How can the crew schedules be modeled?*
- *How can the maintenance schedules be modeled?*
- *How can the expected costs and revenues be modeled?*

To be able to develop a logistic concept suitable for Aeolus Aviation the following sub questions need to be answered.

- *What logistic concepts of the air taxi service can be developed, regarding the strategy of Aeolus Aviation?*
- *What logistic concept fulfills the requirements and scores best on the criteria?*
- *What recommendations can be done, regarding a suitable logistic concept of the air taxi service of Aeolus Aviation*
- *What would be the implementation trajectory for this logistic design of the air taxi service network?*

C. Research methods

In this paragraph the methods to answer all sub questions will be described. For each part of the research the research methods will be described.

1) *Methods to analyze the strategic level of the air taxi service*

To analyze the requirements, assumptions and criteria influencing the air taxi service of Aeolus Aviation the environment of Aeolus will be analyzed in detail. Several requirements, assumptions and criteria are already mentioned in the background analysis. Existing air taxi services or researches about it, in Europe, but in the USA as well, can

provide more information about the requirements, assumptions and criteria.

When analyzing the strategic level of the air taxi service, the first step is to determine the type of potential customers and competitors for the air taxi service of Aeolus and the location where those customers are. The market of the air taxi service of Aeolus Aviation will be analyzed by means of interview with mr. Ghijs. Mr. Ghijs has the knowledge of his potential customers and competitors. Based on that interview the customers and competitors will be analyzed in detail by means of a literature research

Based on the market analysis the most important regions can be determined by means of a research on data of demographic factors. An example for such a data base is Eurostat and interviews with professionals in the business.

Within these regions airports needs to be chosen. Data for this choice will be gathered by interviews with pilots and an analysis of the airport as well on the internet or by phone. The pilots will have knowledge about airports in West Europe. To eventually make a choice between airports this data is used in multi criteria analysis.

2) *Methods to develop a tool to analyze logistic concepts*

To be able to indicate the requirements to model logistic concepts all variables influencing the air taxi operations, which have to be modeled, will be identified. Based on these requirements the decision support tool can be built.

To develop a tool to analyze the logistic concepts simulation software will be used. With the ARENA, simulation software of Rockwell Software, the network of airports and the factors influencing this network can be simulated. The choice for a simulation model is made because in this way the current situation can be analyzed and experiments can be done to distinct the best alternative transport manner. ARENA is developed based on flow oriented simulation. This means that a real situation can be presented as a chain of delays and operations an entity needs to go through [19].

3) *Methods to develop a suitable logistic concept*

To be able to generate logistic concepts the strategic analysis of the air taxi service will be used and interviews with professionals in the business will be done. This will finally end up in several logistic concepts.

These logistic concepts will be modeled to compare them to each other. The output of these models will be analyzed.

This will end up in recommendations for Aeolus Aviation. A part of these recommendations is the implementation trajectory. The implementation trajectory will be determined by means of a literature research and interviews with professionals in the business..

D. *Status of Project*

At this moment the analysis of the strategic level of the air taxi service is almost finished. A questionnaire is send to several parties to ground the choice for important regions of potential customers. Based on demographic data and busy business travel routes the most important cities are determined. Based on these cities and the constraints mentioned in the research proposal the most important airports are selected.

The development of a simulation tool is in progress. This is done with the Rockwell software ARENA. ARENA is discrete simulation software and is based on flow-oriented simulation. ARENA is perfectly able to show the consequences of processes during air taxi operations.

ACKNOWLEDGMENT

I want to thank Ir. Stefaan Ghijs for the opportunity to do the internship at Aeolus Aviation and for the good assistance. I also want to thank the teachers at the University of Technology in Delft for the assistance during my msc Thesis.

REFERENCES

- [1] Eurocontrol, Long term Forecasts, 2006
- [2] Cordeau J.F. et al. 2007, Transportation on demand, Handbook in OR &MS. Vol 14
- [3] Bonnefoy, Philippe A. 2005, Simulating air taxi networks.
- [4] Bonnefoy, Philippe A. 2005, Simulating air taxi networks.
- [5] Gaarlandt et al. 2006. On demand airline, the future of personal transportation
- [6] Bootsma P.D. (1997), Airline flight schedule development –analysis and design tools for European hinterland hubs, Enschede
- [7] Bonnefoy, Philippe A. 2005, Simulating air taxi networks
- [8] Bovy, P. 2006, CT4801 transportation modeling, TUDelft
- [9] Eurocontrol, 2006, Getting to the point: Business aviation in Europe
- [10] Doganis R. (2002), Flying off course, the economics of international airlines, London
- [11] http://nl.wikipedia.org/wiki/Revenue_Management
- [12] JAR OPS 1, 2007
- [13] Ball, M., et al 1985, A graph partitioning approach to crew scheduling, Transportation Science,
- [14] Keskinocak, Martin and Jones, 2003, Optimizing On-Demand Aircraft Schedules.
- [15] Eurocontrol., 2006, getting to the point: Business aviation in Europe
- [16] Gaarlandt et al. 2006. On demand airline, the future of personal transportation
- [17] Ghijs S, Kinds, D, 2006, Business plan
- [18] Ghijs S, Kinds, D, 2006, Business plan
- [19] Verbracke A. et al., 2002, Handboek Arena, Faculteit TBM

Route Preliminary Demand Forecast Model

For All-Business Airlines or Flights

Nicolaas H. Elferink BSc., Stefaan S. Ghijs MSc.

Dept. of Aerospace Engineering, Aerospace Management and Operations

Delft University of Technology

Delft, the Netherlands

n.h.elferink@student.tudelft.nl , s.s.a.ghijs@tudelft.nl

Abstract — From 2005 onward, a number of new low-cost all-business airlines have emerged in the transatlantic market. All of these airlines have aggressive plans for expanding their route network. This paper describes the development of a preliminary forecasting tool, to be used by such airlines in a preliminary profitability study of new or current routes. This means the model can be used to assess both the profitability of a new route of interest and of continuing an existing route. The model in its current state indeed provides this capability. This is shown using a rough validation calculation, carried out using the model and based on the business case of Eos Airlines, one of the recently erected all-business airlines. The model's forecasting accuracy is still fairly limited. To improve this, more research should be conducted on both the refinement of the model as on the data required for using the model.

Keywords – all-business airlines, demand forecast, demand drivers, air travel routes, business travel.

I. INTRODUCTION

Recently, from 2005 onward, a number of new low-cost airlines have been erected aiming specifically for transatlantic business air travelers [1]. These airlines provide services with aircraft equipped solely with business class and/or first class seats. Such airlines include MAXjet (which has already filed bankruptcy at the end of December, 2007), Eos Airlines, both operating between the US and the UK (mainly between London and New York), Silverjet, with destinations London, New York and Dubai, and L'Avion, flying between Paris and New York. These airlines, according to reference [1], all have expansion plans with respect to their routes. For instance, Eos Airlines will open a new route between Paris and New York in 2008 [2] and Silverjet has shown interest in flying to Chicago, Miami, India and South Africa [1].

In order to determine whether these relatively new all-business airlines will be able to compete in the fiercely competitive environment that the airline business is composed of, it is important to forecast demand as accurately as possible. According to reference [3], forecasting of demand is the most important and critical aspect of managing an airline, since so many important decisions are based on it. Obviously, demand forecasting is a very complex issue, as a lot of the determining factors, or drivers, for demand are unknown or cannot be known in advance. In short, one is dealing with the uncertainty of the future. Reference [3] also states that forecasting of demand is

extra difficult when applied to new air routes. In the case of the new all-business airlines, this is particularly difficult because of their lack of long term historical demand data and experience.

The aim of this paper is to develop and describe a basic demand forecasting tool for these new airlines that can be used to quickly determine whether or not on a (new) route of interest enough demand is present and what the requirements are to satisfy this demand. This can subsequently be used to determine whether or not it is worthwhile to open or continue a certain route, i.e. the model's forecast can be used to make operational decisions. In this paper, only the basic framework for the model will be presented. This means that a number of assumptions is made throughout the paper in order to facilitate the ease of model development. These, however, limit the capabilities of the model. The assumptions and limitations are summarized in section VI.

In order to devise a market demand model for all-business airlines, first the factors that drive the demand will have to be found and valued according to their importance and influence. These factors are numerous and a lot of them are interconnected. The initial discussion of this paper will be on these factors and their importance (section II). The next section will express the importance of each driver in a quantitative manner in order to use them for the development of the model. Section IV will deal with the full-scale development of the demand forecasting model. Section V will apply the model based on the route expansion and fleet plans of an existing all-business airline. Section VI will discuss limitations of the model and resulting recommendations for further research and model development. Finally, section VII will conclude this paper.

II. DEMAND DRIVERS

Before fully developing the market demand model, first the determining factors for the demand will be presented. A distinction can be made between five different types of drivers; economic drivers, so-called demand inherent drivers, service quality, promotion and competition. These are the drivers that affect business travelers in the making of their choice for a certain airline and service.

A. Economic Drivers

This category of drivers is quite extensive and the future variation of the drivers in this category is most difficult to predict, let alone the measure to which they affect the market demand. A number of sub-drivers determine the main driver in this category, which is the relative or perceived ticket price. These sub-drivers are the global and regional economic climate, the currency exchange rates between the currencies in the destination areas, and price elasticities for the targeted passenger types. The currency exchange rates are influenced by the global and regional economic climates. Together with the ruling price elasticities, the economic climate, the currency exchange rates and the absolute ticket price influence the relative or perceived ticket price for the passenger. Furthermore, they also are of influence to the intensity of business activities globally and between the regions of interest, which will be discussed in the next subsection. Of course the economic climate also influences the absolute ticket price. These interconnections are shown in Fig. 1. The arrows and lines in this figure merely indicate the influence of the different aspects on each other.

1) Economic Climate and Currency Exchange Rates

“Air transport has experienced rapid expansion since the Second World War as the global economy has grown and the technology of air transport has developed to its present state” [4]. Furthermore, “The world economic climate and the rate of economic growth in particular countries or regions of the world influence demand in a variety of complex ways” [3, p. 196]. They determine both the level and distribution of personal income and, more relevantly, company revenues and nature of international business activities and trade and through this the demand for air transport services.

According to reference [4], air transport has traditionally undergone larger growth than most other sectors in the economy. Furthermore, according to reference [3, p. 196], the demand for the total air travel grown roughly twice as fast as the world GDP. In this paper, the effect of changes in global and regional GDP are assumed to be incorporated in the trade elasticities, which will be discussed later on. This is valid according to reference [3], which states that the effect of changes in income, and thus in economic climate, on demand can be measured through an income elasticity. For business travel a trade elasticity is often used as an income-related variable.

It is well known that the US Dollar has recently followed a devaluating path for quite some time now. On the one hand this will cause goods from the US to be relatively cheaper for European companies and consumers. On the other hand, for US based companies and customers, doing business with European companies will become increasingly more expensive. Similarly, effects will be acting on ticket prices. It is assumed in this report that the net effect of changing currency exchange rates on demand is zero.

2) Price Elasticities

Next to income, a second factor that has a large impact on the financial drivers of air travel market demand is price [3]. The response of market demand with respect to price is expressed through the price elasticity coefficient (E_p).

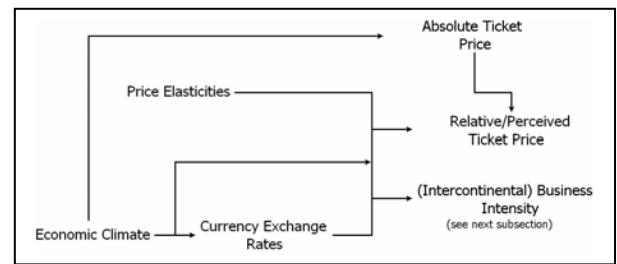


Figure 1. The interconnections between the different economic factors.

This elasticity coefficient is similar to the trade elasticity [3] and is expressed through:

$$E_p = \frac{\% \Delta D}{\% \Delta P} \quad (1)$$

In this equation, ΔD is the change in demand and ΔP is the change in ticket price. Unlike the trade elasticity, price elasticities are always negative, due to the fact that a higher price will induce a lower demand.

Price elasticities are different for different passenger types. Because business travelers usually do not pay their own travel expenses, one would expect them to be less sensitive to changes in ticket prices. Indeed, whereas leisure travelers expose a price elasticity of around -2.0, the price elasticity for business travelers is generally less (absolute sense) than -1.0 [3]. This is affirmed by reference [5], from which an average price elasticity for business travelers of almost -0.8 can be found.

Other reasons for this low price elasticity may be that air travelers have less substitute methods of transport. Furthermore, they have no choice in destinations. Business travelers usually also have a higher valuation of time, so their need for their journey to be as short as possible is greater. Finally, business travelers often have to go straight to work after their flight. In order to better cope with this, a higher level of comfort may get them across less tired and with more energy.

3) Relative or Perceived Ticket Price

The absolute price is simply the price in true units of currency. This is different from the relative, or perceived, price since this is dependent also on overall economic climate and income and revenue standards. Also inflation can play a role. Inflation is, however, neglected in this paper, because prices and revenues are assumed to automatically adapt to this. The relative price is not directly determined in this paper. Rather, the effect on demand is immediately determined through the use of elasticities because this circumvents the need to first determine the perceived ticket price and its influence on demand.

B. Demand Inherent Drivers

1) Trade Elasticity

Besides influencing the relative or perceived ticket price, price elasticities, exchange rates and the economic climate also determine the intensity of global business and thus also between the US and Europe. This effect can be expressed using

a trade elasticity [3], which was already mentioned before. The trade elasticity (E_T) is defined as:

$$E_T = \frac{\% \Delta D}{\% \Delta T} \quad (2)$$

In equation (2), ΔT is the change in trade. Reference [6] states that international business travel for the UK seems to be linked with changes in the volume of trade between the UK and important overseas markets.

With respect to business travel, reference [3] states that the growth of demand for business travel is directly affected by the level of economic activity and trade. According to reference [7], trade elasticities, elasticities where trade is used as an income-related variable, for business travel for Schiphol Amsterdam Airport are between 0.8 and 1.0.

2) *Seasonality and Peak Problems*

As is the case in many industries, the airlines business is exposed to seasonal variations in demand. These variations can be divided in daily, weekly and seasonal variations [3], and can be identified in an airline's flight schedule as peaks and troughs. A high difference between demand in peak periods and in off-peak periods can be severely unbeneficial to an airline. The reason for this is that a lot of extra capacity is needed to be able to satisfy the demand during a peak period. However, this capacity is not utilized or at best underutilized during the off-peak periods.

Peaks and troughs in demand for business travel are mainly dependent on the pattern of annual holidays, such as the Christmas holiday season, and working days and weeks for factories and offices. Daily demand for business travel is usually highest in the mornings and evenings. Similarly, weekly demand is usually highest at the beginning of the week, on Mondays and sometimes Tuesdays, and the end of the week, on Fridays [3]. Thus, certain peak demand variations for business air travel are dependent on each other. The reason for this lies in the scheduling of normal working days and weeks in Western countries.

This also causes the demand for business travel to slack off during the weekends. For all-business airlines, a possibility could be to fill excess capacity during the weekends and holidays with leisure passengers. These airlines could for example attract wealthier British leisure travelers that want to go shopping in New York. Additionally, they could decrease their capacity during the off-peak periods. An optimal way in which to do this would be to schedule maintenance and repair activities during these periods.

A general trend that is visible within the airline industry is that the peak problem is less severe in case there is a lot of business travel on a certain route [3]. The reason for this is that business travel is less prone to seasonality than leisure travel. Most companies simply keep on running during holidays and a lot even during the weekends. This means that business will also keep going. Furthermore, business travel is not dependent on seasonal climate and weather variations, contrary to leisure travel. Still, the peak problem can also be quite a challenge for airlines that transport a lot of business travelers.

C. *Service Quality as Demand Drivers*

1) *Passenger Treatment and Service*

References [8] and [9] state that service quality is defined by the end customer and that the quality the customer perceives is related to the difference between the customer's expectations and perceptions. Factors that are of relevance for the perceived quality of service are on-time performance, (schedule) flexibility, baggage handling, quality of food and beverages, seat comfort, check-in service and in-flight service and treatment [10].

Reference [11] concludes that there are two types of business travelers; "luxury-loving" and "no-frills". Clearly, the one of most interest to an airline having only business class seats is the first type. Reference [12] found service quality to be a major driver for demand for an airline's services.

In case an airline can increase its standards (over that of the competition) without increasing its ticket price, this will surely increase the demand for its services. In contrast, in case of a severe economic downturn, an airline may be forced to lower its standards and increase its ticket price, which would have a negative impact on demand.

2) *Passenger Importance Segments*

According to reference [13] there are five segments to be distinguished in the business traveler market. This segmentation is based on the importance given to different aspects of the travel experience and expenses by the traveler. The five segments are based on:

- Punctuality.
- Comfort.
- Price.
- The Price/Performance ratio.
- Catch-all/Flexibility (all of the above are important).

Which one is of importance to a particular airline is dependent on its exact business model. Whereas price is mostly important to lower and middle management employees, the other four segments are highly valued by middle to senior management executives that are frequent flyers [13]. Furthermore, the price/performance ratio is of more importance to entrepreneurs that have to pay for their flight themselves.

D. *Promotion*

A driver that an airline can actively influence to a large extent is the promotion of its services. This can be done using all kinds of advertising, frequent flyer programs, publications, etc. Through promotion, an airline can also influence its image as perceived by the outside world. To this respect, also performance in respect of safety, service quality and on-time delivery of passengers is critical. Especially for young airlines, promotion can be an important tool to gain market share, as these airlines are still well in the process of expanding and implementing major operations, i.e. they still have to prove their worth in the market.

Reference [14] indicates that advertisement programs aimed for business passengers should put extra emphasis on

high quality ground and in-flight services, connection options and on-time service, since these aspects are of most concern to business travelers. The reason for this is that business travelers usually have inflexible travel plans and do not personally pay the ticket fares [14].

Reference [15] has found that frequent flyer program membership induce the willingness for business passengers to pay significantly extra in order to fly with their carrier of choice. This thus means that frequent flyer programs can increase the demand for an airline's services.

E. Drivers from Competition

With respect to competition, the new all-business carriers are operating in quite a fierce market. The route between London and New York is one of the busiest air transport routes in the world. A large part of this stems from the demand from business travelers, since London and New York are actually among the most important business centers in the world.

Within this highly competitive market, quite a number of rivals is aiming at the business traveler. These include low-cost all-business (LCAB) carriers and regular airlines' business class (RABC). The latter are a major source of competition [1], mainly through their business class and/or first class services. Compared to the recently emerged low-cost all-business carriers, regular airlines' business class can offer more frequency and connectivity, but at the cost of higher ticket prices (roughly up to five times as much as the new low-cost all-business carriers). Furthermore, regular airlines are well established in the market, whereas the new airlines still need to gain trust from the market [1].

Another source of competition is the branch of airlines that provide on-demand business air travel (ODBA), generally called air-taxi services. Also, some large companies may operate their own corporate jets.

Currently, the low-cost all-business airlines have filled a niche in the air travel market. However, economic conditions may change and competition may be overlapping to a much larger extent than it currently is. Therefore, it should be noted that the exact threat from competitors is very difficult to estimate and forecast. It is, however, also not in the scope of this paper to do so.

III. DRIVER QUANTIFICATION

Having described all relevant factors, this section will indicate the amount of influence of each driver on the total demand within the model that will be developed in the next section. This will be done by indicating their specific method of use within the model.

1) Economic Drivers

With respect to the economic drivers, the basic model framework will only take into account the price and trade elasticities. These will constitute the model's back bone, providing a basis for the rest of the model. Thus these elasticities are considered the most important drivers for the demand forecasting model. As was stated in section II, an average price elasticity of -0.8 was found from reference [5].

The trade elasticity was found to typically range between 0.8 and 1.0. The exact value is dependent on the exact route and airline business model considered. It will simply be assumed that the trade elasticity is equal to 0.8 in this paper as a result of a lack of accurate data. When a more detailed demand prediction is needed from the model, more research should be performed on the exact value of both elasticities.

Based on these elasticities and predicted or assumed price and trade changes, a current demand can be extrapolated to a future demand. Obviously, for this method to be accurate, accurate predictions of the future price and trade levels compared to current price and trade levels are needed.

2) Demand Inherent Drivers

Apart from the price and trade elasticities, the model will include a correction for seasonality and peak occurrence. Basically, for this a so-called peak factor will be used, which will indicate the difference between peak and off-peak periods. In fact, any peak pattern can be corrected for using this factor.

Reference [3] finds a peak to trough ratio of 1.54 to 1.00 for the route New York – London. Thus this would mean the peak factor is 1.54 on that route. Note, however, that this peak factor is valid for the total passenger air travel, including all non-business passengers. Once again, for more accurate results from the demand forecast model, more research should be conducted on the true peak factor ruling on the particular route of interest.

3) Service Quality Standards

This driver is actually a lot harder to materialize. In the model, use will be made of an efficiency factor for the service quality standards to gain market share (E_{SQ}). In order to establish accurate values for this variable, the effect of service quality on demand should be further researched based on the particular airline of interest. Already quite some research has been performed on the effects of service quality on demand, some even of a quantitative nature [15], [16], [17]. References [15], [16] and [17] provide a good basis for this further quantitative research.

4) Promotion

With respect to promotion, the model will employ a so-called promotion efficiency factor (E_{PR}), which is similar to the service quality efficiency factor. The promotion efficiency factor is simply expressed as a percentage increase in demand resulting from a certain promotional activity. Further research on a per airline and route basis should establish exact values for the promotion activities' efficiency to induce more demand. A good starting point for such research is provided by reference [12].

5) Competition

While three important sources from competition were found in the previous section, in the model, these sources of competition will be grouped into one competition variable. This is done because this paper focuses on the total model and not on the specific characteristics of competition (or any other variable in that respect). Thus:

$$MS_{COMP} = MS_{LCAB} + MS_{RABC} + MS_{ODBA} \quad (3)$$

In this equation MS_{comp} stands for competition Market Share. Now, any airline's future market share is given by the sum of the current total market ($M_{Tot\ i}$) and the change in total market (ΔM_{Tot}), minus the market share of the competition:

$$MS_{airline\ i+1} = M_{Tot\ i} + \Delta M_{Tot} - MS_{COMP\ i+1} \quad (4)$$

Note that the future is denoted by the subscript $i+1$, while the current is denoted by i .

Thus, it is evident that an airline's market demand is dependent on the market share it can capture, which it can partly influence by increasing its internal capabilities of service quality, its ticket price and its promotion intensity for example, and on the developments in the total market, which it cannot influence. The change in total market is by definition of the trade elasticity determined from the change in trade and the trade elasticity.

IV. DEVELOPMENT OF THE DEMAND FORECAST MODEL

A. Model Back Bone

As was stated, the model's back bone is based on the use of price and trade elasticities. When the current total demand for business air travel is known, as well as a trade elasticity value and an expected growth or decline in trade levels for the future time or period of interest, the total demand for that future time frame can be found through:

$$M_{Tot\ i+1} = M_{Tot\ i} \times (1 + E_T \times \Delta T) \quad (5)$$

Before being able to apply a price elasticity on this, in order to assess the effect of a ticket price change, first the market of the particular airline of interest in the future period ($M_{AL\ i+1}$) must be determined. This is done by multiplying by one minus the competition's market share in the future period of interest:

$$M_{AL\ i+1} = M_{Tot\ i} \times (1 + E_T \times \Delta T) \times (1 - MS_{COMP\ i+1}) \quad (6)$$

What remains now is the total expected demand for the airline in the future, should no changes with respect to price, service quality and promotion and airline image occur. It should be clear that for the model to be accurate, the expected market competition market share should be carefully predicted.

As was already hinted, the next step is to include the effect of ticket price on the demand through the use of the price elasticity. Equation (7) shows the result:

$$M_{AL\ i+1} = M_{Tot\ i} \times (1 + E_T \times \Delta T) \times (1 - MS_{COMP\ i+1}) \times (1 + E_p \times \Delta P) \quad (7)$$

B. Model Refinement

The back bone of the model is now firmly established. The model can now be refined by adding corrections for service quality and promotion and image. Furthermore, also a variable peak factor can be implemented in order to take into account a variable peak and trough pattern in demand. First of all, a correction for service quality changes is added:

$$M_{AL\ i+1} = M_{Tot\ i} \times (1 + E_T \times \Delta T) \times (1 - MS_{COMP\ i+1}) \times (1 + E_p \times \Delta P) + (M_{Tot\ i} \times (1 - MS_{COMP\ i})) \times E_{SQ} \quad (8)$$

Thus, an increase (or decrease) due to changing levels of service quality is added. It is assumed in this that a current improvement (or deterioration) in service quality results in an increase (or decrease) in demand in the future, because an increase in service quality is assumed to only reach the passengers that are currently flying with the airline of interest. Thus, this increase (or decrease) is based on the current time market demand.

Similarly, an addition for the promotion effect on demand can be made. The effect of promotion is acting on the market share the airline of interest can obtain in the future time period. Thus, the factor $(1 - MS_{COMP\ i+1})$ is determined from $(1 - MS_{COMP\ i})$ multiplied by $(1 + E_{PR})$ (9).

$$M_{AL\ i+1} = M_{Tot\ i} \times (1 + E_T \times \Delta T) \times (1 - MS_{COMP\ i}) \times (1 + E_{PR}) \times (1 + E_p \times \Delta P) + (M_{Tot\ i} \times (1 - MS_{COMP\ i})) \times E_{SQ} \quad (9)$$

Finally, a peak factor ($PF(x)$) provides the final refinement to the demand forecasting model. From the notation, it should be clear that the peak factor is a variable. One could either use a continuous peak function or a number of discrete peak factor values, both obtained from experience or general market trends. The final model thus becomes:

$$M_{AL\ i+1} = [M_{Tot\ i} \times (1 + E_T \times \Delta T) \times (1 - MS_{COMP\ i}) \times (1 + E_{PR}) \times (1 + E_p \times \Delta P) + (M_{Tot\ i} \times (1 - MS_{COMP\ i})) \times E_{SQ}] \times PF(x) \quad (10)$$

C. Using the Demand Forecasting Model

There are basically two instances in which this model will provide a preliminary demand forecast. The first situation is when an all-business airline is intending to open a new route

and needs to know its potential. The second case is when an airline is unsure of whether or not to continue servicing a certain route or when it is unsure of how much (extra) capacity will be needed.

1) Opening a New Route

For this case the model is somewhat simplified, because in this case the new entrant on a certain route does not yet have any market share, i.e. $(1-MS_{COMP\ i})$ is equal to zero. In effect, the market share the airline can obtain in the future period needs to be obtained through a prediction. Also, it is assumed that the influence of service quality is negligible, because on the new route nobody has had the opportunity to actually experience the quality of service. Thus, the model becomes as in (11).

$$M_{AL\ i+1} = [M_{Tot\ i} \times (1 + E_T \times \Delta T) \times (1 - MS_{COMP\ i+1}) \times (1 + E_P \times \Delta P)] \times PF(x) \quad (11)$$

This means that the effect of promotion of the airline on the new route is incorporated in the $(1-MS_{COMP\ i+1})$ variable, i.e. it is taken along in the prediction for the market share the new airline on the route can obtain.

2) Continuing an Existing Route

In this case, the model is used in its full extent, as given in (10). The variables that are needed to operate the model are:

- The trade elasticity E_T .
- The expected change in trade between the two ends of the route ΔT .
- The price elasticity E_P .
- The expected change in ticket price ΔP .
- The current total market $M_{Tot\ i}$.
- The current competition market share $MS_{COMP\ i}$.
- The expected gain in market share due to promotion E_{PR} .
- The effect of a change in service quality on demand E_{SQ} .
- A peak function or peak factor $PF(x)$.

V. MODEL APPLICATION

Now that the model has been fully developed, its forecasting ability can be assessed based on specific data for one of the new all-business airlines performing transatlantic operations at the moment. For this, Eos Airlines has been selected for no other reason other than that it is still in operation, contrary to MAXjet. Firstly, an existing route will be investigated using the demand forecasting model. This is Eos Airlines' route between London and New York. Secondly, its

plans to open a new route between Paris and New York will be compared to the model's predictions.

A. The London – New York Route

Because this is an existing route, the full model will be applied. Table 1 summarizes the data applicable to this route. Because no adequate data is available for most variables, most data is (partially) based on assumptions. This is, however, not a problem, because the assumed values are close enough and are capable of demonstrating the use and functioning of the demand forecasting model.

The trade elasticity is assumed from the previously found range to be equal to 0.8. The trade change is an average from trade growth factors that actually occurred between 2004 and 2006 between the US and the EU, obtained from reference [18]. The price elasticity was already found to be equal to -0.8 for business travelers, whereas the change in ticket price is an average determined from reference [19]. All three the current total market, current competition market share and the peak factor have been obtained from analysis of the flight schedules of all carriers operating between London and New York, where average load factors of 50 and 80 percent have been assumed (due to a lack of accurate data) for business class and first class seats respectively. From these schedules all business class and first class seats are seen as part of the total market for Eos Airlines. Note that the current total market is a monthly average over the months December 2007 to February 2008. Finally, the promotion and service quality efficiencies have simply been estimated at 30 and 70 percent respectively. This may seem high, but note that promotion is likely to be quite high for a new airline like Eos Airlines, for a large part also due to word-of-mouth promotion. Similarly, by opening up extra routes and increasing capacity [1], Eos Airlines will drastically improve its service quality with respect to its connectivity and schedule characteristics (more flights a day).

Substituting all these values in the model (10) will result in an average forecasted demand over the months December 2008 to February 2009 (the off-peak period) of 3675 passengers per month. For the peak period, which is from June to August 2008, this is 3909 passengers per month.

Knowing that Eos Airlines will operate 29 one way flights a week in 2008 [2], that it holds 48 seats per aircraft [2] and has an average load factor of 70 percent [1], one can determine that it has a capacity of almost 4200 seats per month.

TABLE I. LONDON-NEW YORK ROUTE DATA

Variable	Symbol	Value	Unit
Trade elasticity	E_T	0.8	-
Trade change	ΔT	9.5	%
Price elasticity	E_P	-0.8	-
Ticket price change	ΔP	7.75	%
Current total market	$M_{Tot\ i}$	~38000	pax/month
Current market share competition	$MS_{comp\ i}$	95.2	%
Promotion efficiency	E_{PR}	30	%
Service quality efficiency	E_{SQ}	70	%
Peak factor	PF	1.064	-

This can be seen to comply quite good with the forecasted maximum demand of 3909 passengers per month, yielding a difference of about 6.4 percent.

B. The Paris – New York Route

Quite similarly, the Paris – New York route can be investigated. Because this is a new route for Eos Airlines, the simplified version of the demand forecasting model is used. Table 2 presents the data for this case.

Note that most variables are the same as for the previous case. The current total market share and the peak factor have been obtained from analysis of the flight schedules of all airlines operating between Paris and New York, where, once again, load factors of 50 and 80 percent have been assumed for business class and first class seats respectively. Furthermore, the future market share of the competition is simply assumed to be equal to that of Eos Airlines' competition on the route London – New York because no adequate means for estimation is available.

Once again, substituting all values into the model, this time (11), will result in a forecasted demand of 680 passengers per month for the off-peak period December 2008 to February 2009. Similarly, for the peak period June to August 2008 this is 903 passengers per month.

Unfortunately, no announcements have been made yet about the number of flights to be operated between Paris and New York. However, when assuming an initially low load factor of 35 percent, as Eos Airlines experienced during its first months of operation between London and New York [20], it can be found that the airline needs to perform 14 (one way) flights per week from London to New York and vice versa. This is in accordance with reference [21], which was created during the airline's initial months of operation.

VI. LIMITATIONS AND RECOMMENDATIONS

A. Forecasting Model Limitations

With the demand forecasting model developed and found to operate seemingly well, its limitations will be shortly discussed below.

First of all, the model in its current state is limited to use applicable to all-business airlines only. Modifications could be made in order to prepare the model for use for all airlines, yet this would require quite some additional research to occur (see below). The difference in the model as it is and the model as it would be to be valid for any airline is mainly making the model as it is simpler. In order to make the model valid for any airline, one would have to incorporate quite some complexities, related to having more than one service class and more than one type of passenger, each type having different specifics and requirements. For leisure travelers for example, also destination becomes important and for them, rather than trade elasticities, income elasticities become important. All these complexities would make the model a lot more elaborate, as well as the research that would be needed to obtain the model.

TABLE II. PARIS-NEW YORK ROUTE DATA

Variable	Symbol	Value	Unit
Trade elasticity	E_T	0.8	-
Trade change	ΔT	9.5	%
Price elasticity	E_p	-0.8	-
Ticket price change	ΔP	7.75	%
Current total market	$M_{Tot i}$	~14000	pax/month
Future market share competition	$MS_{comp i+1}$	95.2	%
Peak factor	PF	1.328	-

Furthermore, the model is only suitable for the short term use, depending on the exact accuracy of the variables introduced in the model, and provides a preliminary demand forecast only (back of the envelope calculations). More detailed demand forecasting may be required.

Finally, the model neglects any interrelations between the service quality, promotion and competition threats (except for the relation between the competition market share and the promotion efficiency).

B. Recommendations for Further Research

Recommendations for further research are numerous, all aimed at improving the forecasting ability of the model or the accuracy of predicting the variable values needed for the model.

1) Recommendations with respect to Model Refinement

The most important recommendation would be for the investigation of the exact quantification of the promotion efficiency and service quality efficiency. In the model's application (section V), quite crude assumptions have been made for these variables, seriously limiting the accuracy of the predictions. Further research could take away this shortcoming. This research will then also yield possible better ways to express these factors in the model. For example weight factors could be used in the model itself or within the factors that are input to the model.

Next, the model uses one variable with respect to the competition, i.e. its market share. The reason why the factor $(1 - MS_{COMP})$ was used is because this forces the model user to carefully assess the threats from the competition, rather than simply estimating a market share that the user feels can be achieved by the own airline. In order to increase the accuracy and predicting power of this model feature, further research into the distinctive competitive threats from different competitions (LCAB, RABC, ODBA) should be performed.

2) Recommendations with respect to Data Accuracy

In order to improve the accuracy of the model's back bone structure, the most stringent recommendation for further research with respect to data accuracy would be to investigate the exact value and possible variation of price and trade elasticities between two destinations of interest. Similarly, also the expected trade and ticket price changes should be investigated.

An important consideration with respect to all parameters used, is their variability and sensibility and their effects on the model results, i.e. the model's accuracy, given a certain variability in the parameters. For this, Monte-Carlo simulations would be well suitable. The reason why these have not yet been

performed is because some of the more “abstract” parameters, such as the effects of promotion and service quality, should first be further investigated. Once the way in which these parameters influence the demand is more clearly determined, Monte-Carlo simulations should best be performed in tandem with further research on how to incorporate these factors.

Finally, as a user of the model, the current total demand should be more accurately deduced. The same holds for the exact peak factor or, better, peak function for the particular airline and route of interest used in the model.

VII. CONCLUSION

The purpose of this paper was to develop and demonstrate a preliminary demand forecasting tool for the newly emerged all-business airlines operating across the Atlantic. The model incorporates a number of demand drivers, being the effects of trade and ticket price changes, the effects of service quality changes and promotion, the effect of competitive threats, and finally the effect of seasonality. This is also the difference of this model with respect to generic demand models as described by reference [3] for example. This model takes into account the “physical” parameters instead of only empirical factors and trade and/or price elasticities.

Using data and estimated data for Eos Airlines, the model has undergone a first validation calculation. Even though the data input was for a large part based on assumptions and estimations, the model’s first forecasting accuracy for Eos Airlines’ existing route between London and New York was found to be around 6.4 percent.

Although quite a lot more future research is needed for the model to perform with accuracy, the model interrelates a lot of previously separate areas of research, such as marketing, service quality, seasonality and trade and price changes and effects. Furthermore, the model does provide an initial framework, suitable for adapting if needed, to build on, as well as the crude forecasting abilities to be used in a preliminary route profitability assessment.

REFERENCES

- [1] M. Morrison, “Premier upstarts,” in *Flight International*, 19-25 June 2007, pp. 66-69.
- [2] Eos Airlines, “Eos Airlines celebrates two years of success by announcing two new routes,” Eos Airlines Press Release, New York: October 18, 2007.
- [3] R. Doganis, “Flying off course: the economics of international airlines,” 3rd ed., London: Routledge, 2002.
- [4] ICAO, “Industry situation and airline traffic outlook,” Montreal: Worldwide Air Transport Conference 2003: Challenge and opportunities of liberalization.
- [5] M. Brons, E. Pels, P. Nijkamp and P. Rietveld, “Price elasticities of demand for passenger air travel: a meta-analysis,” in *Journal of Air Transport Management*, vol. 8, issue 3, May 2002, pp. 165-175.
- [6] S. Maiden, “Heathrow terminal 5: proof of evidence forecasting,” London: British Airports Authority, 1995.
- [7] X. Veldhuis, “Forecasting process at Amsterdam Airport Schiphol,” Geneva: Worldwide Forecasting, 1988-92 Conference, IATA, September 1988.
- [8] A. Parasuraman, V.A. Zeithaml, L.L. Berry, “A conceptual model of service quality and its implications for future research,” in *Journal of Marketing*, vol. 49, no. 4, Autumn 1985, pp. 41-50.
- [9] A. Parasuraman, V.A. Zeithaml, L.L. Berry, “SERVQUAL: a multiple-items scale for measuring consumer perceptions of service quality,” in *Journal of Retailing*, vol. 64, no. 1, Spring 1988, pp. 12-50.
- [10] K.M. Elliot, D.W. Roach, “Service quality in the airline industry: are carriers getting an unbiased evaluation from consumers?” in *Journal of Professional Services Marketing*, vol. 9, issue 2, 1993, pp. 71-82.
- [11] C. Huse and F. Evangelho, “Investigating business traveller heterogeneity: low-cost vs full-service airline users?,” in *Transportation Research Part E: Logistics and Transportation Review*, vol. 43, issue 3, May 2007, pp. 259-268.
- [12] M. Abrahams, “A service quality model of air travel demand: an empirical study,” in *Transportation Research Part A: General*, vol. 17, issue 5, September 1983, pp. 385-393.
- [13] T. Teichert, E. Shehu and I. Wartburg, “Customer segmentation revisited: the case of the airline industry,” in *Transportation Research Part A: Policy and Practice*, vol. 42, issue 1, January 2008, pp. 227-242.
- [14] W.G. Browne and R.S. Toh, “Targeting airline advertising copy,” in *Annals of Tourism Research*, vol. 24, issue 4, October 1997, pp. 1005-1008.
- [15] K. Proussaloglou and F.S. Koppelman, “The choice of air carrier, flight, and fare class,” in *Journal of Air Transport Management*, vol. 5, issue 4, October 1999, pp. 193-201.
- [16] D. Gursoy, M.H. Chen, H.J. Kim, “The US airlines relative positioning based on attributes of service quality,” in *Tourism Management*, vol. 26, issue 1, February 2005, pp. 57-67.
- [17] F.Y. Chen, Y.H. Chang, “Examining airline service quality from a process perspective,” in *Journal of Air Transport Management*, vol. 11, issue 2, March 2005, pp. 79-87.
- [18] World Trade Organization, “International trade statistics 2007,” Geneva: WTO Publications, 2007.
- [19] American Express Business Travel, “Business travel demand will outpace capacity in 2008 and drive rate increases across air, hotel, car rental and meetings,” New York: American Express Business Travel Press Release, October 2007.
- [20] A. Ellson, “Doubts over business class airlines,” in *Times Online*, March 16, 2006
http://business.timesonline.co.uk/tol/business/markets/united_states/article741770.ece, retrieved January 24, 2008.
- [21] Eos Airlines, “Eos Airlines announces second daily NY-London flight,” Eos Airlines Press Release, New York: July 10, 2006.

Track 5

Human Factors & Interfaces

An Infovis approach to compare ATC comets

Comparing visual entities with a theoretical foundation

Christophe HURTER

Direction Technique de l'innovation

DGAC/DSNA/DTI R&D

7, avenue Edouard Belin

31055 Toulouse France

christophe.hurter@aviation-civile.gouv.fr

Vincent KAPP

Direction Technique de l'innovation

DGAC/DSNA/DTI R&D

7, avenue Edouard. Belin

31055 Toulouse France

vincent.kapp@aviation-civile.gouv.fr

Stéphane CONVERSY

ENAC – MI – LII

Laboratoire "Interaction et Informatique"

7, avenue Edouard. Belin

31055 Toulouse France

stephane.conversy@enac.fr

Abstract—Air Traffic Control systems display information with multiple visual entities. The research described in this paper is an initial effort to develop a theory-driven approach to the characterization of visual entities. We enhance the state of the art in data visualization to characterize four “comet” designs. This work helps to understand visualization more precisely and provides a basis to help the designer to formally assess the effectiveness of their work.

Information Visualization, design, taxonomy, graphical coding.

I. INTRODUCTION

In current Air Traffic Control (ATC) environments, air traffic controllers use several visualization systems: radar view, timelines, electronic strips, meteorological views, supervision etc... Each of these visualizations is rich and dynamic: it displays numerous visual entities that move and evolve over time.

The objective of our work is to develop a suitable set of tools based on established theoretical methods, in order to evaluate the effectiveness of visual entities before testing them with users. We will answer the simple question: “what are the displayed information, how are they displayed and how can we compare them?”

Our goal is not to answer the question: “what makes one type of visualization better than another?” This answer is linked to controller activity. First, the user is always able to perceive information that is visually coded, but the cognitive resource varies depending on the nature of this visual information, e.g. the difference of perception between text and color. Second, users’ perceptual skills are linked to their activity (tower controllers or en-route controllers do not need the same information although they might be able to work with the same HCIs). We are not trying to answer the following question either: “how can we help designers improve perception?”

Users’ perception is nevertheless very important. These kind of issues have already been addressed in the Information Visualization field (IV). These tools will help us to give an accurate description of visual entities.

Because the characterization of a full image may be tedious, the paper focuses on one visual entity through four designs: the radar comet. In the ATC field, a comet represents aircraft position. In order to understand each comet design, we will have a look at each software feature that uses this design. Then, we will detail the design of each comet, find out their design properties and their associated semantic.

A. The design issues

The design process is very tricky. It takes time and intuition. Joannes Itten [12] p7, a design teacher and an artist, claims that if you don’t know how to create a satisfying painting it means that you are not (yet) an artist. But you can still draw nice paintings with a theoretical approach. You can learn rules and apply them. Most artists know these rules but did not learn them; they just rely on their genius.

It is very difficult to create a new design based on nothing. We identify four different approaches when building a visual entity:

- Empirical approach : design based on trial and error methodology,
- Historical approach: design based on the continuity of previous work with a concern for adaptation to the given context,
- Ecological approach: design based on the respect of both human physical and perceptual characteristics,
- Technological approach: design based on technological opportunities.

Those four approaches shall not be considered as separate spaces; each design process mixes a bit of the other. Of course, there is no clearly defined boundary between the sources of design, and there is a lot of overlap.

The four sources of design inspiration help to understand and justify design choices. By extension, they will give clues on how to perform an exhaustive characterization.

B. The characterisation issues

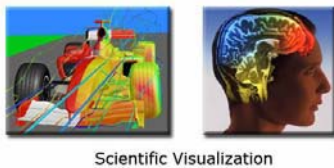
Characterization is a precise and minimal description that unveils differences and allows comparison. Characterization is a very helpful tool for designers along the design process. To perform this characterization we need tools, and the only one

available is the human perception through its eyes and brain. This can't be satisfactory because we can't be certain to perform an exhaustive description of the displayed information. This is the reason why characterization is awkward, and this paper is an initial effort to fill this gap.

II. THE INFOVIS FIELD

As said in the previous section, human perception is involved in the transmission of information. The design and study of human perception of representations is a subfield of the Human Computer Interaction (HCI) field, called Information Visualization, or Infovis (IV). Information visualization is the visual presentation of abstract information spaces and structures to facilitate their rapid assimilation and understanding.

Text-based interfaces require cognitive effort to understand their information content. Humans have remarkable perceptual abilities of graphical entities; they can rapidly and automatically detect patterns and changes in size, color, shape, movement, or texture. Information visualization seeks to present information visually, to offload cognitive work to the human visual perception system.



Scientific Visualization

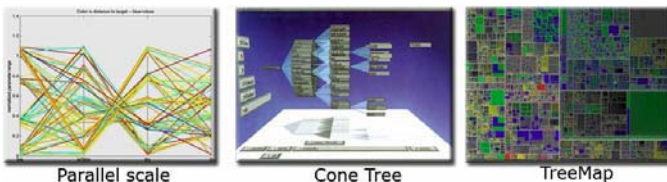


Figure 1 : Scientific and InfoVis

Figure 1, display the difference between visualizations. Scientific visualization display canonical representation of real object and phenomenon. Infovis displays data in an abstract way [11][17][18]. Its goal is to optimize the bandwidth between the displayed data and the perceived data.

The correct perception of visualization has nothing to do with artistic design. There are tools for the IV that can help us to answer the following questions:

- What kind of data is being displayed?
- How is the data processed or updated?
- How can we characterize the suggested visualization?

In the next chapter, we present some of the Infovis Tools which suitable to the analysis of ATC images.

A. Data type

The major distinction we can make for data type is whether their values are:

- Nominal: are only equal or different to other values (e.g. aircraft call sign),
- Ordered: obey a < rule (e.g. an aircraft's number in the landing sequence),
- Quantitative: can be manipulated by arithmetic (e.g. the aircraft speed).

The quantitative type can be split into two parts: Interval and Ratio. Interval can be the gap between values but cannot be null, e.g. the time lapse between 7.00am and 8.00am is the same than 14.00am to 15.00am but we cannot say that 15.00am is twice 7.00am.

The ratio type is the full expressive power of real numbers. The Table 1 summarizes the different terms used in the literature.

Bertin [4]	Stevens [19]	Ware [21]
Nominal	Nominal	Category
Ordinal	Ordinal	Integer
Quantitative	Interval	Real number
	Ratio	

Table 1 : data types

1) Design and Data type

Bertin[4] was the first one to study representation rules. He identified three distinct levels for a visualization analysis: elementary (for a single item), intermediate (for a group of items), and overall (for all the items). He finds out rules to code information in a monosemic way: there can't be any ambiguity in the perception of displayed information.

Afterward, Cleveland[8], McGill[9] and then Mackinlay[15] built scales of expressivity and effectiveness (dependant on the human perceptual capabilities) to assess alternative designs (Figure 2). This scale depends on the data type. The visual property ranked higher in the chart is perceived more accurately than those that are ranked lower in the chart. In the Figure 2 Gray items are not relevant to the concerned type of data.

The quantitative data type ranking has been experimentally verified by Cleveland [9]. Independently of the data type, the best way to represent the data is to code it with a position on a scale. To represent the speed of an aircraft (quantitative data), we can use the length of a line (speed vector). The aircraft position number in the landing sequence (Ordinal) is better coded using the color saturation than length.

Despite the fact that the text involves perceptual and cognitive processing that helps one to decode a graphic in the same way that perceiving color or pattern does, the text entity isn't listed in Mackinlay's perception ranking. "Images are better for spatial structures, location, and detail, whereas words are better for representing procedural information, logical conditions, and abstract verbal concepts." Ware [21]p301-307. Graphical perception is highly parallel which works on visual properties such as position and color, but has limited accuracy.

Text representation is accurate but is limited in capacity. The cognitive workload is very high when we are reading a text. That are the reasons why the text is not integrated in the Mackinlay's perception ranking.

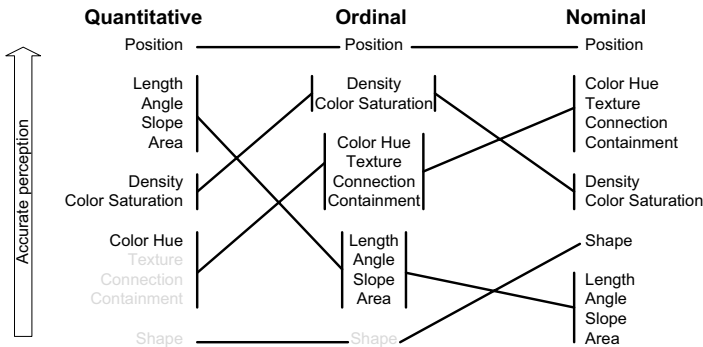


Figure 2 : Mackinlay ranking of perceptual task [15]

This ranking was built for statistical graphs, Air traffic control displays, and other iconic representations of data addressed quite different tasks. Still this approach remains a promising starting point of research to answer the question: "What is the most suitable visual property I can use?"

B. The data flow model

Card, Mackinlay and Shneiderman[6] created a model (Figure 3) which describes visualizations as a data processing sequence from the raw data to the views. The processing is based on structures of intermediate data which is easy to handle by the user. Chi [7] detailed the various stages of this model. This data flow model is still widely used.

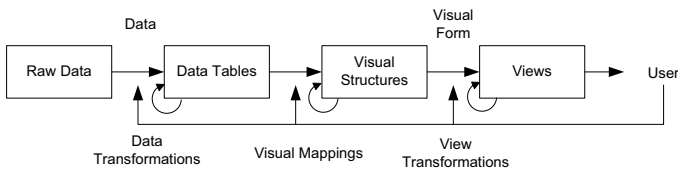


Figure 3: Schematic Dataflow of Information Visualization [6]

This model is based on the management of a data flow. It is used in many toolkits (InfoViz[10], prefuse, VTK, Tulip, Pajek...) and visualization software (SpotFire[1] , ILOG Discovery[2] , nVizN[22]...).

This model formalizes the transformation process from raw data to a screen and is the foundation of a compact and precise characterization.

C. Characterization model

Card and Mackinlay[5] attempted to establish comparison criteria of the images with their work. They propose a table for each function of transformation (Table 2).

				automatic perception					Controles perception		
Name	D	F	D'	X	Y	Z	T	R	-	□	CP

Table 2: C&M representation model

The lines correspond to the input data. The column D and D' indicate the type of data (Nominal, Ordered, and Quantitative). F is a function or a filter which transforms or creates a subset

of D. Columns X, Y, Z, T, R, -, □ are derived from the visual variables of Bertin[4].

The image has four dimensions: X, Y, Z plus time T. R corresponds to the retinal perception which describes the method employed to represent information visually (color, form, size...). The bonds between the graphic entities are noted with '-'. and the concept of encapsulation is symbolized by '□'. Finally a distinction is made if the representation of the data is treated by our perceptive system in an automatic or controlled way. The C&M table is filled with the notations in the Table 3.

L	Line
S	Size
Sh	Shape
f	Function
N, O, Q	Nominal , Ordered, Quantitative
Lon, Lat	Longitude, Latitude
Pt	Point
Orien	Orientation
T	Text

Table 3: C&M Model notations

The previous chapter was a stat of the art of the InfoVis tools. The next chapter deals with the historical design of the comet and its initial use in a non ATC environment.

III. COMETS

The comet visual properties have been used for the first time in the early seventh century by Edmond Halley[16] who coded the trade wind direction on a map [20] p23[21] p 203 . He coded the flow with a stroke.

The comet has accurate design properties; it displays the direction of the shape and its tendency. The comet is composed of a bigger part, its head, and a smaller, its tail. Its head indicates the comet heading. The tendency indicates the future position of the aircraft. The curvature of this shape indicates if it is turning right or left and the amount of steering.

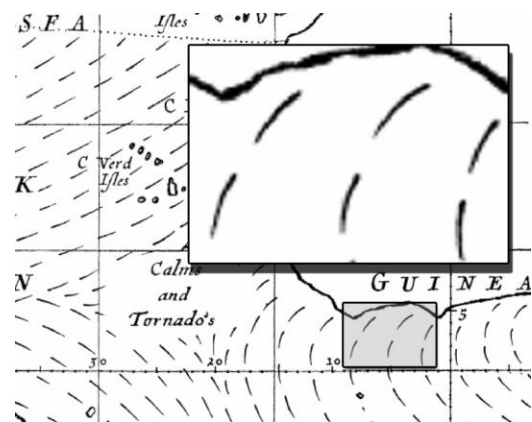


Figure 4 : Detail of Halley's chart of the Trade Winds 1686.

ATC visualization derives some benefit of this comet. To do so, designers use different design options to display the aircraft position with a comet. In the next chapter, we will have detail different ATC systems which use the comet.

A. The ODS comet

ODS is the main French radar view for the air traffic controllers. Its main goal is to display aircraft positions and to help controllers space aircrafts beyond the security minima.

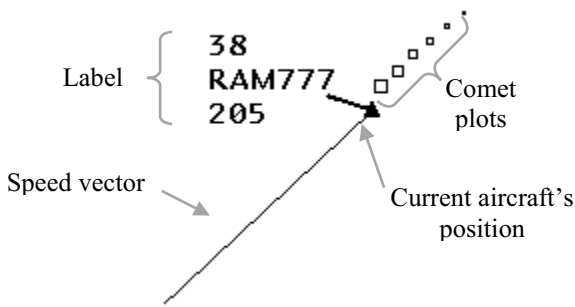


Figure 5 : radar track

Figure 5 displays the terms used to depict a radar track. The radar track presents the aircraft position, its speed, its name, altitude and speed as text. The design of the comet is built with squares, whose size varies with the recentness of the aircraft position: the biggest square displays the last position of the aircraft, whereas the smallest square displays the oldest aircraft position.

The design of this comet is historical. It is not based on the Halley design but on early radar equipment which relied on scope persistency (Figure 6). Old radar scope retained the previous plot position with the fading of the screen phosphor. This kind of design has the same remarkable properties as the Halley comet: it displays the aircraft trajectory's curvature tendencies and shows if an aircraft is turning and the amount of steering.

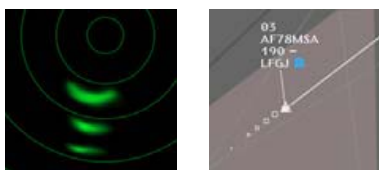


Figure 6 : Spot decreases in intensity over time on a scope (left picture). ODS comet metaphor (right picture).

B. RadarGL

The goal of the RadarGL project is to develop a prospective visualization of the aircrafts position using the latest technologies. This project uses the latest rendering techniques (animation, alpha blending...) and some of the HCI (Human Computer Interaction) techniques for the interaction and the control of the image. RadarGL displays a top view of the aircraft position. The Xscreen is the latitude and the Yscreen is the longitude of each aircraft.

C. ASTER

The ASTER [3] tool was initially designed to assist Air Traffic Controllers in their task on terminal sectors, notably by providing controllers with an efficient way to feed the system with clearance data. In few words, controllers' activity in this context is characterized by the construction of a proper sequencing of arrival flights towards a geographical point called the Initial Approach Fix (IAF, sector exit point) respecting airport capabilities.

The vertical view constitutes one of its specific tools. It allows a better monitoring of the vertical profile. Former studies have proved that controllers tend to be even deliberately blind in the vertical profile in the current environment.



Figure 7 : ASTER comet 1 (left), ASTER comet 2 (right)

In the Figure 7, the aircraft comets show the position of the aircraft in the vertical view, among many other information.

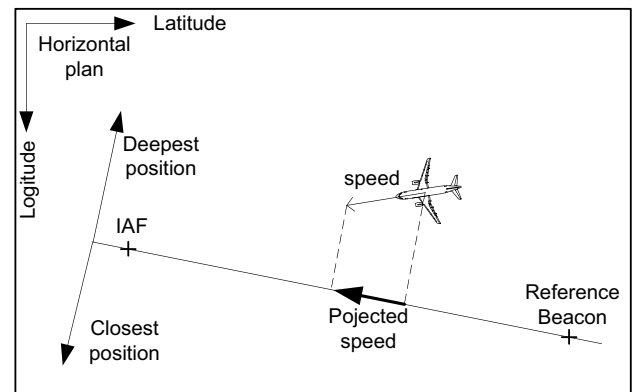


Figure 8 : Aster projection plan

The displayed information in the ASTER project is based on a projection along an axis. The IAF is the first point of this axis and a reference beacon is the second. This axis split the sector into two parts. The aircraft behind this axis are deeper than the aircraft in front. Actually, the aircraft speed representation is the result of the projection of the current speed on this axis, whereas the aircraft position is the distance between the aircraft projected position and the IAF. All the information are summarised in the Figure 8 and in the C&M characterisation (Table 5).

IV. COMET CHARACTERISATION

This section deals with the comet characterisation. Firstly, we will apply the C&M model; secondly we will discover that this model is a partial characterisation. Finally we will characterise comets with a table inspired from the IV tools.

A. Applying C&M characterization



Figure 9 : the comet of an evolving aircraft, the image exhibits direction and acceleration changes

The last positions of the aircraft merge by effect of Gestalt continuity [14], from which a line does emerge with its particular characteristics (curve, regularity of the texture formed by the points, etc). In this case, it is not possible to characterize the radar comet directly using the C&M transformation model. But we can characterize individually the shapes that build the comet (Table 4). With this intention, we introduce the concept of current time (T_{cur}: the time when the image is displayed). The size of the square is linearly proportional to its age.

Name	D	F	D'	X	Y	Z	T	R	-	□	CP
X	Q Lon	f	Q Lon	P				Shape emerge			
Y	Q Lat	f	Q Lat		P						
T	Q	f(T _{cur})	Q				S				

Table 4 : C&M Radar Comet

Name	D	F	D'	X	Y	Z	T	R	-	□	CP
Plot	Lat Lon (QxQ)	f	Q	P				Shape			
Afl	Q	f	Q		P						
Vert. speed	Q	f	Q						O		
speed	Q	f	Q						S		

Table 5 : ASTER Comet characterization

The characterization cannot integrate controllers' analysis of the evolution of aircraft latest positions (speed, evolution of speed and direction). Thus, in Figure 9, the shape of the comet indicates that the plane has turned 90° to the right and that it has accelerated. These data are emergent from the comet design. In other words, they were not directly used to generate the image.

1) ASTER and the Speed Vector

The characterization of the radar speed vector (Table 6) shows that its size (in Bertin's notation, but as it is a line, we can also use the term length), changes with the aircraft's speed.

Name	D	F	D'	X	Y	Z	T	R	-	□	CP
speed	Q	f	Q					S			
direction		f						O			

Table 6 : C&M Speed vector characterisation

In addition, the same information is coded by the length of ASTER comet and by the speed vector of the radar's comet. The ASTER comet is thus equivalent to the radar's speed vector, modulo a translation. It is the characterization and its comparison which allows us to link two visualizations, and thus to give to the designer elements of analysis. This result shows the importance of the work carried out.

2) C&M characterization conclusion

The characterization of C&M does not allow to highlight essential information for end users, and does not allow any exhaustive comparison of different design. The ODS comet is richer than the Aster comet; although the characterization of C&M seems to indicate the opposite. The wealth of information transmitted by each representation is thus not directly interpretable in the characterizations: the model of C&M is therefore not fully adapted.

The next part of this paper will take into account the knowledge of the InfoVis field and apply it to characterize four design of the comet.

B. Alternative characterization

In this part, we present all the available information on each comet. These information are classified into three categories:

- The design process: how to draw the comet ?
- The design properties: what are the design characteristics ?
- The semantic: what are the displayed information ?

Tableau 8 lists all the terms used to characterize the comet design (Tableau 7).

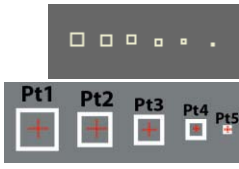
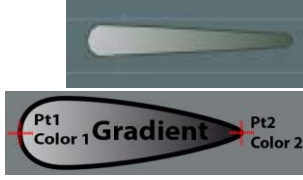
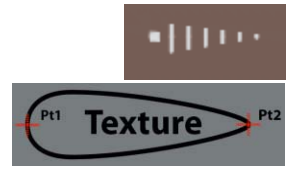
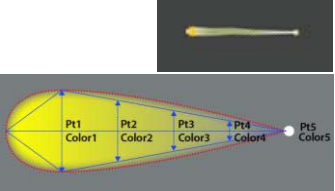
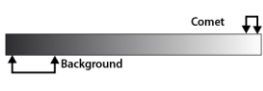
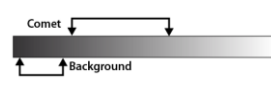
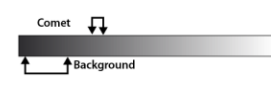
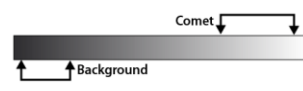
Comet	ODS	ASTER Design 1	ASTER Design 2	RadarGL
				
Design				
Refresh rate	steps	steps	steps	continuous
Enclosure	Same shape and progressive change in the squares size	shape	texture	shape
Design properties				
Zoom invariants	squares size	gradient	texture, color, thickness	gradient, thickness
Background occlusion	partial: holes in the "texture"	full opacity	partial : holes in the texture	partial : (transparency)
Screen depth	yes : priority comet	fake (automatic toolkit Z sorting)	no : texture blending	no : alpha blending
Overlapping resistant	Yes +	Yes ++	Yes +	Yes ++
Background and comet contrast				
Fixed shape	no	yes	yes	no
Semantic				
Acceleration	reflected in the varying distance between the squares	no	no	gradient and dynamic stretching
Depth	not implemented	thickness	no	not implemented
Radar track death	progressive dot fading	fade	fade	lock up and fade
Direction	horizontal plan : the direction given by the tangent at the first point	in the vertical plan : orientation	in the vertical plan : orientation	horizontal plan : orientation curvature
Tendency	curvature	No : unless you perceive the screen refresh	no: unless you perceive the screen refresh	curvature tendency
Speed	horizontal speed : length of the comet	horizontal projected speed : length	horizontal projected speed : length	horizontal speed : length of the comet
Highlight comet head	yes	no	yes	yes
Highlight comet tail	no	no	no	yes
display old positions	yes	no	no	yes

Tableau 7 : description of four comet designs

Refresh rate	Continuous: The aircraft positions are interpolated between two data updates (or frames). Steps: no interpolation, no transition between two frames.
Enclosure	Explains how the comet is perceived as a whole.
Zoom invariants	The visual properties not affected by the zoom variation.
Background occlusion	Is the background still visible through the comet design?
Screen depth	In case of comet overlapping, can we perceive an ordering along the Z axis (screen depth) and how? (Painter Algorithm)
Overlapping resistant	When comets overlap, can we distinguish between comets?
Background and comet contrast	Display the range of luminosity of the background and the comet. Displays if there is a risk of confusion between the background and the comet.
Fixed shape	Does the design always display the same shape?
Acceleration	The aircraft acceleration.
Depth	The depth in altitude (RadarGL and ODS). The depth (ASTER) with respect to the projection axis (Figure 8).
Radar track death	How does the comet show the lost of data?
Direction	The aircraft direction.
Tendency	The evolution of the aircraft direction.
Speed	The aircraft speed.
Highlight comet head	Does the design highlight the current aircraft position?
Highlight comet tail	Does the design highlight the oldest aircraft position?
Display old positions	Does the design display more than the current aircraft position?

Tableau 8 : Comet characterization legend

1) *ODS : historical design*



Figure 10 : ODS comet

The design of this comet has already been depicted. The unity of the resulting shape comes from the Gestalt[14] properties. The progressive change in the square size and the squares spacing glue together the squares. This is helpful when comets overlap: although plots may interfere, we are still able to distinguish between two comets.

2) *ASTER design 1 : ecological design*

ASTER comet thickness codes the position (or depth) of the aircraft compared to the IAF-reference axis (Figure 8). If the aircraft is deeper than this axis, the comet is darker and thicker. This is an ecological design because far object are small and dark. But this design can lead to perception issues with the background.



Figure 11 : Aster design 1 comet

The size of the comet is a function of the ground speed. The vertical speed is coded by the orientation of the comet. The comet length corresponds to one minute flight. The background has an altitude scale. Thus, if we compare the altitude of the comet's end with the altitude of the beginning of the comet we read the vertical speed in Fts/Min (Figure 11).

3) *ASTER design 2 : technical design*



Figure 12 : ASTER comet

The designer wanted a slightly different comet, because ASTER will be used with an ODS screen, and the controller must not confuse the two screens (one code a vertical view, the other a top view).

Due to technological constraints, the first version of ASTER could not code old aircraft positions. Thus the designer used a texture to display the current aircraft position, the length to code the speed and the orientation to code the vertical direction.

4) *GLANCE: empirical, prospective design*



Figure 13 : RadarGL comet

This comet is a shape created with the previous aircraft positions. To draw this comet, five points are needed. Each point has the same color but not the same transparency (alpha). A two pixel width border is added around the shape to smooth the edges. The color choices are empirical, and because the end of the comet blends with the background, the last point of the comet is highlighted with a white dot.

The refresh rate is continuous, the animation is smooth. The acceleration is code with the stretching speed of the comet.

5) Comet Comparison

RadarGI and ASTER first design (ASTER design 1 in the table) have a better occlusion resistance. It means that if they are many overlapping comet you can still figure out each comet? With the ODS and the ASTER second design, the comet is created with several entities (texture with holes and squares). This is a design misleading with the texture form ASTER first design; by analogy with the ODS design, we may think that each line is an old aircraft position.

The comet tendency (direction evolution) in the ASTER design can be seen only if we see the transition between two comets states, which is unlikely. With the ODS design, the old position are always visible and then the tendency.

The ASTER deep of an aircraft (to the projection line) is Quantitative information, but it is coded with the Ordinal luminosity which means you may lose some efficiency.

The comet thickness is not invariant with the zoom, and the thickness code the aircraft deep. This is a software mistake which doesn't interfere with the Air traffic controller activity. In a nominal use of ASTER, the zoom ratio remains the same.

V. CONCLUSION

In this article, we have explored the characterization tools available in the InfoVis field, and applied them to rich and dynamic visualizations. Whereas Card and Mackinley depicted some InfoVis visualizations without explicitly demonstrating how to use their model, we have shown the practical effectiveness of the C&M model in our comparison of the ASTER comet and the ODS speed vector. Although existing characterization tools are evidently valuable, they are not sufficient to characterize emerging data and image dynamics.

In addition, we have built an exhaustive description of four comet designs with the exception of user activity and perception. With this strong constraint, we can still make comparisons, find design justifications and even detect design errors.

This paper describes the first steps toward building a method to describe visual entities systematically. In particular, we try to characterize them, i.e. to find a precise and compact description that unveils differences and allows comparison. We seek to answer the following questions: what information is displayed on the screen? How many information are displayed? How is it displayed? At first sight, it seems that the answer is trivial: the information on the screen is exactly what the designer wanted to put there when he designed the visualization. However, we saw that the answer is more complex, as it does not take into account information built up from our perception system, or from the dynamic aspect of the image. We want to insist on the fact that we do not try to assess the effectiveness of different representation. We only identify what is displayed and not how well a user perceives it. The ability to characterize visualizations would bring several benefits to the design process. It would help designers to assess their designs, reuse existing designs in new contexts, communicate with other designers and write compact and

unambiguous specifications. The research described in this paper is an initial effort to develop a theory-driven approach to the characterization of visualizations.

ACKNOWLEDGMENT

The author gratefully acknowledges the contributions of Sylvie ATHENES, Benjamin TISSOIRES, Gilles TABART and Jean-Luc VINOT. A special thanks to all the members of the DTI for their discussions, comments and suggestions. This work is supported by a PhD scholarship from the DTI R&D and ENAC LII, with the partnership of the IRIT Toulouse, France.

- REFERENCES
- [1] C. Ahlberg, Spotfire: an information exploration environment. SIGMOD, 1996.
 - [2] T. Baudel, Browsing through an information visualization design space. CHI '04. ACM Press, New York, 2004.
 - [3] R. Benhacene, A Vertical Image as a means to improve air traffic control in E-TMA. USA, DASC California, 2002.
 - [4] J. Bertin, Graphics and Graphic Information Processing deGruyter Press, Berlin, 1977.
 - [5] S.K. Card, J.D. Mackinlay, The Structure of the Information Visualization Design Space. In Proc. Information Visualization Symposium '97, 1997.
 - [6] S. Card, J. Mackinlay, B. Shneiderman, Information Visualization Readings in Information Visualization: Using Vision to Think. Morgan Kaufman, introduction p 1-34, 1998.
 - [7] ED. Chi, A Taxonomy of Visualization Techniques using the Data State Reference Model. InfoVis '00. 2000.
 - [8] W. Cleveland, A Model for Studying Display Methods of Statistical Graphics, Journal of Computational and Graphical Statistics, Vol. 2.
 - [9] W. Cleveland, R. McGill, Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. Journal of the American Statistical Association 79, 1984.
 - [10] J.D. Fekete, The InfoVis Toolkit InfoVis'04, Austin, TX, Oct 2004. IEEE Press. pp. 167-174, 2004.
 - [11] A. Inselberg, The plane with parallel coordinates. The Visual Computer, 1984.
 - [12] J. Itten, The Art of Color: The Subjective Experience and Objective Rationale of Color, 1997.
 - [13] B. Johnson, B. Shneiderman, Tree-maps: A Space-filling approach to the visualization of hierarchical information structures. IEEE Visualization '91, 1991.
 - [14] K. Koffka, Principles of Gestalt psychology Routledge 1935.
 - [15] J. Mackinlay, Applying a theory of graphical presentation to the graphic design of user interfaces. UIST '88, 1988.
 - [16] J. Norman, W. Thrower, Edmond Halley as a Thematic Geo-Cartographer, Annals of the Association of American Geographers, Vol. 59, 1969.
 - [17] R. Rao, S.K. Card, The Table Lens: Merging graphical and symbolic representations in an interactive focus plus context visualization for tabular information, in Proc. CCHI '94 ACM, 1994.
 - [18] G. Robertson, J.D. Mackinlay, S.K. Card, Cone trees: Animated 3D visualizations of hierarchical information, 1991.
 - [19] S.S Stevens, On the theory of scales of measurement. Science, 1946.
 - [20] E.R. Tufte, The Visual Display of Quantitative Information, Graphics Press, Cheshire, Connecticut, 1983.
 - [21] C. Ware, Information Visualization, perception for design, Morgan Kaufmann, 2002.
 - [22] L. Wilkinson, The grammar of Graphics. New York: Springer Verlag, 1999.

Functional Analysis of Human-Human Interactions during Collaborative Decision Making in Flight Operation

Matthias Groppe
School of Engineering
CRANFIELD University
m.groppe@cranfield.ac.uk

Marc Bui
LaISC
EPHE Sorbonne
marc.bui@laisc.net

Romano Pagliari
Air Transport Department
CRANFIELD University
r.pagliari@cranfield.ac.uk

Abstract—The objective of this study is to understand the cooperation building process within Human-Human Interaction (HHI) during Collaborative Decision Making (CDM) at a distributed decision making environment across objective functions. It is based upon functional HHI analysis within typical Air Traffic Management (ATM) operation situations.¹

In this paper, different flight and turn-round operation situations are compared and characterized by: (1) a synchronous interaction mode, where all participating operators interact with each other at the same time, and (2) an asynchronous interaction mode, where the participating operators interact with each other at different times. In both situations, only HHI which require cooperation among operators across different locations and objective functions are contemplated. Interactions take place through a written text or speech. Task and decision making for all situations is distributed between operators. The aircraft pilot's perspective and their information requirements during these flight and turn-round situations are used to identify critical information processing during CDM: All situations are usually time constrained, change quickly, and require a highly dynamic information transfer. Thereby, information sharing for decision making can be either homogenous having all operators the same information required or heterogeneous where information is not equally shared among operators.

This study relies on a structural model of team collaboration, developed for analysis on the cognitive mechanisms of CDM, and to handle both synchronous/asynchronous and collocated/ distributed collaboration environments like in geographically distributed and time delayed situations of the military or flight operation.

Index Terms—Air traffic management, asynchronous distributed collaboration, collaborative decision making, human-human interaction

I. INTRODUCTION

UPDATED from earlier projects in United States, the European CDM approach was introduced during field trials at selected European airports with the aim to achieve cooperation at *planning level* via information sharing and common situational awareness (CSA). However, from aircraft pilots' perspective on current air traffic operation, many problems encountered with CDM arise from human-human interactions (HHI) at *action level*; whereby HHI at *action level* refer to interactions with a shorter time span and less abstraction than HHI at *planning level* (Hoc, 2000). Further problems for CDM operation are conditioned on the specific situation of decision making in an *asynchronous, distributed* collaboration environment like it can be found in ATM operational decision making. Operators like aircraft pilots or ground handlers communicate with the operational centers of the airlines, ATC, and the airport through speech (e.g. via phone or radio) or written text (e.g. via ACARS). Hence it will be addressed, how the airport CDM information sharing process is influenced by the following variables:

- Interaction Mode (synchronous versus asynchronous)
- Information Distribution (homogenous versus heterogeneous)

This functional analysis of flight situations also includes *micro level* (neural-cognitive) aspects on CDM. Even little understanding of operators think during CDM in asynchronous, distributed environment exists [25], an analysis of HHI within CDM via the perspective of a single operator (aircraft pilots) is used in order to cope with the still very inadequate mechanisms of collaborative problem solving during operators' decision making. According Ferber [1], HHI situations can be classified as antagonistic, cooperative, or indifferent depending on three main variable components like *aims, resources, and abilities*, hold by each participating operator. This classification is applied in order to understand micro level cognitive aspects of HHI in Airport CDM flight operation situations. The advantage of using aircraft pilots as

¹ This study is conducted with the financial support from EUROCONTROL Experimental Centre and FRAPORT Foundation „Erich Becker“

reference group is that there is a non punishment philosophy against pilots in dealing with punctuality problems: The presently use method of *delay assignment* intends to find out the reason of delay and assign the responsibility to a single operator via delay codes. Usually each operator tries to avoid assignment of a delay because a pay deduction has to be expected.

In this paper, prototypical HHI situations between all operators involved in flight and turn-round operation are introduced. They all take place in a distributed collaboration environment. Four proposed situations concern the turn-round of aircraft, where coordination of processes is necessary; processes include parking, ramp side, land side, and special ground handling processes. Within these situations, cooperative HHIs are mandatory: pilots have to coordinate processes with other operators like representatives of the ground handling companies, airport, airline, air traffic control, and Central Flow Management Unit. Cooperation and decision making is distributed between pilots and other operators: Decision making for the begin of all turn-round process which are in direct relation with the aircraft (e.g. boarding, de-boarding, refuelling, cleaning..), is within responsibility of the pilots: other operators are concerned with decision making for coordination and execution of these processes, and again cooperate with each other. While any delayed process start can result in an overall delay of the subsequent flight, coordination of a standard turn-round (defined as a reference model) is usually predetermined.

During normal turn-round operation, *interactions* between pilots and other operators are usually *synchronous*. Coordination of actions takes place via predetermined key events (*milestones*) [8], organized as a sequence of interactions between operators within the airport operation centre; if a non-standard situation like aircraft change, technical repair, adverse weather operation, etc. occurs, ad hoc coordination of all necessary events *via face-to-face* communication between pilots and ramp agents or via radio/ phone between pilots and other operators coordinating from airport operation centre takes place. The *milestone approach* used for CDM, includes *all* events which are necessary for an uninterrupted turn-round process, whereby some key events take place already far ahead of the turn-round itself. Information distribution during turn-round is mainly *heterogeneous* between participating operators on *action* and *planning* level caused by the information dynamics in the highly dynamic environment of the turn-round operation and the varying tasks in the different domains itself. However, in order to cope with the usually limited time span for turn-round operation, CDM targets *homogenous* information processing to achieve a CSA between all participating operators and to avoid departure delay caused by non-standard operation.

Another four proposed situations concern the *flight*, starting from aircraft leaving the parking position until reaching parking position at destination. Coordination here is also necessary for departure and arrival sequencing with other aircraft, usage of taxiways, airways and airspace/ sectors. It is pilots' responsibility to execute the flight according defined rules under consideration of highest degree of safety possible.

Other operator involved during flight for coordination of traffic is air traffic control (ATC) by keeping safe separation distance between aircraft and managing air traffic flow by issuing clearances to the pilots. The different level of control between pilots and other operators like ATC in this situation is that ATC has authority about assigning the airspace in form of clearances to the pilots and again depend on cooperation from pilots, to adhere to these clearances. Decision making is shared between pilots and ATC within their domain relative to the situational need, but has to be executed under mentioned safety constraints. Other operators like the airline company or CFMU are only marginally involved in decision making during flight operation.

During flight operation, interactions between pilots and air traffic control are *synchronous* or *asynchronous* via radio during flight through one sector or when ATC issue clearances to the pilots; interactions between different ATC sectors can also be *synchronous* or *asynchronous*, resulting in a non-coordinated flight through different sectors; interactions between pilots and other operators during flight are usually *asynchronous* and *distributed*. They are coordinated by *milestone events* within the airport control centre. Information distribution for clearances concerning airspace and routing is always homogenous, while information distribution for e.g. reasons of deviations from clearances can be homogenous or heterogeneous depending of the impartation willingness or time in hand from ATC and aircraft pilots.

Like during turn-round operation, the highly dynamic environment of the flight operation results in high dynamic information content. Some information dynamics like variations in flight progress occur on standard basis and changes are automatically accessible to all participating operators via data link transmission. However, non-standard information dynamics like operational changes or technical issues are transferred by non-synchronized interactions and need to be manually transferred between operators. This requires cooperation among operators' interactions and defines the need to achieve a CSA among all operators.

The resulting objectives for the study are:

- To understand the cooperation building processes of the HHI during day-to-day flight operation which are necessary in context of an distributed collaborative decision making environment across objective functions of all operators.
- To identify the information sharing components which should be employed to optimize the CDM concept in ATM typical standard & non-standard flight situations.
- To understand how agents can support humans in achieving collaborative knowledge during distributed collaborative problem solving.

As a result of study, a new CDM process design will be evaluated during quasi-experiments in the natural CDM setting at Munich International Airport which aims at achieving cooperation of all partners involved.

II. THEORETICAL BACKGROUND

In our context of flight operation, HHI are seen as dynamic relations between pilots and other operators via a number of mutual actions. Each action by one operator has consequences which influence the behavior of the prospective behavior of the operators. Series of actions form events, and a number of events form the turn-round or flight situation (e.g. ATC assigns a parking position for the aircraft to the pilots (event) via mutual communication usually by two-way radio communication (HHI) in a turn-round situation). Ferber [9] defines interaction situations as *a number of behavioral patterns which evolves from a group of agents, who have to act in order to reach their targets and thereby have to regard their more or less limited resources and capabilities*. By using this definition, interaction situations can be described and analysed, because it defines abstract categories like cooperation, antagonism, and indifference via differentiation of observed key commonalities and different interaction situations. The relevant components for classification of interaction situations are the aims and intentions of the different agents, the relations of the agents to available resources, and abilities of the agents in regard to their assigned task. These criteria are used to define different types of interaction situations (Figure 1).

Aims	Ressources	Abilities	Type of Situation	Category
compatible	sufficient	sufficient	Independence	Indifference
compatible	sufficient	insufficient	simple working together	Indifference
compatible	insufficient	sufficient	blockade	Cooperation
compatible	insufficient	insufficient	Coordinated collaboration	Cooperation
incompatible	sufficient	sufficient	pure individual competition	Cooperation
incompatible	sufficient	insufficient	pure collective competition	Antagonism
incompatible	insufficient	sufficient	individual resource conflict	Antagonism
incompatible	insufficient	insufficient	collective resource conflict	Antagonism

Figure 1: Classification of Interaction Situations (Source: Ferber, 2001)

Each type of interaction situation has its own relation towards cooperation: In an *Independence* situation, no interaction takes place and sufficient resources and abilities allow a coexistence of operators without any constraint. This situation has no relevance for ATM at congested airports. A *Simple Working Together* situation defines a collaboration situation which does not require coordination between operators, while a *Blockade*, *Coordinated Collaboration*, *Pure Individual/Collective Competition* and *Individual/Collective Resource Conflict* are situations which are expected to dominate in our contemplated HHI situations. These situations require coordination between operators and, depending on resources, aims, and abilities, can result in cooperative or antagonistic behavior.

During flight operation situations, HHI are usually not binding relations between involved actors and no mutual

influence is exercised between pilots and other operators; therefore social components of the interactions are not contemplated.

According Hoc [12], cooperation can exist within various levels in terms of distance from the action itself: A cognitive architecture of cooperation model classifies cooperation in abstraction level and process time depending on the proximity to the action itself (Figure 2).

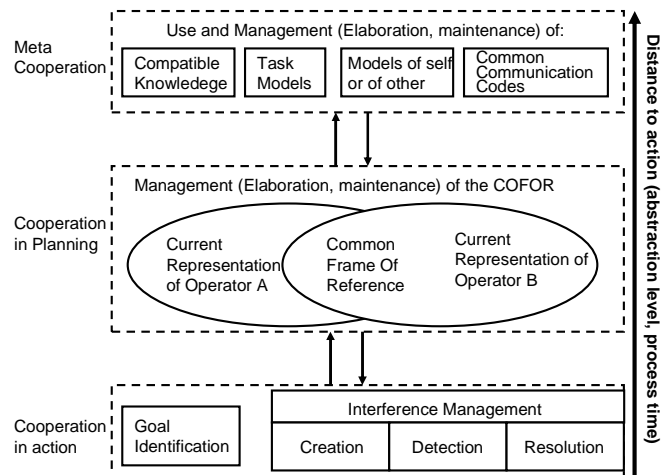


Figure 2: Processing Architecture of Cooperation (Source: Hoc 2000)

For the study of HHI situations, we focus on cooperation (or antagonism, if relevant) on action level. At *action level*, the operators perform operational activities related to their individual goals, resources, and abilities. Hoc [14] has defined four types of activities at action execution level which are interference creation (e.g. mutual control), interference detection, interference resolution, and goal identification (Goal identification also embodies identification of other operators goals). Cooperation at action level has short-term implications for the activity, as opposed to the more abstract type of cooperation at planning level. Interference creation relates to the deliberate creation of interactions; interference detection to the ability of detecting interferences, especially in non-deliberate interference situations; and interference resolution to the actual interaction in order to find a cooperative solution. Mutual domain knowledge is the basis for other operators' goal identification, to facilitate operator's own task, the other's task, or the common task.

At *planning level*, operators work to understand the situation by generating schematic representations that are organized hierarchically and used as an activity guide [13]. Schematic representations include the concept of situation awareness [23], and operators' goals, plans, and meta-knowledge [13]; therefore current approach to CDM operation

in ATM is seen as an approach towards cooperation on planning level. De Terssac and Chabaud [7] use the term COFOR (Common frame of Reference) as a mental structure playing a functional role in cooperation and as a shared representation of the situation between operators likely to improve their mutual understanding [3]. The topmost level in Hoc’s model, the meta-cooperation, as a level developed from knowledge of the other two levels, is not contemplated in the study.

Also Piaget [20] distinguishes between cooperation seen from structural (e.g. network organization) or functional point of view which looks at cooperation as activities performed by individuals within a team in real time. Two minimal conditions must be met in cooperative situations: (1) each actor strives towards goals and can interfere with other actors on goals, resources, and procedures. (2) Each actor tries to manage interference to facilitate individual activities or common task. Both conditions are not necessarily symmetric, because goal orientation or interference management depend on individual behavior or time constraints.

Hoc [12] argues that current air traffic management (ATM) is more concerned with operators’ plans, goals, or role allocation instead of common situational awareness. But Lee [17] determines situational awareness, responsibilities and control, time, workload, and safety constraints as key factors driving collaborative behavior in air traffic control operation: *To have proper awareness of the situation, a controller and/or pilot needs to initiate or be informed of actions taken by other operators.* But time pressure and safety issues have negative effect on communicative behavior and therefore also cooperation or common situational awareness.

Share of responsibility and control are often different but determined through situation (e.g. air traffic controllers issue clearances which have to be executed by pilots). Nevertheless, *the more assistance, the more anticipative the mode of operation in controllers and the easier the human-human cooperation* [13].

Collaborative Decision Making means *applying principles of individual decision making on groups, whereby groups are established with the aim to show collectively a specific behavior* [15]. This implies that cooperation of participating individuals should be beneficial for CDM operation, also in air transport management. But how does cooperative work look like at day-to-day basis? Cooperation has *a wide variety of connotations in everyday usage* [24]. Do people only cooperate, if they are mutually dependant in their work or is mutual dependency sufficient for cooperation to emerge? In context of CDM operation, confrontation and combination of different perspectives of cooperation is an issue: how is pilot’s perspective embedded in the current CDM approach? For Schmidt [24], the multifarious nature of the task can be matched by application of multiple perspectives on a given problem via articulation of the perspectives and transforming/translating information of different domains.

The challenge of CDM operation in ATM is the unique

cognitive mechanisms in a distributed and highly dynamic environment like it can be found in flight operation. Similar situations can be found in military teams with asynchronous, distributed teams for mission planning and mission execution, but in general it is a relatively new area [16]. Other domains which have related aspects to asynchronous distributed collaboration are not contemplated. Warner [7] describes the major factors impacting collaboration which are the *collaborative problem environment, operational tasks, collaborative situation parameters, and team types* (Figure 3).

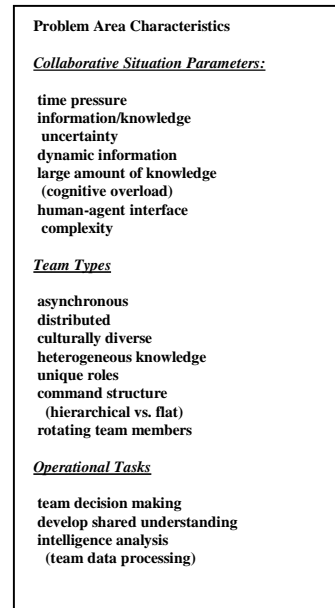


Figure 3: Problem Area Characteristics for Collaboration (Source: Warner, 2003)

His structural model of collaboration focuses on *team decision making, course of action selection, developing shared understanding, and intelligence analysis.* Thereby, various parameters can influence the collaboration performance [26]. The collaborative decision parameters can be adapted to fit the specific environment of CDM in other domains using the respective characteristics under *operational tasks, collaborative situation parameters, and team types.* Werners’ structural model of team collaboration uses the minimum number of unique stages identified in team collaboration literature and the results from a Collaboration and Knowledge Management Workshop (Figure 4).

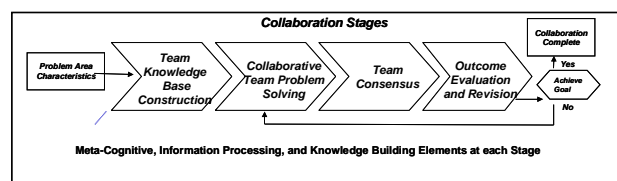


Figure 4: Structural Model of Team Collaboration (Source: Warner, 2003)

This structural model is based on the meta cognitive processes of an information processing and communication approach. For Davidsen [24], meta cognition is the knowledge of one's own cognitive processes in explaining how human cognitive processes are used for problem solving. According Werner, there is 'no generally recognized unified theory of human cognition'. By implementing Ferbers' component approach, a micro level cooperation perspective is applied into the structural collaboration model. This approach allows to adapt the structural model of team collaboration to an distributed decision making environment under consideration of decision making across objective functions (e.g. like Airport CDM).

III. METHODS

A methodological approach is used for the analysis of the cooperative mechanisms within HHI. First, all flight & turn-round situations which are seen as critical for CDM operation in terms of punctuality are determined via in-depth interviews with senior commanders of different airlines. All situations were decomposed in elementary activities and thereafter grouped into event classes. The classes within turn-round situations include the subclasses gate assignment, standard ramp services, standard land-side services, and non-standard turn-round services. Flight situations include the subclasses clearance variations 'ground', clearance variations 'flight', information processing, and information forwarding. Some event classes have only one possible event as problem cause.

For each event class, the collaboration stages analogous Werner's structural model were identified. To understand how participating operators think during each stage, a self-administered questionnaire was developed which aims to get knowledge about information processing (meta-cognitive level) and interaction components (micro cognitive level) between participating CDM operators within distributed collaborative decision making. All questions were designed from the perspective of the airline pilots as members of distributed airport collaborative decision making. (Perspectives of other operators could also usefully be researched). As reported by airline pilots, all event classes have critical elements concerning collaboration. Therefore the questions are designed to find the most problematic stage within the collaboration process.

Team Knowledge Base Construction is the first stage in team collaboration and includes steps like identifying relevant domain information, selecting team members, setting up the communication environment, individual team members own mental model of the situation, and developing individual and team task knowledge. In ATM, information processing is established in day-to-day operation via various communication modes like phone, ACARS, or radio. Whereby, agreed methods of information sharing or filtering are established among all operators within the airport operation centre, while information sharing between distributed operators like airline

pilots or air traffic control is not taking place on an agreed standard. Therefore, an overall cognitive process of how to understand elements, relations, and conditions that compose the emerged problems, is not established among operators (meta cognitive), even some operators may have a mental model of situation parameters and their relationship. The airline pilots as operators of distributed site are asked to state, if information sharing for this stage is seen as sufficient for knowledge building and how important information sharing is for them. No further details like *how* or *which* information should be collected, how to understand problem task, or how communication mechanisms should be established, are analysed at this stage of research.

For the stage of *Collaborative Team Problem Solving* a closed-end question is designed to again catch the overall airline pilots' perspective from distributed participation in CDM. This stage can only successfully be accomplished, if a shared situational awareness of the emerged problem exists, because it builds on the identified and understood problem among all operators. This stage starts after information are processed among operators and has the aim to find a viable problem solution. Since each participating operator has its' unique domain constraints, the definition of a global goal and solution alternatives among all operators is the challenge for this collaboration stage. For the functional analysis, an approach at micro cooperation level is pursued and questions are developed analogous Ferbers' component model, which are *commonality of aims* among participating operators, *amount of resources available*, and *ability of function* to perform assigned task. All three components together form different interaction categories and give evidence about cooperation (or antagonism).

The stages of *Team Consensus* and *Outcome Evaluation and Revision* will not be used for this first phase of research.

The components from Ferber [9] which classify the interaction situations are:

- *Compatibility and Incompatibility of aims*: Effect on cooperation can be negative, if aims are not compatible. Therefore critical activities during turn-round and flight are questioned for possible conflicting goals between pilots and other operators.
- *Availability of resources*: Resources are limited, therefore conflicts can arise which result in disturbances of HHI. Increasing airport congestion and abridged turn-round time of aircraft contribute to possible shortage of resources and result in reduced latitude of action or even individual competition between operators. Questions are out to test, if resources in terms of the time available for ground processes are aligned with the operational and safety requirements. Current approach on CDM operation is an attempt to challenge resource constraints via coordination of actions.

In this context, information is also seen as a *resource* which has to be available to each operator in order to execute individual task: Information has to be shared between pilots and other operators to achieve common situational awareness. Initial data from interviews relate numerous problems regarding to cooperation on failures in information sharing. A number of elementary activities/ events are used to obtain data about possible reason for failed cooperation and effects on flight punctuality caused by information sharing problems. Questions are out to test, if there is a relation between failures in information sharing and delay (departure or arrival delay). Failure is seen, if a part of information is missing or if information is not delivered on-time.

- *Ability of operators in relation to their assigned task:* It can not be assumed that knowledge and abilities of operators are automatically sufficient for executing assigned task. This is of course also appropriate for pilots, but it is unlikely to get realistic results from questioning of pilots, if the person asked has to determine or admits its' own inability. On the other hand it is unlikely that pilots are familiar with all other domains involved and can determine necessary abilities from other operators. Therefore only random questions are out to test pilots' perspective in a few events like failed information sharing or unpunctual process execution.

Finally, semi-structured in-depth individual interviews with further representative commanders will be conducted to clarify the content of the questionnaire results. This is necessary to capture the meaning behind the essential results and to understand operators' attitude towards cooperation.

IV. DEMONSTRATION

Data collection is still ongoing, only primary results are available to demonstrate usefulness and applicability of the survey.

A. The Environment of the Cockpit

Activity analysis on flight decks of commercial aircraft shows two pilots sharing flying and other duties like communication with ATC, monitoring flight instruments and all other tasks necessary. While pilot flying is responsible for steering and navigation of the aircraft, pilot not flying disburdens him with all other duties necessary in order to maximize safety by clear task sharing, since primary responsibility of the pilots is to steer the aircraft from departure to destination airport under maximum possible safety considerations.

In various situations they encounter interactions and interrelations with other actors involved in ATM operation. Flight relevant and operational information is shared with them.

The environment of the aircraft specifies a special case of decision making: the commander of the aircraft has the topmost responsibility of all decision making on board the aircraft. This can be compared as decision making with an individual decision maker and a group of advisers. He can either use his position to listen to his various advocates of different positions or actions or execute a structured analysis by the use of help from experts or advisers (airline company, ATC, ground handlers....). It is his final responsibility to identify key uncertainties in decision making and either adhere to objectives for the organization or his personal goal. Conflicts can arise through levels of authority and responsibility between advisers and the cockpit. Further losses of efficiency in this kind of decision making may result from other players' interactions, lack of information or limited ability of decision making. The advantage from this individual decision making is that a group of advocates is involved and therefore has more resources available [21]. Decision making seen from cockpits' perspective is also distributed since a number of decisions necessary for the flight operation remain in responsibility of the advisers (ATC, airline company, airport....).

Figure 5 shows a typical flow of HHI from cockpit's perspective identified from own experience: Aims, resources, and abilities form the basis for information exchange, decision making, and possible negotiation. Information exchange again is the basis for common situational awareness and coordination among operators. Decision making anticipates information exchange among actors which can also be used for mutual's goal identification.

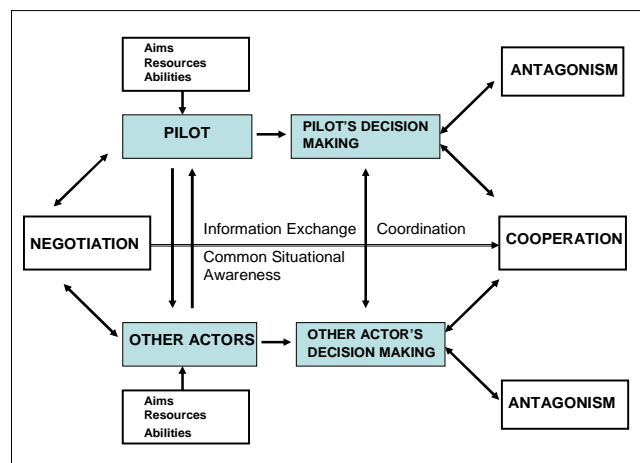


Figure 5: Cockpit's Perspective of Human-Human Interactions in ATM (Source: Own Illustration)

B. The Collaborative Decision Making Approach in ATM

The basic Airport CDM includes *Airport CDM Information Sharing* and the *CDM Turn-round Process* as a requirement for all subsequent airport CDM applications. Information sharing uses existing infrastructure at airports, but combines

data from different sources and operators. Quality of information at each phase of flight is determined by defined rules in order to establish a common situational awareness between all operators involved.

The *Milestones Approach* defines the airport CDM turn-round process which links the flight and ground segments via a set of milestones in the aircraft turn-round process, ranging from planning of the inbound flight until the take off of the flight at the subject airport. Each milestone is monitored and allows participating operators to identify possible deviations from schedule by the use of an alarm system.

Subsequent Airport CDM levels are the *Variable Taxi Time Calculation* and the *Collaborative Management of Flight Updates (Level 2)*. Variable taxi time calculation aims the introduction of a realistic taxi time in order to increase punctuality and slot adherence. Collaborative management of flight updates aims an improved operation and flexibility by slot swapping and slot shifting to take aircraft operators' preferences into account.

A *Collaborative Predeparture Sequence (Level 3)* aims to replace the present 'first come first serve' practice by consideration of aircrafts' and airport operators' requirements. [8].

C. Shared Situational Awareness between Pilots and Other Operators

Technological advances now allow communication and collaboration without being physically together. ATM systems today have adopted these technologies; however the highly dynamic environment of flight operation requires a fast and flexible adaptation to the changed situation. CDM is an answer to tackle problems with the short time span in hand for decision making, but no procedures are established to include interactions with airline pilots as standard CDM process. ACARS, phone, or radio are in place as possible collaboration support tools for synchronous or asynchronous decisions between airline pilots and other operators to contribute to the information sharing/ knowledge building process. Current approach to information sharing/ knowledge building is the issue of a Target Start Up Approval Time (TSAT) and a Target Off Block Time (TOBT). TSAT is a reference start up time for coping with air space constraints, TOBT has airport, airline and turn-round constraints as determining factors. Both should match as close as possible and be communicated to distributed decision makers. Constraining factors from airline pilots are not considered for calculation of TOBT or TSAT.

D. Critical Human-Human Interactions

30 pilots from different airlines were asked during unstructured questioning to name processes with problems in regard to HHI during day-to-day flight and turn-round operation. From all mentioned examples, a number of nine situations were defined and *figure 6* provides an overview of critical HHI as reported by the airline pilots. The situations do not have any statistical relevance in terms of importance or frequency; the aim was to find a wide spectrum of possibly critical HHI.

TURN-ROUND	COOPERATION/ NON-COOPERATION	COOPERATIVE COMPONENT	AIRPORT	FREQUENCY	RELEVANCE
Gate Assignment	Y/N	Aims/Resources/ Abilities	Hub/ Non Hub	Daily/Weekly/ Monthly	Avoidable Yes/No
Ground Handling/ Ramp	Y/N	Aims/Resources/ Abilities	Hub/ Non Hub	Daily/Weekly/ Monthly	Avoidable Yes/No
Ground Handling/ Land	Y/N	Aims/Resources/ Abilities	Hub/ Non Hub	Daily/Weekly/ Monthly	Avoidable Yes/No
Ground Handling/ Special	Y/N	Aims/ Resources/ Abilities	Hub/ Non Hub	Daily/Weekly/ Monthly	Avoidable Yes/No

FLIGHT	COOPERATION/ NON-COOPERATION	COOPERATIVE COMPONENT	AIRPORT	FREQUENCY	RELEVANCE
RWY/DEP Change	Y/N	Aims/Resources/ Abilities	Hub/ Non Hub	Daily/Weekly/ Monthly	Avoidable Yes/No
EAT	Y/N	Aims/Resources/ Abilities	Hub/ Non Hub	Daily/Weekly/ Monthly	Avoidable Yes/No
Clearance Variation	Y/N	Aims/Resources/ Abilities	Hub/ Non Hub	Daily/Weekly/ Monthly	Avoidable Yes/No
Operational Change	Y/N	Aims/Resources/ Abilities	Hub/ Non Hub	Daily/Weekly/ Monthly	Avoidable Yes/No

PILOT	COOPERATION/ NON-COOPERATION	COOPERATIVE COMPONENT	AIRPORT	FREQUENCY	RELEVANCE
Information Forwarding	Y/N	Aims/Resources/ Abilities	Hub/ Non Hub	Daily/Weekly/ Monthly	Avoidable Yes/NO

Figure 6: Critical Information Sharing Situations (Source: Own Data 2007)

E. Responsibility and control allocation between pilots and other actors

'The allocation of functions between humans and machines is a very old topic in human engineering' [14]. Function allocation in terms of responsibilities and control has been identified as key factor for collaborative human-human behavior in ATM [17].

While air traffic controllers are responsible to separate the aircraft during flight, the responsibility of the pilots is the safety of the aircraft. The environment of the aircraft is a special case of interaction mode: final decision making on board of the aircraft is not shared between equitable partners, but is in the hand of the captain of the aircraft. Other actors take the role of advisers for an individual decision maker. Nevertheless all actions and decisions are obligatory on achievement of safety. The captain may either use his position to listen to his various advocates of different positions, or executes a structured analysis by the use of help from experts or advisers (airline company, air traffic control, ground handler...) It is his final responsibility to identify key

uncertainties in decision making and either adhere to objectives for the organization or his personal goal.

The research shows that antagonistic or cooperative behavior can arise through different levels of authority and responsibility between the captain and his advisers [21].

Responsibilities on ground are shared among actors: the flight manager is responsible for boarding and check-in processes, ramp agent for delivery of flight documents and other operational information. To achieve cooperation, all actors should have the same aim [14] which implies that HHI take place between *equitable* partners.

F. Modes of Information Sharing in Pilot – Other Operators' Relationship

Central concern of CDM is information sharing and common situational awareness. Many studies have been devoted to information sharing at the airport control centre of CDM participating airports, but no focus has been made on exchange of information to the cockpit or receiving operational information from the cockpit.

Information sharing between airline pilots and other operators involves human activities, but as far as humans are involved for information provision and creation, failures may occur and have obvious consequences on reliability [19]. Pilots were asked to identify different classes of information failures during all phases of turn-round and flight. From cockpit perspective, the main concern is how information sharing and common situational awareness between flight crews and ground parties is accomplished in order to achieve a predictability and punctuality during flight and ground operation. It has to be addressed:

- How all necessary information is delivered to the cockpit or weather it is jammed at any interface.
- If necessary information delivered on time.
- How the information, forwarded from cockpit, is handled by other actors.
- How much delay is encountered, if information delivery is late or not executed?
- Which information not delivered has the greatest risk of producing delay.
- Which information, forwarded by crew, has greatest risk of producing delay?

G. Time Constraints

Time pressure can have opposite effect on cooperative behavior. During peak traffic and short turnaround, pilot workload is very high for several reasons: Available time for coordination of necessary ground handling processes on ground is short or voice congestion over busy radio frequency demands high attention. Any failures in coordination or any retarded process on ground holds the risk of encountering delay. During flight on busy frequencies, issued clearances by

air traffic control need to be executed promptly and often no time is left for negotiation. Especially during approach, high workload does not leave much time to gain situational awareness. Air traffic controllers' constraints are normally not visible to the pilot, but also controllers' time is very limited during busy approach hours, and therefore not much time is left, to share situational constraints or negotiate with the pilot. Especially in these situations, controllers depend on cooperation from pilots.

V. CONCLUSION

The analysis of micro level cooperative elements represented by the airline pilots' perspective within a distributed collaborative decision making environment across objective functions, is the first attempt to implement individual group members think into a structural decision making model within the domain characteristics of ATM operation. This study is expected to be useful, because the distributed CDM environment shows unique interaction characteristics and therefore requires a focus on operators' thoughts. The airline pilots' perspective is chosen because a non-punishment policy for pilots when causing flight delays is in place, opposite to other operators who have to expect pay deductions. De Ferbers' interaction classification identifies potential non-cooperative flight situations and results from questionnaire will be used for the design of experiments and possible further interactions in form of negotiations between operators. Experiments will also include modification of interaction e.g. via representation models. The unique situation of individual operators' objective function distinguishes decision making in ATM from decision making in other environments like e.g. military.

Further outcomes of the study are expected to include empirically based elements or design characteristics for collaboration in a distributed decision making environment across objective functions including information processing components. Empirically gained data will be used for development of an agent support in airside flight operation situations.

ACKNOWLEDGMENT

I would like to thank all participating pilots from Lufthansa and Lufthansa CityLine, Air Berlin, and Private Air for their collaboration in this study.

I would also like to thank EUROCONTROL Experimental Centre and FRAPORT Foundation 'Erich Becker' for their financial support.

REFERENCES

- [1] Amat, A.L. & Bellorini, A. (1996). The issue of ground/ cockpit integration: results from field studies, *8th European Conference on Cognitive Ergonomics*, ECCE-8, Granada, Spain
- [2] Bellorini, A. & Vanderhaegen, F. (1995). Communication and Cooperation Analysis in Air Traffic Control, *8th International Symposium on Aviation Psychology*, Columbus, Ohio, USA
- [3] Carlier, X. & Hoc, J.M. (2002). Role of Common Frame of Reference in Cognitive Cooperation: Sharing Tasks between Agents in Air Traffic Control. *Cognition, Technology & Work* 4, 37-47, Springer-Verlag London
- [4] Chavalarias, D. (2002). Emergence of cooperation and selection of interactions, *Center of Research of Applied Epistemology*, Ecole Polytechnique, Paris, France
- [5] Cox, G.; Sharples, S.; Stedmon, A. & Wilson, J. (2007). An observation tool to study air traffic control and flightdeck collaboration, *Applied Ergonomics* 38, 425-435
- [6] Dagaëff, Th.; Chantemargue, F. & Hirsbrunner, B. (1996). Emergence-Based Cooperation in a Multi-Agent System, *Computer Science Department*, University of Fribourg, Switzerland
- [7] De Terssac, G. & Chabaud, D. (1990). Référentiel opératif commun et fiabilité In J.Leplat & G.de Terssac (Eds.), *Les facteurs humains de la fiabilité dans les systèmes complexes* (pp. 110-139), Toulouse
- [8] Eurocontrol. (2003). *Airport CDM Applications Guide*, Eurocontrol, Belgium
- [9] Ferber, J. (1995). *Multi-Agent Systems*, Addison-Wesley, Muenchen, Germany
- [10] Hamec, S.; Anceux, F.; Pelayo, S.; Beuscart-Zéghir, M.-C. & Rogalski, J. (2002). Cooperation in Health Care – Theoretical and methodological issues. A Study of two situations: hospital and home care, *Le Travail Humain*, 65, 59-88
- [11] Hoc, J.M. (1988). *Cognitive psychology of planning*, Academic Press, London, UK
- [12] Hoc, J.M. (2001). Towards a cognitive approach to human-machine cooperation in dynamic situations, *Human-Computer Studies* 54, 509-540
- [13] Hoc, J.M. & Lemoine, M.P. (1998). Cognitive evaluation of human-human and human-machine cooperation modes in air traffic control. *International Journal of Aviation Psychology* 8.1-32
- [14] Hoc, J.M. (2000). From human-human interaction to human-machine cooperation. *Ergonomics*, 43, 833-843
- [15] Jennings, N.R.; Norman, T.J. & Panzarasa, P. (2001). Formalizing Collaborative Decision Making and practical reasoning in Multi-Agent Systems, *Journal of Logics & Communication* 12 (1) 55-117
- [16] Keisler, S. & Cummings, J.N. (2002). What Do We Know About Proximity in Work Groups?, In P. Hinds and S.Keisler (Eds). *Distributed Work*, Cambridge, MA, MIT Press
- [17] Lee, P.U. (2005). Understanding Human-Human Collaboration to Guide Human-Computer Interaction Design in Air Traffic Control, *NASA Ames Research Center*, CA, USA
- [18] Millot, P. & Debernard, S. (1993). Men-machines cooperative organizations: methodological and practical attempts in air traffic control, *International Conference on Systems, Man, and Cybernetics*, Le Tourquet, France
- [19] Parasuraman, R., Mouloua, M. & Molloy, R. (1996). Effects of adaptive task allocation on monitoring of automated systems, *Human Factors*, 38, 665-679
- [20] Piaget, J. (1965). *Études sociologiques*, Genève, Switzerland
- [21] Raiffa, H.; Richardson, J. & Metcalfe, D. (2002). *Negotiation Analysis*, The Belknap Press of Harvard University Press, Cambridge, England
- [22] Rogers, Y. (2005). Distributed Cognition and Communication, *Encyclopedia of Language and Linguistics (2nd ed.)*, Elsevier, GB
- [23] Salas, E.; Prince, C.; Baker, D. & Shrestha, L. (1995). Situation awareness in team performance: implications for measurement and training, *Human Factors*, 37, 123-136
- [24] Schmidt, K. (1994). Cooperative work and its articulation: requirements for computer support. *Le Travail Humain*, 57, 345-366
- [25] Terveen, L.G. (1995). An overview of human-computer collaboration, *Knowledge-Based Systems*, 8 (2-3), 67-81
- [26] Warner, N.W.; Vanderwalker, S. & Verma, N. (2002). State of the Art Review on Human-Human Collaboration Research: An Integrated, Multidisciplinary Perspective, *Naval Research Centre*, Arlington VA
- [27] Warner, N.W. & Wroblewski, E. (2003). The Cognitive Process Used in Team Collaboration During Asynchronous, Distributed Decision Making, *Command and Control Research and Technology Symposium*, June 15-17, 2004, San Diego CA
- [28] Groppe, M. & Bui, M. (2008). "Study of Cockpit's Perspective on Human-Human Interactions to Guide Collaborative Decision Making Design in Air Traffic Management," *achi*, pp. 107-113, IEEE Proceedings of First International Conference on Advances in Computer-Human Interaction

Responding to Uncertainty on Approach in Hazardous Situations

Ronish T. Joyekurun

School of Computing
Science,
Middlesex University,
Hendon, London.
United Kingdom.

Ronish.Joyekurun.Ext@eurocontrol.int

Innovative Research
EUROCONTROL Experimental
Centre,
Bretigny Sur Orge,
France.

William Wong Paola Amaldi

School of Computing Science
Middlesex University,
Hendon Campus,
London NW4 4BT.
United Kingdom.

{w.wong; p.amaldi-trillo}@mdx.ac.uk

Abstract— The management of uncertainty is a recurring theme in Air Traffic Management and in understanding the way operators accomplish their objectives in a complex, dynamic environment. The current study reports on the verbal communication processes of crews and controllers during the approach flight phase and faced by uncertain situations. A conversation analysis of six (6) accident transcripts was conducted, with dynamic environmental interactions as a complexity factor. The results are presented in the forms of correlations among factor pairs. Results indicate a large variation (5.46%-32.09%) of the detection of uncertainty across accidents. Air Traffic Control and Ground Services (ATC/Ground) rarely initiated problem-solving exchanges (7%) in uncertain situations, as compared to crews (93%). Crews initiated 80.6% of problem-solving exchanges based on the direct perception of environmental cues while ATC/Ground initiated 19.4% of exchanges based only on indirect cues. Finally, our results indicate that ATC/Ground account for 68.8% of overlapping and 88.9% of compounded verbal exchanges. We conclude that the response to uncertain situations arising from hazardous conditions seems to correlate with a management by crews on approach. The effective transfer of uncertainty cues between crews and controllers does not appear to correlate with collaborative and communicative practices.

Keywords- *Uncertainty; Adaptation; Environmental Hazards; Verbal; Conversation Analysis; Air Traffic Management.*

I. INTRODUCTION

The ability to detect and adapt effectively to uncertainty in naturalistic situations is a crucial requirement for preserving the control of a system faced by dynamically changing environmental factors. In a divided labour setting, the detection and reaction to uncertain situations by an operator, often involves the heedful interrelating of one's own work to that of others [1]. This interaction between system entities occurs differently, depending on the work setting where they are placed – thus, operators working within a proximal physical space can use the full variety of verbal and non-verbal communicative abilities. Conversely, operators located far from each other can communicate verbally through technology, although with perceptually impoverished cues.

Verbal communication between pilots and controllers during the approach flight phase is one of the main perceptual means of interaction. The verbal exchanges consist of task-related information accompanied by coordinative cues to allow both teams to work synchronously [2]. In uncertain situations, the effectiveness of verbal interactions can be negatively affected due to the quality of the radio-link, the informational efficiency of the message and also by the informative effectiveness of the message. While the quality and efficiency of exchanges have been addressed by much research (cf. [3-10]) the role of effective verbal interactions on approach needs to be clarified.

For verbal information to be effective, it needs to relate succinctly with its context of use [11]. Thus, an effective message does not supply more or less information that is required and subsequently, the message sender needs to understand the context within which the information will be integrated. The current study investigates the collaborative effectiveness of crews and controllers by analysing the messages they address to each other during uncertain situations. Due to the wide variety of factors which can lead to uncertain situations, we chose to focus on some specific interactions with the environment – the relationship of weather to other approach factors such as runway choices, checklist executions and plan changes are of interest.

II. HAZARDOUS SITUATIONS IN ATM

The qualification of a situation as being hazardous is considered from two perspectives, namely in hindsight and in foresight. Accidents and incidents are an undesirable state of the ATM system and indicate with a high degree of certainty that a hazardous situation might have been present. In retrospect, the contributory causes of occurrences (accident and/or incident) can be determined by measuring the performance variability of different parts of the system with respect to accepted safety standards. However, the analysis of occurrences tends to provide a picture of performance variations which can readily be deemed hazardous even though the same performances also contribute to effective ATM operations, on a routine basis [12].

In foresight, the determination of a situation as being hazardous cannot be objectively defined using a fixed set of criteria. Risk analysis addresses this problem by mapping known system interactions into complex linear models. However, linear models such as event and fault-tree analysis cannot represent factorial interactions which do not exist a-priori [12]. Therefore, more powerful models which allow dynamic relationships to be represented have been proposed such as the Functional Resonance Accident Model (FRAM) [13]. According to the FRAM safety hazards may arise when separate functions which are designed to tolerate a range of performances, resonate with each other. This interference might be gradual (such as the slow erosion of approach procedures), or abrupt (for instance, a dynamic wind direction change late on approach) but can potentially lead to incidents and accidents.

III. MANAGING UNCERTAINTY ON APPROACH

Two main ways by which organisations which face significant safety-risks handle uncertainty, are described [14]. The first is to try and minimise uncertainty at work and the second tries to teach workers to respond effectively to uncertain situations when they arise. These approaches have very different implications.

A minimisation of uncertainty approach mainly deals with the adequate positioning of safety barriers within the system. Highly procedural work settings such as ATC and piloting consist of an enormous amount of rules, regulations and physical constraints to dampen the variety of the complex environment such that control is preserved over flight operations. Pilots and controllers are formally trained to comply with constraints, primarily as a means of preserving the predictability of the system, and hence its safety. In the event of unexpected situations, pilots are instructed to execute the appropriate remedial procedures, as and if provided in their operations manuals. It is to be noted that such procedures mainly relate to mechanical failures and not psychological issues occurring in the dynamics of the group. It should also be noted that the execution of a procedure entails first of all, an adequate diagnosis which is carried out by the crew members. Hence, given the wrong diagnosis, not only will the remedial measures prove to be inadequate, but the ensuing effects of those measures will be imputable to the captain.

Despite the large amounts of constraints presented by the minimisation of uncertainty approach, expert controller and pilot teams often deviate from procedures as a means of achieving expected performances [15]. The need to remain flexible to an uncertain environment forms the core of the second approach: adaptation to uncertainty. This approach levies some of the system control from the blunt end of organisation and relies on the local competences of operators at the sharp end, to cope with uncertainty. This approach is core to the socio-technical design principle of handling variances at their source (cf. [16]). Figure 1 shows some basic principles for managing uncertainty in organisations. Instead of a choice between the two approaches for managing uncertainty, a compromise or balance forms the target. The feed-forward control approach and the feedback control approach each have their own advantages and disadvantages which need to be

addressed by organisations that wish to remain adaptive at the sharp end although retaining enough control at the blunt end [17].

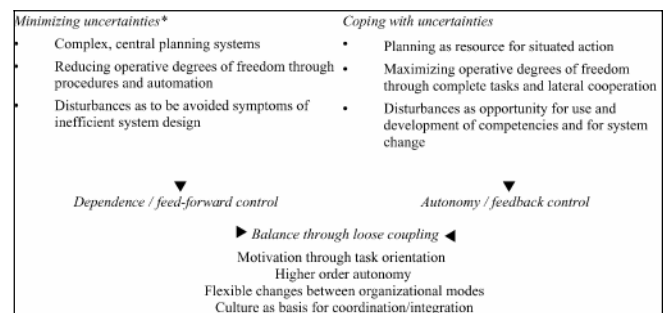


Figure 1. Basic Principles of Uncertainty Management Arising from Organisational Design [17].

IV. ANALYSIS OF VERBAL INTERACTIONS

During the approach flight phase in ATM, the verbal means of communication between crews and controllers is primarily used for co-ordinating actions. The informational structure of verbal exchanges needs to abide by standard aviation phraseology, although exceptions are tolerated [18]. The radio communication protocol is sequential in nature between any two operators, because radio communications are half-duplex – simultaneous voices can still be heard on the radio channel to produce the commonly called ‘party-line’ effect. Verbal exchanges can also be seen as sequential according to turn-taking models of speech – such models regard simultaneous exchanges as a communication breakdown which informs us on the situation at a moment in time [19].

Conversation Analysis (CA) [20-24] builds upon the rich work in Speech Acts [25] to provide a theoretical framework for analysing verbal interactions. CA considers a number of properties of verbal exchanges to hold tacit information about a scenario [26]. The temporal delays which underlie communication handovers among humans, the stutters which are believed to underlie an uncertain mental state and the frequent verbal repetitions which might denote a sense of urgency are some of the verbal properties analysed. The methodology claims to provide detailed analytical accounts, while preserving the context of the verbal data [24, 27].

Neville describes verbal mechanisms which are frequently recurrent across incident/accident investigation reports as providing a useful starting point for analysing the activity patterns of air-ground teams. Neville and Dekker both propose a detailed CA of Cockpit Voice Recorder (CVR) data as a means of understanding the activities of crews and controllers [27, 28]. The limitation of this approach is seen to arise from the provision of original CVR data such that rich sound cues can be transcribed. Indeed, most CVR transcripts exclude subtle cues such as the tonal variations and word accents in sentences which conversation analysts treat as semantically significant.

V. METHODOLOGY

A Conversation Analysis of CVR and investigative data from six (6) aviation accidents was performed. Two aviation

incident and accident databases were chosen as a means of providing our data – the similarity of reporting formats justified this choice. The databases are the National Transportation Safety Board (NTSB), in the United States and the Bureau d'Enquêtes et d'Analyses (BEA), in France. Table I shows the selected sample of accidents.

TABLE I. SELECTED SAMPLED OF ACCIDENT REPORTS

Case Name	Location	Description
5N-MAS	Istres, FR	Loss of engines during flight
F-GRJS	Guipavas, FR	Impact with ground obstacles after landing
N215AA	Arkansas, US	Runway overrun during landing
N497FE	Florida, US	Collision with trees on final approach
N963AS	California, US	Loss of control and impact with Pacific Ocean
N999UA	Colorado, US	Uncontrolled descent and collision with terrain

Although CA explicitly renders the cues which are obfuscated in verbal data, its ability to generate thematic patterns is limited. In this study, conventional qualitative analysis techniques were used to regroup semantically relevant information. Table II shows the qualitative codes used in this study and the categories under which they were regrouped. The references obtained for each code are then counted and a correlation table drawn to build the relationships among coincidental factors. For instance, references to the 'Interruption' category and the 'Information Repair' category might correlate to reveal that further causal investigation of the factor pairs is required.

TABLE II. CODING CATEGORIES, CODES AND DESCRIPTIONS

Code Category	Code	Description
Familiarity	Colloquialism, Exclamation, Expletive.	The level of familiarity expressed by crews and controller.
Cue Perception	Direct Perception Indirect Perception	Environmental Perception. Direct perception is visual. Indirect perception is relayed verbally through ATIS, ATC or cockpit member.
Information Repair	Compounded Exchange Repeated Exchange	Compounded are multiple items of information strung into one exchange, before turn is over. Repeated are requests which have been announced before.
Interruption	Overridden Subject Subverted Subject Overlapping Turn	Overridden is when the subject of an exchange is completely changed. Subverted is when the original turn subject is lost at the end of the conversation. Overlap is when there is less than a second of delay between messages.
Non-Verbal Feedback	Effort Laughter Mechanical	Efforts are cued by heavy breathing noises on the microphone. Laughter is cued by 'haha' variants and 'chuckles' on the microphone. Mechanical feedback are alarms, alerts, clicks, beeps, and any other sound referred as emanating from machines.
Problem	Pilot Flying (PF)	Person initiating a problem-solving

Initiator	Pilot Non Flying (PNF) ATC/Ground	exchange. Machine initiations are situations where alarms initiate a reaction of the persons' exchanges.
Uncertainty	Level of Uncertainty	Level of uncertainty are cued by hesitation marks such as 'err', 'hum' and explicit remarks such as 'I don't know', 'I'm not sure', 'Maybe', etc.

VI. RESULTS

The 6 qualitatively analysed sources generated a total of 1407 coding references. A correlation table was drawn using the initial codes from Table 2 to determine the relationship between different factors. Factor combinations are presented below, with examples of how and when they occurred.

A. Problem Initiator

The system entity initiating a turn can take the form of either a human or a machine – in certain situations, an alarm initiates a problem-solving exchange among the Pilot Flying (PF), Pilot Non Flying (PNF) and controller. In the sources analysed, the PF was seen to initiate the most problem-solving exchanges (38.4%). PNF initiated about 26.1% while ATC and ground operations about 35.5%. Machine initiated exchanges were the least at about 3%. Table III shows the results in summed form.

TABLE III. PARTY INITIATING A PROBLEM-SOLVING EXCHANGE

Party Initiating an Exchange	Sum
Pilot Flying (PF)	78
Pilot Non-Flying (PNF)	53
Air Traffic Control (ATC) / Ground Services	66
Machine	6

The larger number of exchanges relating to problem-solving in the cockpit (67.5%) is an important relationship although the nature of CVR presents more data concerning cockpit conversations than ground exchanges. Further interactions are analysed below to understand problem-solving exchanges and uncertainty management.

B. Uncertainty and Problem Initiator

Verbal exchanges denoting a sense of uncertainty were present across all sources for a total of 258 references. The percentage of uncertainty references was not homogeneous across sources, as shown in Table IV and denotes some fundamental differences in the development and perception of hazardous situations.

TABLE IV. UNCERTAINTY REFERENCES IN SOURCES

Source	Sum	%
N963AS	71	32.09
F-GRJS	23	5.46
N215AA	52	26.86
N999UA	10	6.56
N497FE	28	24.52
5N-MAS	74	10.80

The lowest modes were for F-GRJS (Impact with ground obstacles), at only 5.46% and N999UA (Uncontrolled descent),

at 6.56% of problem initiations with uncertainty. Both of these accidents are characterised by a relatively quiet and non-eventful series of verbal exchanges, until the last minute before the impact.

TABLE V. UNCERTAINTY AND PROBLEM INITIATION

Party Initiating an Exchange	Sum
Pilot Flying (PF)	14
Pilot Non-Flying (PNF)	26
Air Traffic Control (ATC) / Ground Services	3

To understand the role played by operators during uncertain situations, we cross-tabulated the uncertainty information from the sources with the person initiating a problem-exchange. Table V shows the sum of references when an operator was initiating a problem solving exchange and a situation was deemed uncertain. In uncertain situations, the PNF was seen to initiate the majority of problem-solving exchange (60.5%); the PF initiated 32.6% of exchanges. Finally, the controller initiated only about 7% of exchanges during uncertain situations – in many references, these problem-solving attempts by the controller occurred after the crew had displayed a level of uncertainty, verbally.

C. Familiarity and Problem Initiator

TABLE VI. FAMILIARITY AND PROBLEM INITIATOR

Party Initiating an Exchange	Sum
Pilot Flying (PF)	11
Pilot Non-Flying (PNF)	11
Air Traffic Control (ATC) or Ground Services	1

The inter-personal relationship among team members plays a significant role in determining information flows namely during hazardous situations. We observed an equal number of problem-solving exchanges (47.8%) for both the PF and PNF, while addressing each other using familiar terms. Air to Ground displays of familiarity only occurred once throughout all the accidents analysed.

D. Problem Initiator and Cue Perception

TABLE VII. PROBLEM INITIATOR AND ENVIRONMENT CUE

Party Initiating an Exchange	Direct Cue	Indirect Cue
Pilot Flying (PF)	6	6
Pilot Non-Flying (PNF)	11	2
Air Traffic Control (ATC) / Ground Services	0	6

The perception of visual cues is often a verbally described event, namely when the crew is engaged in problem-solving activities. We observed that the PF and PNF combined, perceived about 80.6% of all cues which relate to problem-solving exchanges. Comparatively, ATC and ground services perceived about 19.4% of cues relating to problem-solving – all cues found for controllers were indirect, given their location within approach centres, and in indirect contact with the external environment.

E. Compounded Exchange and Problem Initiator

TABLE VIII. COMPOUNDED EXCHANGE AND PROBLEM INITIATOR

Party Initiating an Exchange	Sum
Pilot Flying (PF)	0
Pilot Non-Flying (PNF)	1
Air Traffic Control (ATC) / Ground Services	8

Compounded exchanges occur when a system entity provides multiple items of information within the same verbal turn. This is thought to lead to longer verbal exchanges. In the sources analysed, we observed that PF and PNF engaged in the least amount of compounded exchanges (11.1%), as compared to ATC and ground services (88.9%), during problem-solving exchanges.

Effectively compounded exchanges can minimise the cost of VHF-occupancy. However, ineffectively compounded information could lead the receiving party to initiate a turn-repair by requesting for the information to be repeated.

F. Overlap and Problem Initiator

TABLE IX. OVERLAP AND PROBLEM INITIATOR

Party Initiating an Exchange	Sum
Pilot Flying (PF)	2
Pilot Non-Flying (PNF)	3
Air Traffic Control (ATC) / Ground Services	11

Overlapping exchanges occur when a verbal turn is taken before the previous one is finished. In the sources analysed, turns were recorded mainly at a 'second' resolution such that overlap could only be inferred when they occurred within a second. However, some instances of human conversation can be less than a second such that two turns are taken at the same recorded time, but are actually sequential. We have taken this limitation into consideration and observed correlations with the entity initiating a problem-solving turn and the number of overlaps. PF and PNF have less overlapping turns (31.25%) than ATC and ground services (68.75%). This is thought to occur mainly because ATC requests are initiated, regardless of the ongoing conversation among the crew, due to the absence of perceptual cues between air and ground.

VII. CONCLUSION

This study showed a number of interactions among human and environmental factors on approach and in hazardous situations. We conclude that a number of those relationships are worth being investigated as potential, causally related factors. The uncertainty displayed verbally within the cockpit varies greatly (5.46%-32.09%) and correlates with a similar variation of cues indicating an uncertain event – however, all those situations are known to have ended in fatal accidents. The relative low frequency of ATC/Ground initiated problem-solving exchanges (7%) as compared to crews (93%) needs clarification since our correlation favours cockpit-occurring conversations. Crews were also seen to initiate 80.6% of problem-solving exchanges based on direct environmental cues

while ATC/Ground initiated 19.4% based only on indirect cues. Finally, ATC/Ground initiated exchanges correlated with 68.8% of overlapping exchanges and 88.9% of compounded verbal exchange.

Finally, the limitations of this study are seen to arise mainly from two factors: the availability of original CVR data and the cockpit-centred perspective of verbal exchanges between air and ground teams. A further study of collaborative practices centred at approach centres and analysing controller problem-solving during uncertain situations is currently under way.

ACKNOWLEDGEMENTS

This study is part of an ongoing doctoral research programme sponsored under financial line (A-B01-C7-120000-61615-INS-0-20-MIPH) at the EUROCONTROL Experimental Centre, Brétigny sur Orge, France. Mentoring and progress tracking are diligently provided by Mr Alistair Jackson and Mr Marc Bourgois.

REFERENCES

- [1] G. Grote and D. E. Zala-Mezö, "The Effects of Different Forms of Coordination in Coping with Workload: Cockpit Versus Operating Theatre," Swiss Federal Institute of Technology Zurich, Zurich 2004.
- [2] R. C. Williges, W. A. Johnston, and G. E. Briggs, "Role of Verbal Communication in Teamwork," *Journal of Applied Psychology*, vol. 50, pp. 473-478, 1966.
- [3] J. Burki-Cohen, "An Analysis of tower (Ground) Controller-Pilot Voice Communications," Office of Research and Development, John A. Volpe National Transportation Systems Center, Washington, DC 1995.
- [4] Eurocontrol, "Air-Ground Communication Safety: Causes and Recommendations," Eurocontrol 2006.
- [5] H. Hering, "Technical Analysis of ATC Controller to Pilot Voice Communication with Regard to Automatic Speech Recognition Systems," Eurocontrol Experimental Centre, Bretigny Sur Orge, France 2000.
- [6] D. Morrow, A. Lee, and M. Rodvold, "Analysis of Problems in Routine Controller-Pilot Communication," *The International Journal of Aviation Psychology*, vol. 3, pp. 285-302, 1993.
- [7] V. Prinzo, "An Analysis of Voice Communications in a Simulated Approach Control Environment," Federal Aviation Administration: Office of Aviation Medicine, Washington 1998.
- [8] J. C. Sperandio, "Analyse des Communications Air-sol en Contrôle D'approche (Orly)," IRIA-CENA 1969.
- [9] G. Van-Es, "Air-Ground Communication Safety Study: An Analysis of Pilot-Controller Occurrences," Eurocontrol Headquarters, Brussels 2004.
- [10] C. E. Billings and E. S. Cheaney, "Information Transfer Problems in the Aviation System," National Aeronautics and Space Administration 1981.
- [11] J. Svensson and J. Andersson, "Speech Acts, Communication Problems and Fighter Pilot Team Performance," *Ergonomics*, vol. 49, pp. 1226-1237, 2006.
- [12] E. Hollnagel, D. Woods, and N. Leveson, "Resilience Engineering: Concepts and Precepts," England: Ashgate, 2006.
- [13] E. Hollnagel, "Capturing an Uncertain Future: The Functional Resonance Accident Model," 2006.
- [14] G. Grote, W. Naef, O. Straeter, and R. Helmreich, "Golden Rules of Group Interaction in High Risk Environments: Evidence based suggestions for improving performance," Gottlieb Daimler and Karl Benz Foundation, Swiss Re Centre for Global Dialogue, Ladenburg and Rüschiikon 2004.
- [15] K. Dismukes, B. Berman, and L. Loukopoulos, "Converging Themes: The Deep Structure of Accidents," in *The Limits of Expertise: Rethinking Pilot Error and the Causes of Airline Accidents*, K. Dismukes, D. Maurino, and S. Dekker, Eds.: Ashgate, 2007, pp. 247-273.
- [16] F. E. Emery, *Characteristics of Socio-Technical Systems*, Document No. 527 ed. London: Tavistock, 1959.
- [17] G. Grote, "Uncertainty Management at the Core of System Design," *Annual Reviews in Control*, vol. 28, pp. 267-274, 2004.
- [18] ICAO, "Manual of Radiotelephony," International Civil Aviation Organisation 2007.
- [19] H. Sacks, E. Schegloff, and G. Jefferson, "A Simplest of Systematics for the Organization of Turn-Taking for Conversation," *Language*, vol. 50, pp. 696-735, 1974.
- [20] G. Jefferson, "Harvey Sacks: Lectures in Conversation: Volume I." vol. Volume 1 Oxford, UK & Cambridge, USA: Blackwell, 1992.
- [21] E. Schegloff, "Sequencing in Conversational Openings," *American Anthropologist*, vol. 70, 1968.
- [22] E. Schegloff, "Conversation Analysis and Socially Shared Cognition," in *Perspectives on Socially Shared Cognition*, L. Resnick, J. Levine, and S. Teasley, Eds. Washington, D.C.: American Psychological Association, 1991, pp. 150-171.
- [23] E. Schegloff, "Overlapping Talk and the Organization of Turn Taking for Conversation," *Language in Society*, vol. 29, pp. 1-63, 2000.
- [24] C. Goodwin and J. Heritage, "Conversation Analysis," *Annual Review of Anthropology*, vol. 19, pp. 283-307, 1990.
- [25] J. R. Searle, *Speech Acts: An Essay in the Philosophy of language*. London, 1969.
- [26] M. Nevile and M. B. Walker, "A Context for Error: Using Conversation Analysis to Represent and Analyse Recorded Voice Data," Australian Transport Safety Bureau 2005.
- [27] S. Dekker, *The Field Guide to Understanding Human Error*: Ashgate Publishing Limited, 2006.
- [28] M. Nevile, "Communication in Context: A Conversation Analysis Tool for Examining Recorded Voice Data Investigation of Aviation Occurrences," Australian Transportation Safety Bureau, Canberra, Australia 2005.

AUTOMATION FOR TASK ANALYSIS OF NEXT GENERATION AIR TRAFFIC MANAGEMENT SYSTEMS

Maricel Medina, BSc, Lance Sherry, PhD
Center for Air Transportation System Research
George Mason University
Virginia, USA
mmedinam@gmu.edu, lsberry@gmu.edu

Michael Feary
NASA Ames Research Center
Moffet Field, CA, USA
Michael.feary@nasa.gov

Abstract—The increasing span of control of Air Traffic Control enterprise automation (e.g. Flight Schedule Monitor, Departure Flow Management), along with lean-processes and pay-for-performance business models, has placed increased emphasis on operator training time and error rates. There are two traditional approaches to the design of Human-Computer Interaction (HCI) to minimize training time and reduce error rates: (1) experimental user testing provides the most accurate assessment of training time and error rates, but occurs too late in the development cycle and is cost prohibitive, (2) manual review methods (e.g. cognitive walkthrough) can be used earlier in the development cycle, but suffer from poor accuracy and poor inter-rater reliability. Recent development of “affordable” human performance models provide the basis for the automation of task analysis and HCI design to obtain low cost, accurate, estimates of training time and error rates early in the development cycle.

This paper describes a usability/HCI analysis tool that this intended for use by design engineers in the course of their software engineering duties. The tool computes estimates of trials-to-mastery (i.e. time to competence for training) and the probability of failure-to-complete for each task. The HCI required to complete a task on the automation under development is entered into the web-based tool via a form. Assessments of the salience of visual cues to prompt operator actions for the proposed design are used to compute training time and error rates. The web-based tool enables designers in multiple locations to review and contribute to the design. An example analysis is provided along with a discussion of the limitations of the tool and directions for future research.

Human Computer Interaction, Usability Analysis, Task Analysis, Probability of Failure to Complete a Task, Trials to Mastery.

I. INTRODUCTION

The evolution of the air transportation system to a “mature” industrial sector has resulted in cost differentiation as a primary means of competitive advantage for airlines. This cost imperative has flowed through the supply chain to aircraft manufacturers and Air Traffic Control. The result has been new business models (e.g. low cost carriers, outsourcing) and incentives for the supply chain vendors to reduce installation

costs and operational costs (e.g. training, operational efficiency, and safety). Air Navigation Service Providers (ANSPs) have embraced this challenge by privatization of Air Traffic Control, pay-for-performance, and the development of large-scale enterprise management and control automation such as Flight Schedule Monitor (FSM), Departure Flow Management (DFM), Surface Management Systems (SMS).

Human Computer Interaction has emerged as one of the ways to reduce costs by streamlining training as well as increase the efficiency of operators. For example, Boeing Commercial Aircraft Group funded a large internal R&D project with the specific design goal of reducing training costs and improving flight deck operational efficiency (Mumaw, Boorman, and Prada, 2006, Castor-Peck, personal communication). Several avionics vendors (Faerber, Vogl, and Hartley, 2007; Jacobsen, Chen, and Widemann, 1999), airlines (Fennell, Sherry, and Roberts, 2006), and NASA’s Exploration Mission Directorate, Human Research Program (NASA, 2008) also have similar initiatives in place.

The most accurate evaluation of the usability of a product is achieved through experimental user testing (Nielsen, 1993). This type of approach is cost prohibitive and can only be conducted at the end of the development cycle when the cost of revisions is highest.

This paper describes a tool based on the Human Computer Interaction Process Analysis model (HCIPA) that this intended for use by software and design engineers in the course of their software engineering duties, to conduct usability analyses. HCIPA attempts to solve two very hard problems in the design of advanced automated systems. The first is capturing the details of operator-system interactions while performing a large number of mission tasks, task analysis. The sequence of operator actions and inferred mental operators for each task is then used to solve the second problem, making useful predictions of time to complete a task, repetitions required to master a task, and the likelihood of failure for failure infrequently performed tasks. This paper presents a web based tool that solves the first problem. Cog Tool (John, et al., 2004) is able to make accurate performance predictions for frequently performed tasks. We are in the process of developing related

methods for predicting repetitions and likelihood of failure. Preliminary versions of models for making these predictions are reported in this paper.

Specifically, the tool enables designers and testers to describe the sequence of operator actions, and rapidly assess the trials to mastery (i.e. time to competence for training) and the probability of failure-to-complete for each task that can be performed by the product under design. The computation of these human performance measures is based on the specification of operator actions and an assessment of the salience of visual cues in the proposed automation user-interface to prompt the next operator action. The web-based tool also provides designers in multiple locations to view and contribute to the design and the usability evaluation.

This paper is organized as follows: Section 2 provides an overview of Human Computer Interaction (HCI) and introduces the Human Computer Interaction Process Analysis (HCIPA) method. Section 3 describes the tasks that can be performed by the functions of the tool. Section 4 provides case studies of usability analysis conducted with the tool. Section 5 discusses the limitations of the tool and directions for future research.

II. OVERVIEW OF HUMAN COMPUTER INTERACTION

Human-computer interaction involves the cognitive, motor, and visual activities of an operator using automation to perform a mission task (Card, Moran & Newell, 1983). The interaction between operator and automation follows a human action cycle of goal formulation, execution, and evaluation (e.g. Norman, 1988). The degree to which the content of the user-interface matches the “semantic space” of the operator determines the usability of the automation (Kitajima, Blackmon, and Polson, 2002).

Several techniques have been used to determine the usability of automation (Nielsen, 1992). The most accurate evaluation of the usability of a product is achieved through experimental user testing. Human subjects perform a list of tasks using the automation under test while observers take notes or record the operator’s behavior. The aim is to identify problems on the product or features that users like and are easy to use. Techniques include “think aloud protocols” and eye tracking. Although quantitative data can be collected by measuring time to learn, speed of performance, and rate of human error; this approach is cost prohibitive and can only be conducted at the end of the development cycle when the cost of revisions is highest (Nielsen, 1994).

Alternative approaches that can be used earlier in the life-cycle, fall into two categories: Manual Inspections and Operator Performance Predictions. Manual inspections, such as participatory design (Muller and Kuhn, 1993), cognitive walkthroughs (Wharton, Rieman, Lewis, and Polson, 1994), heuristic evaluations (Nielsen, 1992), and other forms of expert reviews, have been shown to be effective in certain settings (Dumas, 2003) but are subjective and can be biased by group-thinking (Turner & Pratkanis, 1998). These methods also exhibit poor inter-rater reliability (Hetzum & Jacobsen, 2003) due to differences in granularities of the task definition and the differences in the subjective ratings.

Automated tools, such as CogTool (John, Prevas, Salvucci, and Koedinger, 2004), seek to eliminate these two sources of poor inter-rater reliability by capturing actual end-user button pushes (to eliminate ambiguity in the task definition), and by estimating performance using human performance models such as Keystroke-Level Model (KLM), (Luo & John, 2005). These tools can also be used early in the development cycle.

CogTool, one of the first tools of this class, provides an easy way model skilled users’ performance behavior through storyboards designs. To create the storyboards, CogTool users include the different screen shots on the tool and specify “hot-spots” or widgets on the screen shots to simulate the user interaction. The screenshots are connected through transitions. Once the screens are connected, the user interacts on the screenshots through the widgets, and CogTool generates an executable script of the actions performed by the user that can be processed by an Operator Performance Model such as KLM (Luo & John, 2005), ACT-R (Anderson et al., 1995) or CORE (Vera, Howes, McCurdy, & Lewis, 2004) to compute a prediction of expert time-on-task.

A. The HCIPA Method

HCIPA is a manual task/usability analysis inspection method that was designed to address issues with usability in the aviation and space industries (Sherry et al., 2002, 2006). Specifically, these industries were interested in evaluating usability for trials-to-mastery and probability-to-complete the task.

The HCIPA method has its roots in a model of pilot cognition (Polson, Irving, & Irving, 1994). This method also known as the RAFIV model (Sherry et al., 2002) decomposes tasks into six sequential steps: (1) Identify Task, (2) Select Function, (3) Access Function, (4) Enter data for Function, (5) Confirm and Save Data, and (6) Monitor Function. These steps are illustrated in Figure 1.

The first step is to identify a task based on various external stimuli such as visual cues (menu item, error message), hearing cues (warning sounds), and a request (e.g. checklist) or by remembering (e.g. recall from long-term memory). Operator proficiency is reduced when the user interface does not provide any guidance by salient visual cues (Sherry, Fennell, Feary, Polson, 2006).

Once the user knows what to do, the next step is to decide the right function to accomplish the task, which is to select a function. The function may be the name of a screen, the label on a button, a prompt or any other characteristic that tells the user to initiate the task. The more accessible the function is to the user, the higher the probability is to accomplish the task.

A set of operator actions are performed by a user in order to accomplish the task through the selected function. These operator actions are grouped under Access, Enter, Confirm and Save, and Monitor step.

The Access Step encloses the operator actions needed to access the function on the device. The goal for a designer is to reduce the number of operator actions needed to access the function.

The Enter Step encloses the operator actions needed to successfully execute the function. The operator actions may include data entry, visual data evaluation, and communication with external devices or personnel.

The Confirm and Save Step are all the operator actions needed to trigger the function.

Finally, the Monitor Step encloses the operator actions needed to monitor any change on the system state after the function is triggered.

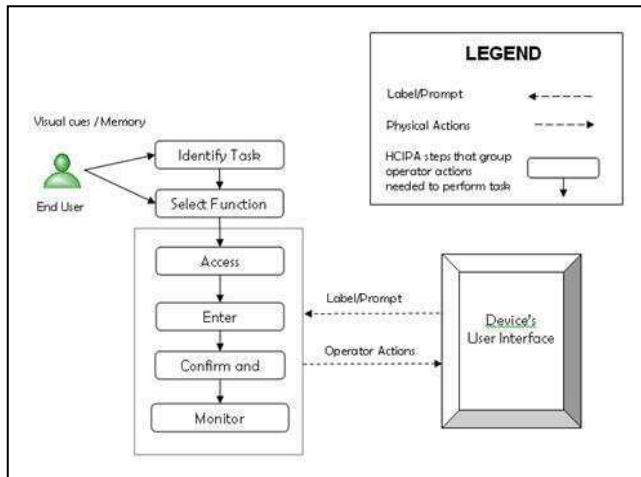


Figure 1. HCIPA Method

There are two basic classes of operator actions: (i) physical actions such as press a button or click on a link, and (ii) decision actions that cannot be viewed externally. A Task is executed by performing operator actions for each of the steps.

HCIPA estimates operator performance based on the minimization of memorized action sequences. When a user interface lacks clear labels, prompts, and/or organizational structure, additional training is required and operators must recall memorized action sequences (Sherry, Polson & Feary, 1998; Fennel, Sherry & Roberts, 2004).

The HCIPA approach has been successfully applied in several applications (Sherry, Polson & Feary, 2002; Sherry, Fennel, Feary, & Polson, 2006). The unguided manual process suffered from several issues: (1) ambiguity of granularity in descriptions of steps, (2) ambiguity in identification of salient visual cues, (3) problems in assessing salience of visual cues, (4) no method to determine trials-to-mastery or probability of failure to complete a task. The tool described in this paper is designed to overcome these shortfalls and includes an affordable Operator Performance Model to compute trials-to-mastery and probability-to-complete the task.

III. THE E-HCIPA TOOL

e-HCIPA is a web based application developed to provide an automated way to apply the HCIPA method. The e-HCIPA is a free accessible web application; therefore, no username or password is required to use the tool. The current version of e-

HCIPA runs only on Mozilla Firefox web browser and provides the following functionalities: Create a Task Analysis, Predict Operator Performance, Edit a Task Analysis, Delete a Task Analysis, and Generate PDF report (Task Analysis Report and User Guideline).

A. e-HCIPA Features

1) *Create Task Analysis*: Allows the user to create a new task analysis by inserting the device name, task name and function name. Once the device, name and function are saved, the labels for all steps are generated and the user can insert the operator actions for each step. Figure 2 shows the main screen of the tool

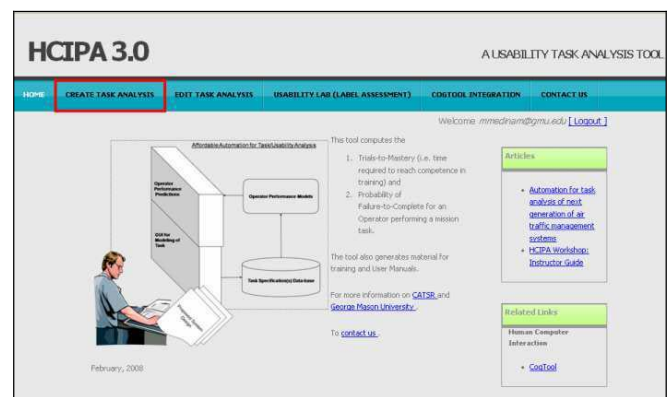


Figure 2. e-HCIPA Create Task Option

The operator actions may involve physical actions (press button, link), visual actions (read data from display field), audio actions (hear warning buzzer) or decision-making actions. Operator actions are automatically generated for the Identify Task and Select Function step based on the information entered on the task name and function name. The operator action for the Identify Task step is always generated as “*Recognize need to:*” concatenated with the task name entered by the analyst. The operator action generated for the Select function step is generated as “*Decide to use function:*” concatenated with the function name entered by the analyst. These two operator actions cannot be deleted by the user. The labels for the steps are created as follow:

- Identify Task Step: *<task name>*
- Select Function: *<function name>*
- Access Step: Access + *<function name>* + function
- Enter Step: Enter data for + *<function name>* + Function
- Confirm & Save Step: Confirm & Save data using + *<function name>* + Function
- Monitor Step: Monitor results of + *<function name>* + Function

The analyst can continue inserting operator actions for the Access, Enter, Confirm and Save and Monitor steps. Figure 3

shows the screen where the operator actions are inserted and the salience assessment takes place.

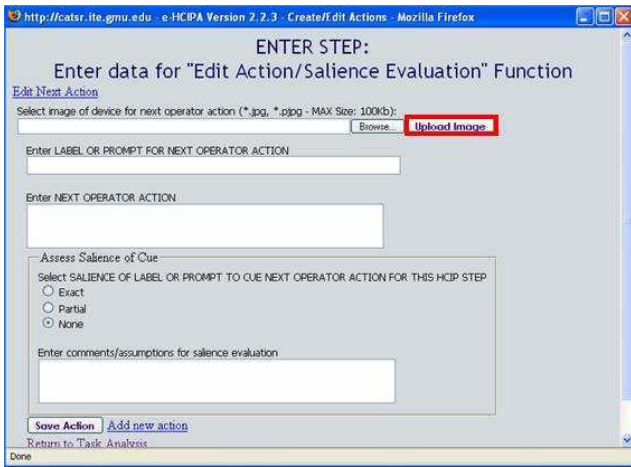


Figure 3. e-HCIPA Enter operator action

2) *Predict Operator Performance:* e-HCIPA calculates the two metrics based on the salience assessment conducted while inserting or editing an operator action: probability to fail a task, and trials to mastery the task. The probability of failure is calculated using (1), while the trials to mastery the task is obtained from (2). In (1), the maximum value used is 1. The values for the operator actions are calculated from the salience assesment using the following values: 0 for Exact, ¼ for Partial and 0 for None. Existing data support the prediction of trials to mastery a task (Bovair, Kieras, Polson, 1990).

$$\text{Probability to Failure} = 0.1753 * \sum \text{Operator actions} \quad (1)$$

$$\text{Trials Mastery Task} = 0.5916 * \sum \text{Operator actions} + 1.9632 \quad (2)$$

3) *Edit a Task Analysis:* e-HCIPA allows to modify any task analysis previously created. The device, task and function name can be changed at any time. If this is done, all the labels for the different steps will change also. The operator actions, including image, operator action description, label and salience assessment can be edited at any time. In order to edit a task analysis, the user must select the desired one from the list of task currently existing in the database.

4) *Delete a Task Analysis:* A task analysis can only be deleted by the person who created the task.

5) *Duplicate a Task Analysis:* A task analysis can also be duplicated. In this case, the system creates a new task with same content and images but it adds the (*Duplicate*) at the end of the Task Description. The person who duplicates the task becomes the creator of the new tasks.

6) *Generate a PDF report:* e-HCIPA allows to generate two .pdf reports. The Task Analysis Report contains all the

operator actions grouped by step, including the trials to mastery and probability to complete the task, a thumbnail image, the label, the salience evaluation, and the salience comments. The User Guideline report contains all the operator actions inserted for the task and ordered sequentially. The User Guideline report can be used for training purposes.

B. e-HCIPA Technical Implementation

e-HCIPA has been developed using PHP 4.4.4 and MySQL database. Figure 4 shows the Entity-Relationship Diagram of e-HCIPA.

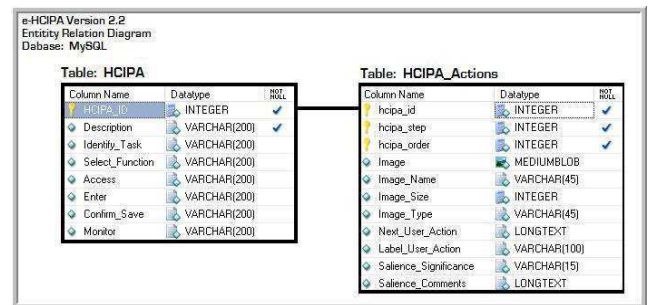


Figure 4. e-HCIPA Entity Relationship Diagram

The database table HCIPA stores the information for the device name, task description and function on fields Description, Identify_Task and Select_Function respectively. Once the user saves a new Task Analysis, e-HCIPA populates the rest of the fields on table HCIPA based on the information stored on the fields Identify_Task and Select_Function. Furthermore, two default operator actions are created: one for the Identify_Task step and the other one for Select_Function.

Table HCIPA_Actions stores all operator actions for the given task. The field hcipa_step is an enumerated field that keeps track of current step for the operator action. The values are: 1 for Identify_Task, 2 for Select_Function, 3 for Access, 4 for Enter, 5 for Confirm and Save and 6 for Monitor.

IV. CASE STUDY

An example HCIP analysis is illustrated below for and Air Traffic Management (ATM) System. The specific task is to run a ground delay program (GDP) at Chicago O'Hare Airport (ORD). Table I shows the input data used through HCIPA to analysis this task.

TABLE I. INPUT DATA FOR A HCIP ANALYSIS ON FMS737

Define Device, Task, and Function	
Device Name	Traffic Management System
Task Name	Run a [Ground Delay] Program at ORD" with General Parameters (Start/End Time/Duration, Arrival Fix, Aircraft Types, Carriers).
Function Name	GDT Setup: General.

TABLE II. HCIP ANALYSIS FOR TASK "MODIFY DEPARTURE RUNWAY"

Step	Operator Action	Label	Saliency Evaluation of Label to Cue Operator Action
Identify Task Run a [Ground Delay] Program at ORD" with General Parameters (Start/End Time/Duration, Arrival Fix, Aircraft Types, Carriers)	<i>Recognize need to: "Run a [Ground Delay] Program at ORD" with General Parameters (Start/End Time/Duration, Arrival Fix, Aircraft Types, Carriers)</i>	Bar Graph ORD Status	None
Select Function GDT Setup: General.	<i>Decide to use function: GDT Setup: General</i>	Tab labeled GDT Setup	Partial GDT stands for "Ground Delay Tool."
Access GDT Setup: General Function	Click on Tab labeled "GDT Setup"	Tab labeled GDT Setup	Exact Assume operator has domain knowledge to interpret menu items
Enter data for GDT Setup: General Function	Select "RBS++" on Program Type Pull-down Menu	Pull-down Menu: Program Type	Exact
	Set Start/End Time (type, move sliders, type duration)	General: Program Time: Start, End (or Duration)	Exact
	Select menu item "All" in pull-down menu "Arrival Fix:"	Pull-down Menu labeled	Exact
	Select menu item labeled "ALL" on pull-down menu labeled "Aircraft Types:"	Pull-down menu labeled "Aircraft Types:" menu item labeled "ALL"	None
	Type "ALL" into text field labeled "Carrier"	Text Field labeled	None
Confirm & Save data using GDT Setup: General Function	<i>None</i>		
Monitor results of GDT Setup: General Function	<i>None</i>		

An ATM specialist must be trained to read bar graph. There are scenarios when a GDP is not run even when the Hourly bars are in excess of the airport capacity (e.g. fog burn-off at SFO, pop-corn thunderstorms). ATM specialist requires significant training to define the parameters of the GDP. The Flight Schedule Monitor used to analyze this task offers no apriori decision-making support related to the parameters.

Table II shows all the operator action needed to complete this task.

The first column includes the HCIPA steps. The second column lists the operator actions. Note, that the operator actions of the Id Task and Select Functions steps are automatically generated by the tool. The third column lists the visual cue (if any) that prompts the next user action. The fourth column is an assessment of the saliency of the cue.

Based on the salience evaluation and distribution of operator actions per HCIPA step, the estimated trials to mastery the task are 3.29, and the probability to fail the task is 0.39. Figure 5 shows the distribution of operator actions by HCIPA step and salience evaluation.

The operator needs to perform eight actions to complete this task. The most critical part is the first operator action. The current label is not obvious to perform the task through the selected function (salience evaluation is None). However, once the operator accesses the function, the visual cues are sufficient to complete the task.

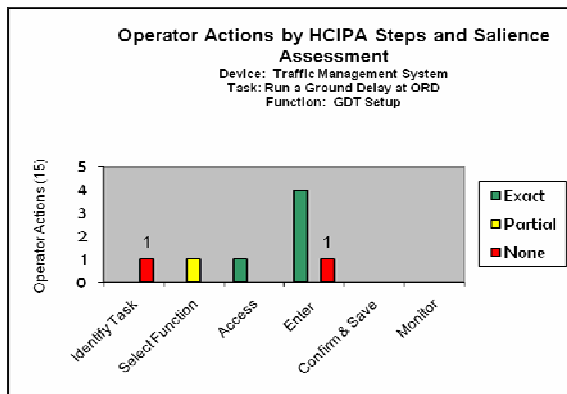


Figure 5. Operator Actions per HCIPA step

The use of HCIPA allows one to identify usability problems on the system for new ATM specialists. It also provides the benefit of generate a user guideline to train new operators on the analyzed task. The Appendix A shows the usability task report generated through HCIPA for this task.

V. FUTURE WORK

This paper describes a tool that this intended for use by software and design engineers in the course of their software engineering duties, to conduct usability analyses. Specifically, the tool enables designers and testers to rapidly assess the trials to mastery (i.e. time to competence for training) and the probability of failure-to-complete for each task that can be performed by the product under design. The computation of these human performance measures is based on the specification of operator actions and an assessment of the salience of visual cues in the proposed automation user-interface to prompt the next operator action. The web-based tool also provides designers in multiple locations to view and contribute to the design and the usability evaluation.

Beta testing of the tool is underway. Future work includes tool implementation, development of new functionalities, improvement of human performance model, and inter-rater reliability of the Assessment of the Salience of the Visual Cues.

- **Tool implementation:** The current version of tool has been tested on Mozilla Firefox and Internet Explorer. It is been developed using PHP 5.2. A security model has been implemented to allow certain functions to be accessed on the creator of the task. In terms of outputs,

the current version only provides two reports on a .pdf format. These two reports will be also available in other format and, as needed; more reports will be developed, including the reports with graphs.

- **New functionality:** (i) hierarchical organization of task that allows to relate other tasks analysis as sub-task, (ii) provide API to enable import/export of models (e.g. with CogTool), (iii) development of a training laboratory by reusing task analysis description and images
- **Operator Performance Model:** The current model is based on empirical data from 4 experiments. Further work is planned to increase empirical data set and leverage existing models such as CORE, ACT-R, etc.
- **Inter-rater Reliability of the Assessment of the Salience of the Visual Cues:** The assessment of the salience of visual cues for prompting the operator's next action is critical for the accuracy of the tool. The current version of the tool relies on the assessment of the salience of the cue by the designer (i.e. None, Partial, Exact). This manual form of assessment suffers several issues. First, the assessment is reliant on the overlap of the designers "semantic state-space" with the end-users "semantic state-space." Recent studies have shown wide variance in semantic state-spaces and large differences between the semantic state-spaces of the designers and end-users. Second, even within a group of end-users and domain experts, the semantic state-space can exhibit a wide distribution. This issue will be investigated in two ways. First it is proposed to add a feature of the tool, loosely named, "Usability Lab." This feature will enable the collation of domain experts' assessment of the salience of the visual cues. Second, several automated techniques exist to automate the salience assessment. Latent Semantic Analysis, LSA (Landauer & Dumais, 1997; Kitajima, Blackmon, and Polson, 2000) and Scent-based Navigation and Information Foraging in the ACT architecture, SNIF-ACT (Pirolli & Fu, 2003) are two of these automated technique that will be researched to evaluate their feasibility to be included in e-HCIPA.

ACKNOWLEDGMENT

Thank you for technical assistance and suggestions from Peter Polson (University of Colorado), Mike Matessa (Alion Inc.), Karl Fennel (United Airlines). Thank you for support of the research to Steve Young (NASA), Amy Pritchett (NASA). This project was funded by grant from NASA – Aeronautics – Intelligent Integrated Flightdeck program, and by internal George Mason University Foundation Funds.

REFERENCES

- [1] The CogTool Project, Tools for Cognitive Performance Modeling for Interactive Devices. <http://www.cs.cmu.edu/~bej/cogtool/>.
- [2] Anderson, J. R., John, B. E., Just, M. A., Carpenter, P. A., Kieras, D. E., & Meyer, D. E. (1995). Production system models of complex cognition.

- In Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society (pp. 9-12). Hillsdale, NJ: Lawrence Erlbaum Associates
- [3] Boorman, Daniel J. , Mumaw, Randall J. (2007) A New Autoflight/FMS Interface: Guiding Design Principles. Air Canada Pilots Association, Safety Conference.
- [4] Bovair, S., Kieras, D.E., & Polson, P.G. (1990). The acquisition and performance of text editing skill: A cognitive complexity analysis. *Human-Computer Interaction*, 5, 1-48.
- [5] Card, S., Moran, T. & Newell, A. (1983). *The Psychology of Human-Computer Interaction*. Hillsdale, NJ: Erlbaum.
- [6] Castor-Peck, Steve (2007) Technical Memorandum to NASA Aeronautics – Intelligent Flight-deck Research Program. NASA-Ames Research Center, Moffet Field, CA.
- [7] Dumas, J. S. (2003) User-based evaluations. In J. Jacko and A. Sears (Eds.), *The Human-Computer Interaction Handbook*. (pp. 1093-1117) Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- [8] Faerber, R.A., Vogl, T.L. , & Hartley, D.E.. (2000) Advanced Graphical User-interface for Next Generation Flight Management Systems. In Proceedings HCI-Aero 2000. September 27-29, Toulouse, France. Pages 107-112.
- [9] Fennell, Karl; Sherry, Lance; Roberts, & Ralph, Jr. (2004). Accessing FMS Functionality The Impact of Design on Learning NASA Technical Report (IH-051; NASA CR-2004-212837).
- [10] Hertzum, Morten & Jacobsen, Niels Ebbe (2003): The Evaluator Effect: A Chilling Fact About Usability Evaluation Methods. In *International Journal of Human-Computer Interaction*, 15 (1) pp. 183-204
- [11] Jacobsen, A. R., Chen, S. S., & Widemann, J. (1999) Vertical Situation Awareness Display. Boeing Commercial Airplane Group.
- [12] John, B., Prevas, K., Salvucci, D. D., & Koedinger, K. (2004). Predictive human performance modeling made easy. In *Human Factors in Computing Systems: CHI 2004 Conference Proceedings*. New York: ACM Press.
- [13] Kitajima, M., Blackmon, M. H., & Polson, P. (2000) G. A Comprehension-based model of Web navigation and its application to Web usability analysis. In *People and Computers XIV*, Springer, 357--373, 2000.
- [14] Landauer, T. K. & Dumais, S. T. (1997) A solution to Plato's problem: The Latent Semantic Analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104, 211-240., 1997.
- [15] Landauer, T. K., Foltz P. W. & Laham D. (1998). "Introduction to latent semantic analysis." *Discourse Processes*, 25, 259-284.
- [16] Luo, L., & John, B. E. (2005) Predicting task execution time on handheld devices using the keystroke-level model, CHI '05 extended abstracts on Human factors in computing systems, April 02-07, 2005, Portland, OR, USA
- [17] Muller, Michael J., & Kuhn, S. (1993) "Participatory Design" In: *Communications of the ACM special issue on participatory design*. 36(4), June 1993.
- [18] Mumaw, R., Boorman, D. J., & Prada, R. L. (2006). Experimental Evaluation of a New Autoflight Interface. Proceedings HCI-Aero 2006, International Conference on Human Computer Interaction, Seattle, WA., September 2006.
- [19] NASA (2008) Human Research Program. Exploration Missions Directorate. <http://www.nasa.gov/directorates/esmd/acd/human.html>
- [20] Nielsen, J. (1992) "The usability engineering life cycle" In *Computer Volume 25, Issue 3, Page(s):12 – 22, March 1992.*
- [21] Nielsen, J. (1994). "Guerilla HCI: Using discount usability engineering to penetrate the intimidation barrier," In: Randolph G. Bias and Deborah J. Mayhew (editors). *Cost-justifying usability*. Boston: Academic Press, pp. 242-272.
- [22] Nielsen, J. & Landauer, T. K. (1993) A mathematical model of the finding of usability problems, Proceedings of the ACM INTERCHI'93 Conference (Amsterdam, the Netherlands, April 24-29), 206-213.
- [23] Nielsen, J. & R.L. Mack (1994) *Usability Inspection Methods*. John Wiley & Sons, New York, NY, 1994.
- [24] Norman, D.A. (1988) "The Design of Everyday Things." MIT Press
- [25] Pirolli, P. & Fu, W.-T. (2003). SNIF-ACT: A Model of Information Foraging on the World Wide Web. In Proceedings of the Ninth International Conference on User Modeling, 2003
- [26] Polson, P. G., Irving, S., & Irving, J. E. (1994). Final report: Applications of formal methods of human computer interaction to training and the use of the control and display unit. Washington, DC: System Technology Division, ARD 200, Department of Transportation, Federal Aviation Administration.
- [27] Sherry, L., Fennell, K., Feary, M., & Polson, P. (2006) *Human-Computer Interaction Analysis of Flight Management System Messages*. *Journal of Aircraft*, Vol 43, No. 5, September-October 2006.
- [28] Sherry, L., Polson, P. & Feary, M. (2002) Designing User-Interfaces for the Cockpit: Five Common Design Errors, and How to Avoid Them. Paper to be presented at the 2002 SAE World Aviation Congress, Phoenix, AZ (November 5 – 7).
- [29] Sherry, L., Polson, P., Feary, M. & Palmer, E. (2002) When Does the MCDU Interface Work Well? Paper presented at the the International Conference on Human-Computer Interaction in Aeronautics, Cambridge, MA (October 22-24).
- [30] Turner, M. E., & Pratkanis, A. R. (1998). Twenty-five years of groupthink theory and research: Lessons from the evaluation of a theory. *Organizational Behavior & Human Decision Processes*, 73(2-3), 105-115.
- [31] Vera, A., Howes, A., McCurdy, M., & Lewis, R.L.(2004). A constraint satisfaction approach to predicting skilled interactive performance. Proc. of the SIGCHI Conference on Human Factors in Computing Systems.
- [32] Wharton, C., Rieman, J., Lewis, C., & Polson, P. (1994). The cognitive walkthrough method: A practitioner's guide. In J. Nielsen & R. L. Mack (eds.). *Usability inspection methods*. New York, NY: John Wiley.

APPENDIX A

Task Analysis Report
Flight Schedule Monitor - "Run a [Ground Delay] Program at ORD" with General Parameters (Start/End Time/Duration, Arrival Fix, Aircraft Types, Carriers)
Estimated TTM: 3.29
Estimated PFC: 0.39

Device: Flight Schedule Monitor
 Task: "Run a [Ground Delay] Program at ORD" with General Parameters (Start/End Time/Duration, Arrival Fix, Aircraft Types, Carriers)
 Function: GDT Setup: General
 Notes: Note 1: Air Traffic Specialist monitors Arrival Rates at an airport and decides that the predicted arrival rate in excess of the capacity for an extended period (e.g. 2 hours). This circumstance initiates the task to "Run a [Ground Delay] Program"
 Note 2: General parameters is the 1st of 4 sets of parameters that must be entered.

1-IDENTIFY TASK:
 "Run a [Ground Delay] Program at ORD" with General Parameters (Start/End Time/Duration, Arrival Fix, Aircraft Types, Carriers)

The screenshot shows a software interface with a bar graph on the right and a list of flight arrivals on the left. The bar graph has several bars of varying heights and colors (red, green, yellow). The list on the left contains text that is partially obscured but appears to be a list of flight numbers and times.

1.1 Recognize need to: "Run a [Ground Delay] Program at ORD" with General Parameters (Start/End Time/Duration, Arrival Fix, Aircraft Types, Carriers)
 [Label]: Bar Graph ORD Status
 [Salience of label/prompt to cue next operator action for this HCIP step]: None
 [Comments/Assumptions]: Note 1: ATM Specialist must be trained to read bar graph. There are scenarios when a GDP is not run even when the Hourly bars are in excess of the airport capacity (e.g. fog burn-off at SFO, pop-corn thunderstorms).
 Note 2: ATM Specialist requires significant training to define the parameters of the GDP. The tool offers no a priori decision-making support related to the parameters (only post entry with the Power Run)

Figure 6. Report generated on e-HCIPA for Task "Run a [Ground Delay] Program at ORD", Flight Schedule Monitor

Human-Centred Innovation: Developing 3D-in-2D Displays for ATC

B.L. William Wong¹, Simone Rozzi¹, Stephen Gaukrodger¹, Alessandro Boccalatte², Paola Amaldi¹, Bob Fields¹, Martin Loomes¹, and Peter Martin³.

¹Interaction Design Centre,
School of Computing Science
Middlesex University
London, UK

²Space Applications Services N.V./S.A.
Zaventem, Belgium

³EUROCONTROL Experimental Centre
Bretigny Sur Orge, France

Abstract—User-Centred Design is extremely useful for improving existing tasks and technologies, however it is less useful for innovation. This paper documents User-Centred Innovation and how it was implemented in the Air Traffic Control domain.

Keywords—User-Centred Innovation, Air Traffic Control, Augmented Reality

I. INTRODUCTION

Human operators will continue to be a major component of air traffic control (ATC) for the foreseeable future. As such, the efficiency and safety of any future ATC system will depend on both the technology developed and the ability of humans to harness that technology. The 3D-in-2D Displays for ATC project is intended to create visualisation tools for air traffic controllers as they face new challenges presented by future ATM programs such as SESAR (Single European Sky ATM Research). The time frame for the realization of these new technologies is 20 to 25 years in the future.

It is difficult to predict work and operational requirements so far into the future with a high degree of certainty. At the time of writing, even the operational concept for SESAR is less than well articulated, although it is clear that the basic ideas underlying SESAR will be fundamentally different from current practice. For example, concepts such as the Reference Business Trajectories and Controlled Time of Arrival within the context of the SWIM (System-Wide Information Management) system network, together with advance on-board avionics, will be integrated to implement the ideas of 4D trajectories. The underlying principles and practices for controlling future flights using such 4D trajectories will be very different from current practice. Also, while some very high level ideas of the system architecture is available, e.g. the SWIM architecture, it is difficult to make estimates about which technologies will be used and the nature of that use. Consequently, the nature of the operational concept, the likely work practices in these yet-to-be-determined work environments and the tools needed to support the work are equally hard to predict.

Our approach to this is to take a Human-Centred Innovation process to invent a number of tools that are intended to

facilitate different aspects of airspace and traffic visualisation. By combining these tools in novel ways, we hope to greatly improve the acquisition and comprehension of information in ATC in these anticipated work environments.

In contrast, a common approach to designing user oriented systems is known as User Centred Design [1]. This approach prescribes an examination of the task being performed, understanding the nature of the demands that the task places on the user, and then to design of the best possible interfaces to improve the performance of the task. User Centred Design assumes that the task exists, but in our dealings with the future ATC system we cannot be certain what the necessary tasks will be, and at this stage, neither can we reliably predict what the technology will be. Indeed it is likely that the capacity of users to deal with the expected increases in air traffic will be a major factor in the design of tasks and technology. This is a “wicked problem” [2], since the process of solving the original problem is likely to reveal other problems, some of which may be more complex than the problem’s original form.

To this end, we engaged in a Human-Centred Innovation process as there are no SESAR ATC tasks or work environments to study work practices and human factors needs. Rather than build on existing operating paradigms of ATC, it is necessary that we take what is known of the abilities of users and attempt to predict which tasks they will best be able to perform and with what technology, and extrapolate that within the context of SESAR. This process of extrapolation will be discussed later in our paper. During our review of the literature we found a limited amount of information on techniques for theory-driven user-focused innovation and exploration of the design space [3]. This paper is partly an attempt to fill this gap in an ATC context.

Despite the uncertainty of what SESAR will demand from the controllers, we are reasonably confident that controllers will need to visualise the airspace they control. In previous visualisation systems for ATC the view of the airspace is often completely 2D, such as the systems in use today, or completely 3D. Three-dimensional rendering of a scene eases integrated attention tasks, in which information is integrated across three or more dimensions. Such tasks may include guiding an aircraft that is making a simultaneous descent and turn. 3D

rendering is also useful for appreciating the positions of aircraft in relation to each other in 3D space. However, well known drawbacks of 3D representations include fore-shortening of distances due to distortions arising from perspective. This makes it difficult to correctly estimate distances and hence aircraft separations. Two-dimensional information is superior for focused attention tasks, such as estimating the horizontal separation of converging aircraft. However, it requires a high degree of experience to re-construct and maintain a dynamic mental picture of the spatial relationships between aircraft in 3D space. An extensive review of the advantages and disadvantages between 3D and 2D has been reported elsewhere [4] and is summarised in Table 1.

In order to maximise performance, it would appear desirable to allow controllers to shift between 2D and 3D views [5]. However, most current rendering systems require the user to switch their attention to a new spatial context, either by shifting their view from a 2D planar radar display to a 3D rendering of it, or re-configuring the 3D display by moving the camera position, to access a top down plan view of traffic. Based on our previous cognitive task analysis and field studies

of air traffic controllers' work, we do not believe that interfaces requiring such shifts in spatial context and visual attention will confer a major advantage. Other techniques such as displaying a smaller window with 3D information overlaying parts of a 2D planar view display may obscure critical information. Such a display design is potentially hazardous and therefore, also not desirable.

We hope to invent novel visualisations that combine 3D and 2D in a way that compensates for the shortcomings of these display techniques, whilst exploiting their complementary potential. We also hope that these new visualisations will increase the controllers information handling capacity, and hence their scope for control.

We will next discuss the human-centred innovation process and describe the designs that emerged from the process, their design rationale and how the concepts developed. This will be followed by a discussion of the findings from an exploratory evaluation of the designs. Then we close with a discussion about our experiences with the human-centred innovation process.

Display type	Advantages (+)	Drawbacks (-)
2D Display	<p>Global traffic picture always available</p> <p>Supports correct distance estimation, focused attention tasks [6-8];</p> <p>Supports improved performances for visual search tasks [9]</p> <p>Easy to orient. User maintain easily orientation awareness</p> <p>Navigation and Selection are easy to achieve</p>	<p>Does not spatially represent altitude information, and requires the controller to read and interpret alphanumerical altitude data to produce and maintain 3D picture</p> <p>Suffers from cluttering. Overlapping labels and blips difficult to read</p>
3D Display	<p>Supports Superior performance for integrated - shape understanding – tasks [6-8];</p> <p>Supports development of accurate mental model of traffic and terrain, effective training tool [10];</p> <p>Supports effective decision making for a/c maneuvering on the vertical plane [11];</p> <p>Supports at glance assessment of consistency of implemented maneuver with the original one as intended by controller [12]</p>	<p>Hampered distance estimation performances due to perspective distortion effects [10];</p> <p>Not possible to oversee global traffic/global sector out of camera view [13];</p> <p>Traffics at the far end of the scene difficult to locate, due to decrease in resolution [14];</p> <p>Navigation and selection difficult, user can get lost when moving the camera</p>

Table 1. Relative advantages and drawbacks of 2D and 3D display (Rozzi et al., 2007)

II. THE HUMAN-CENTRED INNOVATION PROCESS

The Human-Centred Innovation Process combines a focus on the needs of the user with the search for technological advances that can create opportunities for creating innovations in the way people work and in the design of tools to support the work. Focusing purely on the user and their task can be limiting as it is primarily retrospective in nature, investigating work practices that current exist. This could potentially lead us to 'designing for the last war', instead of designing for a scenario of what could be. On the other hand, focusing mainly on technological advances can lead to problems of 'a hammer looking for nails'. While each approach has benefits and shortcomings, we prefer to leverage both, drawing on the notion of the Task-Artifact Cycle [15]. We accept that new

technologies will afford new capabilities that create or enable opportunities for new ways of working and even new forms of work. Taking our case as an example, the future technologies (e.g. the SWIM network) can create opportunities to change the controllers' work from directing and controlling aircraft, to supervising a larger volume of self-managing, self-separating aircraft. Once such opportunities are envisaged, it will place new demands on the future work of the controller, requiring new designs and solutions to support the new tasks.

The Human-Centred Innovation Process we adopted comprised several sub-processes: a review of the literature, and an iterative invention process involving creativity workshops, scenario development, and design exploration [15]. Then in subsequent cycles, the emphasis shifts from creation to the

combination of existing concepts to provide hybrids that compensated for shortcomings in the original concepts.

The innovation process was firstly informed by a review of about 100 papers describing innovative visualisations in ATC, Command and Control, Medical, Geographic, Engineering and Flow imaging. This gave rise to the Combination Display Framework (Table 2). This framework allowed us to hypothesise which types of innovations were likely to be successful and focus our energy accordingly.

The Combination Display Framework is a two-way classification scheme for helping us understand the patterns in existing combinations of 3D and 2D representations, and the basis for their combination. Since the various display formats are largely orthogonal to their display techniques we can use this framework to identify places in the literature where current gaps exist and we can use related research to predict how successful we are likely to be in filling those gaps.

The Display Format, on the horizontal axis of the Combination Display Framework, classifies various ways of combining 2D and 3D visualisations. The broad categories of display format are the Side-by-side, Multiview, Exo-vis and In Place views. In the vertical axis we see different display techniques which implement the principle of Focus and Context, i.e. the need to see information within the context of its data sets or of the task situation. These techniques include rapid zooming, distortion, multiple coordinated views.

In this version of the Combination Display Framework (table 2), we have included some examples of visualisations found in the ATC literature. The Framework also suggests that there are many other possible combinations where exploration may bear fruit. Airspace may be rendered in strict 3D, or combined using differing methodologies.

During the invention cycles, it was necessary to force our participants who were experienced controllers, to break with their traditions and well-established work practices. We used creativity workshops to identify existing assumptions, to discuss and break them, and to consider the technologies likely to be in place in the future, to encourage the controllers to imagine scenarios that were different from the present situation and those of the near future, in order to facilitate the generation of new ideas.

As we tried not to be overly constrained by the technology (and indeed, our lack of knowledge of the what the actual future technologies might be!), we used a low fidelity, paper-based prototyping approach for the design and design space exploration for the visualization tools we were developing. Ideas could be quickly sketched and walked-through, allowing changes and refinements to be made quickly.

Since precise task analysis of the future air traffic control task is not yet possible, we have adopted a strategy for designing these tools that distinguishes between what is visualised, the *content*, and the means and devices that are used to present them, the *container*. These containers allow the combination of various content, such as the magnification of 2D, 3D views of a selected area, integration of 3D in 2D information, or 4D (spatial-temporal) information in numerous ways with different types of containers. These containers can

include “lenses” which can distort or magnify the underlying 2D information; or “scoops” that would allow a controller to scoop out a segment of the airspace. These containers with the content, could then be manipulated and interacted with in various ways. For example, using gestural and touch techniques would allow one to “reach in and grab” or “scoop” segments of the airspace of interest; or using alternative display interaction technologies, we can use point of view tracking with off-axis projection to create “fish tank VR”[16] where what a person sees changes as that person’s point of view changes. The idea in the innovation process is to explore a series of tools that allow us to guide the development of future tasks and technologies.

We will next describe several of the visualisation technologies we have developed. The technologies should be regarded as tools that will contribute to a solution, not solutions in their own right.

III. THE VISUALISATION TOOLS

We used the Combination Display Framework together with our human-centred innovation approach to create new ideas for the visualization tools. 10 readily identifiable novel concepts were developed and were then distilled to the six key tools reported here. It should be noted again that these tools are intended to be extended and used in further combination to form possible future solutions. They should not be seen as stand-alone controller’s tools or systems.

Many of the prototypes described here were developed using the ARToolkit [17], an open source library for developing Augmented Reality (AR) applications. The toolkit uses a camera to detect markers and then augments these markers with additional information. The additional information may be displayed on a screen, via a head mounted display (HMD) or in the environment itself [18]. This is really two separate technologies – registration and display. Registration is the task of finding interesting information in the environment – this is the camera finding the marker. Display is the addition of information in the appropriate place. It should be remembered that the focus is not the registration technique (which could instead be based on magnetic tracking, infra-red tracking, or some new technology) or the display technology (which is also changeable). The only area of concern here is how the container helps controllers visualise the airspace. In many cases the flexible nature of the registration in the ARToolkit has allowed us to create cheap prototypes for technologies that do not need to use the AR Toolkit.

In the following descriptions of the visualization tools, we hope to explain their design rationale, and how these ideas developed. These ideas were based on four considerations: (i) the ability for a controller to reach into a display and select by touch or other means an area of interest for closer examination, (ii) the integration of needed 3D information into 2D to create in-place representations, (iii) the need to support accurate measurement of pertinent information such as distances and altitude differences, and (iv) the need for global awareness of the traffic situation while enabling localized spatial awareness. We will now present these concepts.

DISPLAY TECHNIQUE	DISPLAY FORMAT				
	Strict 3D	2D/3D Combination Displays			
		Side By side	Multiview 2D/3D	Exo-Vis	In Place
Uncorrelated view (no technique)		St. John ('01)			
3D in 2D symbols					Smallman ('01)
<i>Focus + Context Techniques</i>					
Multiwindows	EC-Lin AR ('06)				PiP Display, AD4 ('06)
Rapid Zooming	Ellis ('87) Azuma ('96) Brown ('94) Eddy et. Al ('99)				
Distortion					Distortion Display, AD4 ('06)
Overview Plus detail			Alexander & Wickens ('03)		
Filtering	EC-Lin VR ('05)	Azuma ('00)	Ilog Display		Lens Display, AD4 ('06)
Multiple Coordinated Views	EC-Lin VR ('05)	Azuma ('00) Furtsenau ('06) D3, AD4 ('06)	Ilog Display		

Table 2. Combination Display Framework [19]

A. AR in Your Hand

This concept is based on the notion of using one's hand to reach into a 2D radar display and to grab a segment of the airspace and traffic within that airspace. The AR (Augmented Reality) in your hand concept allows a controller to use his or her hand to select a portion of the 2D radar screen, and then have the 3D (or any other view of the selected area) appear in the palm of his or her hand. In this prototype we use markers to facilitate the process, but in the final implementation, we envisage it would work using just a user's hand.

Figure 1 shows the implemented prototype AR in your hand using the ARToolkit. The AR in your hand display allows a controller to intuitively examine the airspace of interest by rotating their palm to change perspective, and bringing it closer



Figure 1. AR in your hand. A section of the airspace is depicted in 3D. The image is anchored to the marker, so moving the marker moves the 3D information as well. This allows views from multiple directions and collaboration between controllers.

for a closer inspection. As well as allowing a controller to view a volume in detail, it could be used as a collaborative tool for controllers to share information about interesting areas of the airspace or to hand over a situation to another controller.

This localised view could be placed beside a monitor for reference and dismissed at the snap of the fingers. With an HMD or future display techniques it is hoped that the 3D information can be actually viewed in the user's hand.

The key benefits of this interaction technique are that selection and viewing are simple and intuitive, the spatial context of the selection is easily perceived and there is no need for occlusion of relevant spaces. Current display technologies require either a Side-by-Side or Multi-Window format, but this may be mitigated in future.

B. The Tangible Lens

The Tangible Lens is a prototype display technique for smoothing interactions between the physical and the information space. By moving the "lens" over parts of the 2D radar screen, it will augment those parts of the screen with additional information. This information could take the form of textual, graphical or 3D information. Figure 2 shows a picture of the Tangible Lens acting as a magnifying glass, implemented in the ARToolkit.

The lens is envisaged to be a physical tool that will be permanently attached to the screen. When a user needs

additional information the tool will be grabbed, dragged to the relevant area and used to acquire the necessary information. When the information has been obtained, the tool is released and springs out of the way. This tool would greatly reduce the intrusiveness of acquiring additional information.



Figure 2. The tangible lens, implemented with the ARToolkit. The area inside the cardboard "lens" is a magnified view of that region of the airspace.

Additionally, the acquired information would have a high spatial relevance. With the development of multi-touch screens, exemplified by the iPhone, it may be possible to replace the lens with something like a "pinch" touch interface.

This is an example of the In Place display format, since the additional information is replaces part of the 2D view on the main screen. Principle benefits are the rapid acquisition of spatially relevant information, the minimal occlusion period and the intuitive simplicity of the interaction.

C. AR 3D Wall View

The 3D Wall View is an example of the notion of the scoop where a segment of the airspace may be scooped up and placed beside a controller or used for discussion with other controllers. This view uses the Exo-Vis display format implemented using the Side-by-Side display technique. It grew from two earlier concepts, the "local view" and the "precise distance estimation view." While these concepts were not useful alone, in combination they provided a promising new tool.

The 3D view (figures 3 and 4) allows a rapid comprehension of the structure of the airspace, while two 2D representations on the wall (called the 'altitude ruler') give users precise data needed to assess the traffic situation or guide aircraft accurately. Such a view can allow a controller to check whether a given aircraft is able to level off at an assigned flight level after a climb (or descent); visualise the 3D path of an aircraft close to terrains; guide an aircraft through a complex approach, monitor one or more holding stacks at the same time, or access a pilot's point of view during severe weather conditions.

All of these airspace features can be included and magnified in a sub volume such as the one described above. We have implemented this view as both a stack manager and an approach controller. The multiple uses of this view are an example of why these visualisations are regarded as tools as opposed to solutions. The key benefits of this view are the general overview combined with spatially relevant 2D information.

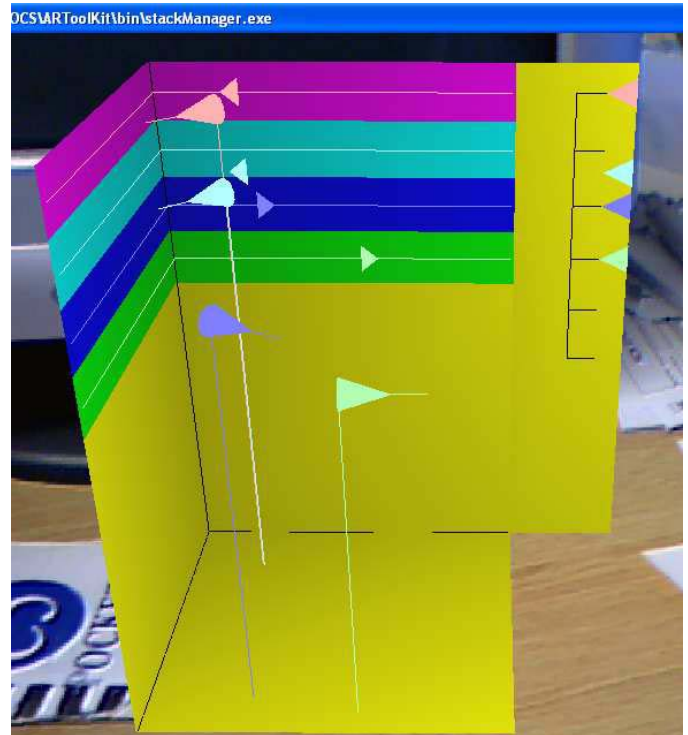


Figure 3. The 3D walls display provides a 3D overview to aid situational awareness and a 2D view to aid precise judgements

This 2D information is much more useful to a controller when making precise judgments such as separation or heading. In the stack manager, it is quite difficult to determine the vertical separation of aircraft from the local view alone. However the wall view allows the controller to rapidly and accurately determine which, if any, aircraft are deviating from assigned flight levels.

In the approach control implementation the complex trajectory of an inbound aircraft can be quickly comprehended with the global view, but the 2D representations on the walls and floor make it easier to determine if the target is on the correct glide path and slope.

A current drawback is the separation of the walled area from its spatial context, but further combination with other displays – such as the Tangible Lens detailed above or the AR Tabletop described below should mitigate or eliminate this problem.

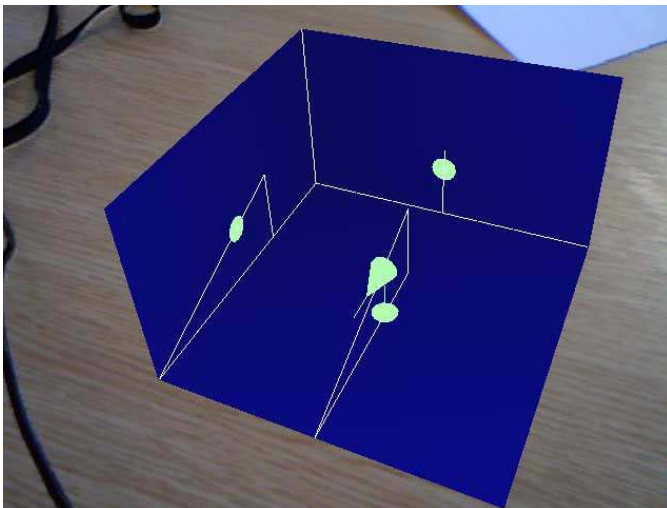


Figure 4. The same 3D walls contain this time with approach control content.

D. Skyscraper and Symbicons

The skyscraper set of display concepts allow a controller to monitor a standard 2D plan view and to selected areas of the 2D space for 3D viewing. The user can see a 3D representation of an area of interest within a 2D overview of the entire airspace. The aircraft in our concept prototype are represented as red-cyan anaglyphs in the selected airspace. When viewed using red-cyan filters (cheap 3D glasses), the selected aircraft appear to stand out of the display, much as if one were looking down at skyscrapers from above. Aircraft at

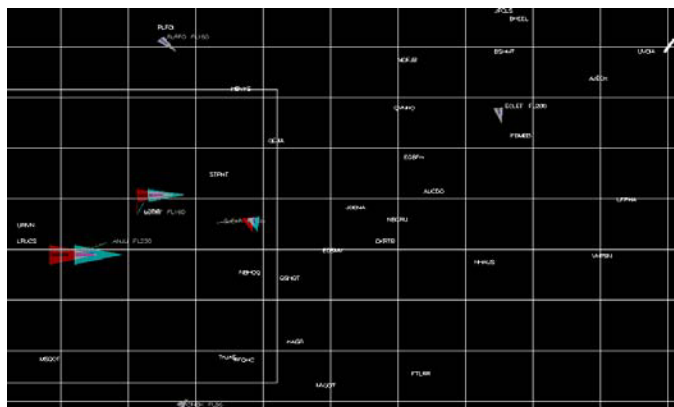


Figure 5. The skyscraper and symbicons display. The red-blue aircraft appear to pop out of the screen when viewed with 3D glasses. The lines on the aircraft indicate whether the target is climbing, descending or flying level.

higher altitudes appear closer to the viewer while aircraft at lower altitudes appear farther away from the viewer (Figure 6).

Such a 3D-in-2D view is useful for de-cluttering since the selected aircraft appear to stand out of the display, gives a controller a quick assessment of the traffic situation, and can provide clear indications of where the aircraft are in 3D space. Note that such a view can be combined in the “AR magnifying glass” lens view, and can be implemented in advanced autostereo display technology that does not require glasses or anaglyphs to simulate visual depth.

The 3D view need not represent actual altitude. For instance, instead of presenting a locality where the third dimension represents altitude, the user could choose to filter aircraft according to airspeed, with the perceived depth designating the relative velocities of the selected aircraft.

Symbicons, or icons with embed symbolic 3D-related information can be used to add information about the aircraft's behaviour, indicating if the aircraft is climbing, descending or banking (figure 6). This is an example of the integration of 3D information into a 2D space.

This symbicon concept has been implemented as a short black line on the axis of the aircraft. If the line is towards the rear of the aircraft (triangle) symbol, it indicates that the aircraft is climbing; and if it is located more forward, it indicates it is descending. If the line is to the left, it indicates that the aircraft is banking left, in a climbing left turn (if to the rear and left), or in a descending left turn (if to the left and forward). Symbicons and the 3D skyscrapers are complementary technologies (figure 5). They allow users to readily perceive information about the intentions and actions of aircraft, and rapidly determine flight levels. In this design, we can preserve the need for global awareness of the overall traffic situation provided by the 2D representation, while decluttering the display by separating the aircraft and their labels in depth and supporting analogical reasoning. Smallman et al. [6] found that symbolic 3D led to better visual identification performance.

Note that symbicon techniques could be added to other displays as well, especially the AR Tabletop described below. The combination of a 2D Display, 3D Filtering and Symbolic information are an excellent example of the utility provided by the Combination Display Framework.

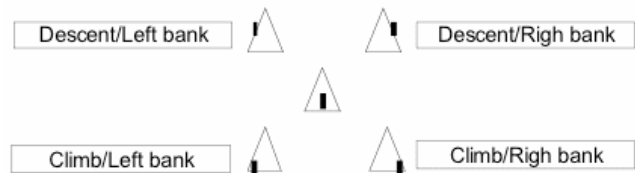


Figure 6. How the symbicons concept works to display attitude information

E. AR Tabletop Display

The Table Top Display (figure 7) is a collaborative workspace rendered in genuine 3D. Multiple controllers can view a single workspace from multiple directions, allowing them to share ideas and solve complex problems. The view allows examination from any angle, and can provide textual information, just like a 2D display. Note that we are able to augment the entire table top – we are not limited by the size of the marker. Practically, there are a number of display technologies to bring this about, especially spatial AR and “fish tank” VR. The display format is a pure 3D environment, but it is easy to see how symbols could augment the display, or how walls could be used to provide precise information in a spatially relevant way (see Figure 13 for an early paper mock-up of how the 3D walls could be implemented on the Tabletop display). This display benefits from a strong spatial context

and easy collaboration, as well as the potential for combination with the other tools described above.

IV. EVALUATING THE CAPABILITIES OF THE VISUALISATION TOOLS

We also carried out an exploratory evaluation of the designs to identify the capabilities afforded by the new 3D/2D representation concepts and their potential uses. In particular, to identify the variety of perceptual and interaction issues that may arise, such as the new user capabilities provided by the tools. Also of interest was the identification of feature interaction effects, i.e. how some of these concepts might in combination lead to more powerful capabilities, or hinder controller performance.

It was not the intention of the evaluation to quantitatively assess the effect of the new designs on a particular set of operational tasks, since the tools were not yet intended to solve particular problems.

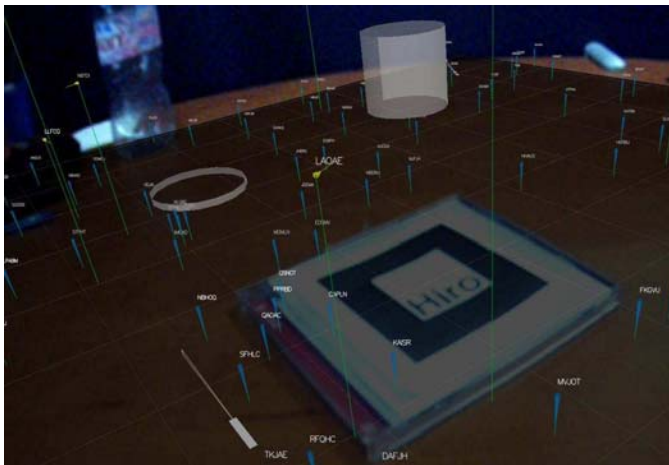


Figure 7. The tabletop AR display uses a single marker to augment an entire workspace. Users can view the display from multiple points of view, and obstacles such as clouds and restricted airspaces are clearly visible.

While the focus of the innovation process was centred on users the design process was, as yet, too immature to make quantitative analysis of the tools productive. Indeed, focusing on specific implementations would have instead hampered the exploration of the tools' full capabilities. The evaluation took place at EUROCONTROL with three non-operational Air Traffic Controllers, each with between 3 to 18 years of operational experience. They were requested to perform exploratory tasks in simplified scenario demos and asked to imagine how the new visualisation could help them to carry out operational tasks. We also used a set of predefined questions to stimulate thinking about alternate uses of the tools.

A. Applications of the 3D-in-2D Concepts in ATC

The study showed in general, that 3D visualisation can help a controller manage holding stacks, busy airspace sectors and

military aircraft interception. Several potential benefits were indicated:

- 3D in combination with 2D appears to offer the potential to reveal 3D positions and trajectories relative to each other.
- Perception of dynamic and static 3D volumetric shapes of restricted airspaces is easier, reducing the cognitive load on controllers.
- Current radar displays can result in nearby aircraft overlapping aircraft on the screen, making it difficult to read call signs. 3D displays can alleviate this.
- The use of 3D might provide military controllers assistance to track high performance aircraft. This supports the finding of a previous work analysis [5]. At this stage however it seems that more operational evidence need to be collected, since none of the subjects involved in the session had military experience.
- In busy airspace sectors, 3D could be useful for disambiguating traffic arriving or departing from a given airport from those arriving and departing from another nearby airport. With the arrival of technologies such as RNAV (Precision Navigation), this could increase the effective use of the airspace. Aircraft could be guided along 3D paths not available today due to uncertainty in navigation. For instance, in between the departure of a fast and a slow aircraft – the former departing first and climbing at higher altitude, the latter departing shortly afterwards and climbing at a lower altitude – it is not possible to put any aircraft today; but with the support of RNAV it would. Such an improved use of the airspace would result in higher density of activities which could be better appreciated in combination with a 3D visualisation.

B. Applications of 3D-in-2D Concepts beyond ATC

Other application areas beyond ATC include airspace planning; procedure design, experimental scenario design, and training.

Three dimensional CAD (Computer Aided Design) applications facilitate planning and design with a combination of both 2D and 3D views. 3D viewing could enhance the planning and design of airspaces. Planners and procedure designers could use a virtual airspace to collect a qualitative understanding of different aspects of the airspace, such as ground clearances or noise distribution.

There is also potential for use in experimental scenario design and performance analysis. A researcher could accurately visualise traffic in 3D space, a task that most researchers will find considerably more difficult than trained controllers.

Trainee controllers could experiment with new airspaces using 3D-in-2D displays. They could interactively explore a localised 3D space using different features to understand its configuration, the distribution of global traffic, and assess the implications of maneuvers, e.g. lack of separation, overshoot or unnecessary travel distance.

These areas may be more suitable for initial airspace design because they are not real-time, operational situations. This means that some limited experimentation is possible. Since airspaces and procedures are carefully designed and examined over long periods of time, the addition of 3D visualisations has the potential to give users a way to improve comprehension, collaboration and efficiency without risking passenger safety.

V. CONCLUSION

The human-centred innovation process was intended to provide new concepts that could aid human performance in the future SESAR ATC context. It was necessary because the domain under consideration consists of loosely defined tasks with new or even future technology. Indeed, this process is likely to inform the development of those tasks and future technologies.

The Combination Display Framework allowed us to reason about which combinations of display format and techniques were likely to produce good results, and focus creative energies in those directions. This aided the transfer of knowledge from the work of others while reducing the risk of duplication.

The iterative component of the Human-Centred Innovation process led to the discovery of potential tool combinations. The skyscraper-symbicons concept and the AR 3D wall view both combined existing concepts to create vastly improved visualisations.

In general, the Human-Centred Innovation process allowed us to creatively develop technology while remaining focused on the needs of users.

VI. ACKNOWLEDGEMENTS

The work reported in this paper was funded by the EUROCONTROL Innovation Research Programme CARE INO 3 Contract Number C06/12399BE. Parts of this paper were originally presented at the EUROCONTROL INO Workshop in December 2008, as part of the Year 1 Final Report. We would like to thank the air traffic controllers at EUROCONTROL who participated in various discussions and evaluations, and who generously gave us their time.

VII. REFERENCES

- [1] Norman, Donald. *Design of Everyday Things*. New York: Basic Books, 2002.
- [2] Rittel, Horst, and Melvin Webber; "Dilemmas in a General Theory of Planning," pp. 155-169, *Policy Sciences*, Vol. 4, Elsevier Scientific Publishing Company, Inc., Amsterdam, 1973. [Reprinted in N. Cross (ed.), *Developments in Design Methodology*, J. Wiley & Sons, Chichester, 1984, pp. 135-144.]
- [3] Naikar, N., Drumm, D., & Sanderson, P. (2003). Designing teams for first-of-a-kind, complex systems using the initial phases of Cognitive Work Analysis: Case Study. *Human Factors*, 45(2), 202-217.
- [4] S. Rozzi, A. Boccalatte, P. Amaldi, B. Fields, M. Loomes, and W. Wong, "D1.1: Innovation and Consolidation Report," EUROCONTROL, Technical Report, Prepared for the 3D-in-2D Display Project June 2007 2007.
- [5] P. Amaldi, B. Fields, S. Rozzi, P. Woodward, and W. Wong, "Operational Concept Report Vol. 1: Approach Control," Interaction Design Centre, Middlesex University, London, Uk, Technical Report OCR1-AD4-WP2-MU, 2005.
- [6] M. St. Jhon, H. Smallman, S., H. Oonk, M., and M. Cowen, B., "The use of perspective view displays for situational awareness tasks," presented at Human performance, situation awareness and automation: user centered design for the new millennium, Savannah, GA, 2000.
- [7] N. Naikar, "Perspective Displays: A review of Human Factors Issues," Departement of Defence DSTO-TR-0630, 1998.
- [8] I. Haskell, D., and C. Wickens, D., "Two and Three Dimensional Displays for Aviation: A Theoretical and Empirical Comparison," *The International Journal of Aviation Psychology*, vol. 3, pp. 87-109, 1993.
- [9] H. Smallman, S., M. S. John, M. Oonk, H., and B. Cowen, M., "Searching for Tracks Imaged as Symbols or realistic Icons: A Comparison Between Two-Dimensional and Three-Dimensional Displays," San Diego, CA 2001.
- [10] D. Wickens, & May, P., "Terrain Representation fo Air Traffic Control: A Comparison of Perspective With Plan View Displays," 1994.
- [11] S. Ellis, R., M. McGreevy, and R. Hitchcock, J., "Perspective Traffic Display Format and Airline Pilot Traffic Avoidance," *Human factors*, vol. 29, pp. 371-382, 1987.
- [12] S. Rozzi, P. Woodward, P. Amaldi, B. Fields, and W. Wong, "Evaluating Combined 2D/3D Displays for ATC," presented at 5th INO EUROCONTROL Innovative Research Workshop, Bretigny sur Orge - France, 2006.
- [13] C. Wickens, M. Vincow, and M. Yeh, "Design Application of Visual Spatial Thinking, The Importance of Frame of Reference," in *The Cambridge Handbook of Visuospatial Thinking*, P. Shah and A. Miyake, Eds. NY: Cambridge University Press, 2005.
- [14] B. Boyer, S. and C. Wickens, D., "3D Weather Displays for Aircraf Cockpits," Aviation Research Laboratory, Savoy, IL Technical report ARL-94-11/NAA-94-4, 1994.
- [15] Carroll, John M. Scenario-based design: Envisioning work and technology in system development. John Wiley & Sons Inc. New York, NY., 1995.
- [16] Deering, M., High Resolution Virtual Reality, *Computer Graphics*. vol. 26, no. 2., July 1992, pp. 195-202.
- [17] Kato, H., Billingham, M. Marker Tracking and HMD Calibration for a video-based Augmented Reality Conferencing System. In Proceedings of the 2nd International Workshop on Augmented Reality (IWAR 99). October, San Francisco, USA, 1999.
- [18] Bimber, O., & Raskar, R. Spatial Augmented Reality: A Modern Approach to Augmented Reality. In Proceedings of Annual Conference on Computer Graphics and Interactive Techniques - SIGGRAPH'05. New York: ACM Press. . 2005.
- [19] S. Rozzi, P. Amaldi, W. Wong, and B. Field, "Operational Potential for 3D Display in Air Traffic Control," presented at ECCE 07, European Conference on Cognitive Ergonomics, British Computer Society, Covent Garden, London, 2007.

Embedded eye tracker in a real aircraft: new perspectives on pilot/aircraft interaction monitoring

Frédéric Dehais

Centre Aéronautique et Spatial
ISAE-SUPAERO, University of
Toulouse
Toulouse, France
Frederic.dehais@isae.fr

Mickaël Causse

Centre Aéronautique et Spatial
ISAE-SUPAERO; INSERM, U825,
University of Toulouse. Toulouse, F-
31000 France
Mickael.causse@isae.fr

Josette Pastor

INSERM, U825, Toulouse, F-31000
France;
University Paul Sabatier, University
of Toulouse, Toulouse, F-31000
France.
josette.pastor@toulouse.inserm.fr

Abstract—Currently, online assessment of the aircrew performance focuses on behavioural data (flight data and pilot's actions) and the detection may intervene too late for coping with the situation degradation. An early assessment of the pilot's "internal state", based on physiological data collected from his autonomous nervous system (ANS) and predictive of his behaviour, is necessary. These data give clues both on the cognitive activity and on the emotional states and stress. The integration of ANS devices in a cockpit presents practical drawbacks and their use is often limited to simulators. In this preliminary study, the pros and cons of the adaptation of a standalone eye tracker in a light aircraft are presented. In spite of a sensitivity to light conditions and a definition of areas of interest limited to a part of the cockpit, the eye tracker has provided interesting behavioural (fixations) and physiological (pupillometry) measures in nominal (from take-off to landing) and degraded (provoke a simulated engine failure and plane down toward the airfield) conditions. The pilots spent less time glancing at the instruments, and focused on less instruments in the degraded condition. Moreover, the pupil size varied with the flight phases in the degraded condition, which reflected the variations of stress and attention levels. These encouraging results open two tracks: the development of new eye trackers able to overcome current technical limitations, and neuroergonomics researches providing guidelines for new man-machine interfaces integrating both flight and crew state vectors.

Keywords: *Neuroergonomics, Eye Tracker, pilot activity, pupillometry, human factors*

I. INTRODUCTION

As long as the human operator (i. e. the pilot) is a key agent in charge of the flight, the definition of metrics able to predict his performance is a great challenge. Currently, online assessment of the aircrew performance focuses on behavioural data measured from the pilot/aircraft interactions. In particular, formal methods are developed to detect human errors thanks to aircraft/pilot behaviour monitoring [1, 2]. A predictive approach based on particle Petri nets [3] is also proposed to anticipate possible pilot-systems conflicts [4] that are accurate

precursors of the pilot's loss of situation awareness [5]. However, all these methods, which rely on the operator's actions, may intervene too late for coping with the situation degradation. An early assessment of the pilot's "internal state", predictive of his behaviour, should tackle the problem. Physiological data collected from the pilot's autonomous nervous system (ANS) are good candidates since they give clues both on the cognitive activity and on emotional states and stress [6] [7]. Arousal, vigilance, emotional states, attentional demand can be derived from heart rate and blood pressure, theta and alpha brain waves, temperature variations, respiration, skin conductance and oculometry [8] [9].

All ANS devices though present practical drawbacks. They are sensitive to the operator's physical state (e.g. sweating perturbs skin conductance, fever changes temperature and heart responses ...) or to the environment (e.g. magnetic and electric fields may create artefacts on electroencephalograph responses), and they are too cumbersome to be easily adapted to a cockpit. For example, oculometry suffers the following limitations:

- As eye fixations "fill up" the total time, not all fixations are relevant to assess the pilot's visual demand. Moreover, a fixation does not necessarily imply perception [10];
- As the pupil diameter varies in function of light intensity to maximise visual capacity, pupillometry cannot be used to assess the level of stress of the pilot under changing light environments ;
- Electro-oculograms and most eye trackers are cumbersome devices and may disrupt pilots' activity since pilots have to wear an electrode close to the eye or special equipment like helmets.

These considerations tend to restrict the utilisation of ANS devices to controlled studies in laboratory (i.e. flight simulator) [11] [12], although they have already been used in real flight conditions [13]. Eye tracking offers a fruitful perspective since

visual perception is a key for pilot to control the flight and oculometry may provide both behavioural and cognitive/emotional physiological measures [7] to assess the pilot's performance:

- The visual-search strategy, or the selective attention to relevant visual stimuli is an index of information needs [14];
- The eye-scanning patterns of pilots in terms of frequency of fixations seem to be related to the instruments' importance. The length of fixations, however, is related to the difficulty in obtaining/interpreting information from instruments [15];
- An increase in workload is accompanied by increased fixation times [16] [17];
- The decrease of the duration and the number of eye blinks are strongly correlated to visual demand [18] [9];
- Low frequency "pupillary oscillations" are linked to fatigue [19];
- The pupillary response is related to mental workload [20] [21] [22];
- In many cognitive processes such as language processing, perception, memory, complex reasoning and attention, the pupil diameter grows with the difficulty of the task [23] [24];
- Pupillary responses also provide clues on the emotional state [25] [26] and pupil size may vary on a continuum according to emotional valence [27] or reflect the emotional activation or arousal [28].

Our long-term goal is to develop an onboard system able to predict the pilot's performance through the analysis of the aircraft state vector, the pilot-aircraft interactions state vector and the pilot's physiological state vector. However, the integration of this latter state vector implies to assess the feasibility and the acceptability of a non intrusive physiological device. Thus, this preliminary study proposes to assess the usability of an on-board eye tracker in real flights and aims at assessing the benefits of this tool for human factor concerns.

II. METHODS

A. Participants

A permit to fly was given by the European Aviation Safety Agency (number 856/2007 – EASA PTF.A07.0232) to conduct the experimentation with the restriction that ISAE flight instructors only were authorised to fly the airplane with the on-board eye tracker. Six ISAE flight instructors, all males, could participate to the experiment. Their mean age was 43 years (range, 35-58). Their mean flying experience was 5896 hours (range, 1480-13000). The six participants were qualified to fly

the Aquila AT01 aircraft (two-seated light airplane, 100 horsepower).

B. Scenario

The flight scenario starts at nightfall at Lasbordes airfield and ends before the beginning of the aeronautical night¹. It is divided into two consecutive sequences (cf. fig 1):

- The first sequence consists in a classical visual traffic pattern : take off (1) – cross wind leg (2) – down wind leg (3) - base leg (4) – last turn (5) – final leg (6) - touch and go (7) . This sequence is the nominal condition;
- The second sequence starts after the previous touch and go (7) and consists in flying back toward Lasbordes airfield at an altitude of 2500 feet. Once over the airfield, the pilot had to chop the throttle so as to perform an engine failure exercise (8) and then to plane down toward the airfield (9-12). This sequence is the degraded condition.

This scenario is presented to the pilot one hour before the beginning of the experimentation. During the briefing, it is clearly exposed to each pilot that he decides the moment of the engine failure exercise and that he may use the throttle at any moment if flight safety is altered.

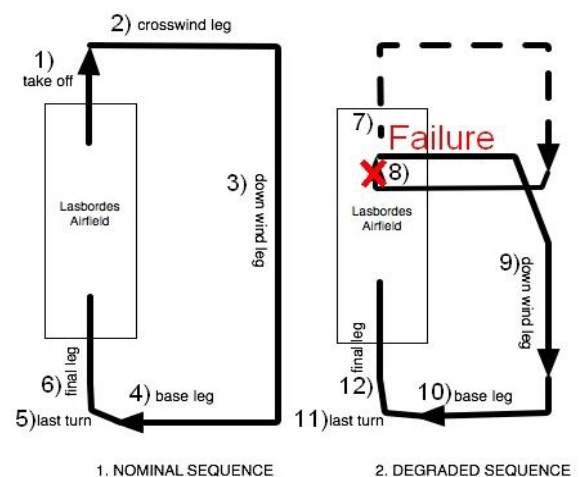


Figure 1: The two flight sequences: the nominal and the degraded one. The nominal sequence ends after the landing (6), the degraded sequence starts just after, with the take off (7)

C. Oculometry

A non intrusive eye tracker Tobii x50 was used for the purpose of the experimentation. This device has 0.5 degree of accuracy and a 50 HZ sampling rates. It also provides instant re-acquisition after extreme head motion. It had to be adapted to be easily set in the Aquila aircraft without any modification of the airplane and without provoking any disturbance for the pilot (e.g. no visual scanning disturbance).

¹ The aeronautical night begins in France thirty minutes after sunset.

The eye tracker was placed below the left part of the instrument panel in front of the pilot's seat (cf. Fig 2). A scene camera was centrally placed under the fix part of the canopy. Data synchronization and processing was done via an analog/numerical video converter, a Tobii external card and a Sony Vaio laptop. These three light devices were situated in the luggage compartment.

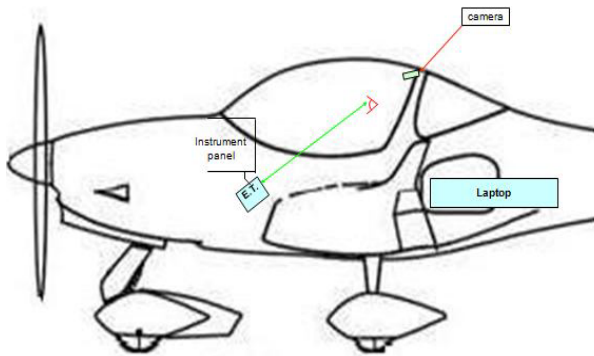


Figure 2: The eye tracker (ET) was fixed under the instrument panel and the data processing system was placed in the luggage compartment. This latter processed and surimposed in real time both the eye tracker data and the video data coming from the scene camera.

The technical characteristics of the x50 eye tracker and its particular location allowed to track the pilot's eye gaze on the left part of the instrument panel where are the primary flight beacons (i.e. airspeed, altimeter, horizon...). As shown in the figure 3 and 4, it is not possible to determine the pilot's eye gaze out of this area (i.e. the eye gaze out of the cockpit, the eye gaze on the beacons situated on the right part).



Figure 3: The blue dot represents the gaze fixation. Note that the pilot is focusing on the airspeed instrument to check the rotation speed (Vr), just before taking-off

1) Area of interests

A dedicated analysis software provides in real time data such as the timestamps, the (x,y) coordinates of the pilot's eye gaze on the left instrument panel and the pupil diameter. Moreover it is possible to determine the number and the duration of fixations in specific "areas of interests". In this sense and in order to study the pilot's ocular behaviour, fourteen areas of interest corresponding to the fourteen beacons of the left instrument panels have been respectively defined: (1)

outside temperature, (2) compass, (3) manifold pressure, (4) alarm, (5) airspeed, (6) horizon, (7) altimeter, (8) tachymeter, (9) bank and turn indicator, (10) directional gyroscope, (11) vertical speed, (12) VOR, (13) switches (14) flaps (cf. fig 4).



Figure 4: The fourteen rectangles define the different area of interests.

2) Pupil diameter variations

As the eye gaze, the pupil size is recorded continuously. In practice, establishing mean physiological values for a group of subjects for an entire task is meaningless because of inter-individual variability. We use delta values (differences between the mean pupil diameter during the concerned flight sequence and the one calculated on the whole experiment) for measuring the pupil variations.

Luminosity measurements were performed. Indeed the pupil regulates the amount of light that enters the eye, and thus, its size variations are highly sensitive to the luminosity changes. Thanks to a lux-meter the ambient luminosity was continuously recorded in order to identify the flight sequences where the light remains reasonably constant. In accordance with Gupta's work [30] on pupil size variations in function of the ambient luminosity, we have limited the pupil variations analysis to sequences where the visible light was inferior to 25 lux.

III. RESULTS

As the sample size was small and the data did not follow normal distributions, nonparametric statistical methods for dependent sample were used. Overall analyses were performed with the Friedman Anova. Wilcoxon signed rank test was used for paired-samples tests. The analyses were performed with Statistica 7.1 (© StatSoft).

A. Behavioural results

The different fixation times, expressed in percentage of the total time spent on the 14 defined instruments during the nominal landing sequence vs. the landing sequence with the

simulated failure (the degraded landing sequence), are presented in figures 5 and 6.

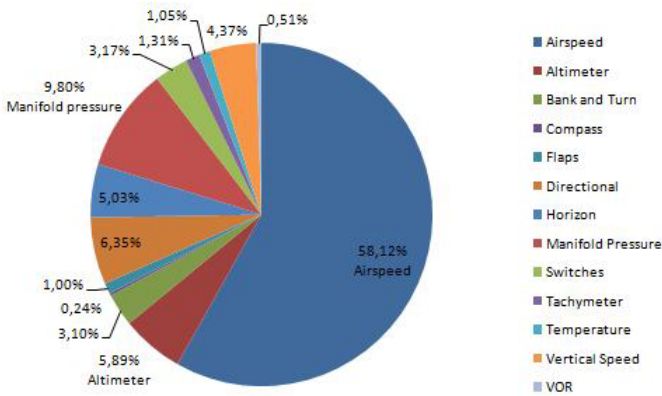


Figure 5: Mean fixation durations in percentages on the areas of interest of the six pilots during the nominal landing, from base leg until the flare (mean total duration = 13.59 sec)

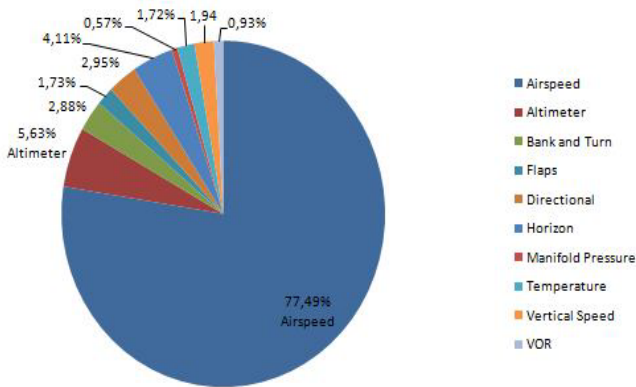


Figure 6: Mean fixation durations in percentages on the areas of interest of the six pilots during the degraded landing, from base leg until the flare (mean total duration = 5.53 sec)

The results showed a reduction of the number of instruments gazed during the degraded sequence in comparison to the nominal one. During the nominal sequence, all the instruments were looked (except for the alarm panel) whereas only 10 instruments were looked during the degraded sequence. More precisely, compass, switches and tachymeter weren't gazed. Moreover, there was an increase of the relative fixation time on the airspeed during the degraded sequence regarding to the nominal (77.49% vs. 58.12%).

Finally, the time spent on instruments appeared to be lower during the degraded sequence (cf. Fig 7), showing that pilots focused more on external information.

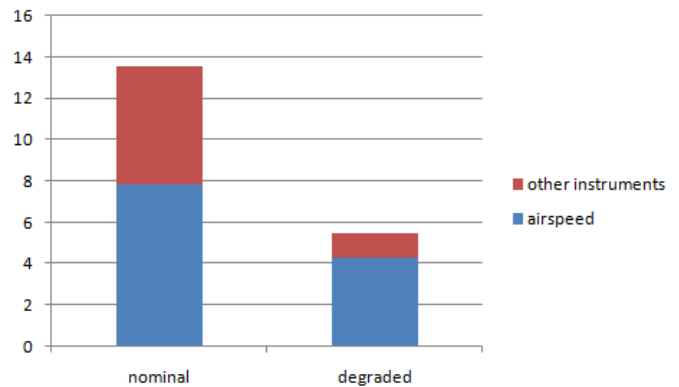


Figure 7: Mean fixation time on the airspeed instrument and all other instruments during the nominal landing and the degraded landing (time in sec)

Below is presented the official "cruise checklist" of the Aquila AT01 (table 1) and the fixation times in percentages (fig. 8) obtained during the cruise check list (generally performed at an altitude of 2000 feet).

TABLE 1: AQUILA ATO1 OFFICIAL "CRUISE CHECKLIST" [31]

Trim	set
Chronometer	Top and estimated
Altimeter	set
Directional	Checked
GPS	Use & Stby state
Engine instruments	Checked
Manifold pressure	25 inches
Tachymeter	2000 RPM

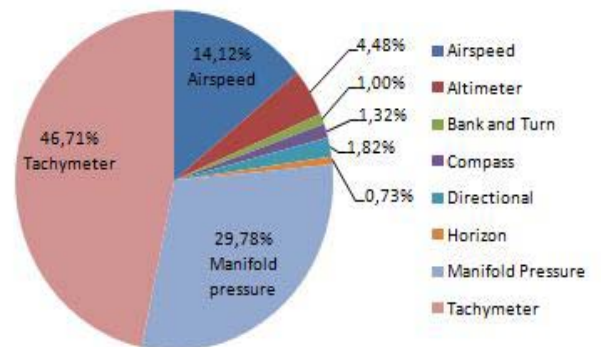


Figure 8: Mean fixation times in percentages on the 14 zones of interest during the "cruise checklist". Note that "airspeed" stands for "airspeed indicator", "directional" stands for directional gyroscope, "horizon" stands for gyro horizon and "bank and turn" stands for "bank and turn indicator" (mean total duration = 7.27 sec)

The analysis shows that the tachymeter is the most looked instrument (46.71%), and then comes the manifold pressure (29.78%) and the airspeed (14.12%).

B. Pupillary response

In spite of the fact that all the experiments were conducted at nightfall, the luminosity variation did not allow to analyse pupil diameter variations during all the sequences (cf. table 2).

Considering this pitfall, only the degraded sequences of four pilots (pilot ID: 1 to 4) were included in the pupil diameter variation analysis. Indeed, mean luminosity variations during the considered sequences were only of 5.75 lux for the four pilots.

TABLE 2: LUMINOSITY MEASUREMENTS FOR EACH PILOT AT THE BEGINNING AND THE END OF THE TWO FLIGHT SEQUENCES (E.G. PILOT ID 1 STARTED THE FIRST FLIGHT SEQUENCE WITH 127 LUX AND ENDED IT WITH 40 LUX ; HE THEN STARTED THE SECOND SEQUENCE WITH 20 LUX AND ENDED IT WITH 7 LUX)

Pilot ID	Luminosity variations (in lux)	
	first sequence	second sequence
Subject 1	127-40	20-7
Subject 2	82-36	8-4
Subject 3	90-35	10-6
Subject 4	93-33	10-12
Subject 5	473-230	220-91
Subject 6	610-600	520-380

The Friedman's ANOVA showed a strong significant difference ($p=0.001$) concerning the delta pupillary diameter among the four flight phases (fig. 9). However, Wilcoxon post hoc paired-samples analysis didn't show any difference.

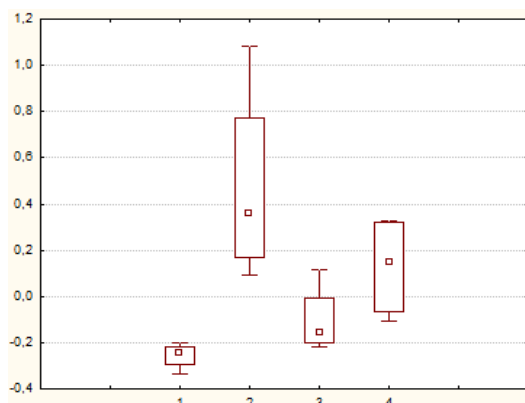


Figure 9: Pupillary diameter changes (in mm) regarding the four flight phases during the second flight sequences, respectively one minute before the failure (1), the failure and the crosswind (2), the base leg (3), from the last turn to the final touch (4)

IV. DISCUSSION

The introduction of a new onboard device for human factors purposes must fulfil three requirements: 1) the device does not disturb the pilot's activity, 2) the device is able to work correctly in real flight conditions, 3) the device improves significantly the human-machine interface, and therefore the flight safety. The preliminary neuroergonomics work that is presented here gives some clues about the capabilities of an eye tracker as a device onboard a small aircraft.

Observation of the six pilots showed that no perturbation was generated by the eye tracker that remained totally unnoticed after the preliminary set up phase. However, this must be confirmed on non expert pilots and on other types of aircrafts.

Concerning the usability of the eye tracker in real flight conditions, the results are more equivocal. Because of light issues, we weren't able to compare pupil dilation during the nominal sequence vs. the degraded one in all pilots. Moreover, AOs can be only defined at places that are constrained by the eye tracker's location in the plane. This can be overcome by helmet eye trackers, however with an intrusiveness of the device in the pilots' activity. Technological progress must thus be accomplished to allow the generalisation of onboard eye tracking experiments.

In spite of these drawbacks, the preliminary results highlight the possibility of deriving interesting measures of the pilot's activity from eye tracker data.

Analysis of the areas of interest captured by the eye tracker is a way to assess its accuracy in real conditions. In particular, analysis of critical events such as the in-flight checklist sequences is a key to evaluate the reliability of this tool:

- The visual scanning is codified by an official procedure in the flight manual than can be used as a model of reference;
- These sequences are very short (less than ten seconds) and flight parameters to be checked by the pilot are vital.

Such constraints lead the pilots to relevant eye fixations during these periods and allow assuming that the areas of interest observed are also a priori the result of a real voluntary attentional activity.

In this perspective, the analysis of the areas of interest during the "cruise checklist" (cf. Fig 5) shows that the visual scanning of the six pilots is limited to eight flight instruments. More precisely the pilots have focused on the tachymeter, the manifold pressure, the airspeed indicator, the altimeter, the directional gyroscope, the compass, the bank and turn indicator and the gyro horizon. These areas of interest are consistent with the ones defined in the official Aquila "cruise checklist": the tachymeter and the manifold pressure have to be set to particular values and the results of these adjustments have to be implicitly checked on the airspeed indicator, the altimeter has to be checked as the "cruise checklist" starts at an altitude

of 2000 feet, and the value of the directional gyroscope has to be checked, which is done thanks to a quick comparison with the value of the compass.

Though it is not explicitly expressed in this checklist, it is totally consistent that flight indicators such as gyro horizon and turn and bank indicator are supervised by the pilots in order to be stabilised perfectly on the three axes (roll, pitch and yaw) to perform an optimal checklist. One may notice that some actions of the checklist are not detected by the eye tracker but:

- The chronometer and GPS settings were not performed as the pilots were not engaged in a complex navigation task but stayed close to the airfield;
- The trim or the engine instruments (e.g. oil pressure indicator) are located on the right part of the instruments panel where no eye tracking could be established due to the limitation in angle of the Tobii x50.

The analysis of the fixation times on the areas of interest during this checklist showed that the pilots focused particularly on the tachymeter (46.71 % fixation time on this instrument during that checklist), the manifold pressure (29.78 %) and the airspeed indicator (14.12 %). Interviews with the six pilots have confirmed that the engine management during this checklist requires a certain amount of attentional demand: very accurate and careful adjustments were needed on the tachymeter, manifold pressure and the speed. The pilots spent less time on checking instruments as the altimeter (4.48%), the directional gyroscope (1.82 %) or the compass (1.32 %). Indeed, discussion with the pilots revealed that the altimeter was rapidly looked to check that a 2000 feet altitude was reached (i.e. to start the cruise checklist). They also just glanced at the directional gyro and the compass: as the pilots were not about to perform a navigation task, the cross-checking of these two indicators were of less importance. In this sense, these findings are consistent with research conducted on the correlation between attentional demand and time fixations [15] [16] [17]: important information implies longer time fixations.

The analysis and the comparison of the pilots' areas of interest during the landing in nominal and degraded conditions (cf. fig 6 and 7) have revealed different ocular patterns. First of all, the total duration of fixations in nominal conditions was more than two times higher to the total duration of fixations in degraded conditions (13.59 seconds vs. 5.53 sec). This suggests that in degraded conditions, the pilots spent more time looking outside the cockpit to assess and adapt their trajectory in reference with the airfield. Another major difference between the two landings relied on the fact that the pilot's areas of interests were less distributed in the degraded conditions than in the nominal one with a particular focus on the airspeed indicator (77.49% of total fixation times in degraded landing vs. 58.12% in nominal landing). A first consideration is to take into account the fact that in the degraded condition, the pilots had no more interest to supervise the tachymeter and the

manifold pressure due to the engine failure. Another consideration is given by the pilots who all agreed that in the degraded condition they had faced troubles to manage their speed as they were surprised by the high lift-to-drag ratio of the Aquila. In this sense, this led them to focus especially on the airspeed indicator and particularly to take a key decision: maintaining the landing or going around.

The analysis of pupil diameter variations shows some evolutions according to the different flight sequences. More precisely, mean delta pupil diameter was of -0.25 mm before the simulated failure, +0.47 mm during the failure and the cross wind leg, -0.10 during the base leg and +0.12 from the last turn to the final touch. These results are consistent with the pilots' interviews that report a high anxiety and cognitive demand due to the management of the aircraft during the few early minutes of the simulated failure, a lower anxiety during the base leg because of the successful stabilization of the aircraft, and finally another increase of anxiety and cognitive demand during the landing sequence because of the required precision and the potential go-around in case of unsafe approach. Furthermore, the literature classically reports a high cognitive demand and a high ANS arousal during the landing [32], which is consistent with the increase of pupil diameter observed during the last turn and the final touch.

During the flying activity, pilots are confronted with numerous stressors that can deplete their performance, such as time pressure, increased anxiety, and unexpected failure. Whereas a growing literature [33] [34] sheds light on the effect of complex flight scenarios or anxiety on pilots' performance and physiological parameters, real flight experiments remain extremely rare. Therefore, on-board eye tracking offers promising perspectives in terms of real condition monitoring of both pilot's actions and physiological states, although this ecological approach shows some technical limitations. Firstly, the analysis of the areas of interest shows the reliability of the tool and its capability to predict behaviours. Indeed, during a takeoff, it has been possible to link the absence of visual scanning on the flaps with an omission of a required action on them later. Moreover, the AOI analysis allows bringing to light differential visual behaviours according to the landing condition (nominal or degraded). Secondly, the measurement of the pupil dilation gives clues on the pilot's emotional state and/or cognitive workload. The pupil response seems to evolve differently during the four flight phases. The pupil diameter appears to be higher just after the simulated engine failure. This observation is coherent with the increase of mental demand and/or anxiety during this particularly critical flight phases. In addition, the occurrence of differential patterns during the failure phases vs. the nominal ones seems to emphasize pupil diameter results. Further work should be conducted by night to totally get rid of the luminosity variations and to attempt to produce more results, in particular concerning the comparison of degraded and nominal conditions.

ACKNOWLEDGMENT

The authors wish to thank all the pilots from ISAE and Fabrice Bazelat, Frank Yvars and Thierry Louvet for the time they spent on setting up the eye tracker onboard the Aquila, Eric Absil for his great job on the eye tracker, and Christian Colongo for his support. The eye tracker Tobii x50 has been kindly lent to ISAE by the firm "Pilot Vision" (Alexander Seger) during the experimentation.

REFERENCES

- [1] M. Heymann, A. Degani, On formal abstraction and verification of human-machine interfaces: the discrete event case. NASA Technical Memorandum, 2001.
- [2] T. Callantine, The crew activity tracking system: Leveraging flight data for aiding, training, and analysis. Proceedings of the 20th Digital Avionics Systems Conference, Daytona Beach, FL, October, 5.C.3-1--5.C.3-12 (CD-ROM), 2001.
- [3] C. Lesire, C. Tessier, Particle Petri nets for aircraft procedure monitoring under uncertainty. 26th International Conference on Application and Theory of Petri Nets and Other Models of Concurrency (ATPN), Miami, Florida, USA, 2005.
- [4] F. Dehais, C. Lesire, A. Goudou, C. Tessier, Toward an anticipating agent to help pilots AAAI Fall Symposium "From reactive to anticipatory cognitive embodied systems", Arlington, VA, USA, 2005.
- [5] F. Dehais, C. Tessier, L. Chaudron, GHOST: experimenting conflicts countermeasures in the pilot's activity. IJCAI-03, Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence, Acapulco, Mexico, 2003.
- [6] HD. Critchley, Electrodermal responses: what happens in the brain? *Neuroscientist*, 8(2):132-42, 2002.
- [7] E. Granholm, SR. Steinhauer, Pupillometric measures of cognitive and emotional processes. *International Journal of Psychophysiology*, 52(1):1-6, 2004.
- [8] M. J. Skinner, P. A. Simpson, "Workload issues in military tactical aircraft," *Int. J. Aviat. Psychol.*, vol. 12, no. 1, pp. 79-93, 2002.
- [9] J.A. Veltman, A.W.K. Gaillard, (1996) Physiological indices of workload in a simulated flight task, *Biological Psychology* ,42, 323-42, Elsevier, 2002.
- [10] D. De Waard, The measurement of drivers' mental workload. PhD thesis, University of Groningen. Haren, The Netherlands: University of Groningen, Traffic Research Centre, 1996.
- [11] R.J. Mumaw, N.B. Sarter, C.D. Wickens, Analysis of pilots' monitoring and performance on an automated flight deck. Proceedings of the 11th biennial meeting of the International Symposium on Aviation Psychology, Columbus, OH, 2001.
- [12] P. Kasarskis, J. Stehwien, J. Hichox, A. Aretz, C. Wickens, "Comparison of expert and novice scan behaviors during VFR flight," 11th International Symposium on Aviation Psychology, 2001.
- [13] J.A. Veltman, A Comparative Study of Psychophysiological Reactions During Simulator and Real Flight. *International Journal of Aviation Psychology*, Volume 12, 2002.
- [14] P.K. Hughes, B.L. Cole, The effect of attentional demand on eye movement behaviour when driving. In A.G. Gale, M.H. Freeman, C.M. Haslegrave, P. Smith & S.P. Taylor (Eds.), *Vision in vehicles-II* (pp. 221-230). Amsterdam: North-Holland, 1988.
- [15] G.F. Wilson, F.T. Eggemeier, Psychophysiological assessment of workload in multi-task environments. In D.L. Damos (Ed.), *Multiple-task performance* (pp. 329-360). London: Taylor & Francis, 1991.
- [16] R.D. O'Donnell, F.T. Eggemeier, Workload assessment methodology. In K.R. Boff, L. Kaufman & J.P. Thomas (Eds.), *Handbook of perception and human performance*. Volume II, cognitive processes and performance. (pp 42/1-42/49). New York: Wiley, 1986.
- [17] R.W Backs, L.C. Walrath, Eye movement and pupillary response indices of mental workload during visual search of symbolic displays. *Applied Ergonomics*, 23, 243-254, 1992.
- [18] L.O. Bauer, R. Goldstein, J.A. Stern, Effects of information-processing demands on physiological response patterns. *Human Factors*, 29, 213-234, 1987.
- [19] B. Wilhelm, H. Wilhelm, Das Pupilverhalten verrät Übermüdung. *Zeitschrift für Verkehrssicherheit*, 41, 116-118, 1995.
- [20] T.I. Shamsi, X. Sam Zheng, B.P. Bailey, Task-evoked pupillary response to mental workload in human-computer interaction. CHI 2004, Vienna, Australia, 2004.
- [21] L.T.M. Hoeks, The pupillary response as a measure of mental processing load: with application to picture naming. PhD Thesis, Nijmegen, The Netherlands: University of Nijmegen, 1995.
- [22] J. Hyönä, J. Tammola, A.-M. Alaja, Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks. *The Quarterly Journal of Experimental Psychology*, 48A, 598-612, 1995.
- [23] J.L. Andreassi, Pupillary response and behaviour. In: Mahwah, N.L.E.A.(Ed.), *Psychophysiology*. Human behav. *Physiol.response*, Vol. 10. Oxford University Press, New York, pp. 218-233, 2000.
- [24] J.R. Jennings, M.W. Van der Molen, Preparation for speeded action as a psychophysiological concept. *Psychol. Bull.* 131 (3), 434-459, 2005.
- [25] J. Hytintk, J. Tammola, A.-M. Alaja, "Pupil Dilation as a Measure of Processing Load in Simultaneous Interpretation and Other Language Tasks". *The Quarterly Journal of Experimental Psychology*, 48A (3): 598-612, 1995.
- [26] M. Causse, B. Pavard, J.-M. Sénard, J.-F. Démonet, J. Pastor, Emotion Induction through Virtual Avatars and its Impact on Reasoning: Evidence from Autonomous Nervous System Measurements and Cognitive Assessment. *Actes de la Conférence VRIC'06*, Laval, France, 18-20 Avril, 2007.
- [27] E. H. Hess, "Pupillometrics". In N. S. Greenfield and R. A. Stembach (Eds.) *Handbook of Psychophysiology*, pages 491- 531. Holt, Richard & Winston, New York, NY, 1972.
- [28] MP. Janisse, "Pupil Size, Affect and Exposure Frequency". *128 Social Behavior and Personality*, 2 (2): 125-146, 1974.
- [29] R. Parasuraman, M. Rizzo, *Neuroergonomics: The Brain At Work*, Oxford University Press, 2007.
- [30] A. Gupta, J. Schwiegerling, J. Straub, Design and use of an infrared Pupilometer for real-time pupil mapping in response to incremental illumination level. *Optical society of America*, 2000.
- [31] A. Schimederer, Aquila AT01 Flight Manual, Technical report AFM-AT01-1010-100-E, AQUILA GmbH, 2005.
- [32] Y-H, Lee, B-S. Liu, Inflight workload assessment: comparison of subjective and physiological measurements. *Aviat Space Environ Med*; 74:1078-84, 2003.
- [33] C. Dussault, J.-C. Jouanin, M. Philippe, C.-Y. Guezennec, EEG and ECG changes during simulator operation reflect mental workload and vigilance. *Aviat Space Environ Med*; 76:344-51, 2005.
- [34] T. Macuda, T. Poolman, P. Keller, Workload Assessment in Flight using Dense Array Eeg Schnell. 25th Digital Avionics Systems Conference, IEEE/AIAA Volume , Issue , Oct. 2006 Page(s):1 - 11, 2006.

Visual Cognition Abilities in X-Ray Screening

Diana Hardmeier* and Adrian Schwaninger †

*Department of Psychology, University of Zurich, Switzerland, and
e-mail: d.hardmeier@psychologie.uzh.ch

†Department of Psychology, University of Zurich, Switzerland, and
Max Planck Institute for Biological Cybernetics, Tübingen, Germany,
e-mail: a.schwaninger@psychologie.uzh.ch

Abstract — The job of aviation security screeners is a highly demanding task. Based on the x-ray image of a passenger bag, a screener has to decide within few seconds only whether the bag is ok or has to be hand-searched. This x-ray screening task includes specific knowledge and visual cognition abilities. The knowledge about which items are prohibited and what they look like in x-ray images of passenger bags have to be learned on the job. In contrast the ability to cope with high bag complexity, superposition and viewpoint of threat items is relatively stable and can only be improved little with on the job training. Whether these abilities can be measured within a pre-employment assessment procedure using different subtests of well established intelligence test batteries was investigated in this study. Results revealed a relationship between the latent variable ability and detection performance in x-ray screening for both samples. However, 4 of the 12 intelligence tests are sufficient to explain detection performance in x-ray screening. The relationship between the latent variable ability, the X-Ray Object Recognition Test and detection performance later on the job was tested additionally.

Abilities, aviation security, visual cognition, x-ray screening

I. INTRODUCTION

In recent years the importance of aviation security has increased enormously. To avoid that passengers bring potential threat items into the security restricted area and on board an airplane, body search and x-ray screening of passenger bags is essential. The x-ray screening task of aviation security screeners is very demanding and includes both specific knowledge and visual cognition abilities. Screeners have to acquire the knowledge about which items are prohibited and what they look like in x-ray images of passenger bags. This job and task specific knowledge and expertise respectively has to be learned after people got employed. Further, considering x-ray images different factors such as bag complexity, superposition and viewpoint of the threat items can influence the detection as well. Studies in this area could show that detection performance decreases significantly if threat items are shown in close-packed bags, if threats are more superimposed by other items and if they are shown in an unusual view. These effects were found for experts and novices. Furthermore, large individual differences could be seen for both, experienced aviation security screeners and novices [1]. Reference [1] defined these factors as image-based

factors in x-ray screening. As they could be found for both groups, they are rather referred to relatively stable abilities than training. Therefore, it can be assumed that job applicants who are able to cope with these image-based factors perform better later on the job. Thus, measuring the ability to cope with image-based factors within a pre-employment assessment should increase detection performance later on the job remarkably.

Therefore, the X-Ray Object Recognition Test (X-Ray ORT), a reliable and valid x-ray screening test that measures image-based factors relatively independent of knowledge was developed [2]. Results could show that test results in the X-Ray ORT correlate significantly with threat image projection (TIP) data which measure detection performance on the job. Further, aviation security screeners who were selected with the X-Ray ORT performed in another x-ray screening test that measures all kind of prohibited items and was applied within the recurrent competency assessment significantly better than screeners who were not selected with this test [3].

However, the image-based factors should also be measurable with general visual cognition tests as these factors can be compared to the visual cognition processes visual search, figure-ground segregation and mental rotation that were investigated in many research studies. Furthermore, it can be expected that other abilities such logical thinking or concentration and vigilance play also an important role. For example the detection of improvised explosive devices (IEDs) which vary widely in shape and form, but share a common set of components differs from the detection of other prohibited items. As not one shape as a hole has to be detected, but the three components power source, detonator and explosive material, this task probably requires rather logical thinking. Moreover, screeners have to be constantly vigilant when performing the x-ray screening task. Therefore, a visual cognition test battery (CTB) including 12 tests that best match the x-ray screening task was applied within the pre-employment assessment additionally. Most tests are part of well established German intelligence test batteries. Four subtests of the Leistungsprüfsystem [4], three subtests of the Intelligenz Struktur Test 2000 (IST 2000) [5], the Raven's Advanced Progressive Matrices [6], the Frankfurter Aufmerksamkeits Inventar (FAIR) [7] and three tests which were developed by the University of Zurich [8] were used.

Tests from the visual CTB were expected to measure the following unobserved latent factors figure-ground segregation, visual search, mental rotation, spatial imagination, logical thinking and vigilance.

In a first step the influence of ability on detection performance in x-ray screening was investigated using the visual CTB. A common factor model was estimated to measure which tests in the visual CTB predict on the job performance best and can therefore be used as pre-employment assessment tool. Further, the common factor model which was estimated to measure the relationship between ability and detection performance in x-ray screening was validated by another sample. In terms of efficiency a possible shortening of the visual CTB was examined. Further, a full structural equation modeling (SEM) was estimated by defining the test results in the X-Ray ORT as additional indicator.

II. METHOD

A. Participants

The two samples used in this study consisted of 169 ($M = 35.10$, $SD = 9.85$; range 20 to 55 years) and 97 ($M = 36.19$, $SD = 11.44$; range 20 to 55 years) respectively job applicants who were employed as aviation security screeners based on their test results in the pre-employment assessment for aviation security screeners. The first sample (2006 Sample) consisted of 66 females and 103 males, the second sample (2007 Sample) of 51 females and 46 males. Part of the pre-employment assessment was the X-Ray ORT, the visual CTB, a German and English language test, a color blindness test, a physical examination test and a job interview. All results except for the visual CTB were used as selection criteria.

B. Measures

1) *Visual Cognition Test Battery (CTB)*: The visual CTB consists of 12 tests which are mostly part of well established intelligence tests. All tests were conducted computer-based and not in the original paper-and-pencil form. To measure the second order factor ability, nine tests were assigned to the four first order factors figure-ground segregation, visual search, mental rotation and spatial imagination conducting Confirmatory Factor Analyses (CFAs). The remaining three tests Raven, Fair and a subtest of the IST 2000 (IST_MF) served as indicators.

a) *Figure-ground segregation*: The latent variable figure-ground segregation was measured with the LPS10 and the Noiser. The LPS10 is a subtest of the Leistungsprüfsystem [4], a major German intelligence test battery. It measures the ability to recognize a shape by ignoring irrelevant other features. Participants have to choose the only simple shape out of five which fits into the complex line drawing. The test includes 40 shapes of increasing complexity. Scored is the number of correct solutions that can be answered within 3 minutes. The Noiser was developed by the University of Zurich [8]. It measures how well people can recognize objects that are not fully visible. The test consists of 80 line drawings of simple objects which are increasingly destroyed (level of

destruction: 75%, 80%, 85% and 90%). Trials are shown for 4 seconds only and then participants have to mark the correct term out of 20 choices. Scored is the number of correct choices.

b) *Visual search*: Visual search was measured with the Letter Search Test (LST) and the Image Comparison Test (ICT) [8]. The LST consists of a total of 60 trials. Participants have to find a lowercase letter within three-dimensional uppercase letters. There are three difficulty levels increasing in the number of uppercase letters. Each trial is presented for 5 seconds only, then participants have to decide whether there was a lowercase letter or not. Only fifty percent of all trials contain a target object. For analysis d' is calculated. The ICT comprises of two almost identical pictures that are presented next to each other. Participants have to mark all 15 differences within 3 minutes. Scored is the number of correct marked differences.

c) *Mental rotation*: The latent variable mental rotation was measured with the LPS7 and the Figureauswahl (IST_FA) that are subtests of two major German intelligence test batteries, the Leistungsprüfsystem [4] and the Intelligenz-Struktur-Test (IST 2000) [5]. In the LPS7 participants have to mark the flipped number or letter in a row of equal but randomly rotated numbers or letters. Participants are given 2 minutes to complete as many trials as possible out of 40. Again, scored is the number of correct solutions. The IST_FA is about rearranging several pieces to one of five possible figures. The test consists of 20 trials that have to be solved within 7 minutes. Scored is the number of correctly answered trials.

d) *Spatial imagination*: Spatial imagination was measured with the LPS8, LPS9 of the Leistungsprüfsystem and the Würfelaufgabe (IST_WÜ) which is again a subtest of the IST 2000. The LPS8 consists of eight trials that have to be completed within 4 minutes. Participants have to mentally fold a leaf of paper into a defined form and determine for several sides which one of the leaf corresponds to the folded form. Again scored is the number of correct answers out of 40. The LPS9 measures spatial ability and asks participants to count the number of sides of three-dimensional geometric objects. Then they have to mark the correct number out of ten choices. Scored is the number of correctly marked numbers. The test duration is 3 minutes and maximum score is 40. Last, the subtest IST_WÜ consists of 20 trials that have to be completed within 9 minutes. Participants have to mentally rotate a cube and decide which of five alternatives match the target cube.

e) *Raven*: Logical thinking was measured using Raven's Advanced Progressive Matrices [6]. This test measures non-verbal deductive reasoning and visual discrimination. Participants have to complete a 3 * 3 matrix of abstract figures whereof the last figure in the lower right corner is missing. They can choose the right figure out of eight alternatives. The total of 47 used matrices increases in difficulty over time and the test duration is set to a maximum of 10 minutes. Again, scored is the number of correct solutions.

f) *Fair*: The Frankfurter Aufmerksamkeits Inventar (FAIR) measures vigilance [7]. The task in this test is to discriminate between very similar looking signs as fast and accurate as possible. The participants are given 6 minutes to attend the test consisting of a total of 640 trials. The number of correctly detected targets as well as correctly rejected non-targets is used for analysis.

g) *Merkfähigkeitstest (IST_MF)*: The IST_MF is as well a subtest of the IST 2000 and measures visual memory capacity [5]. This test that measures performance of short-term memory for figures consists of 13 pairs of symbols that have to be memorized within 1 minute. Then participants have to select the correct counterpart for all 13 symbols out of 5 alternatives within 3 minutes. Scored is the number of correct solutions.

2) *Detection performance in x-ray screening*: The detection performance in x-ray screening was measured with two x-ray screening tests and TIP data. The Prohibited Items Test (PIT) and the Bomb Detection Test (BDT) were part of the recurrent competency assessment which was conducted between 4 and 6 months after employment. Both tests are about recognizing threat items in x-ray images of passenger bags. Images were displayed for 10 and 15 seconds respectively on the screen. Then, participants have to answer whether the bag was OK (included no threat item) or NOT OK (included a threat item) by clicking on the button. Both, the prohibited items and bomb detection test differed in the 2006 and 2007 sample only insofar as other images were used. Results were calculated using d' which is a psychophysical measure and takes into account the hit and false alarm rate [9], [10]. For details about these x-ray screening tests, reliability and validity measures see [3], [11]. TIP is a technology which allows displaying fictional threat items into real passenger bags. That way, detection performance on the job can be measured. Again, d' was calculated and used as detection performance measure. For more information about TIP data see [12].

3) *X-Ray Object Recognition Test (X-Ray ORT)*: The X-Ray ORT is an x-ray screening test which was developed to measure the ability to cope with image-based factors in x-ray screening relatively independent of knowledge. It consists of 256 x-ray images of passenger bags. Half of them contain either a gun or a knife. The other 128 images are harmless bags. Each bag is displayed for 4 seconds on the screen and then participants have to decide whether the bag was OK (no threat item) or NOT OK (a gun or knife) by clicking on the respective button. Detection performance was calculated using the detection performance measure d' . Test construction, its reliability and validity measures can be seen in [2], [3].

C. Procedure

The performance in the visual CTB and the X-Ray ORT was measured within the pre-employment assessment procedure. After employment all screeners had an initial training course which took three weeks. They also received training with the individual adaptive training system X-Ray Tutor (XRT). Screeners worked 4 to 6 months before they passed the first competency assessment which includes three x-ray screening tests and a theoretical exam on computer.

D. Modeling Description

The goal of this study was to test whether results in the single tests of the visual CTB show a relationship to detection performance in x-ray screening later on the job. The model was tested using a step-by-step procedure. First, CFAs were conducted to investigate how well the indicator variables accurately reflect the latent variables. Then, a common factor model was conducted for each group (2006 Sample, 2007 Sample). Second, a possible shortening of the visual CTB was tested. Third, a full structural equation modeling was conducted. As goodness-of-fit indices we report the sample-size-independent comparative fit index (CFI). Its values indicate a good fit the closer they are to one. According to [13] values greater or equal to .90 indicate acceptable model fit. We also report the root-mean-square error of approximation (RMSEA). RMSEA values less than or equal to .05 indicate good model fit. Furthermore, the information theoretical fit measures AIC, BCC, BIC and CAIC are reported because they are less sensitive to small sample size and are not based on statistical inference using probability theory (see [14]). All information theoretical fit measures should be substantially smaller than they are for the saturated model [15].

III. RESULTS

Table 1 shows descriptive statistics for all indicator variables. Table 2 and Table 3 depict the sample correlation matrix for the 2006 sample and for the 2007 sample, respectively (see Appendix).

We first specified a CFA model with the four first order factors figure-ground segregation, visual search, mental rotation, spatial imagination and the three indicators Raven, Fair, IST_MF to measure the second order factor ability. However, results indicate that the second order factor loadings between the second order factor ability and the four first order factors as well as the three indicators were all not significantly different from one. Thus, all 12 indicators load on one factor and there is no need to model separate factors. Furthermore, another first order factor named detection performance in x-ray screening was defined. This factor measured the detection performance in x-ray screening with the three indicators PIT, BDT and TIP. As can be seen in Figure 1, the common factor model includes the two first order factors ability and detection performance.

TABLE I. RELIABILITIES, MEANS, STANDARD DEVIATIONS OF INDICATOR VARIABLES

Indicator variables	Reliability	2006 Sample (N = 169)		2007 Sample (N = 97)	
		M	SD	M	SD
LPS10	.83 ^c / .69 ^c	0.63	0.20	0.61	0.20
Noiser	.95 ^a / .91 ^b	0.18	0.02	0.18	0.02
LST	.73 ^a / .81 ^b	0.36	0.17	0.36	0.14
ICT	.83 ^d	0.65	0.18	0.64	0.16
LPS7	.83 ^c / .61 ^c	0.32	0.18	0.28	0.16
IST_FA	.76 ^a / .79 ^b	0.51	0.20	0.51	0.19
LPS8	.83 ^c / .70 ^c	0.63	0.31	0.62	0.29
LPS9	.83 ^c / .75 ^c	0.58	0.15	0.53	0.15
IST_WÜ	.80 ^a / .86 ^b	0.50	0.19	0.47	0.20
Raven	.93 ^a / .94 ^b	0.34	0.14	0.32	0.14
Fair	> .78 ^b / > .85 ^c	0.34	0.07	0.27	0.09
IST_MF	.92 ^a / .80 ^b	0.54	0.20	0.56	0.24
X-Ray ORT*	> .91 ^a / > .78 ^b	1.74	0.33	1.85	0.22
PIT	> .87 ^a / > .87 ^b	6.02	1.68	--	--
CAT	> .88 ^a / > .84 ^b	--	--	6.58	1.72
BDT1.0	> .80 ^a / > .77 ^b	3.71	2.49	--	--
BDT2.0	> .88 ^a / > .80 ^b	--	--	5.50	1.93
TIP	.58 - .90 ^b	9.00	1.19	8.12	1.02

Note. ^a internal consistency (Cronbach alpha), ^b split-half reliability, ^c retest reliability, ^d parallel test reliability. Split-half reliability for the LPS tests was calculated for the four subtests together. Split-half reliabilities of TIP data vary depending on the image-library used. Values for the CTB are standardized and detection performance measures of all x-ray screening tests except for the X-Ray ORT have been multiplied with an arbitrary constant due to security reasons. * Reliability measures for the X-Ray ORT were based on test results from novices.

The measurement model with the 2006 sample revealed that all factor loadings on the two constructs ability and detection performance were substantial and significant (see Figure 1). The covariance between ability and detection performance was 0.018 ($SE = 0.006$), $p < .01$, corresponding to a correlation of $r = .38$. According to [16], [17] the model fit was good and should not be rejected $\chi^2(89, N = 169) = 125.72$, $p < .01$, CFI = .952, RMSEA = .050, AIC = 187.72 (saturated model 240.00), BCC = 194.24 (saturated model 265.26), BIC = 284.74 (saturated model 615.59), CAIC = 315.74 (saturated model 735.59). As indicated by the goodness-of-fit indices, the model for the 2007 sample reproduced the covariance matrix as well very well $\chi^2(89, N = 97) = 98.99$, $p = .22$, CFI = .981, RMSEA = .036, AIC = 160.99 (saturated model 240.00), BCC = 174.58 (saturated model 292.60), BIC = 238.48 (saturated model 539.98), CAIC = 269.48 (saturated model 659.98). Covariance between the two constructs ability and detection performance was 0.027 ($SE = 0.007$), $p < .01$ and the correlation significant ($r = .57$) respectively (Figure 1). In both

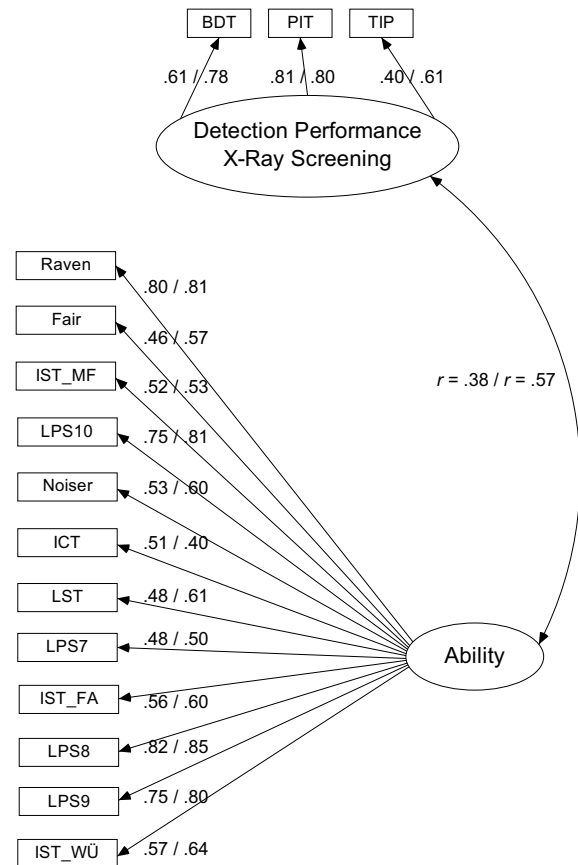


FIGURE 1. Factor model with the two factors ability and detection performance in x-ray screening (circles) and the 15 indicators. For clarity measurement errors are omitted. Standardized loadings are indicated for the 2006 (left) and the 2007 data (right).

models (2006 and 2007) no substantial modifications were required.

To test whether the number of tests can be reduced without losing information, we tested the model with the four indicators Raven, LPS8, LPS9 and LPS10. As can be seen in Figure 1 these tests showed the highest loading on the first order factor ability in both groups. Results evidenced a satisfactory model fit for the 2006 data $\chi^2(13, N = 169) = 24.38$, $p < .05$, CFI = .971, RMSEA = .072, AIC = 54.38 (saturated model 56.00), BCC = 55.88 (saturated model 58.80), BIC = 101.33 (saturated model 143.64), CAIC = 116.33 (saturated model 171.64) and a very good fit for the 2007 data $\chi^2(13, N = 97) = 7.38$, $p = .88$, CFI = 1.00, RMSEA = .000, AIC = 37.38 (saturated model 56.00), BCC = 40.34 (saturated model 61.53), BIC = 74.87 (saturated model 126.00), CAIC = 89.87 (saturated model 154.00). By reducing the number of indicators of ability from 12 to 4 indicators, no difference in the substantive results were found, especially in the prediction of the detection performance in x-ray screening.

In order to test what part of the detection performance can be accounted for by the theoretical variables, we performed a

full structural equation model analysis, but with the four indicators Raven, LPS8, LPS9 and LPS10 only. Besides the latent variable ability of screeners the test result in the X-Ray ORT is expected to account for a part of the detection performance variability. Again, the SEM was first conducted for the 2006 sample and then for the 2007 data. The model fit indicated with $\chi^2(17, N = 169) = 27.38, p = .05, CFI = .976, RMSEA = .060, AIC = 65.38$ (saturated model 72.00), $BCC = 67.53$ (saturated model 76.08), $BIC = 124.85$ (saturated model 184.68), $CAIC = 143.85$ (saturated model 220.68) a good fit¹. Further, results showed a very good model fit for the 2007 sample $\chi^2(18, N = 97) = 10.82, p = .90, CFI = 1.00, RMSEA = .000, AIC = 46.82$ (saturated model 72.00), $BCC = 50.87$ (saturated model 80.10), $BIC = 912.82$ (saturated model 161.99), $CAIC = 109.82$ (saturated model 197.99). Thus, in both groups, ability and the X-Ray ORT display a significant effect on detection performance in x-ray screening.

IV. DISCUSSION

The main goal of this study was to examine which of the 12 general visual cognition tests predict on the job performance best in order to define a reliable and valid pre-employment assessment. Therefore, 12 tests which best match the x-ray screening task were used. To measure on the job performance, test results in the PIT and the BDT as well as TIP data were used. The PIT and BDT are two x-ray screening tests that were part of the recurrent competency assessment. TIP data were measured on the checkpoint and thus on the job performance could be evaluated.

Results revealed that all cognition tests from the visual CTB which are mostly tests from elaborated German intelligence test batteries load on one latent factor ability despite their semantic distinctions. Furthermore, this factor correlates highly with detection performance in x-ray screening for both samples. Reliability of the 2006 sample which was just sufficient may account for the generally worse model fit of the 2006 data compared to the 2007 sample. Our results also suggest that the whole visual CTB which consists of 12 tests can be reduced to four tests without reducing explained variance. Further a full SEM with the X-Ray ORT as additional factor showed that both factors ability and the X-Ray ORT display a significant effect on detection performance. Interestingly, as well the X-Ray ORT which measures the ability to cope with image-based factors in x-ray screening seems to be an important determinant. It has to be considered that the sample used for this study shows relatively small variance as all screeners were already selected based on their ability to cope with image-based factors. Whether ability is even more important is a question that should be answered with a representative sample.

To sum up this study showed that both the ability to cope with image-based factors measured with the X-Ray ORT and the ability measured with the visual CTB play an important role for the x-ray screening task later on the job. The positive relationship between the X-Ray ORT and detection

performance later on the job could also be shown in a previous study by [3]. Thus, the X-Ray ORT as well as the visual CTB can be used within a pre-employment assessment. However, to increase efficiency a reduction of the visual CTB from 12 to 4 tests only should be taken into consideration.

Further analysis should investigate whether other factors relevant for the x-ray screening job could be subjected to SEM, such as training hours, age, personality traits etc. According [20] as well as [18] it could be expected that training hours influence detection performance on the job remarkably. Further studies investigating the effect of age on x-ray screening showed as well a significantly worse detection performance of older screeners compared to younger ones despite their working experience [19].

REFERENCES

- [1] Schwaninger, A., Hardmeier, D., & Hofer, F. (2005). Aviation security screeners visual abilities & visual knowledge measurement. *IEEE Aerospace and Electronic Systems*, 20, 29-35.
- [2] Hardmeier, D., Hofer, F., & Schwaninger, A. (2005). The object recognition test (ORT) – a reliable tool for measuring visual abilities needed in x-ray screening. *IEEE ICCST Proceedings*, 39, 189-192.
- [3] Hardmeier, D., Hofer, F., & Schwaninger, A. (2006a). Increased detection performance in airport security screening using the x-ray ort as pre-employment assessment tool. *Proceedings of the 2nd International Conference on Research in Air Transportation, Belgrade, Serbia and Montenegro, June 24-28, 2006*, 393-397.
- [4] Horn, W. (1983). *Leistungsprüfungssystem (L-P-S). Handanweisung für die Durchführung, Auswertung und Interpretation*. 2., erweiterte und verbesserte Auflage. Göttingen: Hogrefe.
- [5] Amthauer, R., Brocke, B., Liepmann, D., & Beauducel, A. (2001). *Intelligenz-Struktur-Test 2000 R (I-S-T 2000 R), Manual*. Göttingen: Hogrefe.
- [6] Raven, J.C., Court, J., & Raven, J. Jr. (1980). *RAVEN-Matrizen-Test. Manual, deutsche Bearbeitung von Heinrich Kratzmeier unter Mitarbeit von Ralf Horn, Manual*. Weinheim: Beltz Test Gesellschaft.
- [7] Moosbrugger, H., & Oehlschlägel, J. (1996). *Frankfurter Aufmerksamkeits-Inventar: FAIR, Testmanual*. Bern: Huber.
- [8] Marxer, P. (2004). *Visual tests for x-ray screening*. Unpublished master's thesis, University of Zurich, Zurich, Switzerland.
- [9] Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- [10] MacMillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. Cambridge: University Press.
- [11] Koller, S., & Schwaninger, A. (2006). Assessing X-ray image interpretation competency of airport security screeners. *Proceedings of the 2nd International Conference on Research in Air Transportation, Belgrade, Serbia and Montenegro, June 24-28, 2006*, 399-402.
- [12] Hofer F., & Schwaninger, A. (2005). Using threat image projection data for assessing individual screener performance. In C.A.Brebbia, T. Bucciarelli, F. Garzia, & M.Guarascio, *WIT Transactions on the Built Environment (82), Safety and Security Engineering* (pp. 417-426). Wessex: WIT Press.
- [13] Bentler, P. M. (1992). On the fit of models to covariances and methodology to the Bulletin. *Psychological Bulletin*, 112, 400-404.
- [14] Arbuckle, J. L. (2005). *Amos 6.0 users guide*. Chicago, IL: SPSS.
- [15] Byrne, B. (2001). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. New Jersey: Erlbaum.
- [16] Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55.
- [17] Marsh, H. W., Hau, K. T. & Wen, Z. (2004). In search of golden rules: Comment on hypothesis testing approaches to setting cutoff values for

¹ The TIP data for the 2006 sample showed the lowest factor loadings. Nevertheless we tried to integrate this indicator because of his importance to the latent variable detection performance in x-ray screening.

fit indexes and dangers in overgeneralising Hu & Bentler's (1999) findings. *Structural Equation Modelling*, 11, 320-341.

- [18] Koller, S. M., Hardmeier, D., Michel, S., & Schwaninger, A. (in press). Investigating training, transfer and viewpoint effects resulting from recurrent CBT of x-ray image interpretation. *Journal of Transportation Security*.
- [19] Schwaninger, A., Hardmeier, D., Riegelning, J., & Martin, M. (2008). *Use it and still lose it: The influence of age and job experience on detection performance in x-ray screening*. Manuscript in preparation.
- [20] Hardmeier, D., Hofer, F., & Schwaninger, A. (2006b). The role of recurrent CBT for increasing aviation security screeners' visual

knowledge and abilities needed in x-ray screening. *Proceedings of the 4th International Aviation Security Technology Symposium, Washington, D.C., USA, November 27 – December 1, 2006*, 338-342.

ACKNOWLEDGMENT

This research was financially supported by the European Commission Leonardo da Vinci Programme (VIA Project, DE/06/C/F/TH-80403). Thanks to Zurich State Police, Airport Division for supporting this study by supplying screeners and data.

APPENDIX

TABLE II. CORRELATION MATRIX OF INDICATORS FOR 2006 SAMPLE

Indicators	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. ORT	---														
2. LST	.20	---													
3. Noiser	.11	.36	---												
4. LPS10	.15	.35	.45	---											
5. IST_MF	.16	.18	.47	.41	---										
6. Raven	.09	.36	.37	.64	.38	---									
7. Fair	.14	.16	.30	.31	.38	.36	---								
8. LPS9	.05	.33	.33	.62	.34	.60	.36	---							
9. IST_WÜ	.10	.24	.27	.33	.27	.52	.32	.40	---						
10. IST_FA	.05	.36	.26	.41	.25	.43	.18	.47	.41	---					
11. LPS8	.10	.39	.42	.59	.44	.66	.40	.64	.49	.44	---				
12. ICT	.11	.30	.30	.38	.26	.44	.20	.37	.27	.17	.45	---			
13. LPS7	.08	.21	.23	.38	.22	.43	.18	.34	.29	.33	.36	.22	---		
14. TIP	-.17	.09	.14	.19	.16	.28	.21	.26	.12	.15	.21	.13	.20	---	
15. PIT	.30	.34	.20	.19	.17	.13	.09	.19	.08	.18	.21	.12	.18	.34	---
16. BDT1.0	.35	.26	.19	.15	.15	.23	.16	.13	.18	.25	.18	.14	.21	.15	.50

TABLE III. CORRELATION MATRIX OF INDICATORS FOR 2007 SAMPLE

Indicators	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. ORT	---														
2. LST	.23	---													
3. Noiser	.17	.39	---												
4. LPS10	.18	.51	.50	---											
5. IST_MF	.12	.39	.35	.40	---										
6. Raven	.15	.45	.52	.65	.44	---									
7. Fair	.12	.32	.27	.51	.44	.53	---								
8. LPS9	.24	.54	.48	.64	.44	.69	.43	---							
9. IST_WÜ	.17	.25	.29	.49	.26	.54	.31	.52	---						
10. IST_FA	.10	.39	.34	.45	.37	.44	.39	.47	.42	---					
11. LPS8	.27	.54	.50	.70	.39	.67	.45	.65	.64	.54	---				
12. ICT	-.00	.27	.36	.30	.32	.23	.21	.30	.24	.29	.39	---			
13. LPS7	.33	.26	.35	.42	.28	.40	.27	.39	.31	.27	.43	.11	---		
14. TIP	.07	-.05	.13	.30	.07	.33	.25	.26	.24	.12	.26	-.03	.35	---	
15. CAT	.19	.26	.37	.40	.16	.36	.14	.39	.24	.20	.38	.32	.38	.50	---
16. BDT2.0	.13	.29	.26	.34	.23	.40	.21	.33	.35	.22	.42	.10	.33	.48	.62

The Impact of Image Based Factors and Training on Threat Detection Performance in X-ray Screening

Adrian Schwaninger*, Anton Bolfing*, Tobias Halbherr[†], Shaun Helman[‡], Andrew Belyavin[‡] and Lawrence Hay[‡]

*Max Planck Institute for Biological Cybernetics, Tübingen, Germany, and
Department of Psychology, University of Zurich, Switzerland
e-mail: a.bolfing@psychologie.uzh.ch, a.schwaninger@psychologie.uzh.ch

[†]Department of Psychology, University of Zurich, Switzerland, Email: t.halbherr@psychologie.uzh.ch

[‡]QinetiQ Limited, Hampshire, United Kingdom

Abstract—In this study, two experiments are reported which investigated the relative importance of five different image based factors and one human factor (training) in mediating threat detection performance of human operators in airport security x-ray screening. Experiment 1 was based on a random sample of roughly 16'000 records of threat image projection (TIP) data. TIP is a software function available on state-of-the-art x-ray screening equipment that allows the projection of fictional threat images (FTIs) into x-ray images of passenger bags during the routine baggage screening operation. Analysis of main effects showed that image based factors can substantially affect screener detection performance in terms of the hit rate (identification of FTIs). There were strong effects of FTI view difficulty (rotation of FTIs) and superposition of FTIs by other objects in the x-ray image of a passenger bag. The amount of opacity in the x-ray image of a passenger bag had a small although significant effect on detection performance. The two image based factors clutter and bag size did not have a significant effect.

Experiment 2 was conducted using an offline-test in order to provide controlled and more detailed data for analyzing the image based factors from Experiment 1, as well as the human factor of training. In particular the individual factors' main effects on detection performance, main effects of all factors taken together and factor interactions were analyzed. In the test design the following image-based factors were varied systematically: Threat (FTI) category (guns, knives, improvised explosive devices, other threats), view difficulty, superposition, bag complexity (a combination of opacity and clutter) and bag size. Data were collected from 200 screening officers at five sites across Europe. For screener training all five sites use the same computer-based training system. Consistent with the results obtained in Experiment 1, there were large main effects of threat (FTI) category, view difficulty, and superposition. Again consistent with Experiment 1, effects of bag complexity (opacity and clutter) and bag size were much smaller. In addition to Experiment 1, the number of computer based training (CBT) hours was available for each security officer participating in the study. Training turned out to be a key driver to improving threat detection performance in x-ray screening and seemed to mediate the effects of some image based factors.

Possible implications regarding the enhancement of human-machine interaction in x-ray screening are discussed.

I. INTRODUCTION

Screening passenger bags for threat items using state-of-the art x-ray machines is an essential component of airport security. Previous work (Schwaninger, 2003b, Schwaninger, Hardmeier, & Hofer, 2005, and Schwaninger, Michel, &

Bolfing, 2007) has identified image based factors that affect human performance in x-ray screening tasks: object view difficulty, superposition by other objects and bag complexity (opacity and clutter). Recently the question has been raised whether bag size could be another image based factor that affects detection of threat items when visually inspecting x-ray images of passenger bags. In this study we determined effects and interactions of image based factors and human factors (amount of recurrent computer-based training). In addition, with empirically based conclusions regarding the importance of the bag size variable, by itself as well as in relation with other performance relevant factors, this study provided the scientific basis for a political decision making process regarding the improvement of aviation security.

Two experiments are reported. Experiment 1 is based on threat image projection (TIP) data. Experiment 2 is based on an off-line computer based test, which allows investigating the combined effects of image-based factors, effects of training as well as factor interactions. The use of these two methods to answer the same research question will ensure that the overall approach is complementary. Both methods have their own strengths and weaknesses: TIP data give high ecological validity but low experimental control; off-line computer based tests using controlled stimuli allow more experimental control, but less ecological validity. If both methods provide the same answer to the research question, this can be taken as stronger evidence that the findings are genuine, and not simply an artefact of the particular method used.

The two experiments both follow the paradigm using computer algorithms to estimate image based factors that influence threat detection performance in x-ray screening. This paradigm was developed at University of Zurich and presented at ICRAT 2006 in Belgrade (Bolfing, Michel, & Schwaninger, 2006a) and published before (Schwaninger, Michel, & Bolfing, 2007; Bolfing, Michel, & Schwaninger, 2006b; Schwaninger, Michel, & Bolfing, 2005). None of these papers used TIP data for analysis, which ensures high ecological validity. Experiment 2 is based on a much larger data set than the previous studies augmenting reliability. The inclusion of bag size and training as additional factors is completely novel within this paradigm. Since threat detection performance in aviation security x-ray

screening depends on the x-ray images but also on the human screeners—the final decision makers—human factors should not be neglected in a comprehensive model whose goal is to explain the x-ray threat detection process.

A. Image Based Factors

Schwaninger (2003b) and Schwaninger, Hardmeier, and Hofer (2005) have identified three image based factors which affect threat detection by x-ray screeners: view difficulty, superposition, and bag complexity (see figure 1).

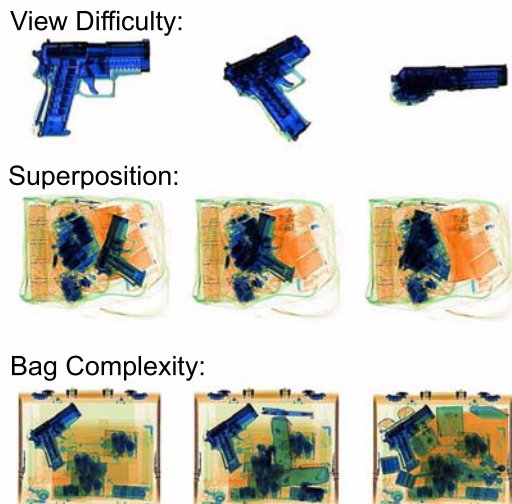


Fig. 1. Illustration of the three basic image based factors suggested by Schwaninger (2003b) and Schwaninger, Hardmeier, and Hofer (2005)

The concepts of these image based factors have been mathematically modeled (Schwaninger, Michel, & Bolfling, 2007, see Bolfling, & Schwaninger, 2007 for the latest version). View difficulty is modeled as a statistically calculable value between 0 and 1 named FTI view difficulty. Superposition and bag complexity are modeled as image processing measurements with bag complexity being split up into clutter and opacity. The introduction of the image based factor bag size in this study necessitated normalization of earlier implementations of clutter and opacity regarding bag size. Formulae and short descriptions of the underlying concepts are specified in Bolfling, & Schwaninger (2007).

II. THREAT IMAGE PROJECTION (TIP) χ^2 ANALYSIS: EXPERIMENT 1

A. Method

1) *Threat Image Projection (TIP) Data:* In order to ensure high ecological validity, we decided to analyze data from threat image projection (TIP). TIP is a software function of state-of-the-art x-ray screening equipment used at security checkpoints in airports, nuclear power plants, navigation docks etc. In aviation security TIP distinguishes between cabin baggage screening (CBS) and hold baggage screening (HBS). In CBS, guns, knives, improvised explosive devices (IEDs) and other threats are subject to identification and

confiscation. In HBS, the focus rests mainly on IEDs and dangerous goods such as gasoline containers or diver lamps. The current investigation is confined to CBS. In CBS TIP, fictional threat items (FTIs) are occasionally projected into x-ray images of passenger bags during the routine baggage screening operation. A sufficiently large sample of TIP events allows statistically reliable measurements of detection performance of human operators (x-ray screeners) on-the-job (Hofer & Schwaninger, 2005) and thus with high ecological validity.

The data basis of this study consists of a random sample of 16'329 TIP events that have been routinely recorded on-the-job with approximately 700 professional x-ray screeners throughout the first half of 2007 at a large European airport. We decided to apply χ^2 analyses to each image based factor separately to measure its impact on detection performance in terms of hit rate (i.e. correctly judging a bag as being NOT OK).

2) χ^2 Analysis: To compare the effects on detection performance of the independent variables¹ FTI view difficulty, superposition, opacity, clutter and bag size, the following procedures were applied to the TIP data described above. A histogram was created for each independent variable (image based factor). For each variable the upper and lower 2.5% of the cases in the data were excluded to remove outlier data from the analysis. Furthermore this made possible the definition of five equidistant bins with at least 100 data points each (TIP events).

Hit rates were calculated for each of the five equidistant bins to run χ^2 tests with the null hypothesis H_0 that the hit rates are equal across bins. Effect size analysis based on Cohen (1988) was used to compare the effect sizes of the different independent variables. For detailed information on χ^2 statistics see for example Coolican (2004).

B. Results

The results below are listed separately for each image based factor introduced above (see Bolfling, & Schwaninger, 2007 for further information and formulae). Each of the following subsections begins with a graphical illustration of the image based factors' effects on the threat detection performance measure hit rate. The x -axes show the five equidistant bins into which the whole data range was subdivided. Low values are on the left, high values on the right. The y -axes show the hit rates of the image based factors' bins. For reasons of confidentiality hit rates cannot be given explicitly, but the hit rate scales are reasonably chosen and kept constant throughout the whole document.

Following the graphical illustrations (figures 2-6), statistical test values are given in tables I-V. χ^2 statistics can be interpreted as follows: the larger the $\chi^2(df, N)$ value the larger the effect. Additionally χ^2 effect sizes w are given.

¹The variables correspond to the continuously represented variables used in the multiple regression analysis in Experiment 2 (see figure 8)

Again, the larger the effect size, the larger the effect. However, please be aware that χ^2 and w values do not state the direction of the effect.

To summarize the χ^2 analysis results a bar plot graphic is provided at the end of this section illustrating the χ^2 effect sizes of the five image based factors on the hit rate (see figure 7). The image based factors are arranged such that their effects decrease in size.

1) *FTI View Difficulty*: Figure 2 illustrates the large impact of FTI view difficulty on human detection performance in terms of hit rate. This is partly due to the fact that objects are more difficult when depicted from an unusual viewpoint (see figure 1). Other factors contributing to this large impact are the threat category of the object and the training of human operators (see Experiment 2).

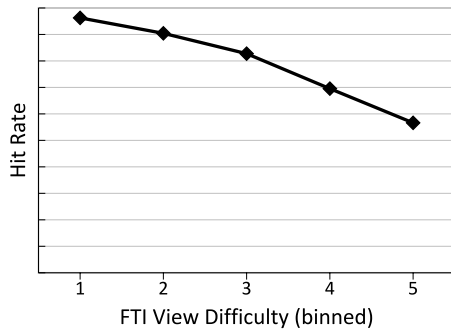


Fig. 2. Illustration of the impact of FTI view difficulty on hit rate.

TABLE I
 χ^2 ANALYSIS RESULTS: FTI VIEW DIFFICULTY

χ^2 value	$\chi^2(4, N = 13'541) = 198.04$
Significance	Highly significant: $p < .001$
χ^2 effect size	$w = .12$

2) *Superposition*: Figure 3 illustrates the large effect of superposition on detection performance.

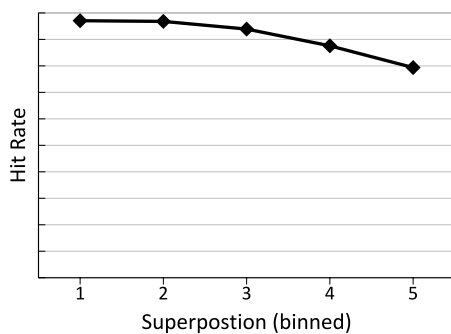


Fig. 3. Illustration of the impact of superposition on hit rate.

3) *Opacity*: Figure 4 shows the significant but relatively small influence of opacity on detection performance in terms of hit rate.

TABLE II
 χ^2 ANALYSIS RESULTS: SUPERPOSITION

χ^2 value	$\chi^2(4, N = 13'713) = 72.98$
Significance	Highly significant: $p < .001$
χ^2 effect size	$w = .07$

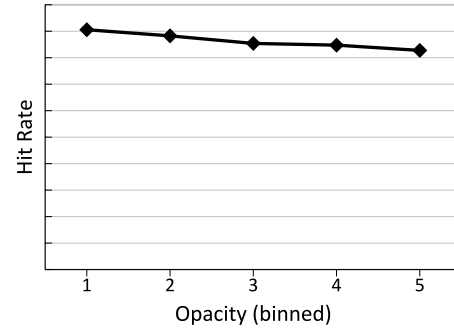


Fig. 4. Illustration of the impact of opacity on hit rate.

TABLE III
 χ^2 ANALYSIS RESULTS: OPACITY

χ^2 value	$\chi^2(4, N = 13'718) = 9.90$
Significance	Significant: $p < .05$
χ^2 effect size	$w = .03$

Here the question arises whether it is opacity as a perceptual concept that does not have much influence on threat detection performance, or whether the image measurement formula of opacity is not properly modeled.

4) *Clutter*: Figure 5 illustrates the hit rates of the five clutter bins. There is no significant effect of clutter on detection performance. As with opacity, the question arises whether it is the concept of clutter that does not influence hit rates in TIP, or whether the computational model of clutter needs to be improved.

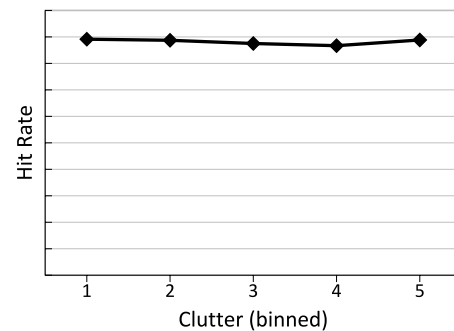


Fig. 5. Illustration of the impact of clutter on hit rate.

5) *Bag Size*: Figure 6 shows the effect of bag size on hit rate in TIP.

As with clutter, the effect of bag size on detection performance does not reach statistical significance.

TABLE IV
 χ^2 ANALYSIS RESULTS: CLUTTER

χ^2 value	$\chi^2(4, N = 13'726) = 0.98$
Significance	Not significant: $p = .913$
χ^2 effect size	$w = .01$

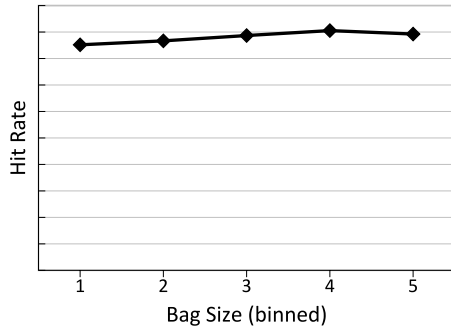


Fig. 6. Illustration of the impact of bag size on hit rate.

TABLE V
 χ^2 ANALYSIS RESULTS: BAG SIZE

χ^2 value	$\chi^2(4, N = 13'758) = 4.45$
Significance	Not significant: $p = .348$
χ^2 effect size	$w = .02$

eliminating certain bags (e.g. small bags rather than large bags) from the TIP image set, and thus reducing their presence. Secondly, it is not always clear how closely aligned TIP scores are with the specific operational situations encountered when threats are deliberately hidden in difficult bags. But most importantly, in Experiment 1 only main effects were analyzed. In order to gain a more complete picture it is important to conduct a more controlled experiment in which main effects in combination and their interactions can be measured reliably. This was conducted in Experiment 2.

III. OFF-LINE COMPUTER BASED TEST:
EXPERIMENT 2

A. Method

1) *Participants*: 200 X-ray screeners from five European sites with varying amounts of training in x-ray image interpretation.

2) *Stimuli*: The stimuli were 1024 complete threat images (CTIs) and 1024 complete non-threat images (CNTIs). CTIs were created by projecting fictional threat items (FTIs) into 1024 X-ray images of bags. FTIs for the study were eight visually similar pairs of each of four types of threat items: guns, knives, improvised explosive devices (IEDs), and 'other' threats. Images of cabin baggage were captured from x-ray machines at a European airport using the auto-archive function. The images were revised by three airport security supervisors to remove inappropriate images (e.g. images containing more than one bag, images containing incomplete bags, bags containing prohibited items or liquids, etcetera). This procedure resulted in 7606 bag images. Additional review by the QinetiQ team resulted in a total of 6659 bag images from which the 1024 bags needed for the study were drawn. The final 1024 bags used for the study were chosen through a process of projecting the relevant FTIs into the bags such that the variables of interest would be orthogonal in the stimulus set. Several full sets of 2048 images (the 1024 images containing the FTIs, and the same images without FTIs) were created. The one with the most desirable properties in terms of variable orthogonality was chosen for use in the study.

3) *Design*: The study employed a 4 (FTI category: guns, knives, IEDs, other) x2 (view difficulty: easy, difficult) x2 (superposition: low, high) x2 (bag complexity: low, high) x2 (bag size: small, large) x2 (image type: FTI, no FTI) within-participants design. Since there were 16 FTIs in each category, this design results in a total of $16 \times 4 \times 2 \times 2 \times 2 \times 2 = 2048$ images which were to be presented to the screeners. The images

6) *Comparison of the χ^2 Effect Sizes*: In figure 7, the effect sizes w are compared. The factor FTI view difficulty has the highest effect size with $w = .12$, while clutter shows the lowest effect size with $w = .01$. The factors opacity, bag size and clutter show small effect sizes. The effects of clutter and bag size did not reach statistical significance.

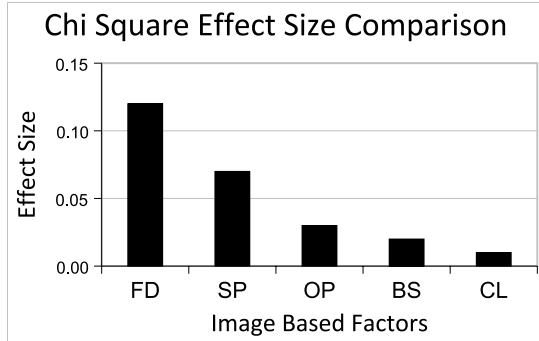


Fig. 7. Comparison of the effect sizes among the image based factor.

C. Discussion

The results obtained in Experiment 1 are consistent with earlier findings. Schwaninger, Hardmeier, and Hofer (2005) found that viewpoint, superposition and bag complexity affect screener performance. Schwaninger, Michel, and Bolting (2007) replicated these results. Using similar image measurements as in Experiment 1, they measured similar effects for FTI view difficulty, superposition, opacity (negatively correlated with transparency in Schwaninger et al., 2007) and clutter. However, several caveats are necessary to qualify the appropriateness of the results obtained in Experiment 1. Firstly, an analysis of auto-archive bags indicated that, as would be anticipated, it is likely that TIP aborts are selectively

were presented to the screeners in a random order in multiple testing sessions of 20 minutes each. As dependent variable the detection performance measure d' (Green & Swets, 1966) was used. This measure provides a more valid estimate of detection performance than the hit rate alone because it takes the hit rate and the false alarm rate into account (see Hofer & Schwanger, 2004 for different measures of x-ray detection performance). Since the off-line test showed each bag once with a threat and once without one, accurate measurements of hit and false alarm rates could be obtained.

B. Results

Data were analyzed in two ways. Firstly, by treating the variables FTI view difficulty, superposition, opacity, clutter, and bag size as continuous, a linear regression was employed to assess the main effects of each image based factor on threat detection performance separately. A multiple linear regression was used to examine the main effects together. Additionally, we calculated a linear regression with hours of recurrent computer based training prior to testing as predictor. In order to examine main effects as well as interactions between the variables, the discrete variables FTI category, view difficulty, superposition, bag complexity and bag size were used in an analysis of covariance (ANCOVA). Training hours served as covariate in the ANCOVA. Figure 8 shows the way in which the continuous and discrete variables are related to each other. Due to a high inter-correlation and a test design that demands independence of its variables, opacity and clutter were encoded into the single discrete variable bag complexity. FTI category and view difficulty were encoded into a single continuous variable because it is not sensible to encode either variable directly into a continuous variable. Instead we defined the variable FTI view difficulty as the difficulty-as measured in threat detection performance (d')-screening officers had in solving a specific threat item in a specific view (easy or difficult) across all other conditions (i.e. superposition, bag complexity and bag size).

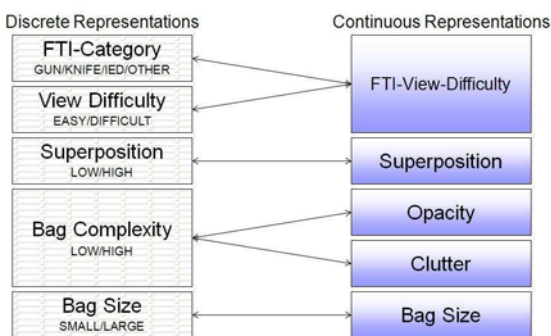


Fig. 8. Illustration of relationship between discrete and continuous representations of variables

1) *Linear Regression and Multiple Linear Regression:* The regression analyses will help us understand the direct relationship between image based factors and d' , as well as training hours and d' . Figure 9 shows the relative effect sizes,

the absolute values of the correlations with the dependent variable d' , for the individual variables. For superposition and training hours a logarithmic transformation was applied. This transformation was necessary in order to achieve a linear relationship between superposition and detection performance d' . With .70, .63 and .58, FTI view difficulty, training hours and superposition all have very high effect sizes. Opacity has a moderate to small effect size with .22, clutter and bag size have very small effect sizes with .05 and .07, respectively. Except for clutter, all correlations are statistically significant.

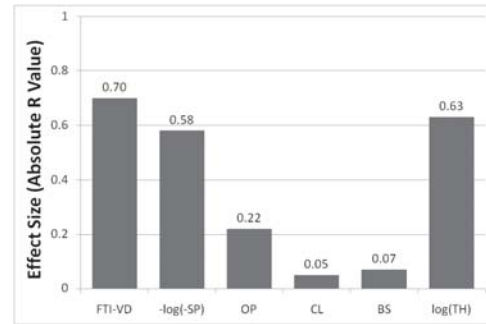


Fig. 9. Illustration of effect sizes R

Figures 10 and 11 show the results of the multiple linear regression with all image based factors: FTI view difficulty, superposition (logarithmically transformed), opacity, clutter and bag size. It shows the overall effect size, again the absolute value of the correlation R, of all the image based factors taken together. With $R = 0.77$ the effect size is very high. The effect size of the only human factor analyzed (hours of recurrent computer based training), with $R = 0.63$, is also large. We can see that in the multiple linear regression model the factor bag size is the only one not reaching statistical significance. Put another way: In the presence of the other image based factors bag size did not lead to a statistically significant change in detection performance in our experiment. As shown in figure 12 adding bag size to the linear model only leads to a minimal increase of its effect size from $R = 0.772$ to $R = 0.773$.

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.773 ^a	.597	.595	.690

^a. Predictors: (Constant), BS, FTIViewDifficulty, OP, -LOG(-SP), CL

Fig. 10. Multiple linear regression overview

2) *ANCOVA:* A repeated measures analysis of covariance (ANCOVA) was conducted to analyze the main effects of image based factors, their interactions and their interactions with training. As can be seen in the main effects summary of figure 13 the repeated measures ANCOVA leads to only a slightly different pattern with regards to effect sizes than the linear regression analyses. These differences are due to the fact that, in contrast to the linear regression models, in the

Coefficients ^a						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-2.232	.314		-7.104	.000
	FTI/ViewDifficulty	.755	.033	.568	23.163	.000
	-LOG(-SP)	-.822	.089	-.227	-9.195	.000
	OP	-10.243	.860	-.366	-11.917	.000
	CL	23.505	3.370	.223	6.975	.000
	BS	1.51E-006	.000	.030	1.304	.193

^a . Dependent Variable: d

Fig. 11. Multiple linear regression details

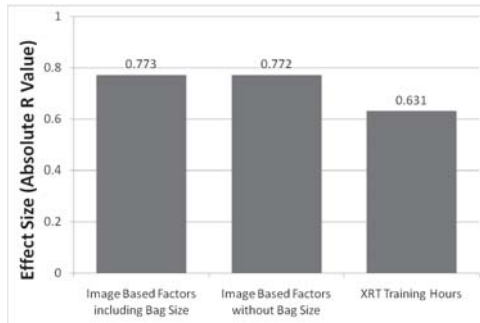


Fig. 12. Combined effect size of image based factors and effect size of training

ANCOVA analysis effects of the covariate training hours are isolated from the effects of image based factors. Furthermore, in the ANCOVA inter-individual differences between screening officers ('screener variance') are taken into account. Superposition shows the largest effect size (η^2), followed by FTI category, bag complexity and view difficulty. The main effect of bag size is clearly smaller than the main effect of any other image based factor. Training hours has noteworthy interactions with FTI category and view difficulty. These interactions make sense, since we know from other studies that training can lead to comparatively larger performance increases for items that are comparatively difficult for novices (Koller, Hardmeier, Michel, & Schwaninger, in press)-for example improvised explosive devices (threat item category) or difficult views (view difficulty). There is also a small interaction of training with bag size, indicating that well trained screening officers are less affected by effects of bag size. Figure 14 gives an overview of the 10 largest interactions in the ANCOVA. All in all over 30 interactions reached statistical significance. Since the effect sizes of most interactions are very small we decided only to report interactions $\eta^2 \geq .07$. The interaction of view difficulty with threat category can at least partly be explained by the fact that detection performance of improvised explosive devices-unlike guns or knives-is largely independent of viewpoint. The interaction of superposition with view difficulty indicates that with difficult viewpoints superposition plays a larger role in determining detection performance than with easy views. The interaction of superposition with threat category indicates that some threat item categories are more sensitive to superposition than others. For example, from the regression analysis above we know that superposition effects are higher with knives than with guns.

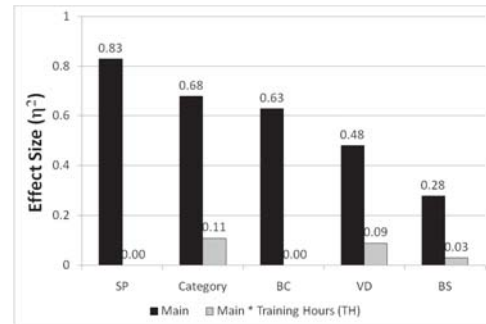


Fig. 13. Illustration of ANCOVA main effects and interactions with the covariate training hours

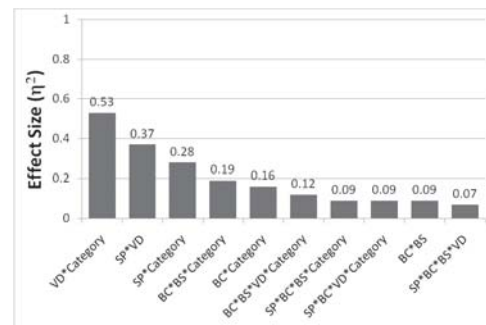


Fig. 14. Illustration of the the ten largest ANCOVA interactions

C. Discussion

With an overall correlation of .77 the linear modeling of detection performance with image based factors has a very high explanatory power. Superposition, although not always with the largest effect size, has shown the most robust effects on detection performance. Interestingly and in contrast to what one might have expected based on the results of the regression analyses, the variable bag complexity (a combination of opacity and clutter) showed a large effect size in the ANCOVA. Apart from this, the ANCOVA results reflect the regression analysis results closely, both in main effects and interactions. Threat category and view difficulty had considerable interactions with the covariate training hours. This shows that training is particularly effective in the case of difficult item categories such as IEDs and for difficult viewpoints. Bag size, although intuitively plausible as relevant factor, turned out to play only a minor role in determining threat detection performance. The same is true for clutter.

IV. GENERAL DISCUSSION

There were large main effects of view difficulty and of FTI difficulty in all of the analyses, as expected. The same was true for superposition and complexity (to a bigger extent for opacity than for clutter). Clearly, these factors need to be taken account of in any future work on performance-relevant image based factors. When looking at the influence on detection performance of all image based factors together, there is no statistically significant effect of bag size. When using a more sophisticated model of data analysis including main effects of

FTI view difficulty, superposition, bag complexity, bag size and the interactions of these variables, there is a small effect of bag size. In Experiment 2 we were also able to examine the effect of the number of CBT training hours on threat detection performance. The key finding from the study is that the effect size for this variable was large, and seemed to mediate the effect of some image based factors on threat detection. Clearly, training is a key driver to improving threat detection performance in x-ray screening, and more work needs to be done to establish exactly which image based factors screeners need to be trained in to give the best improvements in threat detection accuracy.

V. RECOMMENDATIONS FOR IMPROVING

HUMAN-MACHINE INTERACTION IN X-RAY SCREENING

A. FTI View Difficulty and Superposition

The factor FTI view difficulty refers to the fact that the identification of threat objects, as objects in general, is highly dependent on their viewpoint as well as on properties of the very object itself. Current x-ray screening equipment provides only one x-ray image per passenger bag. More recent technology can provide multiple views of a bag. Figure 15 illustrates how such new systems might be able to reduce the detection problems due to view difficulty and superposition. Objects that are superimposed by other objects from one perspective may be clearly visible from another one. Furthermore, training is an important tool in lessening detrimental effects on detection performance of difficult views. Our ANCOVA analysis has supported earlier findings that training leads to particularly large improvements in detection performance for difficult views (Koller, Hardmeier, Michel, & Schwaninger, in press).

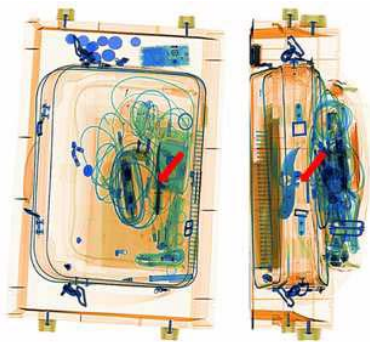


Fig. 15. Illustrative example of how multi-view systems can help improving detection performance in spite of undesirable view difficulty and superposition effects.

B. Opacity

The image based factor Opacity refers to the amount of opaque areas in an x-ray image. X-ray systems with higher penetration have the potential to reduce detection problems due to opacity. In addition, it is possible to implement image measurement algorithms in x-ray equipment that warn the

human operator (x-ray screener) with a "dark alarm", which would be triggered by opaque areas that are deemed too large or dense for unassisted human interpretation. Manual search would follow when a dark alarm was indicated.

C. Screener Selection and Training

A very important approach to face the problem of improving threat detection performance in x-ray screening consists in screener selection and screener training. The psychological literature provides evidence that figure ground segregation (related to superposition) as well as mental rotation (related to view difficulty) are visual abilities that are fairly stable within a person. For example Hofer, Hardmeier, & Schwaninger (2006) and Hardmeier, Hofer, and Schwaninger (2006a) have shown that using computer based object recognition tests in a pre-employment assessment procedure can help to increase detection performance of screeners substantially.

In addition to stable abilities, there are several aspects of visual knowledge relevant to x-ray image interpretation. Knowledge based factors such as knowing which objects are dangerous or prohibited and what they look like in x-ray images are trainable. Training also has beneficial effects on screeners' abilities to deal with certain image based factors. For example, training particularly improves the ability to deal with difficult views. Computer-based training can be a powerful tool to improve x-ray image interpretation competency of screeners (e.g. Koller, Michel, Hardmeier, & Schwaninger, in press; Schwaninger, Hofer & Wetter, 2007; Ghylin, Drury, & Schwaninger, 2006).

ACKNOWLEDGMENT

This research was supported by the UK Department for Transport, the European Civil Aviation Conference Technical Task Force (ECAC TTF), and by the European Commission Leonardo da Vinci Programme (VIA Project, DE/06/C/F/TH-80403). Thanks to Zurich State Police, Airport Division and all other security organizations, companies and airports that supported this study by supplying screeners and data. Special thanks to Olive Emil Wetter and Jonas Sourlier of VICOREG for their valuable contributions in test construction, data analysis, and reporting.

REFERENCES

- [1] S. Koller, D. Hardmeier, F. Hofer, and A. Schwaninger, "Investigating training, transfer, and viewpoint effects resulting from recurrent cbt of x-ray image interpretation," in *Journal of Transportation Security*, in press.
- [2] A. Bolfling and A. Schwaninger, "Measurement formulae for image-based factors in x-ray imagery," in *VICOREG Technical Report, November 26, 2007*.
- [3] A. Schwaninger, S. Michel, and A. Bolfling, "A statistical approach for image difficulty estimation in x-ray screening using image measurements," in *Proceedings of the 4th Symposium on Applied Perception in Graphics and Visualization*, ACM Press, New York, USA, 2007, pp. 123-130.
- [4] D. Hardmeier, F. Hofer, and A. Schwaninger, "The role of recurrent cbt for increasing aviation security screeners' visual knowledge and abilities needed in x-ray screening," in *Proceedings of the 4th International Aviation Security Technology Symposium, Washington, D.C., USA, November 27 - December 1, 2006b*, pp. 338-342.

[5] —, “Increased detection performance in airport security screening using the x-ray ort as pre-employment assessment tool,” in *Proceedings of the 2nd International Conference on Research in Air Transportation, ICRAAT 2006, Belgrade, Serbia and Montenegro, June 24-28, 2006a*, pp. 393–397.

[6] F. Hofer, D. Hardmeier, and A. Schwaninger, “Increasing airport security using the x-ray ort as effective pre-employment assessment tool,” in *Proceedings of the 4th International Aviation Security Technology Symposium, Washington, D.C., USA, November 27 - December 1, 2006*, pp. 303–308.

[7] A. Schwaninger, D. Hardmeier, and F. Hofer, “Aviation security screeners visual abilities & visual knowledge measurement,” in *IEEE Aerospace and Electronic Systems*, vol. 20(6), 2005, pp. 29–35.

[8] J. Bortz, *Statistik für Human- und Sozialwissenschaftler*, 6th ed. Wien, New York: Springer, 2004.

[9] H. Coolican, *Research Methods and Statistics in Psychology*, 4th ed. London: Hodder & Stoughton, 2004.

[10] A. Schwaninger, “Evaluation and selection of airport security screeners,” in *AIRPORT*, vol. 2, 2003b.

[11] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Hillsdale, New Jersey: Lawrence Earlbaum Associates, 1988.

[12] F. Hofer and A. Schwaninger, “Using threat image projection data for assessing individual screener performance,” in *AIRPORT*, vol. 2, 2003b, pp. 417–426.

[13] —, “Reliable and valid measures of threat detection performance in x-ray screening,” in *IEEE ICCST Proceedings*, vol. 38, 2004, pp. 303–308.

[14] D. M. Green and J. A. Swets, *Signal Detection Theory and Psychophysics*. New York: Wiley, 1966.

APPENDIX

For more detailed information on the concepts and mathematical models of the image based factors as well as for examples for a better understanding of the formulae refer to the on-line technical documentation by Bolfig, & Schwaninger (2007) made available at:

http://www.psychologie.uzh.ch/vicoreg/publications/index_byarea.htm

FTI View Difficulty

The general FTI view difficulty equation 1 describes a slight modification of the mean of the inverted detection performance value (DetPerf) over all items (indices N_{OV}) containing the same FTI object (indices O) in the same view (subindices V) as does the item in question. Inverted means here, that the measured detection performance is subtracted from the theoretical maximum detection performance. The slight modification refers to the exclusion of the item in question from averaging.

$$FtiVD_{OVj} = \frac{\sum_{i=1, j \neq i}^{N_{OV}} (max(DetPerf) - DetPerf_{OV_i})}{N_{OV} - 1} \quad (1)$$

For analyzing TIP data the inverted detection performance is the miss rate because usually only bag images containing threat items are recorded. If a large TIP data set is used, the exclusion of the item in question from the averaging can be abandoned due to its very small weight.

$$FtiVD_{OV} = \frac{\sum_{i=1}^{N_{OV}} MissRate_{OV_i}}{N_{OV}} \quad (2)$$

Superposition

Superposition equals the inverted Euclidean distance between the SN images (signal plus noise or threat) and N images (noise or non-threat images) regarding pixel intensity values.

$$SP = C - \sqrt{\sum_{x,y} (I_{SN}(x,y) - I_N(x,y))^2} \quad (3)$$

Clutter

This image based factor is designed to express bag item properties like textural unsteadiness, disarrangement, chaos or just clutter.

The method used in this study is based on the assumption, that such textural unsteadiness can be described mathematically in terms of the amount of high frequency regions.

Equation 4 represents a convolution of the empty bag image (N for noise) with the convolution kernel derived from a high-pass filter in the Fourier space. I_N denotes the pixel intensities of the harmless bag image. \mathcal{F}^{-1} denotes the inverse Fourier transformation. $hp(f_x, f_y)$ represents a high-pass filter in the Fourier space. BS represents bag size (see equation 6). Cut-off frequency f and transition d (the filter’s order) were set to $f = 0.03$ and $d = 11$. The pixel summation on the high-pass filtered image was restricted to the bag’s area.

$$CL = \frac{\sum_{x,y} I_{hp}(x,y)}{BS} \quad (4)$$

where $I_{hp}(x,y) = I_N * \mathcal{F}^{-1}(hp(f_x, f_y))$
 $= \mathcal{F}^{-1}(\mathcal{F}(I_N \cdot hp(f_x, f_y)))$

and $hp(f_x, f_y) = 1 - \frac{1}{1 + \left(\frac{\sqrt{f_x^2 + f_y^2}}{f}\right)^d}$

Opacity

Opacity reflects the extent to which x-rays are able to penetrate objects in a bag. These attributes are represented in x-ray images as different degrees of luminosity. Equation 5 simply implements the number of pixels being darker than a certain threshold (e.g. 64) in the numerator relative to the bag’s overall size (denominator). BS represents the formula of the image based factor bag size (see equation 6).

$$OP = \frac{\sum_{x,y} (I_N(x,y) < 64)}{BS} \quad (5)$$

Bag Size

The bag size formula below is applicable to grayscale images represented by pixel luminosity values between 0 (black) and 255 (white). All pixels with luminosity lower than 254 (near white) are counted and summed up.

$$BS = \sum_{x,y} (I_N(x,y) < 254) \quad (6)$$

Track 6

Airport Operations

Comparison of Data Envelopment Analysis Methods Used in Airport Benchmarking

David Schaar and Lance Sherry, Ph.D.
Dept. of Systems Engineering and Operations Research
George Mason University
Fairfax, Virginia, USA

Abstract—Airport efficiency has been shown to contribute to the overall performance efficiency of the air transportation network. Airport performance efficiency is benchmarked annually and widely published. These benchmarks use several techniques, including several Data Envelopment Analysis (DEA) methods.

This paper examines the differences in results using three DEA methods (Cooper-Charnes-Rhodes (CCR), Banker-Charnes-Cooper (BCC), and Slacks-Based Measure of efficiency (SBM)) on data from 45 airports from 1996 to 2000. The results from the three DEA methods yielded wide variation in results. For example, the CCR analysis showed that efficiencies degraded from small to medium to large airports. The BCC analysis showed no significant difference in efficiency among the three classes of airports. The SBM analysis yielded degraded efficiency from large to medium to small airports. The implications of these results on the use of DEA in benchmarking and the need for guidelines for selection of DEA models and the interpretation of DEA results is discussed.

Keywords—Data Envelopment Analysis; CCR; BCC; SBM; airport efficiency

I. INTRODUCTION

Airport congestion is growing and traveler satisfaction is dropping [1], yet demand is expected to grow at major hub airports [2].

In this environment, the operational efficiency of airports is one of the important determinants of the system's future success. An important component to achieving operational efficiency is to use performance measurement to understand which airports are performing well and which are underperforming and also to understand what the drivers are behind good and bad performance, all with the ultimate goal of improving performance in the areas that are deficient.

As we will show in this paper, much has been written in the academic literature about measuring airport efficiency but several opportunities for improvement exist in this work.

One such opportunity relates to the theoretical foundations for evaluating performance: A number of studies

have been conducted where airports are ranked against each other. However, in these studies there has been limited focus on the selection of the evaluation models used to identify good and bad performers. Several choices exist in this area and a study of each model is necessary to understand which is the most appropriate in a given situation.

This paper will highlight this issue through the use of an example. We will re-run an earlier analysis of airport performance and highlight the impact of the choices in the areas listed above.

II. BACKGROUND

A. General Background

Performance benchmarking as a management tool is frequently cited as having been pioneered by Robert C. Camp at the Xerox Corporation in the late 1970's [7],[8] where he used this tool to identify shortcomings of that company's photocopier production and distribution. He used benchmarking to identify several key practices from the photocopier industry as well as from other industries, all serving to ultimately improve the company's photocopier business. The performance measurement and benchmarking techniques have all evolved significantly since that time, but the ultimate goals for any benchmark remain the same: To identify performance gaps and to find practices that will help close that gap.

Benchmarking has since been applied both in industry and academia in numerous studies. Dattakumar and Jagadeesh [6] conducted an extensive literature review and found more than 350 publications on the topic as of June 2002.

Airport benchmarking studies began appearing in the literature in the early 1990's with one of the first such studies being that of Tolofari, Ashford, and Caves [9] in which airports operated by the British Airport Authority were compared against each other. A series of benchmarking studies have since appeared with a variety of geographic foci and covering a diverse set of performance parameters. A few examples of these studies are listed in Table I. TABLE I.

TABLE I. REPRESENTATIVE AIRPORT BENCHMARKING STUDIES

Study Title	Inputs (resources)	Outputs	Analytical Technique
Measuring airport quality from the airlines' viewpoint [10]	Airport charges; minimum connecting times; number of passenger terminals; number of runways; distance to nearest city center	Level of satisfaction from the airline users of each airport	DEA
Size versus efficiency: a case study of US commercial airports [11]	Operational costs; non-operational expenses; number of runways; number of gates	Passenger throughput; aircraft movements; aeronautical revenue; non-aeronautical revenue; percentage of on-time operations	DEA
Relative efficiency of european airports [12]	<u>Air transport movement study:</u> Airport surface area; total length of runways; number of aircraft parking positions at terminals; number of remote aircraft parking positions	Aircraft movements	DEA / Parametric TFP
	<u>Passenger movement study:</u> Terminal size; number of aircraft parking positions at terminals; number of check-in desks; number of baggage claim belts	Passenger throughput	DEA / Parametric TFP
Developing measures of airport productivity and performance [13]	<u>Terminal efficiency study:</u> Number of runways; number of gates; terminal area; number of employees; number of baggage collection belts; number of public parking spots	Passenger throughput; cargo throughput	DEA
	<u>Aircraft movement study:</u> Airport surface area; number of runways; runway area; number of employees	Aircraft movements	DEA
Performance Based Clustering for Benchmarking of US Airports [14]	Operational costs; number of employees; number of gates; number of runways	Operational revenue; passenger throughput; aircraft movements; cargo throughput	DEA
Measuring airports' operating efficiency: a summary of the 2003 ATRS global airport benchmarking report [15]	Number of employees; number of runways; number of gates; terminal area; purchased goods, materials, and services (outsourcing)	Passenger throughput; cargo throughput; aircraft movements; non-aeronautical revenue	Parametric TFP
Measuring total factor productivity of airports - an index number approach [16]	Capital expenditure	Aeronautical revenue; non-aeronautical revenue	Index number TFP

As is clear from Table I, besides being used in the analysis that will be examined here, Data Envelopment Analysis (DEA) is an analytical technique that is frequently used in benchmarking studies. DEA is a non-parametric methodology used to assess the efficiency of a Decision-Making Unit (DMU – e.g. an airport) in converting a set of inputs into outputs.

A number of different versions of the basic DEA model have been developed to address a series of potential shortcomings of the original DEA model. We will examine three variations of the DEA methodology and study the differences in assumptions of the three as well as compare the different outcomes of each. The purpose of this analysis is to understand whether the choice of methodology may have an impact on the outcome of an analysis.

B. CCR model

In all variations of the DEA models, the DMU(s) with the best inherent efficiency in converting inputs X_1, X_2, \dots, X_n

into outputs Y_1, Y_2, \dots, Y_m is identified, and then all other DMUs are ranked relative to that most efficient DMU.

For DMU 0, the basic DEA model (so-called CCR after Charnes, Cooper, and Rhodes [5]) is calculated as follows:

$$\max h_0 = \frac{\sum_r u_r y_{rj_0}}{\sum_i v_i x_{ij_0}}$$

$$\text{subject to } \frac{\sum_r u_r y_{rj}}{\sum_i v_i x_{ij}} \leq 1 \text{ for each unit } j \quad (1)$$

$$u_r, v_i \geq 0$$

The interpretation of u_r and v_i is that they are weights applied to outputs y_{rj} and inputs x_{ij} and they are chosen to maximize the efficiency score h_0 for DMU 0. The constraint forces the efficiency score to be no greater than 1 for any DMU. An efficiency frontier is calculated, enveloping all datapoints in a convex hull. The DMU(s) located on the frontier represent an efficiency level of 1.0, and those located inside the frontier are operating at a less than full efficiency level, i.e. less than 1.0.

The above fractional program is executed once for each participating DMU, resulting in the optimal weights being determined for each DMU.

Before solving the problem, the denominator in the objective function is removed and instead an additional constraint is added. Also, the original constraint is manipulated in order to convert the fractional program to a linear program. These two steps result in the following linear program:

$$\begin{aligned} \max h_0 &= \sum_r u_r y_{rj_0} \\ \text{subject to } \sum_r u_r y_{rj} - \sum_i v_i x_{ij} &\leq 0 \\ \sum_i v_i x_{ij_0} &= 1 \\ u_r, v_i &\geq 0 \end{aligned} \tag{2}$$

In simpler notation, this is written as:

$$\begin{aligned} \max(v,u) &= uy_0 \\ -vX + uY &\leq 0 \\ \text{subject to } v x_0 &= 1 \\ v \geq 0, u \geq 0 \end{aligned} \tag{3}$$

Finally, before solving, the linear program is converted to its dual for computational efficiency reasons:

$$\begin{aligned} \min(\theta, \lambda) &= \theta \\ \theta x_0 - X\lambda &\geq 0 \\ \text{subject to } Y\lambda &\geq y_0 \\ \lambda &\geq 0 \end{aligned} \tag{4}$$

With the addition of slack variables, the dual problem becomes:

$$\begin{aligned} \min(\theta, \lambda) &= \theta \\ \theta x_0 - X\lambda &= s^- \\ \text{subject to } Y\lambda &= y_0 + s^+ \\ \lambda \geq 0, s^+ \geq 0, s^- &\geq 0 \end{aligned} \tag{5}$$

The slack variables can be interpreted as the output shortfall and the input overconsumption compared to the efficient frontier.

C. BCC model

The CCR model is designed with the assumption of constant returns to scale. This means that there is no assumption that any positive or negative economies of scale exist. It is assumed that a small airport should be able to operate as efficiently as a large one – that is, constant returns to scale. In order to address this, Banker, Charnes, and Cooper developed the BCC model [3].

The BCC model is closely related to the standard CCR model as is evident in the dual of the BCC model:

$$\begin{aligned} \min(\theta, \lambda) &= \theta \\ \theta x_0 - X\lambda &= s^- \\ \text{subject to } Y\lambda &= y_0 + s^+ \\ e\lambda &= 1 \\ \lambda \geq 0, s^+ \geq 0, s^- &\geq 0 \end{aligned} \tag{6}$$

The difference compared to the CCR model is the introduction of the convexity condition $e\lambda = 1$. This additional constraint gives the frontiers piecewise linear and concave characteristics.

D. SBM model

Finally, the second adjustment to the basic CCR model is the Slacks-Based Measure of efficiency (SBM), proposed by Tone [18]. The motivation for the development of this model is the observation that while both the CCR and the BCC models calculate efficiency scores, neither is able to take into account the resulting amount of slack for inputs and outputs. Consequently, the purpose of this model is to minimize the input and output slacks, resulting in this fractional program, which is converted to a linear program before solving:

$$\min(\lambda, s^+, s^-) \rho = \frac{1 - \frac{1}{m} \sum_{i=1}^m s_i^- / x_{i0}}{1 + \frac{1}{s} \sum_{r=1}^s s_r^+ / y_{r0}}$$

$$\begin{aligned}
 & x_0 - X\lambda = s^- \\
 \text{subject to } & Y\lambda = y_0 + s^+ \\
 & \lambda \geq 0, s^+ \geq 0, s^- \geq 0
 \end{aligned} \tag{7}$$

III. METHOD OF ANALYSIS

A. Introduction

The analysis presented here is based on a re-run and extension of the analysis performed by Bazargan and Vasigh [11].

B. Inputs and Outputs

Input and output data was collected from a series of sources in order to assemble the same inputs and outputs as in the Bazargan and Vasigh study for the same airports and years (1996 – 2000). Table II lists the airports covered in the analysis and Table III lists the input and output variables and the source from which they were collected.

TABLE II. AIRPORTS INCLUDED IN STUDY

Large	Medium	Small
ATL	BNA	ALB
DEN	CLE	BHM
DFW	DAL	BOI
DTW	IND	COS
EWB	MCI	DAY
IAH	MDW	ELP
JFK	MEM	GEG
LAS	MSY	GSO
LAX	OAK	GUM
MIA	PDX	LIT
MSP	RDU	OKC
ORD	SJC	ORF
PHX	SJU	RIC
SFO	SMF	ROC
STL	SNA	TUL

TABLE III. AIRPORT DATA SOURCES

Inputs	Source
Operating expenses	FAA Form 127
Non-operating expenses	FAA Form 127
Number of runways	NFDC 5010 database
Number of gates	Airport websites
Outputs	Source
Aeronautical revenue	FAA Form 127
Non-aeronautical revenue	FAA Form 127
Portion of on-time operations	BTS Form 41
Total enplanements	TAF
Number of air carrier operations	TAF
Number of other operations	TAF

C. Model Selections

In the Bazargan and Vasigh study, the model selected was the CCR model. No discussion was provided regarding why this model was selected ahead of others; presumably it was selected on the premise of being the “default” DEA model.

One of the purposes of the present study was to examine the impact of different model selections. Hence, efficiency scores were calculated using three different, basic DEA models: CCR, BCC, and SBM. These are considered to be among the standard DEA models by Cooper, Seiford, and Tone [19]. These models were all executed in output-oriented mode, meaning that efficiency scores were calculated by analyzing which levels of outputs should be “possible” to generate if an airport were operating at full efficiency given the current levels of inputs. These scores were calculated using the “Learning Version” of the DEA-Solver software [4].

D. Model Configuration

One of the particular choices made in the Bazargan and Vasigh study, driven by practical considerations, was to create an artificial, “super-efficient” airport by assigning it the minimum input values present in the dataset and the maximum output values. The reason for this was that when the CCR model was initially executed with only the basic list of airports included, a very large portion of the airports were ranked as fully efficient (1.0). The cause of that outcome was the relatively small number of airports (45) as compared to the number of parameters considered (10). In general, this will frequently be a phenomenon in analysis having a low ratio of DMUs to parameters.

Bazargan and Vasigh also describe the theory of adding a condition requiring all weights u and v to be $\geq \epsilon$ (an infinitesimal value) in order to avoid setting all but one input and one output’s weights to a non-zero value. However, it is not indicated if this was used in the study and if it was, there is no specification of the value used for this infinitesimal value is provided. Consequently, the model will in this case be run without this additional constraint.

IV. RESULTS AND DISCUSSION

A. Comparison of CCR Model Runs

The results of the CCR model runs are shown in Table IV, Table V, and Table VI. These values are in general lower than the values calculated in the Bazargan and Vasigh study and the reason for this can not be ascertained since the raw data of the previous study were not available. It can be postulated that the difference is due to differences in the inputs; in particular if the values assigned to the artificial, “super-efficient” airports were different in the two studies, since that would cause a general offset between the two.

What is noteworthy, however, is that the general trend from the Bazargan and Vasigh study is confirmed: Large airports are in general exhibiting lower efficiency scores than medium-sized airports, and medium-sized airports are

exhibiting lower scores than small airports. These results are displayed in Figure 1. and Figure 2. The Kruskal-Wallis test was performed to determine whether the differences in efficiency ranks among the groups were significant, and that was found to be the case. The results from the Kruskal-Wallis test are presented in Table VII.

TABLE IV. RESULTS OF CCR RUNS FOR LARGE AIRPORTS

Airport	1996	1997	1998	1999	2000
ATL	0.1910	0.1990	0.2000	0.2000	0.2000
DEN	0.1515	0.1667	0.1667	0.1435	0.1281
DFW	0.1216	0.1231	0.1380	0.1199	0.1251
DTW	0.1405	0.1310	0.1256	0.1400	0.1432
EWR	0.2335	0.2595	0.2614	0.2511	0.2884
IAH	0.2202	0.2141	0.2149	0.2145	0.2238
JFK	0.2500	0.2491	0.2485	0.2500	0.2500
LAS	0.2034	0.2088	0.1977	0.2005	0.1946
LAX	0.2953	0.2995	0.2950	0.2698	0.2719
MIA	0.2500	0.2500	0.2500	0.2500	0.2500
MSP	0.2064	0.2015	0.1932	0.2155	0.2192
ORD	0.1667	0.1667	0.1667	0.1667	0.1667
PHX	0.2723	0.2765	0.2587	0.2674	0.2607
SFO	0.1784	0.1962	0.1783	0.1946	0.1804
STL	0.1937	0.2083	0.2032	0.2117	0.2134
Average	0.2050	0.2100	0.2065	0.2064	0.2077

TABLE V. RESULTS OF CCR RUNS FOR MEDIUM SIZED AIRPORTS

Airport	1996	1997	1998	1999	2000
BNA	0.2140	0.2140	0.2179	0.2143	0.2186
CLE	0.2054	0.2132	0.2142	0.2085	0.2183
DAL	0.5603	0.5087	0.5257	0.5319	0.5262
IND	0.2811	0.3043	0.3044	0.3026	0.3067
MCI	0.2674	0.2816	0.2812	0.2836	0.2843
MDW	0.2361	0.2298	0.2296	0.2260	0.2296
MEM	0.2206	0.2070	0.2065	0.2256	0.2289
MSY	0.2765	0.2799	0.2869	0.2845	0.2916
OAK	0.4802	0.4704	0.5455	0.5147	0.5455
PDX	0.2518	0.2764	0.2778	0.2855	0.2767
RDU	0.4480	0.4152	0.4027	0.3749	0.3807
SJC	0.3022	0.3156	0.3051	0.3057	0.2951
SJU	0.3374	0.3967	0.4215	0.4155	0.4378
SMF	0.4121	0.4252	0.4171	0.4205	0.4019
SNA	1.0000	1.0000	0.9914	1.0000	0.9882
Average	0.3662	0.3692	0.3752	0.3729	0.3753

TABLE VI. RESULTS OF CCR RUNS FOR SMALL AIRPORTS

Airport	1996	1997	1998	1999	2000
ALB	0.4772	0.5101	0.5047	0.4895	0.5099
BHM	0.5345	0.5400	0.5482	0.5461	0.5698
BOI	0.8089	0.5714	0.7842	0.5392	0.8086
COS	0.5191	0.5561	0.5672	0.5725	0.5625
DAY	0.3472	0.3708	0.3674	0.3620	0.3652
ELP	0.6856	0.6810	0.6863	0.6876	0.6757
GEG	0.4460	0.4662	0.4726	0.5016	0.4695
GSO	0.8082	0.8553	0.8350	0.8115	0.8459
GUM	0.5714	0.5714	0.5714	0.5714	0.5714
LIT	0.8031	0.8385	0.8493	0.8590	0.8710
OKC	0.5854	0.6006	0.6053	0.6155	0.5961
ORF	0.3994	0.4555	0.4217	0.4122	0.4306
RIC	0.4214	0.4470	0.4403	0.4326	0.4439
ROC	0.4341	0.4572	0.4570	0.4321	0.4374
TUL	0.5098	0.5332	0.5373	0.5349	0.5472
Average	0.5568	0.5636	0.5765	0.5578	0.5803

Figure 1. Results of Bazargan and Vasigh study

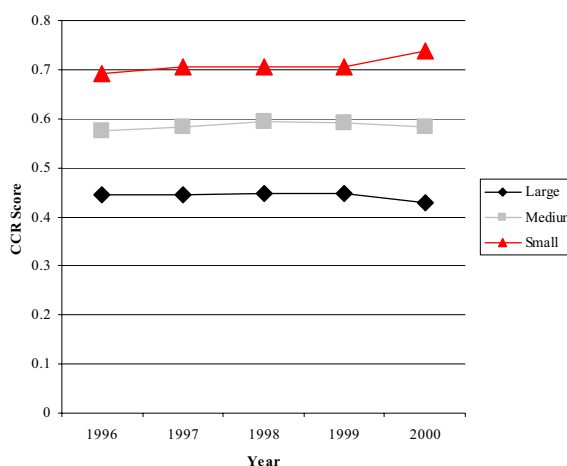


Figure 2. Results of CCR re-run of Bazargan and Vasigh study

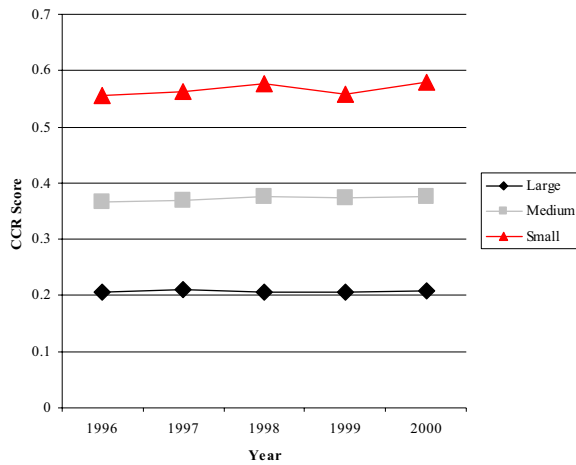


Figure 3. Results of BCC analysis

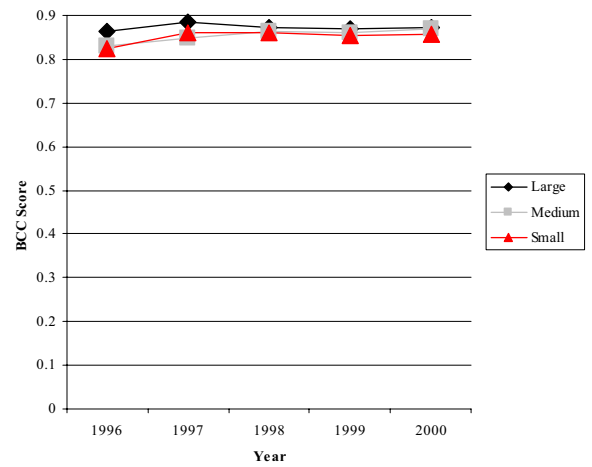
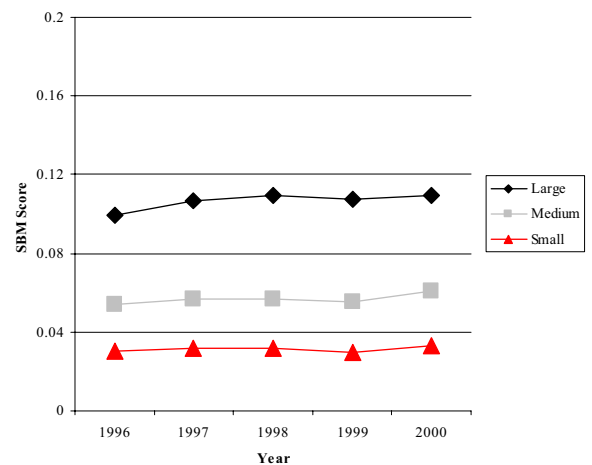


TABLE VII. RESULTS OF KRUSKAL-WALLIS TEST ON CCR RANKS

Year	Mean efficiency rank			Chi square	Asymptotic significance
	Large airports	Medium airports	Small airports		
1996	36.13	22.40	10.47	28.691	0.000
1997	36.07	23.07	9.87	29.848	0.000
1998	36.33	22.60	10.07	30.200	0.000
1999	36.33	22.53	10.13	29.876	0.000
2000	36.27	22.60	10.13	29.716	0.000

Figure 4. Results of SBM analysis



B. Comparison of CCR, BCC, and SBM Results

In applying the three DEA models (CCR, BCC, and SBM) to the same dataset, some striking results emerged: While the CCR analysis showed that small airports were more efficient than both medium sized and large airports, and that medium sized airports were more efficient than large airports, in the BCC model no significant difference among the three groups was observed, and in the SBM dataset the difference was in fact reversed. These findings are displayed in the following figures and tables, where the average efficiency scores are displayed and the results of the Kruskal-Wallis tests on the efficiency ranks.

TABLE VIII. RESULTS OF KRUSKAL-WALLIS TEST ON BCC RANKS

Year	Mean efficiency rank			Chi square	Asymptotic significance
	Large airports	Medium airports	Small airports		
1996	18.53	23.47	27.00	3.149	0.207
1997	22.33	25.67	21.00	1.006	0.605
1998	23.33	24.97	20.70	0.808	0.668
1999	23.40	23.37	22.23	0.077	0.962
2000	21.40	22.63	24.97	0.572	0.751

TABLE IX. RESULTS OF KRUSKAL-WALLIS TEST ON SBM RANKS

Year	Mean efficiency rank			Chi square	Asymptotic significance
	Large airports	Medium airports	Small airports		
1996	9.67	23.60	35.73	29.589	0.000
1997	9.67	23.67	35.67	29.449	0.000
1998	9.47	23.40	36.13	30.939	0.000
1999	9.33	23.33	36.33	31.710	0.000
2000	9.47	22.93	36.60	32.010	0.000

These results point very clearly to the fact that the choice of DEA model has a strong impact on the outcome of this study and that selecting the appropriate model is paramount. While the selection of the appropriate model is key, and previous studies have pointed this out, the literature does not provide many examples of or guidance for selecting the most appropriate model.

In comparing which inputs and outputs are assigned significance in the three different models, the following two observations can be made:

- The same weights were assigned to outputs in the BCC as in the CCR model, and these weights were all focused on a single output for each airport. However, input weights differed between the two models
- In comparing the CCR and the SBM models, the same input variables are assigned non-zero weights, although these weights are different. On the output side, the SBM model assigns non-zero weights to a much broader set of variables than does the CCR model, which usually assigns non-zero values to only a single variable. It can be argued that this leads to the SBM model covering a more comprehensive and thereby accurate scope, rather than only considering a single output variable as in the CCR case.

Notably, running the BCC model, which considers variable returns to scale, resulted in either constant or *decreasing* returns to scale. While impossible to depict graphically in a multi-dimensional model, the concept of decreasing returns to scale is reflected in the shape of the convex hull spanning the DMU(s). In practical terms, this means that rather than observing any economies of scale in terms of efficiency, the model resulted in either no difference or actually a negative effect on airport efficiency as scale grew.

C. Observations on the Weakness of Using Artificial, "Super-Efficient" DMUs

In running the CCR analysis using the artificial, "super-efficient" DMU, one serious problem that arises is the fact that for nearly every non-efficient DMU, the model selects just one single input and a single output to have a non-zero weight while all others are assigned a zero weight. That

means that the resulting efficiency score is highly skewed and does not give an accurate representation of a DMU's efficiency.

In the analysis, Bazargan and Vasigh make reference to the use of an infinitesimal value in a constraint to avoid setting any of the weights to 0, but make no reference as to whether this was used and if so, what values were used. As was observed in this analysis, using such a technique is important in order to achieve an accurate assessment of all DMUs across all input and output parameters rather than focusing the analysis on just a subset of parameters.

An alternative approach to handle the issue of setting weights to 0 is the so-called cross-efficiency score, as originally proposed by Doyle and Green [17]. This method calculates the optimal weights for each DMU in the typically manner for CCR, BCC, or SBM, but then applies each DMU's weights to all other DMUs, and then computes an average score for each DMU. While this does aid in mitigating the effects of one DMU's score being calculated on the basis of only a small set of inputs and outputs, it also can be argued to run counter to one of the fundamental principles of DEA; the assumption that a management tradeoff is made between performing well across the given inputs and outputs for an individual DMU, and therefore the model should permit each DMU to appear in the "best possible light" by selecting its optimal weightings.

V. CONCLUSIONS AND FUTURE WORK

The analysis in this paper has shown that depending on the DEA model chosen, radically different results may be obtained. Consequently, any study of airport efficiency needs to begin with a thorough examination of the models available and a motivation for why a particular model was selected. Without an upfront analysis of this kind, a study's final results may be called into question.

In order to simplify such analysis, a thorough study of the implications of each of the most commonly used models is necessary. Such a study needs to examine the implications and interpretations of each model in an airport context since the characteristics of each model may take on different meanings depending on the application area.

Finally, a further, broader area worthy of additional analysis relates to the selection of inputs and outputs to airport efficiency benchmarks: Before proceeding to calculating the airports' efficiency scores, any study needs to present a discussion of what the true goals of an airport are – for example maximizing passenger throughput, maximizing aircraft movements, minimizing delay, maximizing profits, minimizing costs, etc. Only after determination of the airports' goals can the appropriate inputs and outputs be selected. Many of the analyses to-date have largely omitted this discussion and appear to have selected inputs and outputs merely on the availability of data.

REFERENCES

- [1] Bureau of Transportation Statistics, "Flight Delays, Mishandled Bags, Consumer Complaints increase in 2007",

- http://www.bts.gov/press_releases/2008/dot017_08/html/dot017_08.html, accessed 10 February 2008.
- [2] Federal Aviation Administration, "Terminal Area Forecast Summary 2007 - 2025," HQ-08371, http://www.faa.gov/data_statistics/aviation/taf_reports/media/TAF2007-2025Summary.pdf, accessed 10 February 2008.
- [3] R.D. Banker, A. Charnes, and W.W. Cooper, "Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis," *Management Science* vol. 30, iss. 9, pp. 1078-1092, 1984.
- [4] K. Tone, "DEA-Solver-LV" software, version 3.0. <http://www.saitech-inc.com>
- [5] A. Charnes, W.W. Cooper and E. Rhodes, "Measuring the Efficiency of Decision Making Units," *European Journal of Operational Research* 2, pp. 429-444, 1978.
- [6] R. Dattakumar and R. Jagadeesh, "A review of literature on benchmarking" *Benchmarking: An International Journal*, vol. 10, iss. 3, pp. 176-209, 2003.
- [7] J. Spendolini, *The Benchmarking Book*. New York, NY: Amacom, 1992.
- [8] M. Yasin, "The theory and practice of benchmarking: then and now," *Benchmarking: An International Journal*, vol. 9, iss. 3, pp. 217-243, 2002.
- [9] S. Tolofari, N. Ashford, and R. Caves, "The cost of air services fragmentation," Loughborough University of Technology, Department of Transport Technology Working Paper TT9010.
- [10] N. Adler, J. Berechman, "Measuring airport quality from the airlines' viewpoint: an application of data envelopment analysis," *Transport Policy*, vol. 8, iss. 3, pp. 171-181, 2001.
- [11] M. Bazargan, B. Vasigh, "Size versus efficiency: a case study of US commercial airports," *Journal of Air Transport Management*, vol. 9, iss. 3, pp. 187-193, 2003.
- [12] E. Pels, P. Nijkamp, P. Rietveld, "Relative efficiency of European airports," *Transport Policy*, vol. 8, iss. 3, pp. 183-192, 2001.
- [13] D. Gillen, A. Lall, "Developing measures of airport productivity and performance: an application of data envelope analysis," *Transportation Research Part E: Logistics and Transportation Review*, vol. 33, iss. 4, pp. 261-273, 1997.
- [14] J. Sarkis, S. Talluri, "Performance based clustering for benchmarking of US airports," *Transportation Research Part A: Policy and Practice*, vol. 38, iss. 5, pp. 329-346, 2004.
- [15] T. H. Oum, C. Yu, "Measuring airports' operating efficiency: a summary of the 2003 ATRS global airport benchmarking report," *Transportation Research Part E: Logistics and Transportation Review*, vol. 40, iss. 6, pp. 515-532, 2004.
- [16] P. G. Hooper, D. A. Hensher, "Measuring total factor productivity of airports – an index number approach," *Transportation Research Part E: Logistics and Transportation Review*, vol. 33, iss. 4, pp. 249-259, 1997.
- [17] J. Doyle and R. Green, "Efficiency and Cross-Efficiency in DEA: Derivations, Meanings, and Uses," *The Journal of the Operational Research Society*, vol. 45, iss. 5, pp. 567-578, 1994.
- [18] K. Tone, "A slacks-based measure of efficiency in data envelopment analysis," *European Journal of Operational Research*, vol. 130, iss. 3, pp. 498-509, 2001.
- [19] W. W. Cooper, L. M. Seiford, K. Tone. *Data Envelopment Analysis*. New York, NY: Springer, 2007.

Improving Aircraft Turn Around Reliability

Specific Aircraft Body Design Parts hamper Ground Handling and Airport Performance

Hartmut Fricke

Institute of Logistics and Aviation
Technische Universität Dresden
Dresden, Germany
fricke@ifl.tu-dresden.de

Michael Schultz

Institute of Logistics and Aviation
Technische Universität Dresden
Dresden, Germany
schultz@ifl.tu-dresden.de

The Airport, and specifically the turn around time (TAT) of aircraft at the gate or a remote position from the terminal have been recognized as crucial element to ATM system performance. Currently, the TAT ranges around 30 min for short/medium range aircraft. For the 2020 Single European Sky, SESAR claims as performance target for a cut down to 15 min while also increasing process reliability. There are several reasons, why the turn around is still remarkably uncertain, mainly caused by shared responsibility for the individual ground handling processes, a frequent distortion of gate occupancy schemes at the airport and still deficient interfaces with the aircraft body. All this leads to only a limited predictability of the “Earliest Off Block Time”, this being an important time constant to trigger the departure and consequently the arrival sequence. This paper reveals the current data quality as found during a large field study at German Airports, derives the reasons for largely varying process times both on the technical and procedural level and shows the potential for improved TAT reliability through aircraft interface optimization.

Aircraft Turn Around, Aircraft Body, Ground Handling, Boarding, Critical Path, Monte Carlo Simulation

I. INTRODUCTION

The Turn Around Time (TAT) of aircraft has been defined by IATA's Aircraft Handling Manual (AHM) 810 [1] as the time period an aircraft occupies a stand or a gate at the airport. More specifically this period is framed by two activities: The positioning and removal of the aircraft wheel chocks, respectively named as On and Off Block Time. As this time is directly impacting the airport / terminal capacity, there is a vital interest in predicting exactly the Gate Occupancy Time (GOT) by means of stable ground handling processes from the Airport Operator's point of view and a similar one in minimizing that time from an Airline's perspective since block time is commercially lost time. The GOT has therefore become a central planning parameter for Airport / Apron Terminal Design Processes [4]. The largest component during on block is the boarding and de-boarding of passengers, as field measurements clearly show [2]. Hence, it could also be learned, that ground handling events are characterized by a remarkable diversity in processes, occurrence of such processes and their service depth, making every aircraft turnaround

somehow a unique procedure in terms of required times, interfacing and services. Our referred field study was performed in summer 2006, to learn about these constraints and to gather relevant data with the aim to exploring ways how to improve the reliability of and also shorten the time needed for a Turn Around. This with SESAR's performance target in view, to cut down the TAT for domestic flights to 15 min, and 30 min for international flights [5].

Chapter II presents the results of that field study; chapter III discusses the management of the gathered data in a relational database to allow systematic analysis and dissemination of the results to the various stake holders. Chapter IV reveals the statistical data modeling to determine the level of reliability in today's Turn Around operations by means of Monte Carlo (MC) Simulations, Chapter V creates an improved aircraft interface scenario and derives the expected increase in Turn Around reliability, again by applying MC modeling.

II. CURRENT TURN AROUND PRACTICE

A. Overview

The Turn Around has been described in several studies such as [6], many of them putting emphasis on the boarding and de-boarding processes, [7], [8], [9] and is also part of many standard documents such as the *Aircraft Characteristics For Airport Planning Manual* of any modern aircraft [3]. It is generally represented as a bunch of processes, from which a subgroup may run in parallel, and others are required to run sequentially, e.g. due to legal or logistical requirements such as limited space around the aircraft, tool availability, or legal constraints such as to prohibit Fueling with Passengers onboard.

The collection of sequential processes consuming the maximum time during turn around is called its *critical path*. As stated, typical process members of the critical path are boarding, de-boarding, fueling, loading, unloading and service processes such as cleaning or catering [2], [3].

The following picture depicts the typical process and dependencies:

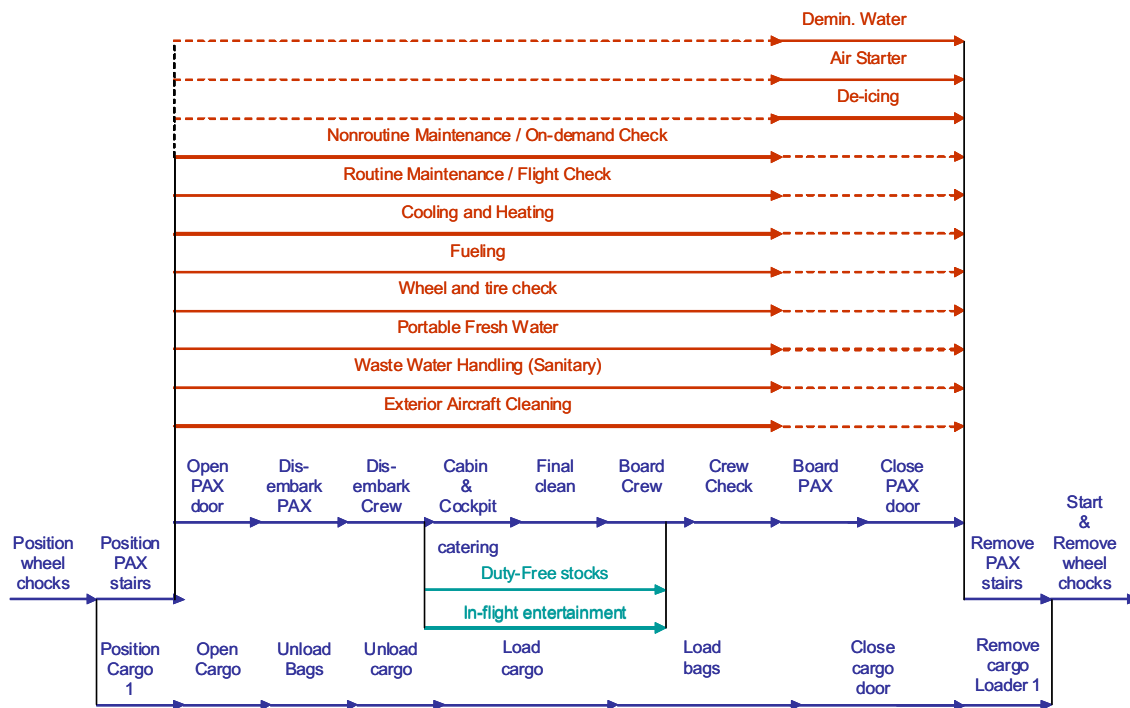


Figure 1. Parallel and sequential Turn Around Processes

B. Set up of the Field Study

To identify possibilities to reduce individual process times and their expected remarkable variance in time especially on the critical path, the field investigation aimed at exploring potential changes to the aircraft-design and arguments for innovative technologies to be installed in the aircraft or applied on ground in order to smoothen and accelerate the ground services resulting in shorter turnaround times with increased reliability.

This activity shall be seen in concert with e.g. the physical implementation of supervision technologies such as process scanning and tracing devices or e.g. data networking at the apron, as e.g. [6] is calling for, and modification of current legal regulations (see Chapter III).

The datasets collected had to cover nearly all models and types of Airbus-manufactured aircraft and at least some representative Boeing aircraft. It was further anticipated to consider different airport types, such as the Hub-Airport Munich (MUC) and mid size airports such as Leipzig/Halle (LEJ) and Dresden (DRS) in Saxony. The data gathered also covered interviews with ground handling staff as well as detailed time sequence measurements to allow determination of specific process indicators (fuel quantity, passenger rates, etc., see TABLE I.). Further, process interruptions due to limited or distorted logistics were noted.

These data did form the platform for various process analyses looking at optimization potential. Based on the Aircraft Handling Manual [1] providing a detailed framework on how to cluster and categorize ground handling services, then

extended for the boarding related activities, the following collections of services were gathered (see also Figure 1.):

- Aircraft Servicing (covering exterior services, such as water and lavatory service, e.g. attachment of stairs and passenger bridges);
- Loading and Unloading (on- and offloading of belly payload, container and bulk);
- De-boarding (differentiating in remote and gate positions as well as number of doors used);
- Catering (Handling at Ramp);
- Interior Cleaning (including cabin and crew rest compartment cleaning);
- Crew Change and Cabin Preparation (including duties to be performed by cabin crew);
- Boarding (hence without distinguishing into different boarding procedures, this later recognized as important data and being subject to further studies at TUD); and
- Fueling (number and location of fuel valves used).

Based on that structuring, the measurements performed in the field did include timestamps assigned at the lowest observable level, so linked to a process, task or element as respective sub-processes. A distinct process start and end was required to be observed. As such, those timestamps were tested on their traceability and practicability before they were implemented into the process templates (see Figure 2.).

The field studies needed for this research were kindly supported by Airbus Deutschland, Dresden, Leipzig/Halle and Munich Airport, and the Ground Handling Agents Mucground and Port Ground.

At some processes, intermediate timestamps were defined in order to allow gathering information about process interruption times, waiting times or other intermediate points (milestones). Besides collecting time sequences, the sampling of the following additional information affecting the turnaround was performed:

- Manpower (amount of personnel performing a specific process, task or element)
- The equipment types used;
- The equipment quantities;
- Load figures (e.g. Passenger figures, Seating, Baggage-, Cargo-, Mail-figures etc.);
- Aircraft layout information (if possible: amount of lavatories, galleys, etc.);
- Transfer volumes (e. g. fuel quantity figures).

The various data collection templates designed for the field activity are depicted in the following figure e.g. for the loading case¹:

A/C-type: _____		Date: _____			
Flight-Nr.: _____					
Load Distribution Message (LDM) - Operations					
Pax in	#	(MVT inbound)			
Pax out	#	(MVT inbound)			
#seats (split up)	F:	B:	Y:		
#cabin crew	#				
Container/Pallet Distribution Message (CPM - inbound) - Operations					
Weights req in case of bulk loaded A/C					
UNLOAD	Weight mail/ULD aft in [kg]	#ULD:	or	W:	
	Weight cargo/ULD aft in [kg]	#ULD:	or	W:	
	Weight mail/ULD fwd in [kg]	#ULD:	or	W:	
	Weight cargo/ULD fwd in [kg]	#ULD:	or	W:	
Weight mail bulk in [kg]				W:	
Weight cargo bulk in [kg]				W:	
Loading Instruction / -report (- outbound) - Foreman Loading or Operations					
Weights req in case of bulk loaded A/C					
LOAD	Weight cargo/ULD aft out [kg]	#ULD:	or	W:	
	Weight mail/ULD aft out [kg]	#ULD:	or	W:	
	#bags/ULD aft out [kg]	#ULD:	or	W:	
	Weight cargo/ULD fwd out [kg]	#ULD:	or	W:	
	Weight mail/ULD fwd out [kg]	#ULD:	or	W:	
	#bags/ULD fwd out [kg]	#ULD:	or	W:	
Weight cargo bulk out [kg]				W:	
Weight mail bulk out [kg]				W:	
#bags bulk out [kg]		#:	or	W:	
???					
#bags/ULD aft in [kg]	?				
#bags/ULD fwd in [kg]	?				
#bags bulk in [kg]	?				
#trolleys/boxes fwd catering	?				
#trolleys/boxes mid catering	?				
#trolleys/boxes aft catering	?				
#trolleys/boxes belly uplift catering	?				

Figure 2. Data collection template, the loading case

C. The Collected Data: A Web Application

To allow efficient handling of the remarkable amount of data collected, a relational MYSQL Database was implemented. Further a PHP application was designed so forming a powerful

¹ ULD = Unit Load Device, Standard cargo container. Typical dimension: 317cm length, 243 cm width, 299 cm height. Various subclasses do exist according to IATA coding definitions.

web application to create easy access to the data, and providing a clear structure and documentation platform:

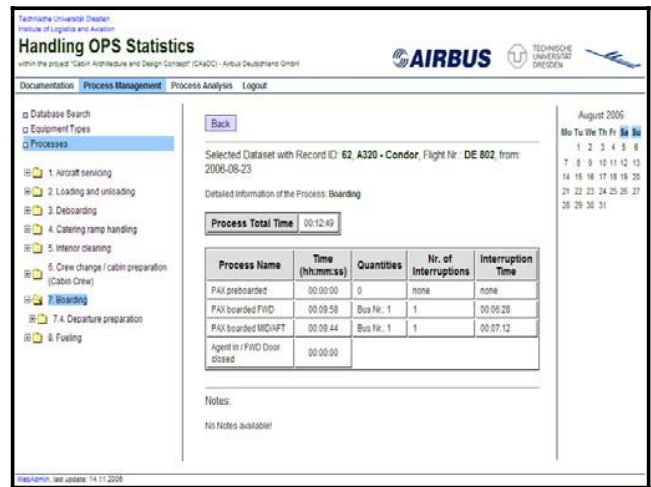


Figure 3. Web Front end to access data of Turn Around Database

D. Data preparation for Turn Around comparison means

For each of the analyzed processes, and sub-processes, dedicated values were derived from the field measurements in order to allow comparing the individual turnarounds. The following parameters were defined for the individual processes [2]:

TABLE I. SPECIFIC PROCESS PARAMETER

Specific Process Data generated (extract)	
Process	Value Subhead
(De-)Boarding	A/C (De-)Boarding Rate [PAX/min]
	A/C (De-)Boarding Time [min]
	Avg. Boarding Flow Interruption Time [min]
Catering	Catering Time Split up AFT Galley [%]
	Catering Time Split up FWD Galley [%]
	Total Catering Vehicle retention time [min]
Cleaning	Cleaning rate [per seat]
	Cleaning Time Cabin, Lavatory, Galley [min]
Fueling	Average Fuel Flow rate [l/min]
	Starboard/Portside Fueling Split up [%]
	Tanker-Dispenser Split Up [%]
(Un-)Loading	(Un-)Loading AFT/FWD [ULD/min]
	(Un-)Loading Bulk [kg/min]
Servicing	A/C Service Vehicle Retention Time [min]
	Pushback Waiting and Standby Time [min]
	A/C Servicing Equipment Split up [%]

These specific data could be sampled for over 120 complete turn around events, measured at the three sites MUC, DRS and LEJ.

III. STATISTICAL PROCESS DATA ANALYSIS

A. Legal and Procedural constraints

In accordance with ICAO Doc 9626, the Manual on the Regulation of International Air Transport [10], ground handling consists of those processes as listed previously and implemented in our database. Further, the ICAO Doc 9562 Airports Economics Manual [11] separates ground-handling services into terminal handling (passenger check-in, baggage and freight handling, flight plan processing) and ramp handling (aircraft handling, cleaning and servicing). For all of these processes, operational interdependencies do exist, limiting the parallelizing of these processes, as the area around the aircraft is very limited in light of the various service equipment used during ground handling. Further, legal requirement call for granted maneuvering capability for the fire brigade on one side (typically located right hand to the aircraft) and passenger related activities (stairs typically installed left hand).

Obviously, Airlines and Ground handling companies are permanently investigating into possibilities to reduce process constraints, since further parallelizing of processes besides mainly fueling, catering, cleaning and servicing nowadays as found during the field study and shown in Figure 4. is seen as mandatory to substantially reduce the TAT. This aspect will be focused in subsequent papers.

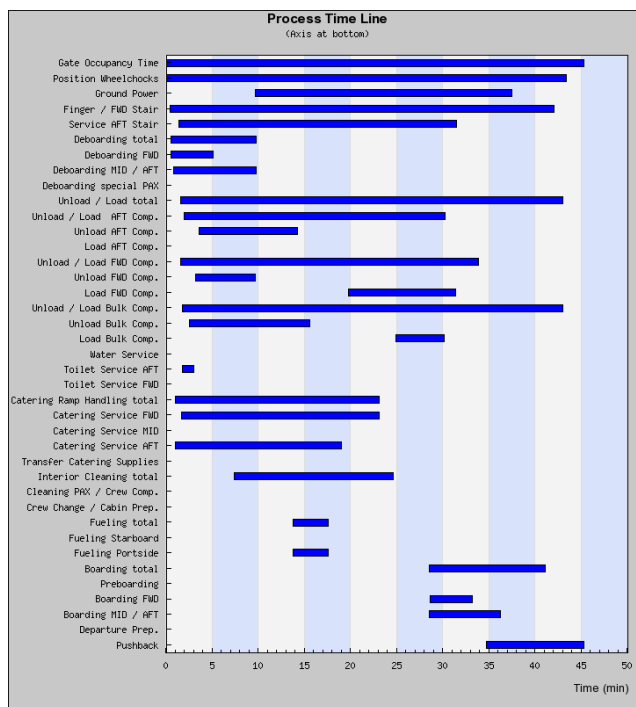


Figure 4. Sample Parallel Processing during turnaround – field study

Candidates for (increased) parallelized processing so remain

- loading/unloading and
- fueling.

Further,

- shortening and
- increasing the reliability of the boarding and deboarding time

do represent a second field for improvement, being subject of this paper. Investigation was consequently directed into the technical aspects of these processes in the field for getting a clear picture of all operational requirements and systematic deficiencies linked to them, presented in the next section.

B. Technical constraints

Based on the interviews done with ground handling personnel in the field and observations gathered during the monitoring phase, the following representative process distortion cases were found [2]:

TABLE II. TECHNICAL DEFICIENCIES AT TURNAROUND

Technical Deficiencies at Turnaround (extract)	
Process	Deficiency
(De-)Boarding	Installation of second passenger bridge on gate position does not result in the time saving of the two passenger stairs on remote position due to the marginal distance between the two doors.
Boarding	Boarding is influenced by the mutual obstructions of the passengers.
Catering	The catering door is opened by the purser after all passengers left the aircraft and the catering vehicle arrived on position. Usually, the employee signaled his arrival by knocking at the door which is occasionally not realized by the crew and result to a delay of the beginning of the process.
Cleaning	Leaving of the vehicle affected by obstructions because of loading equipment especially on smaller aircraft with bulk load.
	High quantity of waste produced because of in-flight services. The lack of disposal units leads to defilement of the aircraft and time consuming removal by the cleaning staff.
Fueling	Obstructions between the service vehicles result to delayed start of the fueling process and a non-compliance of the safety requirements
(Un-)Loading	Simultaneous fueling on both wings enhances the flow rate not to the double of a single side fueling
	The fuel computer has problems calculating the exact filling quantity due to temperature variations. So, the centre tank is opened too late and fuelled with a flow rate lower than for the other tanks.
Servicing	Cargo Door: Panel for opening hatches of the A300, A330 and A340 attached to high to be reached manually by all personnel
	Conveyer belts used for (un)loading bulk. Height of aircraft influenced by the current weight of Payload and Fuel. Permanent Altitude adaptation needed inducing repositioning of the equipment
Servicing	Cockpit misses the information about the removal of the GPU by the loading crew. However, clear commitment necessary before disconnection.
	Reasoned by weight of the adapter, negative effects can occur for the employee by attaching the unit as well as for the holding of the plug connection itself.



Figure 5. Physical Deficiencies : Displaced Control Panel for cargo door (left) / Manual positioning of load compartment (right)

To conclude, quite a bunch of technical deficiencies were found during the field study that effectively hamper the manual and also partly automated activities during Turnaround in terms of delays and consequently reduced reliability for process times.

The following section looks into the turnaround database and determines the statistical behavior of the processes.

C. Determination of Process Stability

To determine the nature of the generated process descriptors as listed in TABLE I. typical density distributions were compared to the data. Since the data has a non normal distribution character, a WEIBULL density function $f(x)$ with α and β as shape parameter was finally selected for probing the data fitting, since it is most commonly used in life data analysis and due to its flexibility. It can mimic the behavior of other statistical distributions such as the normal and the exponential. A Weibull distribution may generally written in the form

$$f(x, \alpha, \beta) = \frac{\alpha}{\beta^\alpha} x^{\beta-1} e^{-(x/\beta)^\alpha} \quad (1)$$

Where $x \geq 0$ and $f(x) = 0$ for $x < 0$
 x to substituted by $(x + x_offset)$
 α = scale parameter
 β = shape parameter

For all processes found to be timely critical during today's turn around (see Figure 4), data distributions were analyzed, statistically tested on its significance against chi-square. The probing procedure is shown exemplarily for the de-boarding process, referring to the process parameter passenger rate per door [PAX/min] to allow normalizing of the varying seat load factors. It may be noticed that for some processes, an offset is required to adopt the data range to the Weibull distribution by x_offset . For the sample case presented below, this offset is at 10 PAX/min:

- De-boarding:**
 x-Offset: 10 [PAX/min]
 Number of counts: 97
 Minimum flow rate: 4,1 PAX / min
 Maximum flow rate: 38,9 PAX / min
 Number of classes: 10 (see TABLE III)

Class with: 4 [PAX/min]

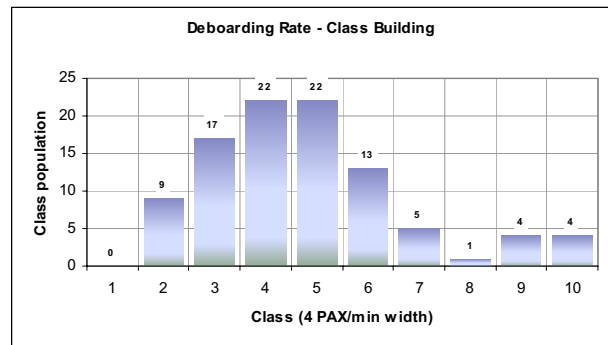


Figure 6. Class Building for the De-Boarding Case

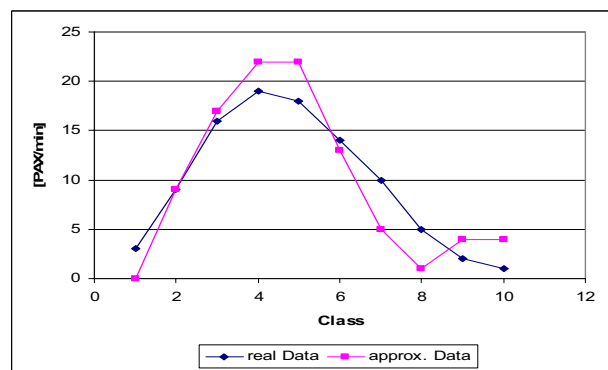


Figure 7. Approximated versus Real Data Distribution – De-boarding case

TABLE III. STATISTICAL FIT AND PROOF DATA – DE-BOARDING CASE

Statistical Fit and Proof Data	
(De-)Boarding [PAX / min]	Value
Class number logic	5 log(number of counts)
α	2.23619
β	19.2339
x offset	10
Mean	17,0354
Standard Variation	7.92729
Variance	0.465343
Density (Weibull)	$0.0030067 e^{-(0.00134457 \cdot x2.23619)} x^{1.23619}$
Distribution	$1 - e^{-(0.00134457 \cdot x2.23619)}$
Chi ²	10,3228
a	0,05
Proof	12,5916
Pass Status	TRUE

For some other timely critical processes, the following distributions are depicted below, showing the remarkable magnitude in the distribution variations:

TABLE IV. REAL DISTRIBUTIONS – CRITICAL TA PROCESSES

Real Distributions – Critical TA Processes	
(De-)Boarding [PAX / min]	
α	2.23619
β	19.2339
x offset	10
Boarding [PAX / min]	
α	2.29263
β	19.5224
x offset	0
loading AFT&FWD [bags / min]	
α	3.13022
β	165.289
x offset	0
Unloading AFT&FWD [bags / min]	
α	4.05691
β	215.208
x offset	0

These parameter sets do allow sorting the processes against their stability, as with α (scale parameter) tending to smaller values and β (shape parameter) tending to larger values do express increasing variance of the distribution. It comes to the following ranking:

1. Unloading
2. Loading
3. Boarding
4. De-boarding

As observations did further show, the reason for this inter-process uncertainty trend refers to the varying standardization and automation level of the individual processes as well as the quality of technical support means, as described above.

IV. UNCERTAINTY EFFECT ON THE TURN AROUND

When accumulating the critical processes along the line, the agglomeration of all partial process uncertainties will obviously induce a total uncertainty for the turn around. To determine the magnitude of that total variation, analytically modeling was replaced in favor of statistical random probing. This was achieved by applying the Monte Carlo method, relying on repeated random sampling of all critical processes, using the calibrated distributions as found above. The results are presented in the next section.

A. MC Simulation for overall turnaround uncertainty determination

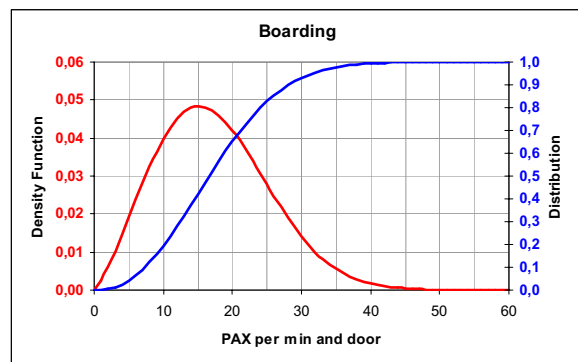
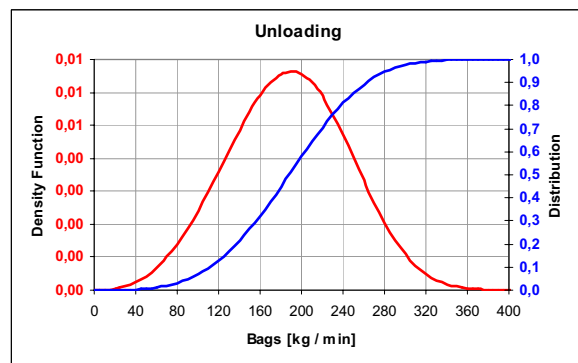
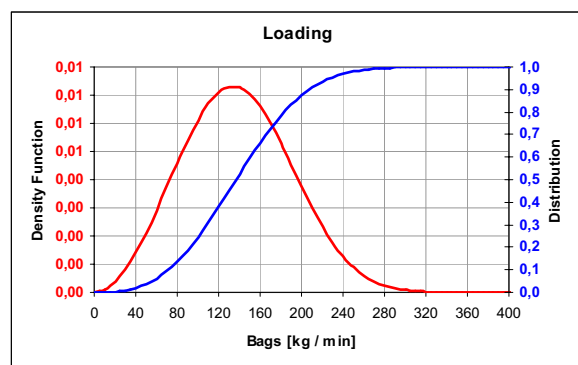
The processes

- boarding, including the emphasis on the door closing procedure after the last Passenger did board, de-boarding, obviously running in sequence;
- loading and unloading, obviously running in sequence, too, and

- either fueling, catering or cleaning, all running in parallel

have been identified as critical processes according to Figure 4. Concerning the fueling process, process interruptions were recorded on a regular basis, resulting from the fact, that fueling is generally initiated by the responsible personnel without knowing the exact quantity beforehand, to be ordered by the cockpit crew. So a typical volume is being fueled in a first step, than for most of the tracked cases, a break occurs within which communication takes place and the exact quantity will be told. This effect was also included in the process modeling.

The following density and distribution shapes do result from functional fitting as presented in Chapter III, reflecting also the x offset applied to the de-boarding case:



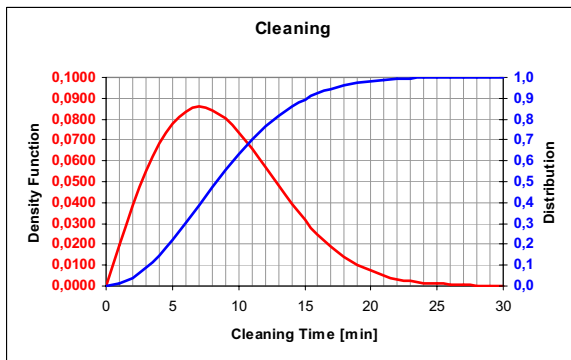
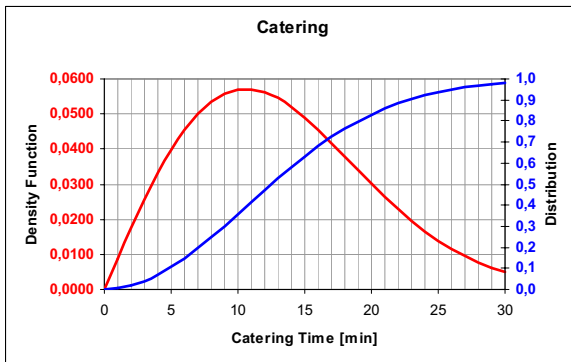
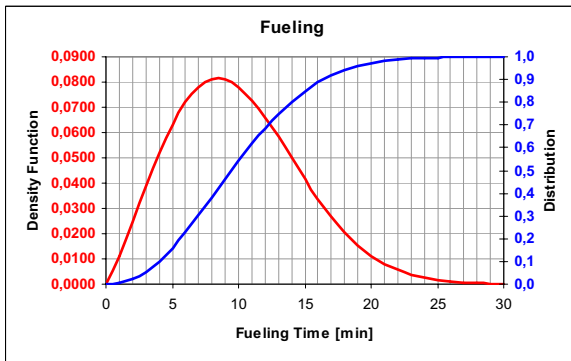
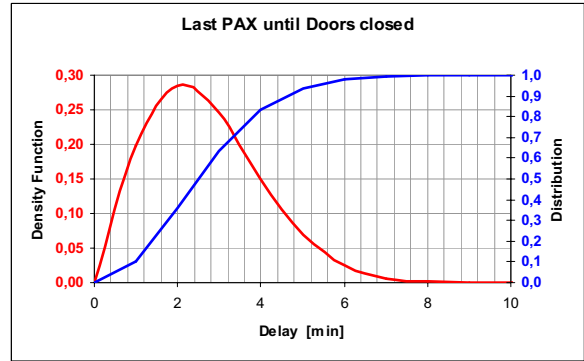
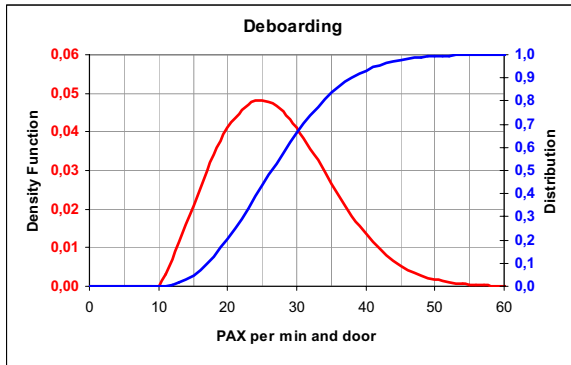


Figure 8. Modeled distributions for all time critical processes

The MC method is then applied to the collection of these processes, linked according the prescribed constraints, and executed for 10^4 simulation runs. It comes to the following TAT distribution:

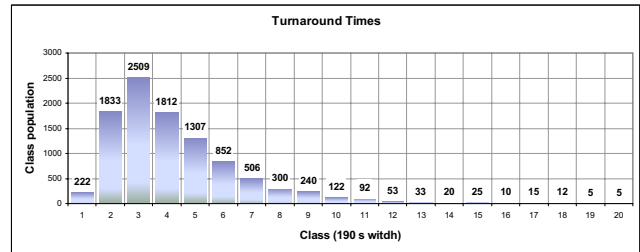


Figure 9. Turn Around Time Distribution

The shape of the TAT Distribution so follows again a WEIBULL behavior with the following over all process characteristics:

- Minimum Time: 1135 s (19 min)
- Maximum Time: 12486 s
- Mean: 1872 s (31,2 min)
- Standard Deviation: 23 s

Obviously, the central limit theorem according to which under certain conditions such as referring to independent and identically-distributed data with finite variance, the sum of a large number of random variables is approximately normally distributed, may not applied without unacceptable errors for this data case.

Instead, it is shown, that all type of optimization of single (and critical) processes does lead to a turn around time distribution with Weibull character. This to be proven by the different spot case analyzed in chapter V.

The 3 critical processes running in parallel in between deboarding/boarding and unloading/loading were limiting with the following distribution, based on 10^4 simulation runs, too:

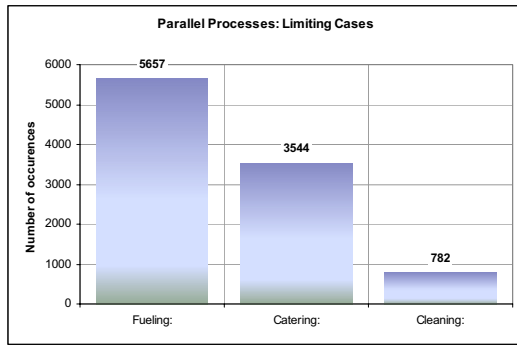


Figure 10. Distribution of limiting process linking de- boarding and boarding

So fueling was the limiting process out of the parallel processes for most of the cases (56%), followed by catering (36%).

V. POTENTIALS TO INCREASE TA RELIABILITY

The field study has shown a set of technical deficiencies which lead to uncertain, non standard processing in many cases. Based on representative interviews with ground handling personnel, individual impact effects were affiliated to the specific aspects. These have to be transferred into adopted process shape distributions, those in turn being subject of further MC simulations to derive the overall potential gain on the turnaround.

A. Addressing time burden to technical deficiencies

With reference to TABLE II, the following potential for reducing the impact effect to uncertainty of the individual deficiencies was found to be applicable according to expert judgment in the field (bold values mean consideration for the following MC simulations as time critical processes):

TABLE V. ADDRESSED TIME EFFECTS TO TECHNICAL DEFICIENCIES

Fighting Technical Deficiencies at Turnaround		
Process	Deficiency counter measure	Judged Variance reduction potential
(De-)Boarding	Second passenger bridge on gate position	10%
Boarding	Innovative Boarding Concept / New seat arrangement.	20%
Catering	Catering door signaling technically supported	10%
	Enhanced vehicle logistics lowering obstruction effects	15%
Cleaning	Reduced waste produced	25%
Fueling	Enhanced vehicle logistics lowering obstruction effects	15%
	Optimized Simultaneous fueling on both wings	20%
	fuel computer upgrade to calculate FOB precisely	5%

(Un-)Loading	Optimized Cargo Door: Panel	10%
	Auto adjustment for Conveyer belts & equipment	30%
Servicing	Cockpit Info granted about GPU removal	10%
	Light-weight adapter for GPU	10%

It shall be noticed that the expected gains shown in the above table may be seen as (realistic) indicators only, and subject of alteration when performing interviews at other airports. Nonetheless, the following section will clearly show how optimization potential and TA reliability do correlate.

B. Conversion into Adopted Distributions

This potential is to be transformed into process distribution behavior, respectively the modeling of reduced process uncertainty to be modeled through

- α (scale parameter) tending to larger values, and
- β (shape parameter) tending to smaller values.

The approximation for the new shape and scale parameter follows the following relationship between the function parameters α , β , and the variance, using the Gamma Function Γ :

$$\Gamma = \int_0^{\infty} t^{x-1} e^{-t} dt \tag{2}$$

$$Var(x) = \alpha^{-2/\beta} \left(\Gamma\left(\frac{2}{\beta} + 1\right) - \Gamma\left(\frac{1}{\beta} + 1\right)^2 \right) \tag{3}$$

This leads us to the following adopted shape parameters for the time critical processes:

TABLE VI. ADOPTED DISTRIBUTIONS – CRITICAL TA PROCESSES

Adopted Distributions – Critical TA Processes	
(De-)Boarding [PAX / min]	Value
α	2,534656
β	19,193186
x offset	10
Boarding [PAX / min]	
α	3,4680421
β	19,23082
x offset	0
loading AFT&FWD [bags / min]	
α	3,252597
β	154,5100
x offset	0
Unloading AFT&FWD [bags / min]	
α	4,047813
β	206,81878
x offset	0

These adopted distributions are now being re-applied to the MC simulation set up to explore the effect of every single counter-measure on the turn around time. With another 10⁵ simulation runs, it comes to the following distribution:

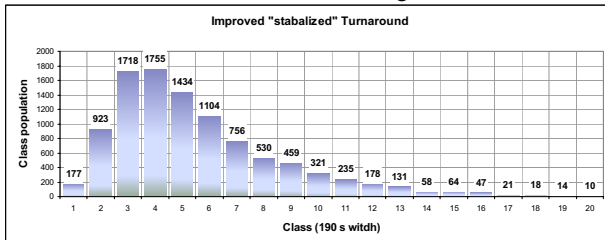


Figure 11. Improved "stabilized" Turn Around Time Distribution

The shape of the improved TAT distribution now follows the over all process characteristics:

- Minimum Time: 1.222 s (20 min)
- Maximum Time: 7.790 s
- Mean: 1.784 s (29,7 min)
- Standard Deviation: 19s

When comparing both scenarios, the following table depicts the potential for stability increase of the turn around:

TABLE VII. POTENTIAL FOR INCREASED PROCESS RELIABILITY

Expected Benefits of technical counter measures for increased process reliability			
Figure	Baseline	Optimized	Gain / Loss
min	1135 s	1222 s	7,67%
max	12486 s	7790 s	-37,61%
mean	1872 s	1784 s	-4,70%
Variance	535 s ²	395 s ²	-26,17%
SD	23 s	19 s	-17,4 %

The 3 critical processes running in parallel for this improved case follow the distribution as depicted below:

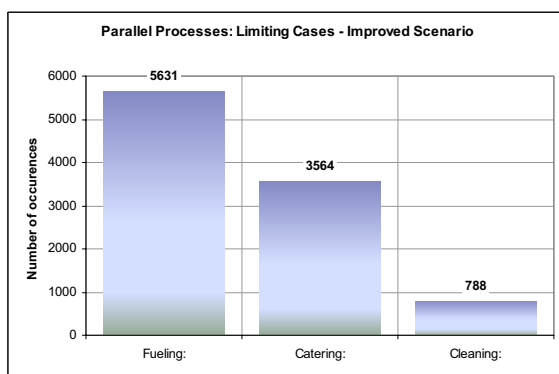


Figure 12. Distribution of limiting process linking – improved scenario

So equal to the baseline scenario, fueling did constrain the parallel processes for 56%, followed again by catering (36%).

C. Reduced process uncertainty affecting TA reliability

Two effects can be noticed when comparing the reference scenario with the improved: Although the mean of the TA time is only decreasing by roughly 4% its stability, indicated by the variance is increasing by roughly 25%. So obviously, installing efficient technical upgrades at the aircraft for the sample case only would significantly improve the turn around reliability achieved through stable (deterministic) processes.

TABLE VII indicates a potential for increasing the turn reliability by roughly 25% when retrofitting the aircraft body accordingly, and optimizing the manual procedures for the processing. This is seen as an important contribution which obviously is worse to be investigated further.

Hence, it also came clear that this type of counter measures will in no way allow complying with the SESAR performance targets for the turn around. This one can so only be achieved by parallelizing Fueling and hereafter catering with the critical processes (de-)boarding and (un-)loading.

For the boarding processes, an in-depth study has recently been completed at TUD, exploring the effects of different strategies (random, inside-out..) onto the process time.

ACKNOWLEDGMENT

The author thanks Ms. Susann Lehmann and Ms. Denise Hentschel, who conducted the time consuming field evaluations even at really bad weather conditions. Thanks also to Christoph Thiel to support the set up of various statistical analyses and Daniel Fiedler for setting up the Web portal.

REFERENCES

- [1] IATA, Airport Handling Manual AHM 810, January 2003
- [2] H. Fricke, S. Lehmann, Cabin Operation – Service Handling within Final Status Report within Phase II, "Cabin Operation – Service Handling" within the project "Cabin Architecture and Design Concept" (CAADC), December 2006
- [3] Airbus Industrie, A320, Airplane Characteristics For Airport Planning, AC, Chapter 5 Terminal Operation, July 1995
- [4] ICAO, Airport Planning Manual, Doc 9184-AN7902 Part 1, 1987
- [5] SESAR Consortium, SESAR Definition Phase, D2 Report "The Performance Target", Page 10, Brussels, November 2006
- [6] Wu, C. L. Monitoring Aircraft Turnaround Operations — Framework Development, Application and Implications for Airline Operations. Transportation Planning and Technology, 2007 (under review)
- [7] Tom Caswell, Kyle Story, and Rafael Frongillo. 2007. The Airplane Seating Problem. Mathematical Contest in Modeling, Consortium for Mathematics and Its Applications (COMAP), 2007
- [8] Michael Bauer, Kshipra Bhawalkar, Matthew Edwards. 2007. Boarding at the Speed of Flight. Mathematical Contest in Modeling, Consortium for Mathematics and Its Applications (COMAP), 2007
- [9] Bachmat, Elkin, Bounds on the Performance of back-to-front airplane boarding policies, July 2006
- [10] ICAO, Doc 9626 - Manual On The Regulation Of International Air Transport (2nd Edition), August 12, 2005, Montreal, Canada
- [11] ICAO, Doc 9562 . Airport Economics Manual, (2nd Edition), Montreal, Canada

Runways sequences and ground traffic optimisation

Raphael Deau

Jean-Baptiste Gotteland

Nicolas Durand

Direction des Techniques et de l'Innovation

7, Av. Edouard Belin - BP4005

F31055 Toulouse Cedex 4

Email: gotteland@tls.cena.fr

Telephone: (+33) 5 62 17 41 68

Fax: (+33) 5 62 25 95 99

Abstract—At the airport level, the new systems involved in the A-SMGCS (Advanced Surface Movement Guidance and Control System) give the possibility to take advantage of some innovative decision support tools bound to the optimisation of the ground traffic management.

In this article, two different tasks assumed by airport controllers are analysed and modeled: the runway sequencing process and the application of runways sequences at the ground level. An existing ground traffic simulator is adapted to measure the potential improvements that could be expected by the use of some optimisation methods applied on these two modeled problems.

I. INTRODUCTION

Airport congestion is still a key point to be studied for the next years: the evolutions that are expected concerning the future management of the ground traffic situations (A-SMGCS: Advanced Surface Movement Guidance and Control System) at the airport control level have obviously several environmental and economical issues.

On major airports, these evolutions can be technically provided by taking advantage of some new developed systems such as the surface radars, the D-GPS (Differential Global Positioning System) associated with the ADS-B (Automatic Dependant Surveillance mode B) and numerous other coordination tools like AMAN (Arrival MANager) and DMAN (Departure MANager).

In this context, this article focuses on the possible optimisation of two major aspects of the airport controllers tasks: the aircraft sequencing at the runway level and the conflict resolution between taxiing aircraft. An airport simulation tool (ATOS: Airport Traffic Optimisation Simulator) is adapted and used to measure the delay reduction that could be expected.

II. RELATED WORK

A. Projects

Two main approaches called AMAN (Arrival MANager) and DMAN (Departure MANager) deal with the aircraft sequencing problem:

- AMAN are decision support tools that provide the controllers with information on arrival flows, including calculated arrival runways sequences. These informations are regularly updated with the actual positions of aircraft in the approach sectors.

- DMAN are planning tools developed to improve the departure flows at airports by optimizing the departure runways throughput.

Some projects as PHARE (Program for Harmonised ATM Research in Eurocontrol) [1] have defined some DMAN and/or AMAN systems but do not coordinate both of them at the airport ground traffic level. PHARE was a project instaurated by Eurocontrol. It was a collaborative research program that investigates an air traffic management concept.

Gate to Gate [2] is a European project which takes into account aircraft from their departure gates to their arrival gates. It mostly improves an AMAN project by managing the air traffic problem. A DMAN is used for the mixed runways and shares out the main informations about the departures. The AMAN has to set the arrivals with the information provided by the DMAN.

The NLR develops a concept which schedules the aircraft on the airport by using constraint relaxation [3]. A lot of parameters are considered, as the runway separations, SID routes, exit points, ... DMAN and AMAN are not explicitly described but implicitly defined.

OPS [4] defines a new DMAN to schedule departures. This project is based on more human interactions (from pilots and controllers for example) but is also more flexible (the users can setup a lot of values). They look forward to define a real-time decision support tool.

Total Airport Management (TAM) [5] tries to merge together as many concepts as possible defined by Eurocontrol, concerning AMAN, DMAN in the ATC in order to optimise the airport capacity and improve the predictability of airports traffic.

CADM (Coordinated Arrival Departure Management) [6] is a concept that mixes a DMAN and an AMAN but does not consider precisely the taxiing times. It uses fuzzy inference mechanism to determine rules to use to set the aircraft sequence.

All these projects focus on the definition and/or the prediction of airports runways sequences for arrivals and/or departures, trying to share as efficiently as possible all the available informations given by the approach sectors and the airport systems. However, the taxiing phases of the flights are still the ones that are the most difficult to predict with a good accuracy: the tested DMAN systems are still not so satisfying

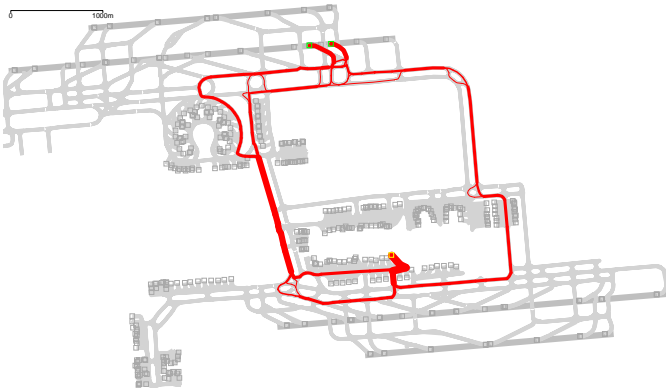


Fig. 1. Roissy map

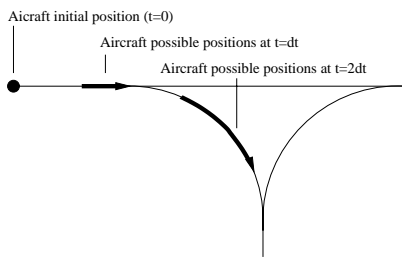


Fig. 2. Speed uncertainties

for ground controllers, as the predicted informations remain uncertain.

B. Background

Previous publications [7], [8], [9], [10] studied the ground traffic optimisation on Roissy Charles De Gaulle and Orly airports. The ATOS (Airport Traffic Optimisation Simulator) simulator was developed and used to compare the efficiencies of several optimisation methods applied to different traffic situations.

The simulator uses a detailed description of the airport taxiways, gates, push-backs, runways and the existing constraints (one-way taxiways for example) to calculate a set of possible paths for each aircraft (see figure 1). The whole traffic is simulated using the real airport flight plans demand of a day of traffic. The flight plans contain information such as the aircraft type, the gate position, the landing or take-off time, the runway used . . .

Using these informations, and for each traffic situation, the possible paths for each aircraft are calculated on a defined time window T_w , taking into account uncertainties on taxiing speeds and a trajectory prediction done: in such a prediction, the future aircraft position is not a point but a set of possible points (a line segment) on the taxiways used (see figure 2).

The problem to solve consists in assigning a path to each aircraft, with holding points if necessary, in order to solve every *conflict* with other aircraft within the time window T_w . Two aircraft are in *conflict* each time the separation standards defined by the operating rules of the airport are violated.

These rules are modelled as follows:

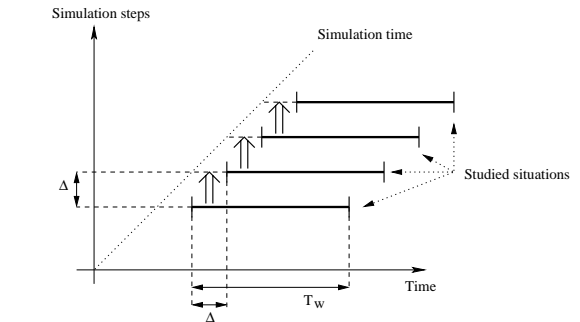


Fig. 3. Shifted windows

- parked aircraft are supposed to be conflict free;
- a minimum separation distance is required between each taxiing aircraft pair;
- time separations are required between aircraft on runways, depending on the aircraft types and their wake turbulence categories.

Among all the possible solutions, the best conflict free trajectory is search according to a global criterion taking into account the delay due to holding points and longer paths chosen.

The problem is very combinatorial because of the number of possible paths and holding points for each aircraft. When the number of aircraft involved increases, the problem can become very difficult to solve using exact methods. Different optimisation methods have been studied by the DSNA/DTI R&D POM team to solve it:

- A sequential deterministic approach consists in first ordering aircraft, give the shortest path to the aircraft with the highest priority, and optimise the $n + 1^{th}$ trajectory solving conflicts with the n previous already optimised trajectories. Optimising the trajectory of one aircraft, solving conflicts with n other known trajectories is quite simple and can be done with an A^* or branch & bound method [HT95], [11].
- A global approach using stochastic optimisation based on genetic algorithms [12], [13] can be used to find the best paths combination. To increase the efficiency of the algorithm, the partial separability of the problem can be used to define *crossover* and *mutation* operators able to optimally recombine current solutions during the convergence process [14].

When a traffic situation is solved at time t , the paths obtained are applied to the moving aircraft during a time Δ ($\Delta < T_w$) called time shift window, to create the updated situation at time $t + \Delta$ (figure 3). A whole day of traffic can thus be simulated using this shifting window modeling, taking into account the uncertainties on aircraft taxiing speeds.

C. Obtained Results

The first simulations results obtained with this modeling have shown some characteristics on the traffic and can help quantifying some parameters of the problem:

- When "realistic" uncertainties on taxiing speeds (from 20% to 50%) are used, the time window must be reduced to 10 minutes to be able to solve the problem. This emphasizes the difficulties encountered on the DMAN concept because with a 10 minutes advance notice, the final take-off time for an aircraft on the runway is to lately known.
- The method using a genetic algorithm consider globally each traffic situation (without classifying aircraft with priority orders) reduces the delays from 1 to 2 minutes per aircraft during peak hours at Roissy Charles De Gaulle Airport, which shows that the possible time saved by optimising the taxiing phase of a flight is quite significant on such airports.
- During peak periods, knowing precisely the paths followed by each aircraft is necessary to manage correctly the CFMU¹ slots: an optimised simulation can help knowing the delay due to congestion and allows to anticipate departures from gates if necessary.

D. Conclusions

These last results show that on large airports such as Roissy Charles De Gaulle, the performance of systems such as AMAN and DMAN depends on an optimised taxiing management. On the one hand, the calculation of runway sequences must take into account the taxiway paths and holding points given to aircraft. On the other hand, it might be better to take into account the full runway sequence to optimise the gate time departures and the aircraft holding points and paths.

As it was shown that the acceptable time window adapted to realistic uncertainties is less than 10 minutes, it is proposed in this article to split the resolution process of each traffic situation in two steps:

- First, the best runways sequences compatible with the current aircraft positions and the known arrival flows should be computed, with a large anticipation time (about 30 minutes if possible). This point is described in part III.
- Then (part IV), the aircraft paths and holding points should be optimised in order to fit as close as possible to these targeted runways sequences, with an adapted anticipation time (less than 10 minutes)

III. RUNWAYS SEQUENCES OPTIMISATION

A. Goals

This part, focuses on finding some optimal runway sequences, respecting a given traffic situation and considering the arrival flows.

A system merging the AMAN and the DMAN informations at the airport level is defined: on runways shared by arrivals and departures, this is the only way to optimise correctly the sequence, and as far as arrivals and departures have to share the same airport infrastructure, they have to be managed together at the ground level.

¹Central Flow Management Unit

B. Problem modeling

To meet these goals, the problem is defined as follows:

- The variables of the problem are the slots that must be assigned to each aircraft;
- The main constraints are the landing times, the minimal remaining taxiing time of each departure, the runway separation rules and the CFMU² slots allocated to some departing aircraft;
- The criterion to minimise measures the departures delays and the deviations from the CFMU slots;
- The prediction time should be close to 30 minutes.

1) *Constraints*: At the airport level, the arrival sequence cannot be substantially modified: the ordering of arrivals is fixed by approach sectors and it is reasonable to consider that each arriving aircraft cannot be delayed more than a reduced time λ ($\lambda < 1$ minute) if these kind of decisions can be taken in advance enough.

Concerning the departures sequences, the minimum possible taxiing time to the runway becomes a constraint for aircraft leaving parking positions: to obtain a feasible sequence, each aircraft's shortest time to runway is considered as a constraint in the sequence search.

The most important constraints used in the sequence optimisation is the separation due to the wake turbulence. The minimum time between two aircraft depends on their weight category. For example: a "low weighted" aircraft cannot take off less than 180 seconds after a "heavy weighted" aircraft has taken off. Three categories of aircraft (and associated wake turbulence) are defined: "low", "medium" and "heavy". The following table shows each separation time (in seconds):

1st acft →	A. L	A. M	A. H	D. L	D. M	D. H
A. L	60	120	180	60	120	180
A. M	60	60	90	60	60	120
A. H	60	60	90	60	60	90
D. L	60	120	180	60	120	180
D. M	60	60	60	60	60	120
D. H	60	60	90	60	60	90

L = Low, M = Medium, H = Heavy
A = Arrival, D = Departure

Other constraints concern aircraft which are assigned some CFMU slots: these aircraft have to be inserted in the sequence between arrivals and classical departures, respecting their fixed CFMU slots. According to the CFMU official acceptance, a CFMU slot is respected if the aircraft takes off from 5 minutes before to 10 minutes after the schedule. However, the actual off-gate times does not always allow to respect all the CFMU slots (as for example when a departure leaves the gate later than its assigned slot). In consequence, the only constraint that is defined as such is the interdiction to take off more than 5 minutes before the slot and the other CFMU requirements will be integrated in the criterion.

²CFMU: Central Flow Management Unit

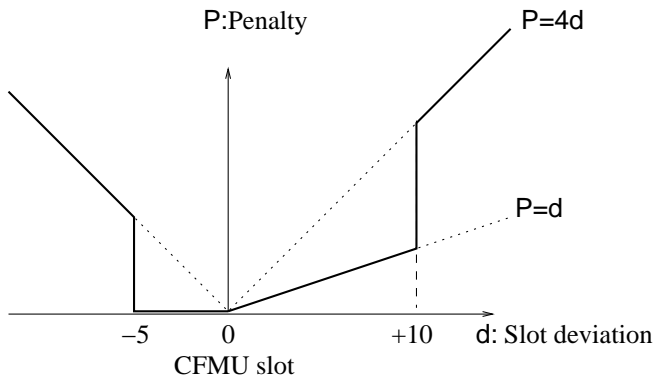


Fig. 4. Criterion for CFMU slots

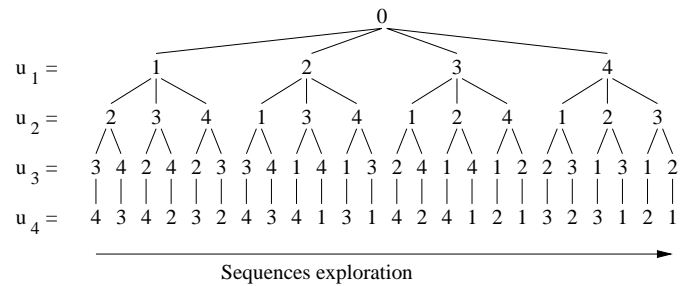


Fig. 5. Sequences tree

2) *Criterion:* In the estimated sequence, the penalties relative to the deviation from each CFMU slot (see fig. 4) and the delay regarding the minimal runway access time for each other departures are computed. The aim of the optimisation is to minimise the sum of those values: the more the CFMU slots are respected and the shorter is the time spent by aircraft on taxiways, the better the sequence is evaluated.

On the other hand, arrivals sequencing delays are not taken into account in the criterion as these little delays (less than λ seconds for each arrival) are not penalising. Of course, sequences which would cause one landing to be delayed more than λ seconds are considered not valid regarding the constraints that are defined and cannot be accepted.

3) *Prediction time:* The sequence optimisation takes into account all the aircraft that may appear in the sequence during the next T_s minutes. The choice of this prediction time is influenced by several factors: as the sequence optimisation problem is very combinatorial (for n aircraft, the complexity of this scheduling problem is proportional to $n!$), the prediction time should be short enough to keep the problem size small enough. On the other hand, it seems logical that the larger the prediction is, the better the sequence will be optimised on the whole day.

C. Resolution

The sequence optimisation problem is a classical scheduling problem that can be solved with deterministic Constraint Satisfaction Problem algorithms.

1) *Problem modeling:* The problem is to find the best sequencing for a given list of aircraft considering the almost fixed arrivals and the separation time between two aircraft. To find the best solution, each permutation of the aircraft list has to be explored.

2) *Branch & bound algorithm:* To solve this problem, a classical branch & bound algorithm is used: each branch of the sequences tree (see fig. 5) is potentially explored and at each node of the tree, the delay generated by the already assigned slots is calculated. Once a solution has been found (i.e.: when one of the leafs of the tree is joined), the cumulated delay obtained updates the "best current solution found" for the next

steps: when a node generates a higher delay than this "best current solution found", the branch is cut.

This algorithm can be summarised as follows:

Mainloop: For each non yet inserted aircraft 'a':

- Insert 'a' in the current (partial) sequence
- Calculate the new resulting penalty
- If this penalty is acceptable then:
 - If some aircraft still remain to insert, explore the next node of this branch (call back the **Mainloop**)
 - Otherwise, mark the current sequence as the best solution found
- remove 'a' from the sequence

To obtain the best cuts in the tree exploration, the initial order plays a major role. It was observed that defining the initial order in accordance to the "ideal" time on the runway (i.e: the minimal runway access time for non-CFMU departures and the CFMU slots for the others) was a good strategy.

Moreover, some other cuts must be implemented to minimise the number of explored nodes. These cuts are relative to some specific characteristics of the problem:

- There is obviously no need to explore a branch in which some arrivals are not in the right order.
- There is no need to explore a branch which swaps two equivalent aircraft in the sequence. The equivalence between aircraft is defined according to their wake turbulence category, their type (arrival or departure) and their CFMU profile(with or without a CFMU slot).

D. Results

In order to measure the efficiency of the proposed runway sequence optimisation method, ground traffic simulations were carried out with a traffic sample relative to a heavy day at Roissy Charles De Gaulle, when the fourth runway was not yet in operation (this period presents the advantage to provide a mixed runway 09-27 shared by both departures and arrivals).

At each simulation step (every $\Delta = 2$ minutes), an optimal sequence is computed for each runway, with different anticipation times (from 20 minutes to 50 minutes), and the resulting theoretical delay for departures in these optimal sequences are recorded (this theoretical delay is measured by the difference

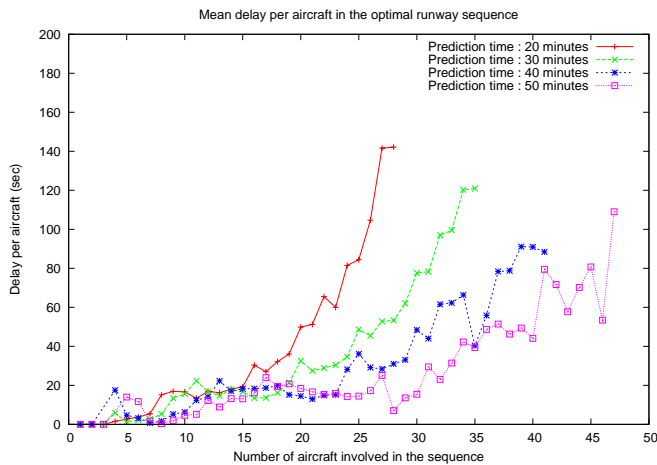


Fig. 6. Mixed runway 27 at Roissy

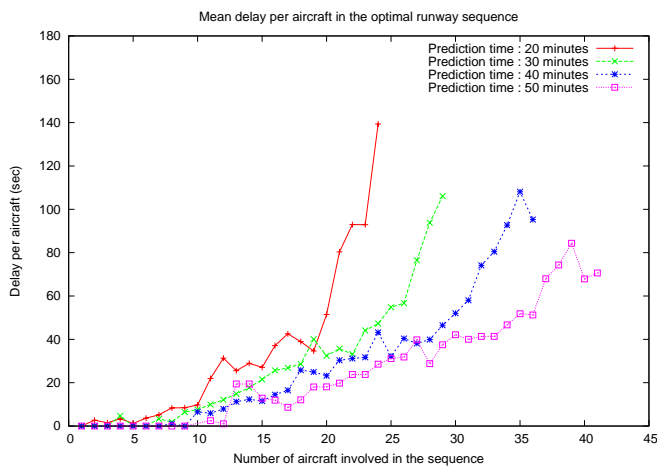


Fig. 7. Runway 26R at Roissy

between the proposed slots in the sequence and the “ideal” slots for aircraft.)

The figures 6 and 7 give the mean delay per aircraft as a function of the number of aircraft involved in the runway sequence, for the two runways used for departures: runway 27 (shared with arrivals) and runway 26R (only used for departures.)

On these figures, the influence of the prediction time window on the quality of the runway sequence can be observed and confirms the expected conclusion: the highest the prediction time is, the best the runways sequences are. Moreover, this relation can be quantified: when the prediction time grows from 20 minutes to 50 minutes, the mean departures delay decreases from 140 seconds to less than 100 seconds in heavy traffic situations.

Of course these delays measures are theoretical and could only be realised if the calculated runway sequences were exactly applied. This is the subject of the next part, in which the retained prediction time window that is considered for the runway sequences is 30 minutes, as it seems to be the best

compromise between what can be expected from AMAN and DMAN systems and what can be treated by the optimisation process.

IV. APPLICATION TO THE GROUND TRAFFIC SITUATIONS

A. Goals

In this part, the runway sequences previously optimised are considered as a target for the ground traffic solver: the objective is to fit as close as possible to the predefined sequences while solving aircraft conflicts on taxiways and gates.

Many options can be studied in this framework: it can be interesting, for example, to favor the earliest departures of each runway sequence and to delay in priority aircraft which might arrive on the runway in advance according to their allocated slot. It can also be interesting to strategically sequence departures before they leave their gate position, by assigning them an initial delay.

These different concepts will be studied and compared to the results obtained without the runway sequence optimisation in the last part of this article.

Therefore, the different optimisation methods must be adapted to take into account the new objective (i.e. fit to the optimised runways sequences) and not only minimise the aircraft delay: this part details the modification that were done for each resolution method.

B. Sequential resolution method

The sequential resolution method deals with a simplified problem, in which aircraft are initially sorted and then considered one after an other (first considered aircraft have priority on last considered ones). As a consequence, this resolution method can be easily adapted to fit some given runways sequences, as these sequences will directly provide the aircraft classification to be considered: each aircraft a can be associated with its slot t_a in the concerned runway sequence, and the sequential method will be applied in the order given by (t_a) .

In most of the cases, this process ensures that on each runway, departures always take off in the order defined by the runway sequence (except in some very particular cases relative to the limited aspect of the prediction time window.) However, on runways shared both by departures and arrivals, the order between arrivals and departures is not ensured: if a departure planned just before an arrival reaches the runway area too late, it will be forced to take off after the arrival (in this case, the resolution method has to modify the aircraft classification to find a conflict free solution).

Another point has to be considered, concerning the departures constrained by CFMU slots: generally, the CFMU slots correspond to some delayed take-off times. When such aircraft are involved in a runway sequence, the classification of departures is not the only factor to assume: the exact take-off time of concerned aircraft must also precisely correspond to their given CFMU slots, in agreement with the official CFMU acceptance (no more than five minutes before the

slot, nor ten minutes after). During low traffic periods, these departures could take off much earlier while respecting the runway sequence order. For these reasons, a minimal departure time (from the gate) has to be assigned to aircraft [9] and the resulting delay must be propagated over the following departures, in order to keep a consistent sequencing of departures.

Therefore, an *initial wait* w_d is calculated for each departure d , as a function of:

- its minimal runway access time $t_{\min d}$,
- its optional CFMU slot t_{cd} ,
- the official acceptance for CFMU slots ($\delta_c = 5$ minutes),
- and the initial required wait w_p of the prior aircraft p in the sequence (all aircraft are sorted in the order given by the sequence).

For the first aircraft d_0 :

- If d_0 has a CFMU slot t_{cd_0} :

$$w_{d_0} = \max\{0, t_{cd_0} - \delta_c - t_{\min d_0}\}$$

- Otherwise,

$$w_{d_0} = 0$$

For the following departures d_i ($i > 0$):

- If d_i has a CFMU slot t_{cd_i} :

$$w_{d_i} = \max\{0, t_{cd_i} - \delta_c - t_{\min d_i}, w_{d_{i-1}} + t_{\min d_{i-1}} - t_{\min d_i}\}$$

- Otherwise,

$$w_{d_i} = \max\{0, w_{d_{i-1}} + t_{\min d_{i-1}} - t_{\min d_i}\}$$

The best trajectory for a departure d_i (looked forward by the sequential branch & bound algorithm) is the one corresponding to the shortest path allowing a delay as near as possible to the requested wait d_i .

C. Genetic algorithm solver

The conflict resolution method based on a genetic algorithm consider globally each traffic situation, without assuming any classification between aircraft: thus, the logical way to adapt this method to the new problem (which is the application of some predefined runways sequences) consist in modifying the global criterion to minimise: in this way, the predefined sequences will be considered as a goal but not as a constraint, which is necessary to ensure that some acceptable solutions still exist, even when the traffic situation does not allow to carry out the targeted sequences.

Concerning acceptable (conflict free) solutions, the global criterion is defined as the sum of each specific criterion relative to each aircraft: for a departure, this specific criterion must be refined, in order to estimate the difference between the take-off time that would result from the proposed solution and the take-off time targeted in the optimal runway sequence.

Different definitions of such a criterion can be considered. The main difficulty is obviously relative to the difference between the anticipation time used for the ground conflicts resolution and the one used to compute the runways sequences: on large airports such as Roissy Charles De Gaulle, departures

taxiing times can easily exceed the prediction time window, so that the take-off times that will result from the proposed paths and holding positions are uncertain on the long range.

Moreover, looking forward to hold the departures when their positions seem to be in advance compared to their targeted take-off slots is not appropriate, as such a ground traffic management would clearly risk to propagate every form of ground delay to the whole airport.

As a consequence, the proposed criterion is still proportional to aircraft delay, but a balance is applied, in order to penalise more the delay of an aircraft when its minimal runway access time becomes closer to its targeted slot.

With this kind of criterion (based on delay), the same treatment as before has to be considered concerning the management of the CFMU slots (see IV-B): an initial wait w_d has to be computed for each departure d , in order to ensure the correct insertion of these particular departures in the rest of the traffic. Obviously, this initial wait also affects the definition of the criterion.

Finally, the penalty $P(a)$ to be minimised for each aircraft a is estimated as a function of the delay dl_a of the aircraft (including assigned wait and/or path lengthening), the minimal runway access time $t_{\min a}$ and the targeted slot t_a of the aircraft, as follows:

- For an arrival a :

$$P(a) = dl_a \text{ (unchanged)}$$

- For a departure d with a CFMU slot t_{cd} which is late ($t_{\min d} > t_{cd} + \delta_c$):

$$P(d) = 20 * (dl_d + t_{\min d} - t_{cd})$$

- For a departure d with a CFMU slot t_{cd} which is in advance ($t_{\min d} < t_{cd} - \delta_c$) and which required wait is w_d :

$$P(d) = 10 * (|dl_d - w_d| + t_{cd} - t_{\min d})$$

- For each other departure d which required wait is w_d :

$$P(d) = 5 * (|dl_d - w_d| + \max(0, t_{\min d} - t_d))$$

In these definitions, the balance that is applied to departures delay is defined in order to favor as often as possible departures against arrivals.

V. RESULTS

To measure the efficiency of the proposed optimisation methods, four simulations are carried out (with the same traffic sample as in III-D): the two different ground traffic solvers (i.e. the sequential method and the genetic algorithm solver) are tested on two scenarios: in the first one, there is no runway sequence to target, while in the second one, the optimal runways sequences are computed and targeted as explained before.

These four simulations are compared by the generated delay, the slots deviations for concerned departures, and the differences between the targeted slots and the final ones.

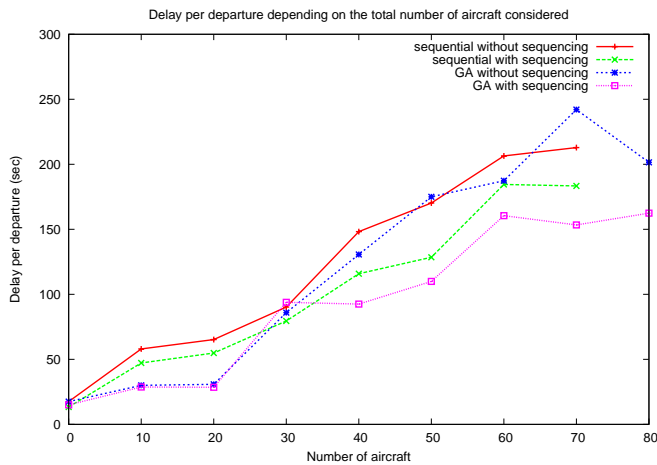


Fig. 8. Mean delay per departure (seconds)

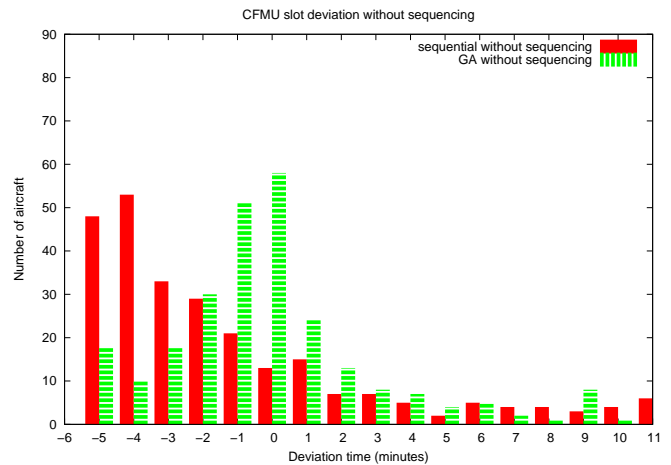


Fig. 10. CFMU deviations without sequencing

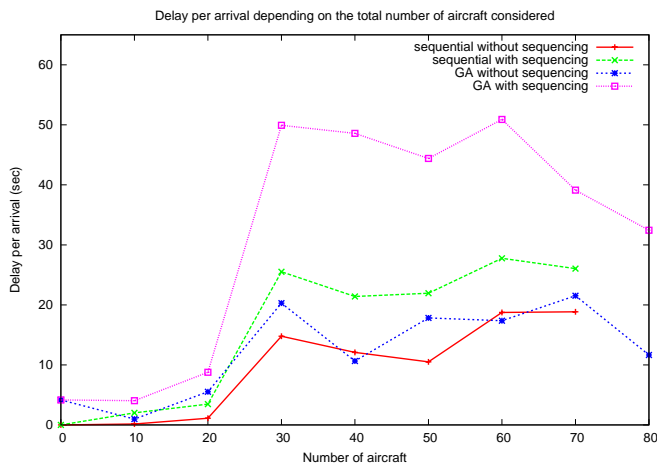


Fig. 9. Mean delay per arrival (seconds)

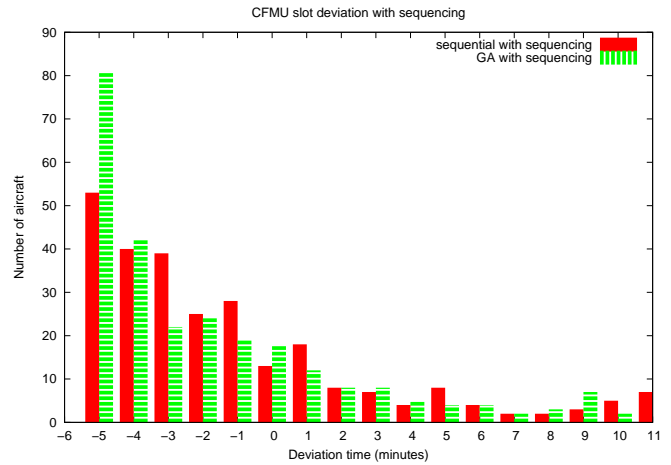


Fig. 11. CFMU deviations with sequencing

A. Delay comparisons

The following table gives the global results of each simulation concerning the delay of “classical” departures (i.e. the departures that are not constrained by a CFMU slot) and arrivals:

		Aircraft delay	
		Without	With sequencing
Sequential method	Dep.	16h42 2min.10'/acft	13h43 1min.45'/acft
	Arr.	1h56 10'/acft	3h12 16'/acft
GA	Dep.	14h46 1min.55'/acft	11h26 1min.25'/acft
	Arr.	2h12 11'/acft	5h33 28'/acft

Aircraft total and mean delay

These delays can also be measured as a function of the number of taxiing aircraft on the airport, as shown on figure 8: the mean delay is calculated for each period of 10 minutes

of the day and is put in relation with the corresponding number of taxiing aircraft in this period.

As one can see, the runway sequencing optimisation process enhances the departures results of both solvers, and the sequential solver becomes almost as efficient as the genetic algorithm one. Globally, the difference between a “basic” management of taxiing aircraft (i.e. the sequential method without runway sequencing) and the final genetic algorithm solver is really significant: 45 seconds per aircraft are saved on the whole day, and more than 1 minute per aircraft can be saved during traffic peaks.

Of course, the arrivals delay shown on figure 9 follow an opposite progression, especially with the final genetic algorithm solver (for which the defined criterion voluntarily give priority to departures). In an operational point of view, the 20 seconds of delay added per arrival in compensation to departures management enhancement should be profitable.

B. Deviations to CFMU slots

The figures 10 and 11 shows the distribution of the CFMU slots deviations observed for the concerned departures.

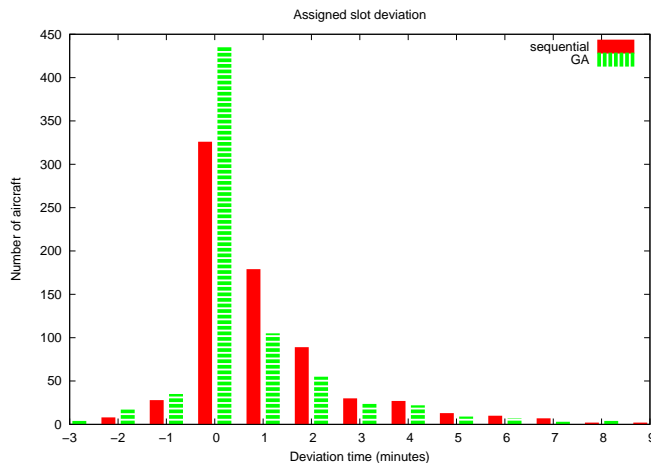


Fig. 12. Deviations to targeted slots (minutes)

These results show that the CFMU slots are quite respected (in fact, the only exceptions concern departures that are ready to leave their gate too late to catch their slot, in accordance with the CFMU requirements).

As no penalty was defined from 5 to 0 minutes before the CFMU slot, the major slots deviations observed are globally spread in this time period. As a consequence of the criterion defined for the genetic algorithm solver, the runway sequencing process allows to concentrate the departures take-off times at the beginning of the allowed period.

C. Deviations to targeted slots

The difference between the calculated runway sequences and the final ones is measured: the figure 12 shows the number of aircraft concerned by each value of slot deviation formulated in minutes.

This figure shows that the runway sequences generated with the genetic algorithm solver are closer to the targeted ones, but there are still a lot of aircraft that don't succeed in catching correctly their initial assigned slot. This result shows that the management of the taxiing aircraft is a very critical step and should really take advantage of some appropriate optimisation methods, rather than only be defined with some basic circulation rules.

VI. CONCLUSIONS AND FURTHER WORK

In the first part of this article, a classical (and exact) method has been implemented and tested to compute some optimal runways sequences at Roissy Charles De Gaulle airport, taking into account each traffic situation with precision, adding the expected arrival flow fixed by approach sectors to the actual ground positions of all taxiing aircraft. The simulation results show that the runways sequences can be largely enhanced if they are well organised with a sufficient anticipation time (the largest possible anticipation time greater than 30 minutes). As a consequence, the delay (and then the taxiing times) assigned to departures at the airport level could be significantly reduced

if the surface management of the airport allows to perform these targeted optimal runways sequences.

The second part of this article explores and proposes some concepts that should be developed at the airport control level to manage correctly the taxiing aircraft, while targeting the computed runways sequences. This concepts can decrease significantly the departures delay, but the simulations carried out also confirm that this task is very complex during traffic peaks, as far as the speed of the taxiing aircraft is not precisely known and as far as the conflicts that have to be solved between these aircraft affect the feasibility of the runways sequences. The ground management of aircraft becomes again more complex for runways shared by both departures and arrivals, as departure delays can sometimes totally change the optimal sequence to target.

Further work will consist in refining the way to target a runway sequence at the airport level, by considering for example some new methods to enhance the estimation of the appropriate decisions that must be taken concerning the taxiing aircraft, in order to keep the ground traffic situations consistent with the targeted runways sequence.

Another development will concern the generalisation of the constraints defined to perform the runways sequences optimisation, taking into account more operational issues, like the sequencing rules used by the airport controllers relative to aircraft SID (Standard Instrument Departure) and STAR (Standard Terminal Arrival Route).

REFERENCES

- [1] Eurocontrol *PHARE presentation* <http://www.eurocontrol.int/>
- [2] Eurocontrol *Gate to Gate project* <http://www.eurocontrol.int/> 2004
- [3] P. van Leeuwen and N. van Hanxleden Houwert *Scheduling aircraft using constraint relaxation* NLR: Nationaal Lucht & Ruimtevaartlaboratorium (National Aerospace Laboratory), 2003
- [4] H.W.G. de Jonge and E.E. Tuinstra and R.R Seljée *Outbound Punctuality Sequencing by Collaborative Planning* NLR: Nationaal Lucht-en Ruimtevaartlaboratorium, (National Aerospace Laboratory), 2005
- [5] Günther, Y., A. Inard, B. Werther, M. Bonnier, G. Spies, A. Marsden, M.-M. Temme, D. Bohme, R. Lane and H. Niederstrasser *Total Airport Management (Operational Concept and Logical Architecture)* 2006
- [6] D. Bohme, R. Brucherseifer and L. Christoffels, *Coordinated Arrival Departure Management*, German Aerospace Center DLR, Institute of Flight Guidance, Braunschweig, Germany, 2007.
- [7] Gotteland, J-B. *Optimisation du trafic au roulage sur des grands aéroports* PHD Thesis, EDIT, 2004.
- [8] Gotteland, J-B. and Durand, N. and Alliot, J-M. *Genetic Algorithms Applied to Airport Ground Traffic Optimization* Proceedings of the Congress on Evolutionary Computation, Canberra, 2003
- [9] Gotteland, J-B. and Durand, N. and Alliot, J-M. *Handling CFMU slots in busy airports* Proceedings of the Air Traffic Management R&D Seminar, Budapest, 2003
- [10] B. Pestic, N. Durand and J-M. Alliot. *Aircraft Ground Traffic Optimization using a Genetic algorithm* Proceedings of the Genetic and Evolutionary Computation Conference, San Francisco, 2001.
- [HT95] R. Horst and H. Tuy. *Global Optimization: Deterministic Approaches* Springer-Verlag, 1996.
- [11] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin. *Network Flows, Theory, Algorithms and Applications* Prentice Hall, 1993.
- [12] D-E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning* Addison-Wesley, 1989.
- [13] Durand, N. and Alliot, J-M. and Noailles, J. *Automatic aircraft conflict resolution using genetic algorithms* Proceedings of the Symposium on Applied Computing, Philadelphia, 1996.
- [14] N. Durand and J. M. Alliot. Genetic crossover operator for partially separable functions. *Genetic Programming Conference*. 1998.

Track 7

Airline Operations

Accuracy of Reinforcement Learning Algorithms for Predicting Aircraft Taxi-out Times

(A Case-study of Tampa Bay Departures)

Poornima Balakrishna, Rajesh Ganesan, Lance Sherry
Center for Air Transportation Systems Research
Department of Systems Engineering and Operations Research
George Mason University
Fairfax, U.S.A
pbalakri@gmu.edu, rganesan@gmu.edu, lsherry@gmu.edu

Abstract—Taxi-out delay is a significant portion of the block time of a flight. Uncertainty in taxi-out times reduces predictability of arrival times at the destination. This in turn results in inefficient use of airline resources such as aircraft, crew, and ground personnel. Taxi-out time prediction is also a first step in enabling schedule modifications that would help mitigate congestion and reduce emissions. The dynamically changing operation at the airport makes it difficult to accurately predict taxi-out time. In this paper we investigate the accuracy of taxi out time prediction using a nonparametric reinforcement learning (RL) based method, set in the probabilistic framework of stochastic dynamic programming. A case-study of Tampa International Airport (TPA) shows that on an average, with 93.7% probability, on any given day, our predicted mean taxi-out time for any given quarter, matches the actual mean taxi-out time for the same quarter with a standard error of 1.5 minutes. Also, for individual flights, the taxi-out time of 81% of them were predicted accurately within a standard error of 2 minutes. The predictions were done 15 minutes before gate departure. OOOI data available in the ASPM database maintained by the FAA was used to model and analyze the problem. The prediction accuracy is high even without the use of detailed track data.

Keywords-taxi-out delay; prediction; reinforcement learning.

I. INTRODUCTION

Flight delays have a significant impact on the nation's economy. The United States National Airspace System (NAS) is a complex system consisting of several components including the administration, control centers, airlines, aircraft, and passengers. Flight delays propagate over the NAS and

increases with time over the length of the day due to the cascading effect. Stakeholders, particularly the ground and tower controllers, are overwhelmed during peak hours when the number of departures and arrivals increase; at times beyond capacity. Taxi-out delay is a major component of flight delays. Predictability of taxi-out time would help ease congestion and mitigate delays via better gate departure planning. Taxi-out time of a flight is defined as the time between gate pushback and time of takeoff. Increased predictability of taxi-time at the departure airport will also increase planning efficiency at arrival airports.

Delays are caused by several factors. Some of these include increased demand, weather, near-capacity operation of major hub airports, and air traffic management programs such as Ground Delay Programs (GDPs) and Ground Stops (GS). The delay phenomenon is continuously evolving and is both stochastic and elastic in nature. The stochastic nature is due to the uncertainties that lie at the local level (such as the local control tower, arrival/departures movements on ground, and human causes), system level (such as GDP), and in the environment (weather). The elastic behavior is due to the fact that delay could be adjusted (positively or negatively) by flying speed, taking alternate routes, turnaround time on the ground, and position in the departure clearance queue especially during busy hours of the airport. In order to minimize the taxi-out delay component of the total delay, it is necessary to accurately predict taxi-out under dynamic airport conditions. This information in turn will allow the airlines to better schedule and dynamically adjust departures, which minimizes congestions, and the control towers will benefit from smoother airport operations by avoiding situations when demand (departure rates) nears or exceeds airport capacity.

II. LITERATURE REVIEW

Several research attempts have been documented to understand the departure process at airports. These include both simulation models and analytical formulations. The Departure Enhanced Planning And Runway/Taxiway Assignment System (DEPARTS) [1] developed at MITRE Corporation attempts to reduce taxi times by generating optimal runway assignments, departure sequencing and departure fix loading. Results of their analysis also indicate that pushback predictability could influence all phases of flight and traffic flow management.

A simulation based study of queueing dynamics and "traffic rules" is reported in [2]. They conclude that flow-rate restrictions significantly impact departure traffic. The impact of downstream restrictions is measured by considering aggregate metrics such as airport throughput, departure congestion, and average taxi-out delay. Other research that has focused on departure processes and departure runway balancing are available in [3,4]. Many statistical models that consider the probability distribution of departure delays and aircraft takeoff time in order to predict taxi-time have evolved in recent years [5,6].

In [7] a queueing model for taxi-time prediction is developed. They identify takeoff queue size to be an important factor affecting taxi-out time. An estimate of the takeoff queue size experienced by an aircraft is obtained by predicting the amount of passing that it may experience on the airport surface during its taxi-out, and by considering the number of takeoffs between its pushback time and its takeoff time. However, this requires prior knowledge of actual takeoff times of flights and hence may be unsuitable for planning purposes. The model is valid for a specific runway configuration since the runway configuration at the future time of taxi-time prediction is unknown. Suggested extensions to the model include a runway configuration predictor. A queueing model based on simulation to test different emissions scenarios related to duration of taxi-out was developed in [8]. Some of the scenarios that are considered are redistribution of flights evenly across the day, and variation in number of departures under current capacity. The study showed that lower taxi-out times (and thus lower emissions) are experienced by airlines that use less congested airports and don't rely on hub-and-spoke systems. Other research that develops a departure planning tool for departure time prediction is available in [9-13].

Direct predictions attempting to minimize taxi-out delays using accurate surface surveillance data have been presented to literature [14,15]. Recent work using surface surveillance data presented in [16] develops a bivariate quadratic polynomial regression equation to predict taxi time. In this work data from Aircraft Situation Display to Industry (ASDI) and that provided for Northwest Airlines for DTW (Flight

Event Data Store, FEDS) were compared with surface surveillance data to extract gate OUT, wheels OFF, wheels ON, and gate IN (OOOI) data for prediction purposes.

A Bayesian networks approach to predict different segments of flight delay including taxi-out delay has been presented in [17]. An algorithm to reduce departure time estimation error (up to 15%) is available in [18], which calculates the ground time error and adds it to the estimated ground time at a given departure time. A genetic algorithm based approach to estimating flight departure delay is presented in [19].

It is useful to keep in mind that a lot of the data is proprietary and the different attempts in the literature use different data sources depending on accessibility.

III. RL METHODOLOGY

In this research a machine learning approach is used for the task of taxi-out time $a \in A$ prediction, where A denotes the action space. The evolution of system state $x \in X$ is modeled as a Markov chain, where X denotes the system state space. The decision to predict the taxi-out time based on the system state is modeled as a Markov decision process (MDP). For the purpose of solving the MDP, it is necessary to discretize X and A . Due to the large number of state and action combinations (x,a) , the Markov decision model is solved using a machine learning (reinforcement learning (RL), in particular) approach.

The purpose of the RL estimator is to predict taxi-out time given the dynamic system state. The input to RL is the system state and the output of the learning process is a reward function $R(x,a)$ where $a \in A$ is the predicted taxi-out values. The utility function (reward) $R(x,a)$ is updated based on the difference between the actual and predicted taxi-out values $r(x,a,j)$. We define reward $r(x,a,j)$ for taking action a in state x at any time t that results in a transition to state j , as the absolute value of error $r(x,a,j) = |Actual\ Taxi-out - predicted\ Taxi-out|$ resulting from the action. The transition probability in a MDP can be represented as $p(x,a,j)$, for transition from state x to state j under action a . Then the prediction system can be stated as follows. For any given $x \in X$ at time t there is a prediction a such that the expected value of error (*Actual - predicted Taxi-out*) is zero. Theoretically, the action space for the predicted taxi-out could have a wide range of numbers. However, in practice, for a non-diverging process, the action space is quite small, which can be discretized to a finite number of actions.

For practical implementation since transition probabilities $p(x,a,j)$ are not known, we use the reinforcement learning

version of the Bellman's optimality equation [21] to update $R(x,a)$ as follows

$$R^{t+1}(x,a) = (1-\alpha)R^t(x,a) + \alpha[r(x,a,j) + \beta \min_{b \in A} R^t(j,b)] \quad x, j \in X \quad a \in A$$

where α is a learning parameter that is decayed over time, and β is the discount parameter.

The state variables $x = \{x_1, x_2, x_3, x_4, x_5\}$ for the taxi-time prediction problem were determined by analyzing the available data. Analysis of the data suggests that for a specific aircraft that is scheduled to pushback, the number in queue at the runway (x_1), the number of departure aircraft co-taxiing (x_2), and the number of arrival aircraft co-taxiing (x_3) are the major factors that influence taxi-out time. In addition, taxi-out time changes gradually over the day. The taxi-out time during a given quarter was found to depend on the taxi-out times of the previous two quarters. So the average taxi-out time of the previous two quarters was considered as a factor influencing taxi-out time (x_4). Along these lines, the time of day (x_5) was also included as a factor. Thus, there are 5 variables that comprised the state vector.

Several measures of performance such as discounted reward, average reward, and total reward can be used to solve a MDP. At the beginning of the learning process, the R-values are initialized to zeros. When the process enters a state for the first time, the action is chosen randomly since the R-values for all actions are zero initially. In order to allow for effective learning in the early learning stages, instead of the greedy action (action with lowest R-value) the decision maker, with probability P_r , chooses from other actions. The choice among the other actions is made by generating a random number from a uniform distribution. The above procedure is commonly referred to in RL literature as exploration. The RL based functional block diagram is shown in Fig. 1. Theoretical details of the RL algorithm can be obtained from [20-24].

A. Obtaining Predicted Taxi-Out Time

Once learning is completed, the R-values (reward) provide the optimal action choice for each state. At any time t as the process enters a state, the action a corresponding to the lowest non-zero R-value indicates the predicted taxi-out time a . In what follows we present the steps of the RL algorithm in the implementation phase. The RL estimator was coded in MATLAB®.

B. Steps in RL

- **Step 1:** Once the states, actions, and the reward scheme are set up, the next step is to simulate the $t+45$ look-ahead window. Assume 15 minute decision (prediction) epochs *i.e.* prediction was done for flights in a moving window of length t to $t+15$ minutes. This means that for each departing flight in the 15 minute interval from current time, the airport dynamics

was simulated for 30 minutes from its scheduled departure time.

- **Step 2:** Simulate the first 15 minute window. For each flight in the window obtain the system state x . To calculate average taxi-out times before current time t , actual flight data between t and $t-30$ are used. Initialize $R(x,a)$ to zeros.

- **Step 3:** If exploration has decayed go to step 4, else choose arbitrary actions (predictions from set A). The window is then moved in 1 minute increments and all flights in the window are predicted again. This means that every flight, unless it leaves before scheduled time, has its taxi-out time predicted at least 15 times. Simulate the new window of 15 minutes. Find the next state j for each flight. Compute $r(x,a,j)$. Update reward $R(x,a)$ using the fundamental Robbin-Monro's stochastic approximation scheme [25] that is used to solve Bellman's optimality equation [21] provided earlier.

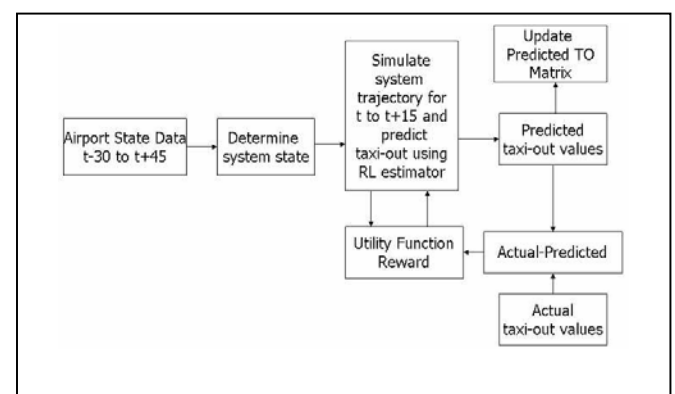
- **Step 4:** If learning phase is in progress, choose greedy action a from set A (action corresponding to the lowest R-value). The window is then moved in 1 minute increment and all flights in the window are predicted again. Simulate the new window of 15 minutes. Find the next state j . Compute $r(x,a,j)$. Update $R(x,a)$.

- **Step 5:** Continue learning by simulating every 15 minute interval, until 45 minutes have been completed. Next, move the window of width 60 minutes by a fixed time increment (say 15 minutes) and repeat learning by going to Step 2.

- **Step 6:** Continue learning with several months of ASPM data until a stopping, or a near-optimal criterion is reached such as $|R^{t+1}(x,a) - R^t(x,a)| \leq \epsilon$ where ϵ is a very small number.

- **Step 7:** Once learning is complete, the optimal prediction for a given state is the one that corresponds to the minimum R-value for that state.

Figure 1. Reinforcement Learning Based Functional Block Diagram for Taxi-Out Time Prediction



IV. RESULTS

A. Data Source

OOOI (Out,Off,On,In) data was obtained for Tampa International Airport (TPA) from the ASPM (Aviation System Performance Metric) database maintained by the FAA (Federal Aviation Administration). Data from June 1st 2007 up to August 25th 2007 was used to train the RL based taxi-time estimator, and data for August 26th to August 31st was used for testing the accuracy of prediction. OOOI data provides the following information for each recorded flight – Scheduled pushback time from the gate, Actual pushback time from the gate, Actual Wheels Off time, Actual Wheels On time at the arrival airport, and Actual In time which is recorded when the aircraft reaches the gate after the taxi-in process. In addition, the ASPM database also provides an airline (not individual flight) specific seasonal average for the nominal or unimpeded taxi-out time and taxi-in time. In this research we assume that if an aircraft completes the corresponding nominal taxi-out time, it joins a runway queue.

It is possible that an aircraft pushes back from the gate and for varying reasons may have to return to the gate and pushback again. It is unclear as to whether the actual pushback time reported by the airlines indicates the first pushback or the second pushback. The Bureau of Transportation Statistics (BTS) recently issued a directive [26] requiring all airlines reporting data to ensure that the first pushback time be recorded as the actual gate-out time. This clearly influences the measured taxi-out time.

B. Observations

We adopt two methods to analyze the results. First we compute the taxi-out time prediction accuracy for individual flights on a given day. Second, we evaluate the prediction accuracy of average taxi-out times in 15 minute intervals of the day. Table 1 below summarizes the prediction accuracy for individual flights for six days in August 2007.

TABLE 1. A comparison of prediction accuracy for individual flights across days of August 2007 for TPA.

Day (August 2007)	26th	27th	28th	29th	30th	31st
Mean Actual Taxi-Out time	11.53	11.73	11.17	11.21	11.06	16.06
Mean Predicted Taxi Time (min)	10.82	11.20	10.09	10.38	10.74	11.96
Std. Dev. Actual Taxi Time (min)	3.80	4.76	3.30	4.02	6.29	8.93
Std. Dev. Predicted Taxi Time (min)	1.86	4.23	1.29	1.76	1.93	3.26
Median Actual Taxi Time (min)	10.8	10.2	10.2	10.2	10.2	12
Median Predicted Taxi Time (min)	10	10	10	10	10	12
% of Flights with RMSE of 2 min	87.5	81.62	83.15	79.18	84.48	69.83

The results in Table 1 indicate that the mean of predictions for a given day are comparable to the mean of actual taxi-out times. We note that the standard deviations of predicted taxi-out time values are not very closely matched with the standard deviations of actual taxi-out times. A possible reason for this is that we consider a flight to enter the runway queue if it has not taken off by the end of its nominal or unimpeded taxi-out time. The nominal taxi-out time data available for this research is however a seasonal average specified for each airline, and not for each individual flight. This is also our only measure in some sense of gate to runway distance. This will undoubtedly introduce further uncertainty in our predictions since factors such as runway configuration are not captured in this average. Also, predictions of taxi-out times are made 15 minutes prior to scheduled pushback of flights. In what follows, we analyze an alternative method to compare the prediction results by considering the following four cases.

Case 1: Consider all flights *scheduled* to pushback in a specific quarter (15 minute interval) of the day. Plot their corresponding mean predicted and mean actual taxi-out time with respect to the same quarter. Note that all flights that are scheduled to pushback in a certain quarter may not take off together (around the same time). This is because taxi-out time is defined as the time elapsed between pushback from the gate and take-off time; and thus depends on several other factors influencing individual flights such as distance of gate from runway, enforcement of downstream restrictions such as Ground Delay Programs (GDPs), Ground Stops (GS) and Miles-In-Trail (MIT).

Case 2: Consider all flights that *actually* took off in a specific quarter of the day. Plot their corresponding mean predicted and mean actual taxi-out time with respect to the same quarter. In this case the flights being considered would have pushed back from the gate at different times spread over different quarters.

Case 3: Consider all flights that were *predicted* (by the algorithm) to take off in a certain quarter (predicted off time can be computed by adding the predicted taxi-out time to scheduled gate-out time). Plot their corresponding mean predicted and mean actual taxi-out times with respect to that quarter.

Case 4: Consider all flights that were predicted to take off in a certain quarter and plot their corresponding mean *predicted* taxi-out times. Now, consider all flights that actually took off in that same quarter. Plot their corresponding mean *actual* taxi-out times. Here we note that the flights that *actually* took off in the quarter being analyzed may not exactly match the set of flights that were *predicted* to take off in that same quarter – this is an inherent limitation of the data available in the ASPM database. Information regarding downstream

restrictions affecting individual flights is not available. Hence we cannot account for passing of aircrafts on the taxiway. Plots of the four cases are provided in Fig. 2-5 for 26th August 2007.

It is easy to see that cases 1-3 represents the average accuracy of prediction for a specified group of flights per quarter, while case 4 discusses average accuracy of prediction for a specified time interval of day which indicates behavior of the airport. In this study, Case 4 is extremely useful in predicting average airport taxi-out time trends approximately 30 minutes in advance of the given time of day (specifying the take off quarter).

For each of the days and for each of the four cases, the accuracy of prediction was measured as the percentage of the time for which the mean predicted taxi-out time per quarter matched the mean actual observed taxi-out time for the same quarter with a standard error of 1.5 minutes. The results are tabulated in Table 2.

TABLE 2. A comparison of prediction accuracy of averages across days of August 2007

Day	Case 1	Case 2	Case 3	Case 4
(August 2007)	% <i>accuracy</i> <i>Standard</i> <i>Error of</i> <i>1.5 min</i>	% <i>accuracy</i> <i>Standard</i> <i>Error of</i> <i>1.5 min</i>	% <i>accuracy</i> <i>Standard</i> <i>Error of</i> <i>1.5 min</i>	% <i>accuracy</i> <i>Standard</i> <i>Error of</i> <i>1.5 min</i>
26 th	97.1014	97.1014	100.0000	95.6522
27 th	88.4058	84.0580	88.4058	95.6522
28 th	88.4058	92.7536	78.2609	92.7536
29 th	89.8551	92.7536	92.7536	92.7536
30 th	98.5507	98.5507	98.5507	95.6522
31 st	89.8551	86.9565	94.2029	89.8551

V. CONCLUSIONS

The analysis using the artificial intelligence methodology (reinforcement learning) indicates that on an average, for a given day, the accuracy of mean predicted taxi-out time per quarter in comparison to the actual taxi-out time for the same quarter is approximately 93.7% (case 4) with a standard error of 1.5 minutes. It is to be noted that the prediction was done 15 minutes before scheduled departure for individual flights which were then averaged in quarter time intervals. Prediction accuracy for individual flights was also tested, and on average, 81% of flights were predicted within a root mean square error value of 2 minutes.

It is expected that control tower operations, surface management systems, and airline scheduling can benefit from this prediction by adjusting schedules to minimize congestion,

delays, and emissions, and also by better utilization of ground personnel and resources. Especially, with airport dynamics changing throughout the day in the face of uncertainties such as weather, prediction of airport taxi-out time averages combined with individual flight predictions, could help airlines manage decisions such as incurring delays at the gate as opposed to increasing emissions due to longer taxi times. Air Traffic Control would also benefit from this knowledge when making decisions regarding holding flights at the gate or ramp area due to increased congestion. This could improve the performance of air traffic flow management both on ground and in air across the entire NAS in the US and worldwide. It can be integrated to support the futuristic Total Airport Management concepts beyond Collaborative Decision Making [27] that envisions automation of several airport operations.

As part of future work, accuracy of predications will be improved by incorporating runway direction. This is because runway configurations change during the day which could alter the gate to runway distance. Also a sensitivity analysis of the learning parameters will be conducted. Further analysis to capture seasonal trends and incorporation of runway and gate assignments could improve prediction accuracy. Also study of other majors hubs are part of this ongoing research.

ACKNOWLEDGMENTS

The authors would like to thank FAA for their permission to obtain data from the ASPM database.

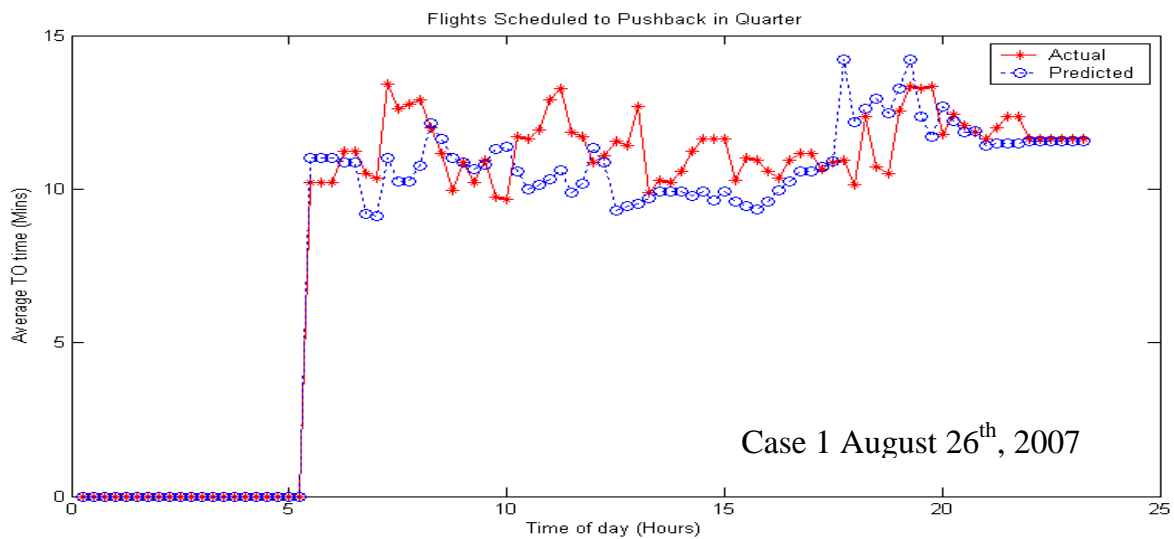


Figure 2. Case 1: Plot of Actual Taxi-Out Time vs. Predicted Taxi-Out Time

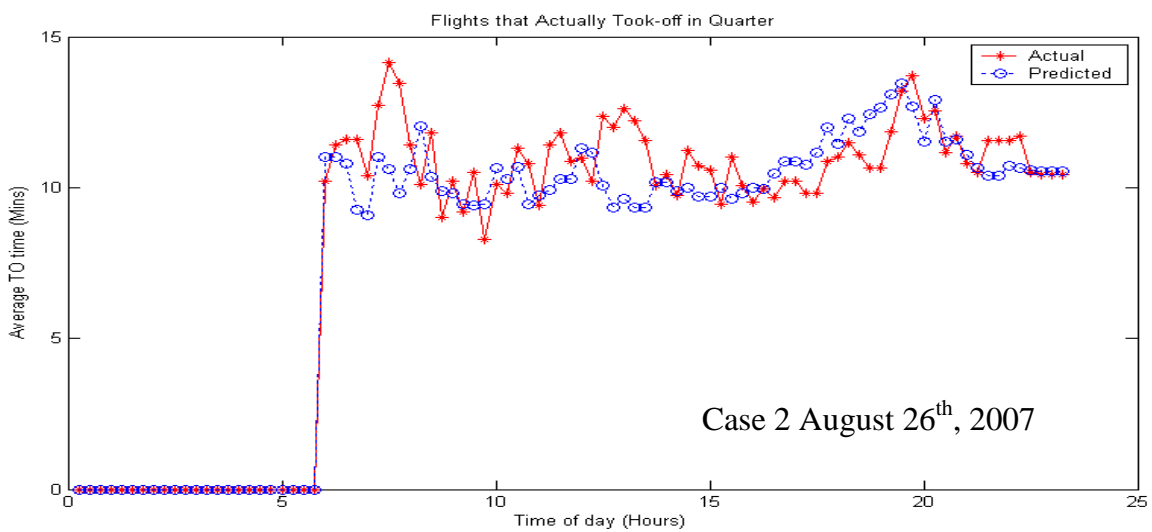


Figure 3. Case 2: Plot of Actual Taxi-Out Time vs. Predicted Taxi-Out Time

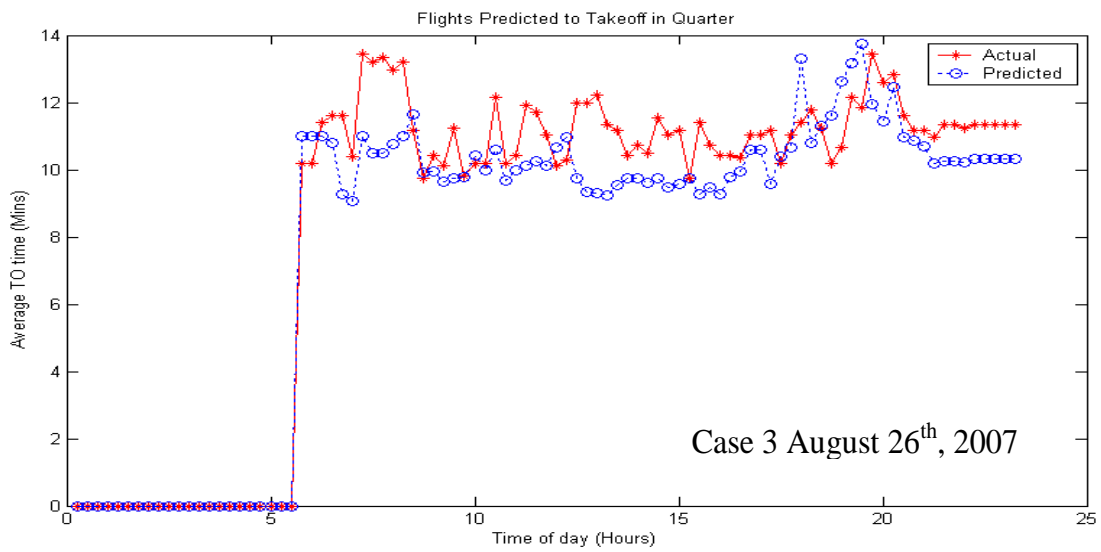


Figure 4. Case 3: Plot of Actual Taxi-Out Time vs. Predicted Taxi-Out Time

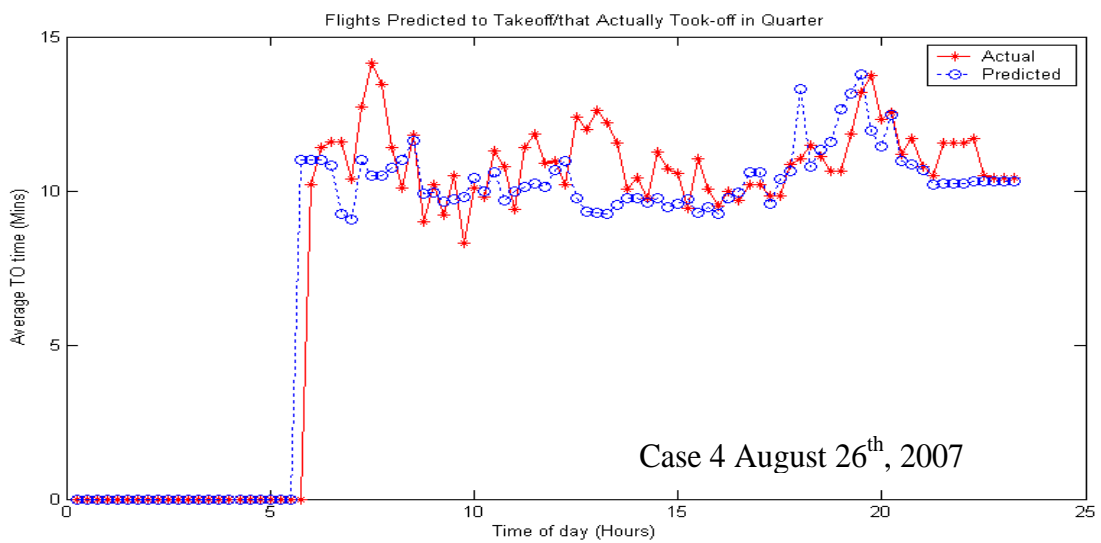


Figure 5. Case 4: Plot of Actual Taxi-Out Time vs. Predicted Taxi-Out Time

REFERENCES

- [1] Cooper, W.W. Jr., E.A. Cherniavsky, J.S. DeArmon, J.M. Glenn, M.J. Foster, S.C. Mohleji, and F.Z. Zhu. *Determination of Minimum Push-Back Time Predictability Needed for Near-Term Departure Scheduling using DEPARTS*. The MITRE Corporation, 2001, url: http://www.mitre.org/work/tech_papers/tech_papers_01/cooper_determination/cooper_determination.pdf
- [2] Carr, F., A. Evans, J-P. Clarke, E. Feron. Modeling and Control of Airport Queueing Dynamics under Severe Flow Restrictions. *Proceedings of the American Control Conference*, AACC, Anchorage, AK, 2002.
- [3] Atkins, S., and D. Walton. Prediction and Control of Departure Runway Balancing at Dallas Fort Worth Airport. *Proceedings of the American Control Conference*, 2002.
- [4] Idris, H., I. Anagnostakis, B. Delcaire, J.P. Clarke, R.J. Hansman, E. Feron, and A. Odoni. Observations of Departure Processes at Logan Airport to Support the Development of Departure Planning Tools. *Air Traffic Control Quarterly*, Vol. 7(4), pp. 229-257, 1999.
- [5] Tu, Y., M.O. Ball, and J. Wolfgang. *Estimating Flight Departure Delay Distributions - A Statistical Approach with Long-Term Trend and Short-Term Pattern*. Robert H. Smith School Research Paper, RHS 06-034, 2005, url: <http://ssrn.com/abstract=923628>
- [6]
- [7] Shumsky, R.A. *Dynamic Statistical Models for the Prediction of Aircraft Take-off Times*. Ph.D. Thesis, Operations Research Center, MIT, Cambridge, MA., 1995.
- [8] Idris, H., J.P. Clarke, R. Bhuvu, and L. Kang. Queuing Model for Taxi-out Time Estimation. *Air Traffic Control Quarterly*, 2002.
- [9] Levine, B.S. and O.H. Gao. Aircraft Taxi-Out Emissions at Congested Hub Airports and the Implications for Aviation Emissions Reduction in the United States. CD-ROM. Submitted to the TRB 2007 Annual Meeting, 2007.
- [10] Barrer, J.N., G.F. Swetnam, W.E. Weiss. *The Feasibility Study of using Computer Optimization for Airport Surface Traffic Management*. The MITRE Corporation, MTR89W00010, 1989.
- [11] Idris, H.R., B. Delcaire, I. Anagnostakis, W.D. Hall, N. Pujet, E. Feron, R.J. Hansman, J.P. Clarke, and A.R. Odoni. Identification of Flow Constraints and Control Points in Departure Operations at Airport System. *Proceedings AIAA Guidance, Navigation and Control Conference*, AIAA 98-4291, Boston, MA, 1998.
- [12] Anagnostakis, I., H.R. Idris, J.P. Clarke, E. Feron, R. J. Hansman, A. R. Odoni, and W.D. Hall. A Conceptual Design of a Departure Planner Decision Aid. Presented at the *3rd USA/Europe Air Traffic Management R&D Seminar*, Naples, Italy, 2000.
- [13] Shumsky, R.A. Real Time Forecasts of Aircraft Departure Queues. *Air Traffic Control Quarterly*, Vol. 5(4), 1997.
- [14] Lindsay, K., D. Greenbaum, C. Wanke. Pre-departure Uncertainty and Prediction Performance in Collaborative Routing Coordination Tools. *Journal of Guidance, Control, and Dynamics*, Vol. 28(6), 2005.
- [15] Clow, M., K. Howard, B. Midwood, and R. Oiesen. *Analysis of the Benefits of Surface Data for ETMS*. Volpe National Transportation Systems Center, VNTSC-ATMS-04-01, 2004.
- [16] Welch, J., S. Bussolari, S. Atkins. Using Surface Surveillance to Help Reduce Taxi Delays. *AIAA Guidance, Navigation & Control Conference*, AIAA-2001-4360, Montreal, Quebec, 2001.
- [17] Signor, D.B., and B.S. Levy. *Accurate OOOI Data: Implications for Efficient Resource Utilization*. Sensis Corporation, 25th Digital Avionics Systems Conference, 2006.
- [18] Laskey, K.B., N. Xu, C-H. Chen. *Propagation of Delays in the National Airspace System*. Technical Report, 2006, url: http://ite.gmu.edu/~klaskey/papers/UAI2006_Delay.pdf
- [19] Futer, A. Improving ETMS' Ground Time Predictions. 25th Digital Avionics Systems Conference, 2006 IEEE/AIAA, pp.1-12, 2006.
- [20] Tu, Y., M.O. Ball and W. Jank. Estimating Flight Departure Delay Distributions -- a Statistical Approach with Long-Term Trend and Short-Term Pattern. Technical Report, U. of Maryland, 2005.
- [21] Gosavi, A. *Simulation Based Optimization: Parametric Optimization Techniques and Reinforcement Learning*. Norwell, MA: Kluwer Academic, 2003.
- [22] Bellman, R. *The Theory of Dynamic Programming*. Bull. Amer. Math. Soc., vol. 60, 1954, pp. 503-516.
- [23] Howard, R. *In Dynamic Programming and Markov Processes*. MIT Press, Cambridge, MA, 1960.
- [24] Bertsekas, D., and J. Tsitsiklis. *In Neuro-Dynamic Programming*. Athena Scientific, Belmont, MA, 1995.
- [25] Puterman, M.L. *Markov Decision Processes*. Wiley Interscience, New York, 1994.
- [26] Robbins, H., and S. Monro. A Stochastic Approximation Method. *Ann. Math. Statistics*, vol. 22, 1951 pp. 400-407.
- [27] http://www.bts.gov/programs/airline_information/accounting_and_reporting_directives/technical_directive_15.html
- [28] Meier, C., P. Eriksen. *Total Airport Management: A Step Beyond Airport Collaborative Decision Making*. 2006, url: http://www.eurocontrol.int/eec/public/standard_page/EEC_News_2006_3_TAM.html.

Research of the relation between the sustainability of hourly capacity at Schiphol airport, KLM arrival punctuality and the percentage of KLM transfer passengers at risk of losing their connection

Dragana Mijatovic

MS/R&D/Research Department
Air Traffic Control the Netherlands (LVNL),
Schiphol Airport, the Netherlands
e-mail: d.jovanovic@lvnl.nl; (D. Mijatovic-Jovanovic)

Marleen Meert

OCC - Operations Development
KLM, Royal Dutch Airlines
Schiphol Airport, the Netherlands
e-mail: marleen.meert@klm.com

Abstract— Schiphol airport has become one of the major airline hubs in Europe as a result of KLM's growth strategy in the 1980s and 1990s. To provide a high customer satisfaction level, it is of great importance for KLM to provide a highly reliable network. In this paper two performance indicators (PI) are identified to express this reliability: the arrival punctuality according to the KLM timetable and the percentage of KLM transfer passengers at risk of losing their connection.

This research addresses the effect of the performance of the Dutch air navigation service provider (Air Traffic Control the Netherlands – LVNL) on the arrival punctuality and the percentage of KLM transfer passengers at risk of losing their connection. The contribution of LVNL is expressed in hourly inbound capacity and delays caused by its reduction as well as in terms of the LVNL performance indicator – sustainability of hourly capacity.

The study has shown that there is a linear relation between the sustainability of hourly capacity, KLM arrival punctuality and the percentage of KLM transfer passengers at risk of losing their connection. The relations are derived from the historical data and they are limited to KLM European inbound flights in the second bank of the day (based on the 7 banks system KLM is operating with) when two landing runways were in use. These are made based on sustainability value calculated for the given declared capacity. The derived models can be used for making estimations of arrival punctuality and the percentage of KLM transfer passengers at risk of losing their connection for a given sustainability. The estimate values have been compared with the actual ones and have shown that the actual values lay well within the confidence interval of the models demonstrating the accuracy of the models.

Furthermore, the effect of a reduced hourly inbound capacity on arrival punctuality and the percentage of KLM transfer passengers at risk of losing their connection has been researched. When the capacity forecast value is higher than demand at Schiphol, LVNL induced delays are low. Values of arrival punctuality and the percentage of KLM transfer passengers at risk of losing their connection are almost constant for small values of the LVNL influenced delays and therefore do

not fluctuate much when the capacity forecast is higher than demand.

The relationships found lay the basis for decisions support models and tools for optimizing and further developing the KLM network operations at Schiphol airport.

Keywords: Schiphol airport, KLM, ATC the Netherlands, LVNL, arrival punctuality, transfer passengers, sustainability of hourly capacity, hourly inbound capacity.

I. INTRODUCTION

KLM has about 11 million passengers each year arriving at Schiphol as one of the major European hubs. About 70% of them are transfer passengers. To serve these transfer passengers, KLM must assure high quality connections at the Schiphol hub, providing a quick but reliable connection onto the next flight. To offer an attractive travel schedule to its passengers KLM designed its timetable to minimize the travel time between origin and destination through short connection times.

In order to maximize the number of high quality connections at Schiphol airport, KLM's timetable has been constructed into several arrival and departure waves or banks (see Fig. 1). The duration between scheduled arrival time and scheduled departure time of a connection should be at least the minimum connecting time (MCT). If it is shorter than the MCT, passengers cannot make a connecting flight and tickets for such connections cannot be sold.

As air travel demand increases, KLM faces the challenge to optimize and expand its hub operation at Schiphol airport as part of the competition between global airline alliances. To maintain and expand its market share KLM must offer more attractive connections through its hub at Schiphol. This growth of connectivity requires an increasing number of flights in arrival and departure banks. This increased peak demand leads to a saturation of airport capacity at the expense of increased

arrival-delays, which in turn affects the reliability of connections which is vital to KLM's operation.

The development of KLM's network thus is determined by the balance between airport capacity and network demand. This capacity – demand balance governs both KLM's strategic¹ and tactical² decision making process. To optimize this decision making process it is necessary to relate factors which play a role in KLM's performance (such as arrival punctuality and the percentage of transfer passengers at risk of losing their connections) to LVNL – factors (Air Traffic Control the Netherlands – factors, such as capacity forecast and sustainability) (see Table I).

This paper researches relations between performance indicators (PIs) of KLM and LVNL. The relationships established in the study are used as models for forecasting values of KLM PIs (arrival punctuality and the percentage of KLM transfer passengers at risk of losing their connection (percentage of the KLM sub-MCT passengers, see Table I) based on a given sustainability of hourly capacity (LVNL PI). This paper is the follow-up study of the paper from Ref. [1] and forms a basis for the tools which can help KLM in developing and optimizing its network at Schiphol.

II. DEFINITIONS

Table I gives a list of the definitions used in this paper.

TABLE I. LIST OF DEFINITIONS USED IN THIS RESEARCH.

	Definition
Arrival punctuality	The percentage of flights that arrived on time or earlier compared to the scheduled arrival time, i.e., SBA (at the gate).
Arrival delay	The difference in time between scheduled on-block arrival time (at the gate) according to the timetable (SBA) and the actual on-block arrival time (ABA), i.e., ABA-SBA. A flight that arrived earlier than or at the SBA is considered to be "on time". One minute late or more is considered to be "delayed".
FIR delay	A delay of a flight within the Dutch Flight Information Region (FIR). It is calculated as a difference between the actual time an aircraft flies ^a and taxis from FIR boundary until gate (ABA-FIR) and a sum of nominal flight and taxi times from FIR entry through specific sector to specific runway and from there to gate. [FIR delay= (actual flight time from FIR + actual taxi time from runway) – (nominal flight time from FIR ^b - nominal taxi time from runway ^c)]
ATFM delay	A delay of the aircraft at the outstation caused by the reduced capacity at Schiphol airport.
Transfer passenger	A passenger with a ticket for the connecting flight, arriving at Schiphol on a KLM inbound flight and leaving Schiphol on an outbound flight of any carrier.
Minimum connecting time (MCT)	A minimum transfer time between the inbound and outbound flights a passenger ^d needs to make a connection to. Tickets can be sold only for the connecting flights with scheduled transfer time of at least MCT. For Schiphol, the MCT between both

¹ Strategic: referring to KLM's network development process.

² Tactical: referring to the network management process on the day of operation.

	European inbound and outbound flight is 40 minutes, other connections have MCT of 50 minutes.
Sub-MCT passenger	A transfer passenger having an effective transfer time of less than MCT. ^e
Effective transfer time	The time difference between actual arrival time of the inbound flight and schedule departure time of the connecting outbound flight.
Capacity forecast	A number of landings per hour, which can be realized based on the expected availability of the runways (due to the weather conditions, maintenance of the runways and/or available staff). It is issued 4 times a day for the next 6 hours by LVNL (ATC the Netherlands). ^f
Demand	A number of planned landings (of all airlines) on Schiphol filed a day before their actual landing. ^g
Delta capacity	A difference between the capacity forecast and demand.
Declared capacity	A number of landings per hour that can be handled by the LVNL. It is determined for a longer period of time (year) and slot allocations are based on it. Declared capacity is 68 landings per hour for years 2006 and 2007.
Sustainability	A percentage of time a declared capacity is indeed realized by the providers (see Fig. 2).

^a Actual flight time can be extended significantly due to the tactical flight management: vectoring (stretching the flight path) and/or use of holdings.

^b Nominal flight time is calculated as a median value of the difference between the touch down time (TDT) and time of the FIR entry (FIR): Nominal flight time = (TDT-FIR)median. It is calculated for an undisturbed flight from each ACC sector entry until a particular runway.

^c Nominal taxi time is calculated as a difference of the time the aircraft crosses the red line (the border between the platform and maneuvering area - it is considered in that case that the aircraft reached the gate) and the touch down time: Nominal taxi time = (time (crossing the red line) - TDT)aver. Only average taxi times are available at LVNL.

^d It is assumed that the baggage of the passengers is transferred together with the passenger to the connecting flight.

^e It is assumed that departure flights depart on time.

^f It may differ from the actual capacity, but rarely and therefore gives a reliable indication of the capacity at Schiphol.

^g Data are received from the Amsterdam Airport Schiphol (AAS) and KLM.

At this point it is important to underline that in this research the percentage of the sub-MCT passengers used differs from the actual percentage of passengers who cannot make their connections. Actual percentage of passengers who cannot make their connection is lower than the percentage of the sub-MCT passengers, because even if the passengers arrive with the effective transfer time smaller than the MCT, in some cases it is still possible to make their connections (such as: gates are close to each other, departing flight was delayed, passenger can reach the gate faster than assumed in the MCT value etc.). The percentage of the sub-MCT passengers is used here instead of the actual percentage of passengers who cannot make their connection to avoid a number of effects at the airport that cannot be influenced by the LVNL (such as: delays of the departing flights, delays in opening the aircraft doors etc.). Actual number of passengers who miss their connection is a result of a daily (tactical) handling of flights. It is not suitable for this research, because the results of the research will be used for the strategic development at Schiphol.

III. RELATIONS BETWEEN THE HOURLY INBOUND CAPACITY, LVNL- INFLUENCED DELAY, SUSTAINABILITY OF HOURLY CAPACITY, ARRIVAL PUNCTUALITY AND PERCENTAGE OF THE SUB-MCT PASSENGERS

The aim of this paper is to research the relation between the important performance indicators (PIs): sustainability of hourly capacity (as an LVNL PI) and arrival punctuality according to timetable and percentage of the sub-MCT passengers (as the KLM PIs) (see Fig. 3(a)). Hence, it is needed first to better understand the relation at a lower aggregation level, i.e., between the LVNL -factors, arrival punctuality and percentage of the sub-MCT passengers (see Fig. 3 (b)).

As LVNL-factors are considered:

- Delays influenced by the LVNL (Dutch ATC), i.e., the delays at outstation caused by the reduced capacity at Schiphol airport (ATFM delays) and delays within the Dutch FIR (FIR delays) (see Table I) and
- Hourly inbound capacity. Hourly inbound capacity consists of the capacity forecast and delta capacity (see Table I).

Research is limited to:

- KLM European inbound flights connecting to outbound flights of KLM or any other carrier. European inbound flights are chosen because they can be regulated by the Air Traffic Flow and Capacity Management (ATFCM) restrictions and therefore the impact of LVNL is higher on them;
- summer 2006 and winter 2006/2007 (7 banks system) (see Fig. 1 (a));
- arrival bank 2 (see Fig. 1 (a)). This bank has been chosen, because demand of the flights in the core of the bank is close to the maximum inbound capacity. Moreover, this bank does not suffer from the snowballing effect (flights in the later banks can be delayed as a consequence of the accumulated delays during the day);
- the periods when two landing runways were in use and one runway was used for the departures (from about 7.30 until 9.15 LT);
- Sustainability values calculated by using the declared capacity value of 68 landings per hour.

The exact values on the axis in the graphs that follow are not given, because of commercial sensitivity of data for the parties involved in the research.

Fig. 4 shows delta capacity vs. capacity forecast for bank 2. From this graph the capacity demand (capacity forecast value when the demand is equal to the capacity forecast) is determined and it is the same in both seasons. Moreover, critical capacity and corresponding critical delta capacity for each season are denoted, because above critical capacity value, the percentage of the sub-MCT passengers will not change significantly (as will be discussed below).

Fig. 5 shows that as long as the capacity is higher than the demand, ATFM+FIR Delay stays rather low and nearly constant, which is in about 80-90% of time (depending on the season). In summer the percentage of time with capacity higher than critical one is higher due to the better weather conditions (and therefore fewer restrictions) than in winter. However, when the capacity drops below the critical value an increase in the ATFM+FIR Delay occurs. The increase shows stronger slope for the winter season, which can be attributed to more frequent capacity reduction. These graphs are not fitted, because the numbers of measurement points in the region below critical capacity are not enough to obtain a reliable fit.

Arrival punctuality and percentage of KLM sub-MCT passengers vs. ATFM+FIR Delay has been plotted in Fig. 6. It can be observed that the increase in the ATFM+FIR Delay causes the decrease in the arrival punctuality and increase in the percentage of sub-MCT passengers. For small ATFM+FIR Delay arrival punctuality as well as the percentage of sub-MCT passengers stays almost constant. The scale of both x-axis is the same. It can be noticed that for the same value of the ATFM+FIR Delay, the percentage of sub-MCT passengers increases rather slow whereas arrival punctuality decreases rather fast. This can be explained by the following: If the aircraft is even one minute delayed compared to the schedule, it is considered as delayed flight and it contributes to the reduction of arrival punctuality. However, a number of transfer passengers connecting to different outbound flights could be on this inbound flight. Therefore, the delay of the inbound flight does not mean that all passengers are under the risk of losing their connecting flights. Only the passengers that arrive with the effective transfer time less than MCT are at that risk.

Fig. 7 shows arrival punctuality and percentage of the KLM sub-MCT passengers vs. capacity forecast and delta capacity. The arrival punctuality does not change significantly above the capacity demand value and for positive delta capacity (see Fig. 7 (a), (b)). Below the capacity demand value and when the demand is larger than the available capacity, arrival punctuality drops rather fast. Percentage of sub-MCT passengers behaves more inertly (see Fig. 7 (c), (d)): when the capacity is above the critical capacity and related critical delta capacity (see also Fig 4), the percentage of the sub-MCT passengers stays almost the same. Note that the unit on y-axis in Fig. 7 (d) is half of the values of the unit on y-axis in Fig. 7 (c). The explanation of this behaviour of the percentage of sub-MCT passengers is the same as for Fig. 6.

Percentage of sub-MCT passengers vs. arrival punctuality exhibits almost linear behaviour (see Fig. 8). Higher arrival punctuality results in low percentage of the sub-MCT passengers. Each data point corresponds to the arrival punctuality and percentage of the sub-MCT passengers for a period of 30 days using method of sliding window with a timeframe of one week.

Relationships between dependences of the arrival punctuality and percentage of sub-MCT passengers vs. sustainability of hourly capacity are given in Fig. 9 (see also Fig. 3 (a)). Linear fits and the confidence interval are given and used as models to obtain the values for arrival punctuality and percentage of sub-MCT passengers for expected sustainability value. The models were tested for the bank 2 of summer 2007, because the same limitations to models apply to this period (see description of the limitations given above in this Chapter). Sustainability value for summer 2007 is estimated from the data available at LVNL [2]. Arrival punctuality and percentage of sub-MCT passengers values are estimated (forecasted) from the linear fit (see Fig. 9) for the estimated sustainability value and compared to the actual values for summer 2007. As it can be observed from the graphs, actual values for arrival punctuality and percentage of sub-MCT passengers are very close to the forecasted values and very well within the confidence interval. Hence, it can be concluded that this model can be used for the estimation of the arrival punctuality as well as percentage of sub-MCT passengers values and gives good results as long as the limitations to the model (given above in this Chapter) apply. In case that some of the conditions change (for instance, KLM changes the banks system from 7 to more, the model is applied to another bank of the day, sustainability is calculated for another value of declared capacity etc.) a new model has to be built.

IV. CONCLUSIONS AND OUTLOOK

Research has been conducted to relate important performance indicators (PIs) for KLM: arrival punctuality according to timetable and percentage of sub-MCT passengers, and LVNL performance indicator – sustainability of hourly capacity.

This high level research framework identified that a correlation is deducible between these PIs. Linear relations between them are found. The empirical models have been made for the KLM European inbound flights arriving in bank 2 when two landing runways were in use and when KLM has been operating with 7 banks system (summer '06 and winter '06/'07). The models were tested for summer '07. It can be concluded that the actual values of arrival punctuality and percentage of sub-MCT passengers for estimated sustainability of hourly capacity for summer '07 are close to the forecasted values and well within the confidence interval of the model. Hence, these models can be used for forecasting as long as the mentioned limitations apply.

Additionally, when the capacity forecast at Schiphol is higher than demand, low LVNL-influenced delays occur. Arrival punctuality and percentage of sub-MCT passengers are almost constant for small LVNL-influenced delays. It is shown that arrival punctuality and percentage of sub-MCT passengers do not fluctuate much when the capacity at Schiphol is higher than demand. However, decrease in arrival

punctuality and increase in percentage of sub-MCT passengers occur when demand exceeds capacity forecast values.

This research has been limited to bank 2. To apply it to the whole day, more research has to be done. This can be rather complex and results may not be easy to compare due to the snowballing effect of delays during the day and the difficulty of defining their exact cause.

The follow-up study will focus on the potential optimization of the scheduled flying time for each KLM flight by minimizing the effect of LVNL induced delays.

ACKNOWLEDGMENT

This paper presents results of a project "Research of the relation between the KLM No Connection (NOC)-rate and sustainability of hourly capacity at Schiphol airport" conducted in the Knowledge and Development Centre Mainport Schiphol (KDC).

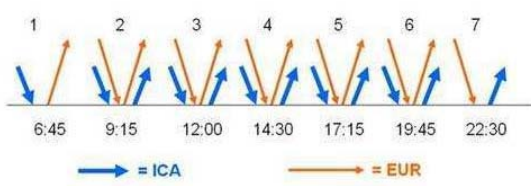
Authors thank to the M.Sc. students from TU Delft: J. Wanga and K. El-Bachraoui for their dedicated work on this project. Authors thank also to the project team members: L. Hoogerbrugge (LVNL), R. Rooij (KLM), W. van Miltenburg (KLM), R. Frijns (KLM), H. Erens (LVNL), B. Gimberg (AAS), J. Goedhart (KLM) and A. Geebelen (AAS); and the steering group members: B. van der Weyden (KLM), J. Kerckhoff (KLM), S. Lentze (LVNL), P. Cornelisse (KLM) as well as Y. de Haan (KLM), E. Westerveld (LVNL), T. Dortmans (KLM), L. Oudkerk (KLM), Y. Obbens (LVNL), F. Dijkstra (LVNL), F. Bloem (LVNL), G. Vale (LVNL), M. Opbroek (LVNL) and G. Gardner (TU Delft) for the valuable discussions.

REFERENCES

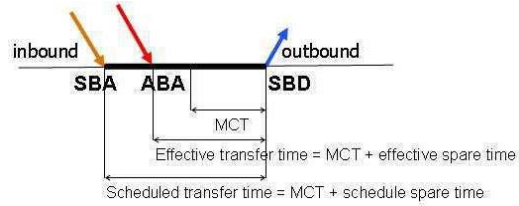
- [1] D. Mijatovic, M. Meert, K. El-Bachraoui, J. Wanga, "Research of the relation between the hourly inbound capacity at Schiphol airport and the number of KLM transfer passengers at risk of losing their connection," Proceedings of the 6th EUROCONTROL Innovative Research Workshop Bretigny, France, 4-6 december 2007.
- [2] VEM Performance Standard Summer 2007 D/R&D/023, LVNL, Schiphol Oost, The Netherlands. (2007)

LIST OF ABBREVIATIONS

AAS	Amsterdam Airport Schiphol
ACC	Area Control Centre
ATC	Air Traffic Control
ATFCM	Air Traffic Flow and Capacity Management
ATFM	Air Traffic Flow Management
FIR	Flight Information Region
KLM	Royal Dutch Airlines
LT	Local Time
LVNL	Luchtverkeersleiding Nederland (Air Traffic Control the Netherlands)
MCT	Minimum Connecting Time
PI	Performance Indicator



(a)



(b)

Figure 1. (a) Inbound and outbound KLM flights are organized at present in 7-banks system at Schiphol. Times shown are the local times (LT). Blue lines denote intercontinental flights (ICA) and the orange ones European flights (EUR); (b) Schematic representation of connecting flights: inbound flight is represented by the orange arrow (SBA-scheduled on-block arrival, i.e., the time a flight is scheduled to arrive at the gate, according to the timetable); delayed arrival is given by the red arrow (ABA-actual on-block arrival; i.e., the time the flight actually arrived at the gate); outbound flight is denoted by the blue arrow (SBD-scheduled off-block departure; i.e., the time a flight is scheduled to depart from the gate, according to the timetable). Scheduled transfer time is the time between the scheduled on-block arrival of the inbound flight and scheduled off-block departure of the outbound flight the passenger is transferring to (calculated as SBD-SBA). Scheduled transfer time consists of the MCT and the scheduled spare time. Effective transfer time is the time between actual arrival time (gate) of the inbound flight and the scheduled departure time (gate) of the outbound flight the passenger is transferring to (calculated as SBD-ABA). Effective transfer time consists of the MCT and effective spare time.

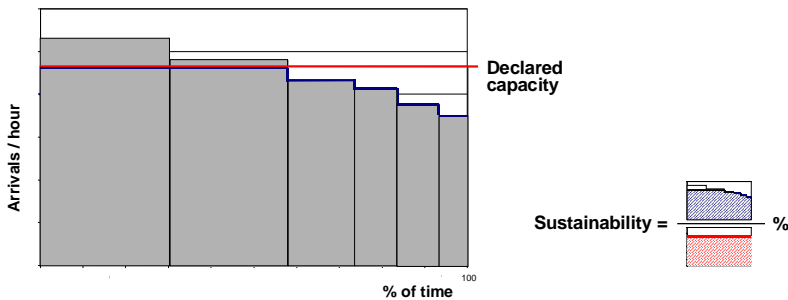
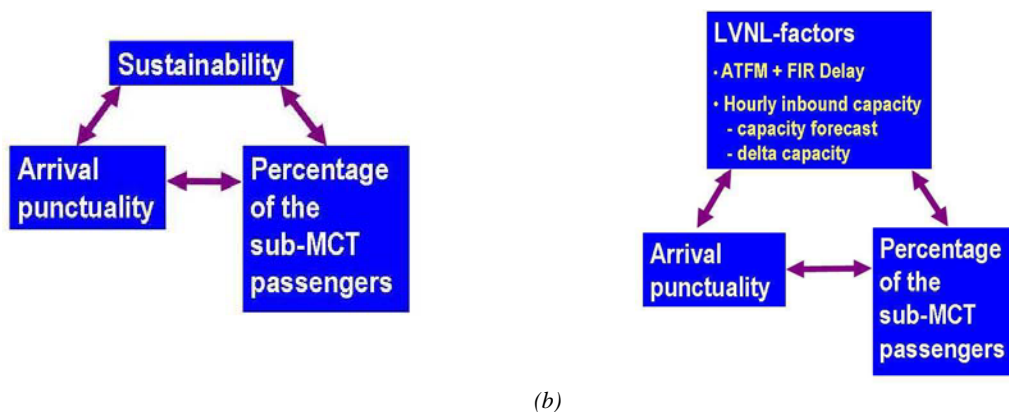


Figure 2. Determination of the sustainability of hourly capacity. On x-axis percentage of time for a certain available capacity forecast (given on y-axis) is presented. Sustainability is a ratio between the gray area under the red line and the whole area under the red line (declared capacity).



(a)

(b)

Figure 3. (a) Research framework at the higher aggregation level; (b) Research framework on the lower aggregation level.

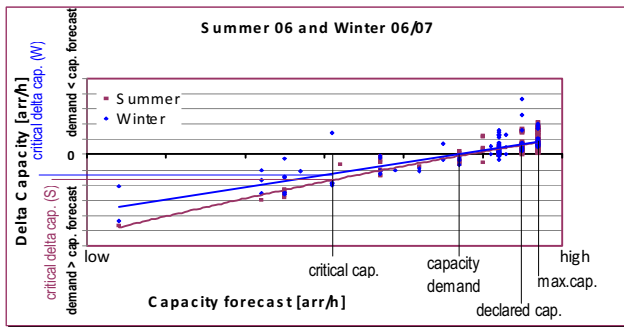
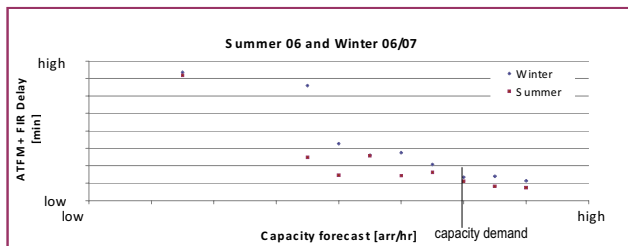
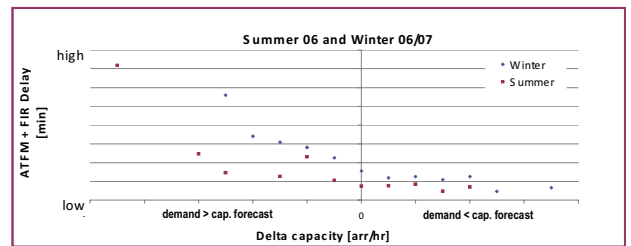


Figure 4. Delta capacity vs. LVNL capacity forecast.

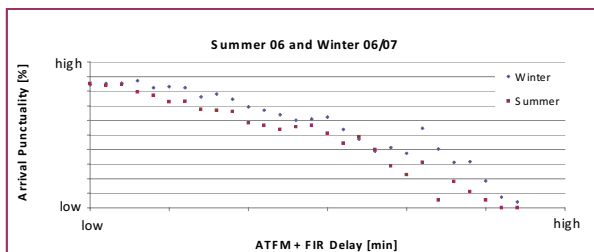


(a)

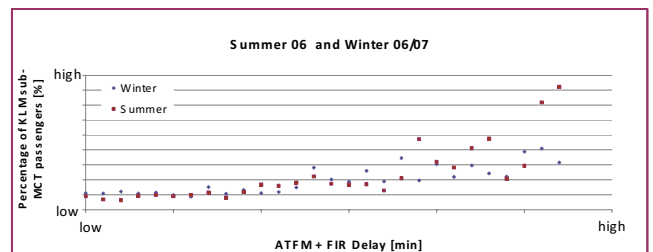


(b)

Figure 5. (a) ATFM+FIR Delay vs. LVNL capacity forecast; (b) ATFM+FIR Delay vs. delta capacity. Each point represents the ATFM+FIR Delay calculated as an average value for a given capacity forecast or delta capacity value.



(a)



(b)

Figure 6. (a) Arrival punctuality vs. ATFM+FIR Delay; (b) Percentage of the KLM sub-MCT passengers vs. ATFM+FIR Delay. Each point represents the arrival punctuality and percentage of the KLM sub-MCT passengers, respectively, calculated for each ATFM+FIR Delay.

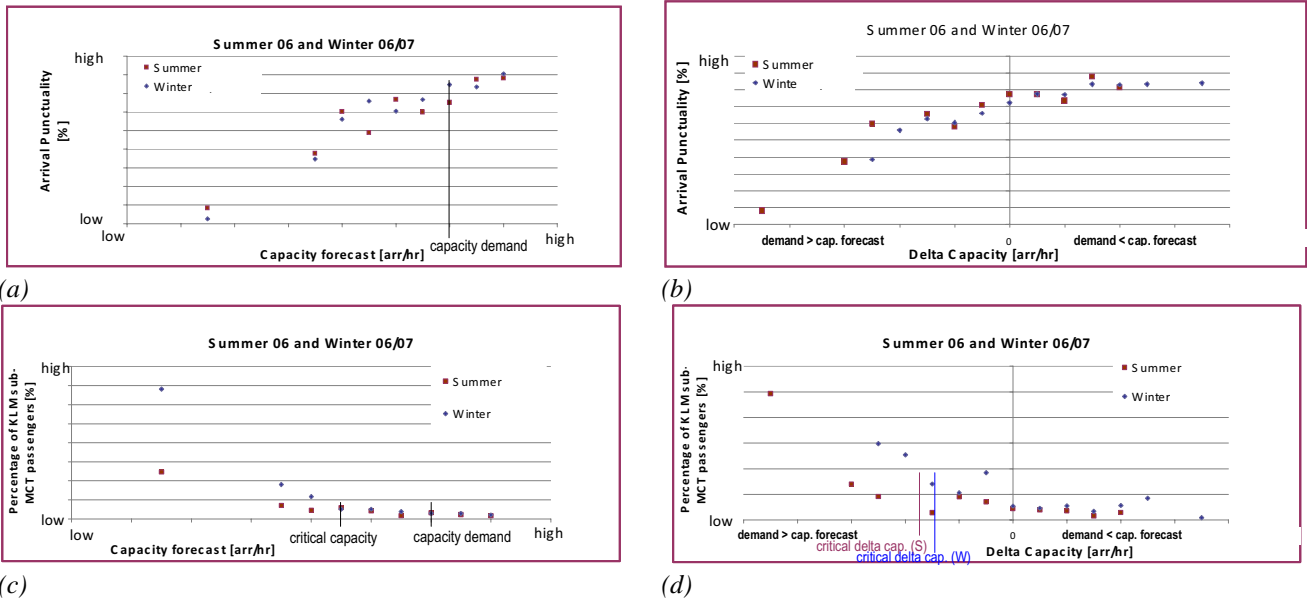


Figure 7. (a) Arrival punctuality vs. capacity forecast; (b) Arrival punctuality vs. delta capacity forecast; (c) Percentage of the KLM sub-MCT passengers vs. capacity forecast; (d) Percentage of the KLM sub-MCT passengers vs. delta capacity. Each point represents the arrival punctuality and percentage of the KLM sub-MCT passengers, calculated for each capacity forecast and delta capacity, respectively. These graphs are not fitted, because the number of measurement points in the region below critical (delta) capacity is not enough to obtain a reliable fit.

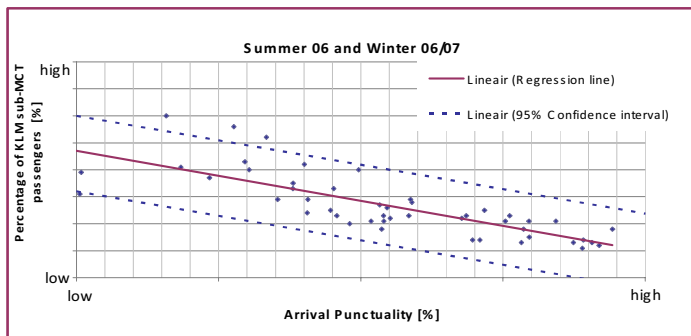
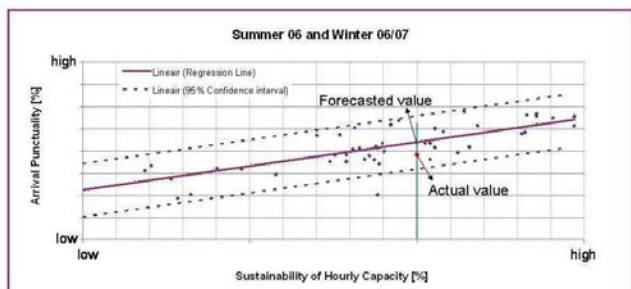
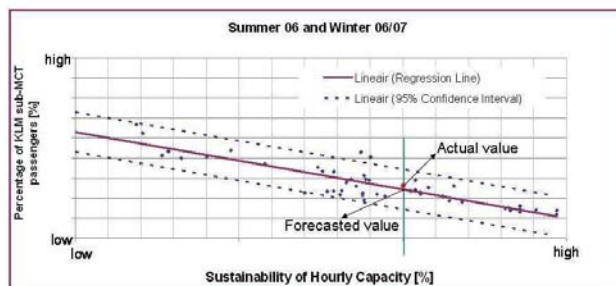


Figure 8. Percentage of the sub-MCT passengers vs. arrival punctuality. Each data point corresponds to the arrival punctuality and percentage of the sub-MCT passengers for a period of 30 days starting at the first day of the summer season (March 26th, 2006) and ending at the last day of the winter period in 2007 (March 24th, 2007). Use is made of sliding window with a window of 30 days and rolling with a timeframe of one week. Hence the first data point corresponds to the arrival punctuality and percentage of the sub-MCT passengers of the flights between the period of March 26th – April 24th. The second data point corresponds to the period from April 2nd – May 1st etc.



(a)



(b)

Figure 9. (a) Arrival punctuality vs. sustainability of hourly capacity; (b) Percentage of the sub-MCT passengers vs. sustainability of hourly capacity. Each data point is calculated as in Fig. 8.

Efficiency of Aircraft Boarding Procedures

Michael Schultz

Institute of Logistics and Aviation
Technische Universität Dresden
Dresden, Germany
schultz@ifl.tu-dresden.de

Christian Schulz

Institute of Logistics and Aviation
Technische Universität Dresden
Dresden, Germany

Hartmut Fricke

Institute of Logistics and Aviation
Technische Universität Dresden
Dresden, Germany

Abstract—Efficient boarding procedures are the basis for fast turnaround times. The boarding is an essential part of the critical path of the turnaround process, so time savings directly advance the overall process. Previous research results pointed out that the boarding time can be significantly reduced by using adapted boarding procedures. In this paper we present a comprehensive analysis of boarding procedures (A320-200, 174 passengers) considering different seat load factors, passenger acceptance of chosen boarding order, and arrival rates. The results of the analysis yield a lower boundary for an efficient boarding of approx. 40% acceptance rate, 50% seat load factor and an arrival rate of 7 passengers per minute. Furthermore, the use of the rear door has a substantial effect regarding the boarding efficiency. An enhancement of approx. 25 % is reached, without the disturbing influences of the strategy acceptance rate.

Boarding; Critical path; Efficiency; Turnaround

I. INTRODUCTION

To manage future challenges in aviation the Advisory Council for Aeronautical Research in Europe (ACARE) provided the Strategic Research Agenda 2 in 2004 [1]. Herein, the ACARE asks for efficient procedures and processes, new standards for service, safety, security and quality, as well as decreasing operational costs at all levels. To achieve these objectives High Level Target Concepts (HLTC) are defined, whereas the safety regulations always have major importance. In this context, boarding processes have to be high time efficient, i.e. short turnaround times (see fig. 1).

Following Airbus' definition for the turnaround (fig. 1) the turnaround time is defined as the aircraft parking time, between on-block and off-block. While the aircraft is at the position (at the gate or apron) processes like (un-) loading, catering, cleaning, refueling, and (de-)boarding are executed. Due to safety regulations and logistic requirements some processes run parallel to others and others have to be executed sequentially. The overall turnaround time is defined by the termination of the last process. According to fig. 1 the moving of passenger bridges, the boarding and the refueling are part of the critical path. Shortening the processes on the critical path implies a shortening of the overall turnaround process as well. Reduced turnaround times achieved by improved operational procedures have several positive effects. The airline reduces the ground idle time and saves ground costs while the airport benefits from the reduced gate (apron) occupancy time.

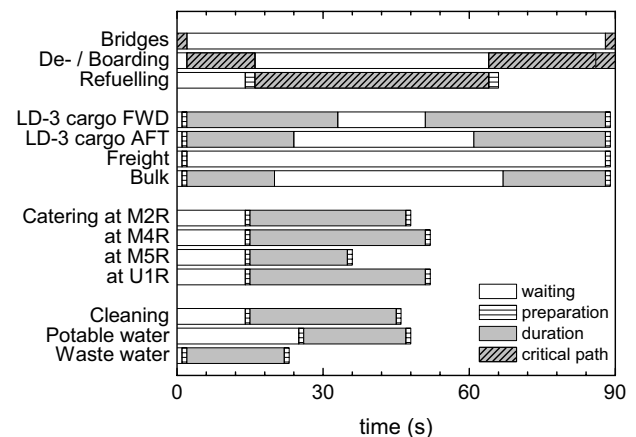


Figure 1. Turnaround time schedule of A380 (90 min, baseline) [2]

Various research studies were performed on the field of efficient boarding procedures. They as a typical reference apply analyses for single aisle aircraft, such as the A320. Airbus and Boeing expect a business volume of single aisle aircraft of 40 % and 42.5 % respectively until 2026. Both aircraft manufacturer plan to deliver approx. 17000 single aisle aircraft each (68 %, 62 % of production) [3, 4]. These aircraft often come into operation for low cost airlines, where the market pressure forces the airlines to be highly competitive and to achieve high efficiency at all operational levels. In this context the optimization of the boarding procedures could be one deciding competitive factor.

There are different disturbances during the passenger boarding process. Landeghem and Beuselinck [5] divided the disturbances into three operational parts: calling passengers, boarding pass control at the gates, and passenger installation within the aircraft. An adequate strategy for reducing the boarding time is to split the passengers into groups, whereas these groups are separately called to enter the aircraft. Due to the high quantity of possible parameter variations, such as block size, block sequence or block affiliation Marelli et al. [6] highlight the importance of model driven evaluations to optimize the boarding procedures. These boarding evaluations provide an insight into the associated mechanisms. However, a simulation environment is only capable to run pre-defined scenarios, but it does not provide autonomous algorithms for developing the most efficient strategy (van den Briel et al. [7]).

II. SIMULATION APPROACH

Our research project mainly focused on disturbances occurring during passenger installation, namely the congestion in the aisle, the storage of hand baggage, and number of occupied seats between the aisle and the assigned passenger seat. However, disturbances based on passed rows are not taken into account.

A. Aircraft

For the simulation environment an A320-200 aircraft seating layout is chosen, which is used within the airline Air Berlin [8]. The aircraft is a regular single aisle aircraft with three seats on each side of the aisle and with a seating capacity of 174 seats in 29 rows (see fig. 2).

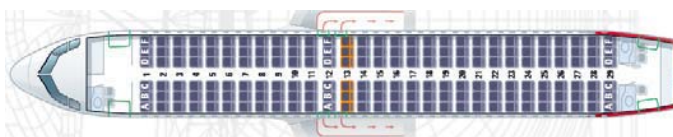


Figure 2. Aircraft seat layout

B. Model

In contrast to the mixed integer linear program approach introduced by Bazargan [9] or the multi-parameter discrete random process from Bachmat et al. [10], our simulation model is based on the so called asymmetric simple exclusion process (ASEP). The ASEP was successfully used for road traffic investigations. The boarding can be described as a stochastic forward directed, one dimensional, and discrete (time and space) process as well [11-13]. For this purpose the aircraft layout will be transferred into a regular grid as shown in fig. 3. The regular grid consists of cells with a size of 0.4 x 0.4 m. Each cell can either be empty or contain exactly one passenger.

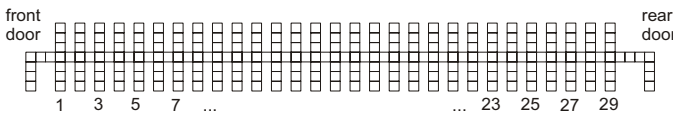


Figure 3. Grid based simulation model

To move forward the passenger can enter an empty cell at one timestep only. If the cell in front of is occupied the passenger has to wait in this timestep (probability to overtake passengers is set to zero). Assuming a maximum speed of 0.8 ms⁻¹ at the aisle (60 % of maximum passenger speed), the timestep has a width of 0.5 s. At each timestep during the simulation the position of all passengers is updated via a sequential shuffled update procedure [14, 15].

The passengers enter the aircraft at the front (rear) door and move from cell to cell along aisle until they reaches the assigned seat row. For each simulation run the arrival time at the aircraft is determined as a constant factor (*n* passengers per minute (PPM)). Before entering the aircraft all arriving passengers join the aircraft queue. If the queue is empty, they proceed directly into the aircraft; while otherwise they have to wait until all passengers arrived earlier have entered the aircraft. If both, front and rear door are used for boarding, the

passengers from seat row 1-15 use the front door and passengers with seat row 16-29 use the rear door.

Additionally to the general ASEP model, we assumed that a passenger leave this one dimensional process (walking at the aisle), if he has reached the assigned seat row. The time *t*, which the passenger needs to take his seat, depends on several factors. First, *t* depends on time of baggage storage *t_B*, (related to the number of baggage) as well as on the time for handling occupied seats *t_S* and on the response time *t_R* of all involved persons. For all of the time components a statistical probability (triangle distribution) is defined (compare [5, 11]). The distribution values are determined in tab. I.

TABLE I. PROBABILITY DISTRIBUTIONS

Distribution	Time values (s)				
	min. value	modus value	max. value	mean Value	standard deviation
store one piece of baggage	5.0	10.0	20.0	11.67	13.23
time for one seat movement	1.8	2.4	3.0	2.40	1.04
Reaction	6.0	9.0	20.0	10.5	12.77

The response time *t_R* can be directly calculated from the given probability distribution without any further input data. The storage time *t_B* is calculated by adding a random value for each piece of baggage, generated with the determined baggage storage distribution function. To determine the number of baggage pieces the distribution in tab. II is consulted.

TABLE II. BAGGAGE DISTRIBUTION

ratio (%)	Number of baggage pieces			
	0	1	2	3
	0	60	30	10

To determine *t_S* the character of the seat row state has to be clarified. At the chosen layout with a 3-3 seat configuration four different kinds of seat row states are possible:

- Seat access without any disturbances, (state A)
- Blocked aisle seat, (state B)
- Blocked middle seat, and (state C)
- Blocked aisle and blocked middle seat. (state D)

This disturbance list is sorted by the degree of arrangement complexity, by meaning of increasing time consumption. For example, to take a seat at the window with a blocked middle seat, the passenger at the middle seat has to move to the aisle seat and from there to the aisle itself (the aisle is blocked during the whole seating process). Now the window-seated passenger enters the seat row followed by the middle seat passenger (7 movements in total).

However, the number of required movements to ensure the aisle availability is lower than 7, because following passengers can pass the row 2 movements earlier: passenger one (on middle seat) need 2 moves to the aisle, passenger enters the

row and reaches the middle seat (2 moves), at this moment passenger one clears the aisle by entering the seat row as well (1 move). The further seat row arrangements (get to the corresponding window seat and middle seat) will be proceed without influencing the aisle passengers.

In the simplest case (state *A*) a passenger needs only 1 move to enter the row, *B* requires 4 movements, the third state *C* consumes 5, and the most complex row state *D* requires not less than 9 movements. Finally, t_S is calculated as a product of required movements and a random number given by the probability distribution (tab. I). In order to speed up the boarding process, it seems obvious to eliminate the most time consuming disturbances first.

C. Boarding strategies

For prearranging the passengers before entering the aircraft a call-of-system is used at the boarding counter. To determine the efficiency of boarding strategies, four different strategies are chosen:

- *Random*: the passengers get into the aircraft without a special order.
- *Outside-In*: passengers with window seats enter the aircraft first, followed by passengers with middle seats, and passengers with aisle seats.
- *Back-to-Front*: the aircraft is parted into blocks, whereas the block with the highest distance is boarded at first.
- *Block* boarding (best sequence): the aircraft is parted into blocks, whereas the fastest sequence of the blocks is used for boarding.

The *random* strategy is used as a baseline scenario to allow a target-performance analysis. Former studies pointed out, that the *outside-in* procedure is one of the fastest and suitable boarding strategies (see van den Briel et al. [7]). Therefore it is used to mark the upper limit of the boarding time. The *back-to-front* method is often determined as an unfavorable procedure, because the effort for arranging passengers is disproportionate to the expected time savings.

Finally, the common *block* boarding (fig. 4) is part of the analysis, although Landeghem [5] and Ferrari et al. [11] showed that *block* (or *half-block*) strategies are not significantly faster than *random* boarding procedures. However, a first evaluation with our simulation model yields different results, even considering several block sizes, block sequences, acceptance rate of boarding procedure, and seat load factors. Additionally, the use of the rear aircraft door is taken into account. An example of the block classification (6 blocks) is given at the following figure (fig. 4).

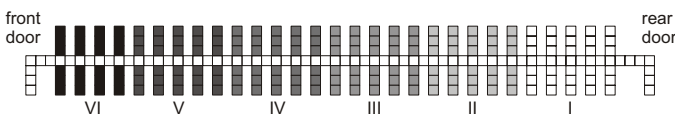


Figure 4. Block classification at grid model

Attention should be paid to the numbering sequence at fig. 4, which starts at the end of the aircraft. Consequently, the

back-to-front procedure is equivalent to a *block* boarding procedure with the sequence 123456.

D. Simulation runs

The parameters AR (acceptance rate), SLF (seat load factor) and PPM (arrival rate - passenger per minute) are varied within the boarding simulation environment. The simulation scenarios are generated by the combination of the following factors, whereas the default values are declared in braces.

- SLF and AR from 20% to 100% {85%}
- Arrival rate from 1 to 40 (PPM) {85%}
- 4 different boarding procedures {random}
- One and two door configuration {one door}

Each scenario is simulated 10000 times, to allow a significant statistical analysis of the results.

III. RESULTS

Waiting times arising from boarding disturbances are primarily caused by suboptimal seat row states. At *random* boarding the probability of seat row state *A* (no blocked seats) is about 66% (tab. III), whereas the *outside-in* boarding increase the quota to nearly 91%. Even the state *C* could be reduced to a marginal quantity of 1%. Thus, the change-over from the *random* to the *outside-in* boarding procedure results in system enhancements up to 20% (see fig. 11, 12).

TABLE III. SEAT ROW STATE FOR BOARDING PROCEDURES

Procedure	Seat row state (%)			
	A	B	C	D
Random	65.6	20.3	6.1	8.0
outside-in (AR=0.85)	90.8	5.2	2.9	1.1
outside-in (AR=1)	100	0	0	0

The influence of the row disturbances is not limited to the local row. Depending on the number of passengers which are not able to pass this critical row the local disturbance affects the whole boarding process and therefore the boarding time. In the next picture (fig. 5) the overall waiting time with respect to aisle position is shown, where $x = 0$ is marked as the aircraft door. If a passenger is not able to move forward a marker is left on this particular position at each timestep.

After the finishing the simulation all markers of each aisle position are counted. With increasing aisle length the waiting time declines. In the vicinity of door the waiting time is very high, indicating that the passengers could not move forward due to indirect disturbance in front of them. With increasing rate of arrival, this effect has greater influence. A large waiting time at the door is connected to a high queue length. If the gradient of the waiting time is nearly linear the optimal system load is reached. For the *random* boarding configuration with one door, AR = 0.85, and SLF = 0.85 the optimal system load is achieved at an arrival rate of approx. 9 PPM.

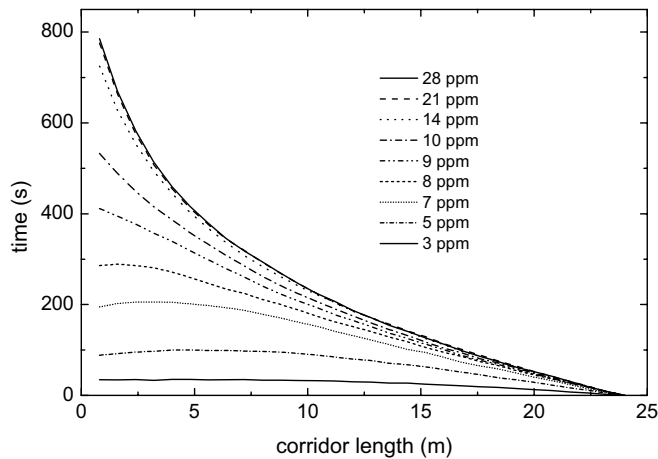


Figure 5. Waiting time

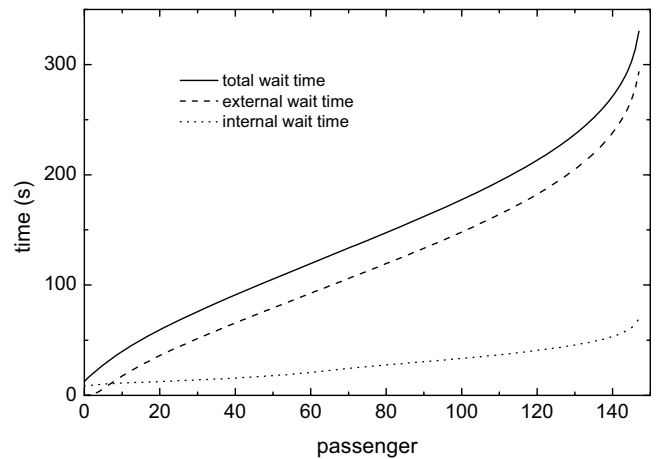


Figure 7. External, internal and overall waiting time

To solve the increasing door-handed waiting time distribution the second door is used for the boarding process as well (fig. 6). Passengers with seat rows from 15 to 29 could leave the front door queue and are directly guided to the rear part of the aircraft, without disturbing the passengers from seat row 1 to 14. Due to the queue shifting and the enhanced passenger segmentation the optimal system load is increased from 9 PPM to approx. 14 PPM in the *random* configuration. The small discrepancy between the left and the right shape of the curve in fig. 6 is caused by the different assigned row numbers for each door (front: 15 rows, back: 14 rows).

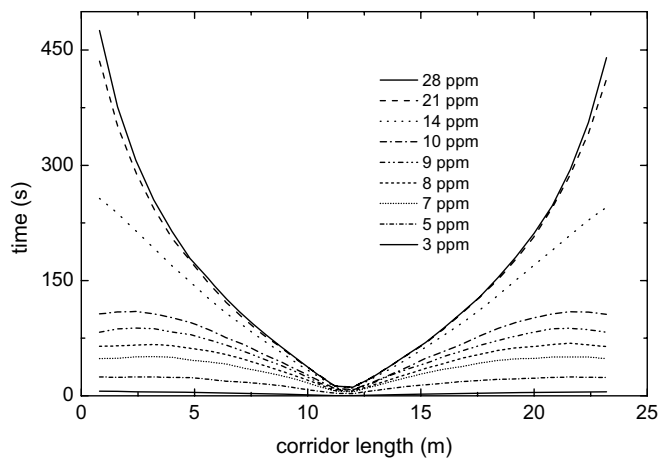


Figure 6. Waiting time (2 doors)

Generally, the passenger waiting time inside the aircraft can be separated into a direct and an indirect contribution. The direct part is only attributed to the number of baggage and the individual time for stowing the baggage. The indirect part depends on the seat row state and the aisle blocking time due to other passenger's activities (stowing baggage, waiting for seating, or waiting for passing). In fig. 7 the accumulated waiting time characteristics is shown. The internal waiting time has only a small impact on the passenger itself. Due to the passenger interactions the external waiting time is the main contribution to the overall waiting time. (46 passengers wait 99.8 s or less, whereas the direct part has a size of 16.9 s and the indirect part a size of 73.9 s.)

During the boarding the number of passengers without a seat is continuously decreasing. In fig. 8 the center line represents the mean value and outward lines are the 75th percentile, 90th percentile and the maximum (25th percentile, 10th percentile and minimum, respectively). Depending on the stochastic model assumptions the overall boarding time varies between approx. 925s and 1550s (see fig. 8 at 0 passengers without seat). The shape of the boarding time corresponds to a normal distribution with $\mu \approx 1191$ s and $\sigma \approx 83.8$ s. From 100 s to 900 s simulation times a nearly constant ratio of approx. 7.2 s per passenger seating rate is observed.

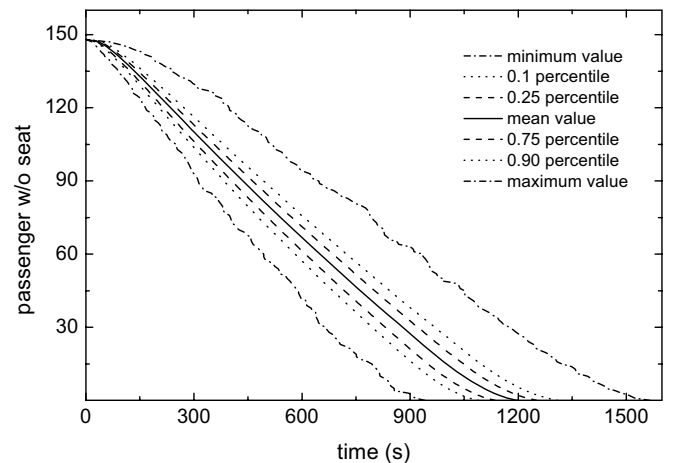


Figure 8. Passengers without seat during boarding process

After analyzing one single scenario, we focused on the comparison of different boarding strategies and parameter variations in the next paragraphs.

A. Block boarding and optimal block sequence

To determine the efficiency of *block* boarding two parameters have to be defined. The first parameter is the block size, which is similar to the number of rows which are boarded at the same time. Regarding to the A320 layout (fig. 4), the restriction to integer values, and the almost equal block size, the block number have to be element of {1,2,3,4,5,6,7,9,15}. In the following table (tab. IV) the simulation results for all

possible 3-block-sequences are shown in relation to the *random* boarding. Due to the fact that the *back-to-front* boarding (sequence 123) is defined separate boarding procedure, it is separated from the *block* sequences. The simulation analysis yields in no significant benefit of the *block* sequences over the *random* boarding procedure.

TABLE IV. BOARDING TIME FOR ALL 3-BLOCK SEQUENCES

Sequence	Mean value (s)	Standard deviation (s)	Efficiency (%)
1-2-3	1173.9	81.9	+ 1.4
2-1-3	1246.5	89.5	- 4.6
1-3-2	1332.4	96.4	-11.8
3-1-2	1378.8	100.6	-15.7
2-3-1	1419.8	96.0	-19.1
3-2-1	1612.6	103.3	-35.4
random	1191.0	83.8	0

The characteristics of the best sequence *block* boarding shown in fig. 9 points out a significant relationship between block size and boarding efficiency. The creation of two separate blocks could improve boarding procedure by 3.9 %, whereas the efficiency decreases by using three blocks to 1.4 % for *back-to-front* and reaches even negative values of -4.6 % for *block* procedure.

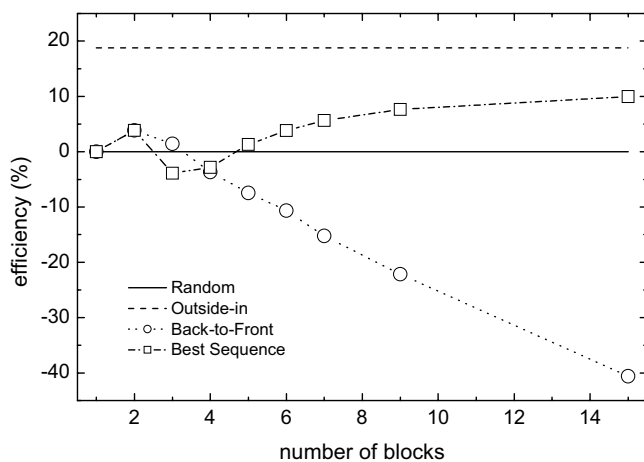


Figure 9. Efficiency of different block numbers

Using an appropriate block size of 6 blocks (approx. 5 rows per block with 30 passengers) the efficiency of *back-to-front* drops to -10.7 %, but increases for *block* to the prior level of 3.9 %. A further seat row segmentation finally results in efficiency measurements of -40.6 % and 10 % respectively. To evaluate the best *block* sequence, all possible sequences were tested; an n block configuration produces $n!$ specified sequences. The 720 sequences for a 6 block configuration are shown in fig. 10, whereas the variance is exemplary highlighted by error bars. Obviously, the sequence 246135 (compare fig. 4) with $\mu = 1133.5$ s, $\sigma = 72.52$ s is significantly faster than all other sequences and the sequence 654321 is the slowest sequence ($\mu = 2005.4$ s, $\sigma = 122.4$ s). Even though the sequence 135246 should be as fast as 246135, the impact of the reduced row number at block 6 (only 4 rows instead of 5) results in different values of $\mu = 1141$ s and a $\sigma = 78$ s.

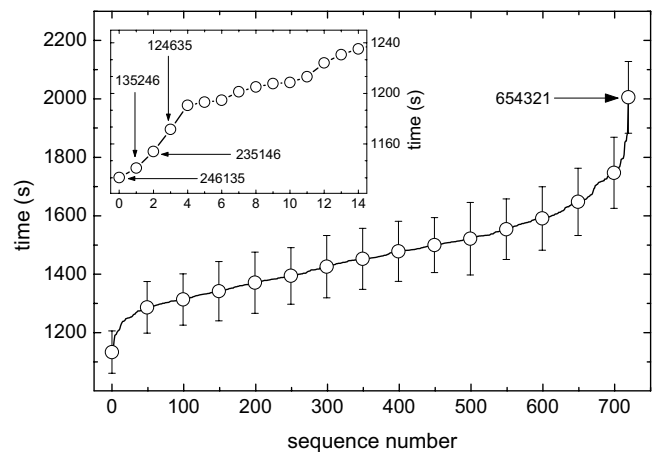


Figure 10. Sequences (720) for 6-block boarding procedure

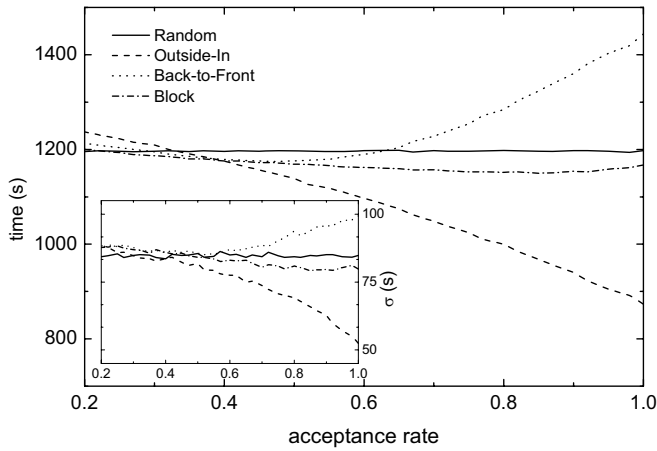
The simulation runs for the different block sizes points out, that alternating block sequences are much faster than other sequences. If the passenger group of the first block enters the aircraft they are queued in the aisle segment of the prior block. This prior block with occupied aisle should not be used for boarding and therefore one block is skipped in the *block* boarding sequence. Furthermore, the most efficient sequence starts always even numbered (246...) and followed by the odd numbered blocks (135...). In our further analysis the 6 block classification is used. In this context the *back-to-front* and the *block* nomenclature represent the block sequence 123456 and 246135 respectively. For the two door configuration this nomenclature has to be adapted. The blocks 123 are boarded through the rear door and 456 are boarded at the same time through the front door. Hence, the sequence for *back-to-front* is 342516 and for *block* 253416. In contrast to the one door boarding passengers, the effective block size is reduced to 3, because the passenger from blocks 123 do not disturb passengers from blocks 456.

B. Comparison of boarding procedures

To analyze the different boarding strategies one parameter (SLF, AR, PPM, number of doors) varies and the other parameters are kept constant at their default values defined in section II.D. For the comparison of the different boarding procedures the *random* procedure acts as a baseline indicator. This procedure is always marked with a solid line in the following figures. The investigation starts with a one door configuration, but the results of the two door configuration are already shown on the opposite. Due to the use of the same scale gradations an overall evaluation of the efficiency can be ensured. The inserted diagrams show the corresponding standard deviation characteristics.

With the increment of the acceptance rate (AR) from 0.2 to 1.0 (see fig. 11, 12) the boarding time of the *outside-in* procedure decreases at an average of 44.8 s per 0.1 acceptance rate for one door configuration and 23.4 s for two door configuration. At AR = 0.32 the *outside-in* procedure reaches the breakeven point. As expected, the *random* boarding times are constant, whereas the two door configuration shows an improvement of 25.9 % regarding to the boarding time and a reduced standard deviation of 28.4 s.

1) One-Door Boarding



2) Two-Door Boarding

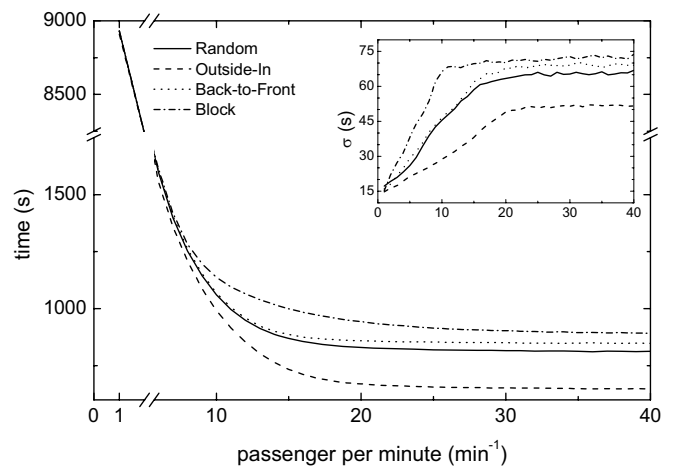
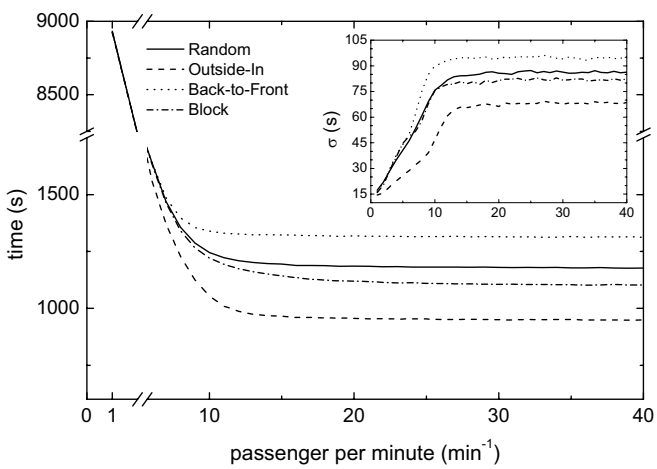
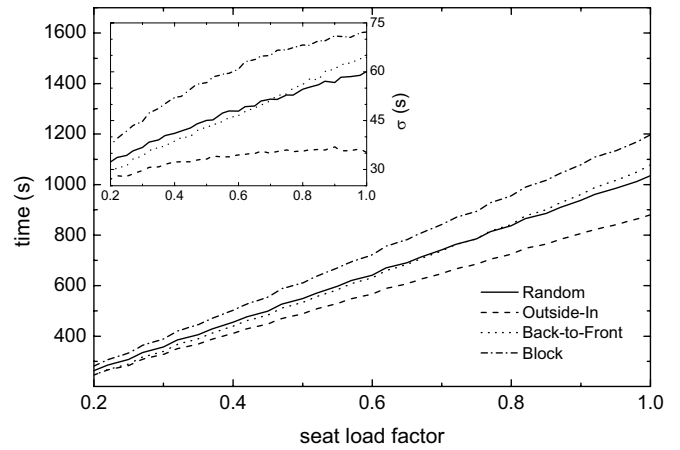
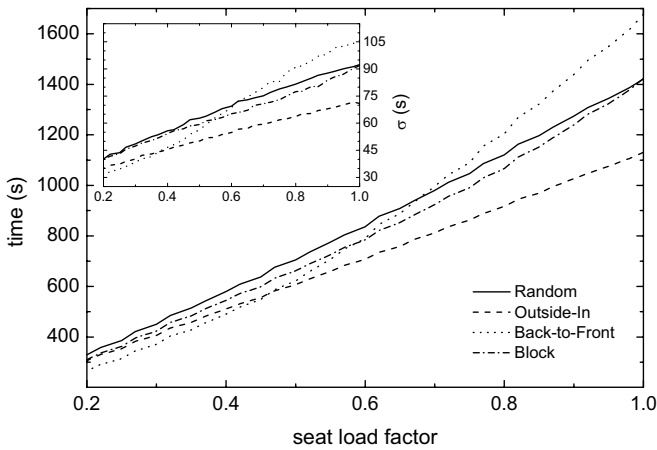
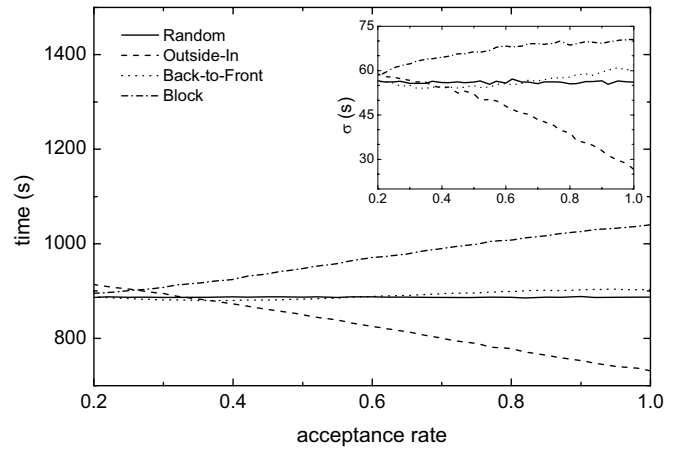


Figure 11. Boarding results using one door (acceptance rate AR , seat load factor SLF , and passenger per minute PPM vs. time)

Figure 12. Boarding results using two doors (acceptance rate AR , seat load factor SLF , and passenger per minute PPM vs. time)

The *block* boarding reaches the maximum efficiency of 3.9 % at AR = 0.85 (one door). In contrast, the *block* boarding is not efficient at the two door configuration. If the AR exceeds the 0.62 level at the one door configuration the *back-to-front* procedure gets inefficient, and induces no significant efficiency enhancements at the two door configuration. This inverse effect of *block* and *back-to-front* procedure was already point out in section III.A (see fig.9). The analysis of the SLF variation yields nearly linear correlations between the SLF and the boarding time, except the *back-to-front* procedure in the one door configuration. Analogue to the acceptance rate the boarding time gets inefficient at a certain point (SLF = 0.68), whereas the standard deviation already indicates this trend at SLF = 0.37.

The analysis of the increasing arrival rate provides no additional information about the comparison of different strategies. However, the direct comparison of the one door versus a the two door configuration shows that the arrival rate of approx. 11 PPM (one door) and 16 PPM (two door) assign an upper value for the arrival rate regarding to the boarding efficiency. From this point of view a further increment of the arrival rate will only have a marginal influence on the boarding time. This result corresponds to the waiting time analysis at the beginning of section III.

Finally, the comparison of the one door versus the two door configuration yields the results shown in tab. V. The parameter AR, SLF and PPM are kept at their default values of 0.85 %, 0.85 % and 14 PPM respectively. The standard deviation percentage refers to the procedure regarding mean value, whereas the efficiency refers to the *random* boarding with a one door configuration.

TABLE V. COMPARISON OF ONE DOOR VS. TWO DOOR CONFIGURATION

	procedure	mean (s)	standard deviation		Efficiency (%)
			(s)	(%)	
1 door	random	1191.0	83.8	7.0	0.0
	outside-in	968.3	65.8	6.8	18.7
	back-to-front	1324.3	94.8	7.2	-11.2
	block	1151.7	80.8	7.0	3.3
2 door	random	886.8	55.6	6.3	25.5
	outside-in	764.3	35.8	4.7	35.8
	back-to-front	901.2	57.7	6.4	24.3
	block	1018.8	69.2	6.8	14.5

The utilization of the second aircraft door results in an enhanced efficiency of 25.5 %, without even considering particular boarding procedures. In comparison to the *outside-in* procedure (one door) an additional improvement of 7 % and reduction of the standard deviation by 0.5 % is realized. Furthermore, the result points out that *back-to-front* and *block* boarding are not recommended procedures. A marginal efficiency value of 3.3 % with a nearly unchanged standard deviation does not legitimate the application of the *block* boarding at the one door configuration. Looking at the two door configuration, the *outside-in* procedure achieves the best efficiency of 35.8 % with the smallest standard deviation of 4.7 %.

IV. CONCLUSION

This paper presents a simulation model to evaluate different boarding procedures and the influence of the variation of the corresponding input parameter (seat load factor, passenger acceptance rate of boarding procedure and arrival rate). The results of the simulation runs show the expected high efficiency of the *outside-in* boarding procedure and the marginal advantage of adjusted *block* procedures. With the utilization of the second aircraft door further enhancements are achieved. Airlines with apron-parking aircraft could easily use a second door for boarding. To achieve an efficiency enhancement of approx. 25% (see tab. V), the passengers have to split up in two groups only, regarding their assigned seat row.

REFERENCES

- [1] Advisory Council for Aeronautical Research in Europe, Strategic Research Agenda 2, Volume 1+2, 2004.
- [2] Airbus, A380-Airplane Characteristics for Airport Planning AC, 2005.
- [3] Airbus, Global Market Forecast (GMF), 2007.
- [4] Boeing, Boeing Current Market Outlook (CMO), 2007.
- [5] H. van Landeghem and A. Beuselinck, "Reducing passenger boarding time in airplanes: A simulation approach", *European Journal of Operations Research*, 142, pp. 294-308, 2002.
- [6] S. Marelli, G. Mattocks, and R. Merry, "The role of computer simulation in reducing airplane turn time", *Boeing Aero Magazine*, Issue 1, 1998.
- [7] M. van den Briel, J. Villalobos, G. Hogg, T. Lindemann, and A.V. Mule, "America West develops efficient boarding strategies", *Interfaces*, 35, pp. 191-201, 2005.
- [8] Air Berlin, Timetable Winter 2007/2008, p.27 www.airberlin.com
- [9] M. Bazargan, "A linear programming approach for aircraft boarding strategy", *European Journal of Operational Research*, 183, pp. 394-411, 2007
- [10] E. Bachmat, D. Berend, L. Sapir, S. Skiena, and N. Stolyarov, "Analysis of aeroplane boarding via spacetime geometry and random matrix theory", *Journal of physics A*, 39, 453-459, 2006.
- [11] P. Ferrari and K. Nagel, "Robustness of efficient passenger boarding in airplanes", *Transportation Research Board Annual Meeting*, paper number 05-0405, Washington D.C., 2005.
- [12] B. Urban, "Untersuchungen zur Optimierung des Boarding-, Deboardingverhaltens von Passagieren unter Nutzung eines individuen-basierten Simulationssystems", unpublished thesis, Technische Universität Dresden, 2007.
- [13] A. Kirchner, H. Klüpfel, K. Nishinari, A. Schadschneider, and M. Schreckenberg, "Simulation of competitive egress behavior: comparison with aircraft evacuation data", *Physica A*, 324, pp. 689-697, 2003.
- [14] H. Klüpfel, "A cellular automaton model for crowd movement and egress simulation", PhD thesis, Universität Duisburg, 2003.
- [15] M. Wölki, A. Schadschneider, and M. Schreckenberg, "Asymmetric exclusion processes with shuffled dynamics", *Journal of Physics A: Mathematical and General*, 39, pp. 33-44, 2006.

A Preliminary Evaluation of Potential Cargo Demand for Very Light Jets

Yue Xu

*Department of Civil and Environmental Engineering
Virginia Polytechnic Institute and State University,
Blacksburg, VA, 24060*

Antonio A. Trani

*Department of Civil and Environmental Engineering
Virginia Polytechnic Institute and State University,
Blacksburg, VA, 24060*

Abstract— this paper presents a research effort to study future air cargo demand using new generation Very Light Jets. Cargo demands are generated at county and airport level using T100D and Woods & Poole demographics data. At airport level, a growth factor based FRATAR model is applied to distribute air cargo demand among cargo airports up to year 2025. Historical trends of all-cargo carriers load factors are analyzed. An economics model is built to study Very Light Jet cargo transport cost. Cases studies are conducted to assess the competitiveness of the VLJ in terms of transport time and cost. Throughout our analysis, air cargo is further categorized into freight and air mail as they have different characteristics.

Keywords-component *Very Light Jet, Air Cargo, Growth Factor, Demand Forecast*

I. INTRODUCTION

This paper is composed of two parts. Air cargo demand at airport and county levels is described in the first part. A Very Light Jet (VLJ) cargo transport cost model and case studies are presented in part two. The traditional four-step-model is applied and this paper addresses demand generation, distribution and mode split.

The first part uses Bureau of Transportation Statistics Air Carrier Statistics Domestic Market Database (T100D) as the primary data source. Air cargo demand is generated at around 900 cargo airports based on socio-economics. A growth factor based FRATAR model is applied to distribute predicted air cargo demand. In addition, demands at special transfer locations such as Memphis Airport (hub for FedEx) and Louisville Airport (hub for UPS) are redistributed.

In the second part of the paper, a cargo aircraft economics model is proposed to estimate cargo shipping cost using VLJ. The first Very Light Jet, Eclipse 500 from Eclipse Aviation is used as the prototype vehicle in the analysis. Various cost components including hourly variable cost, annual fixed cost, periodic cost, personnel cost, and facility costs are integrated in the model. The cost estimates are derived from estimates published by Business and Commercial Aviation and ARG/US Operations Planning Guide [1]. The model generates life cycle cost metrics including cost per hour, cost per mile, and cost per pound-mile, and a summary of annual cost components. Air cargo size and load factor are also analyzed.

II. LITERATURE REVIEW

Freight demand modeling has emerged as a major issue in the transportation industry. State and regional cargo demand modeling have received extensive attention in recent papers [2][3][4][5][6]. These models usually concentrate on rail and truck operations. State-of-the-Practice freight databases are well documented by A. Mani and J. Prozzi [7].

There are several nation-wide freight models in the U.S. and Europe. Freight Transportation Research Associates (FTRA) developed an input/ output model that forecasts freight demand in the U.S. The model collects data from 1965 through 2004 at 3-digit Standard Transportation Commodity Code (STCC) level and assigns shipments to rail, truck, pipeline and water [8]. Another national freight model, Global Insight's North American Trade and Transportation Data (TRANSEARCH INSIGHT) represents a more detailed multi-modal model. The model utilizes Global Insight's quantitative economics model to forecast freight volume up to 2030. It estimates inbound/outbound shipments at BEA (Bureau of Economic Analysis) and county level in terms of 4-digit STCC commodity [9].

In addition to proprietary databases, the Federal Highway Administration is developing a Freight Analysis Framework (FAF) that integrates data from various databases to estimate freight flow among states, regions, and major international gateways [10][11]. The first generation of FAF, FAF¹, is based on Bureau of Transportation Statistics (BTS) Commodity Flow Survey (CFS) 1997. CFS is one of the most commonly used freight flow databases in the literature. It collects sample data through survey forms from a universe of about 800,000 establishments [12]. However, CFS is estimated to cover about 70-80 percent of the U.S. cargo movement [13]. Therefore, the FAF model incorporates complimentary data sources to complete missing elements of CFS records. The first version, FAF¹ estimates shipment in 1997 and provides forecasts for 2010 and 2020. The new product, FAF², derives inputs primarily from CFS 2002 and predicts freight shipments through 2035 in 5 year increments [13]. It is comprised of three categories of data including CFS within-scope data, auxiliary data and CFS out-of-scope data. CFS within-scope data comes directly from CFS 2002. Auxiliary data represents complementary databases and is

used in a log-linear model and iterative proportional fitting to estimate CFS out-of-scope data. The most detailed Origin-Destination (OD) table is a region to region OD table by mode by 2-digit SCTG code.

National freight movement modeling attracts extensive attention in Europe as well [14]. In Great Britain, a national multi-mode freight model, Great Britain Freight Model (GBFM), has been developed to predict freight demand and policy impacts [15]. The model estimates international cargo movements between counties within Great Britain and Europe as well as internal county to county and postcode to postcode cargo flows. Policy impacts can be evaluated via a set of operating cost models and transport route characteristics. A logit model is applied for mode and route choice. Meanwhile, there are several similar models such as Sweden SAMGODS [16][17], Norway NEMO [18], and Belgium WFTM [19]. All of them represent network models with cost functions and freight flows on modes and routes [20]. Besides network models, discrete models are widely used. Italy's SISD and the Netherlands's TEM models belong to this family [20]. In addition to TEM, the Netherlands has two other freight models that employ different philosophy [21]. The first one is SMILE, a model that introduces logistic segmentation and intermodal transport chains into a multimodal network [22]. The second model, MOBILEC, applies casual relationships between economy, mobility and dynamic growth [23]. France has two national freight models. Simulation techniques are used in the first model to forecast demand and assess policy impacts whereas the second model (named Transalpine) model is based on transport costs [20].

Apart from multi-mode cargo database, there are several dedicated air cargo databases. The Official Airline Guide (OAG) Cargo Database produces annual worldwide cargo schedules between origin and destination [24]. OAG Cargo covers cargo data from 138 carriers (as of Nov. 2006) who report their statistics to OAG. However, data from mainline cargo shippers such as FedEx, UPS and Airborne are not included. A more complete database, the FAA's Air Carrier Activity Information System (ACAIS) All-Cargo Activity Report covers 492 domestic carriers in 2007 [25]. It is designed for Federal funding allocation but not for direct cargo demand modeling because it only records the maximum possible cargo landing weight at airports, i.e. the maximum landing weight of scheduled cargo flights no matter how much payload is carried. Another public database, BTS Form 41 T100D Domestic Market database, comprises fewer carriers (64 all-cargo carriers in 2005) but includes mainline shippers and actual payload records. It consists of monthly data reported by certified U.S. air carriers on passengers, freight and mail transported [26]. Furthermore, it provides critical information on available capacity and seats, aircraft type, service class, aircraft hours, etc. Therefore, it is used as the main data source for airport-to-airport cargo flow estimation.

III. MODELING

In the beginning, multi-modal public databases such as FAF and CFS are explored to estimate air cargo share. However, due to the absence of detailed mode specific data, especially the air mode, it is difficult to quantify air cargo components using FAF and CFS. Consequently we turned to the air transportation oriented database T100D as our main data source. Figure 1 shows the flowchart for Very Light Jet cargo demand generation and distribution.

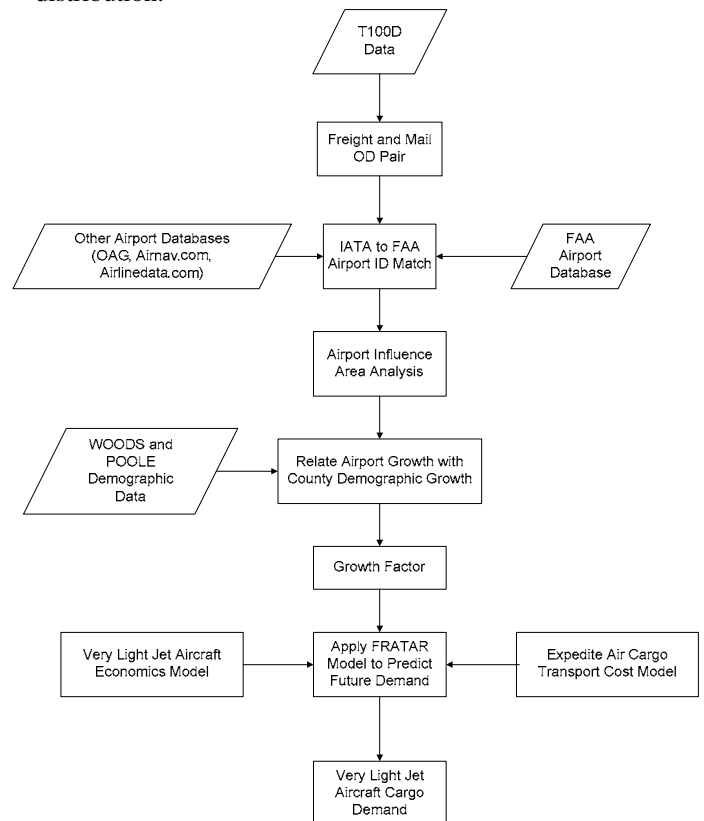


Figure 1: Very Light Jet Cargo Demand Generation and Distribution Flowchart.

T100D is used as the primary database and the baseline OD flows are derived for freight and mail. An airport influence area study is conducted to associate airport growth with neighboring county demographic growth. A growth factor based FRATAR model is used to distribute demand among OD pairs. The following sections will elaborate on each subject.

A. AIRPORT TO AIRPORT CARGO DEMAND

T100D consists of monthly data reported by certified U.S. and foreign air carriers on passengers, revenue freight and mail. Our analysis emphasizes on the cargo components of this database - revenue freight and mail. Conventional four-

step model is applied to estimate VLJ cargo market share and demand. Domestic airport to airport cargo flows, including freight and mail, is extracted from origin and destination airports in the demand generation step. A demand prediction up to 2025 is obtained using growth factors. This paper covers demand generation, demand distribution and mode split.

Freight and mail demands are generated at airports using T100D database (year 2003). Different airport identifier naming conventions are discovered between T100D and FAA. An effort is made to convert T100D airport ID to FAA airport ID using auxiliary data source such as airnav.com and airliners.com in order to obtain necessary airport facility data offered by FAA Landing Facility Database.

1) AIRPORT SERVICE RADIUS ANALYSIS

VLJ is capable of operating at small to medium community airports to save transport time. To explore the potential cargo transport demand of the VLJ, it is necessary to redistribute the cargo demand at current cargo airports to surrounding counties and then re-evaluate the airport choice.

In order to relate airport with counties demographics, a service radius analysis is conducted. First, a population based county centroid is located at each county based on census tract level population data. Then a correlation study is made using ArcGIS. Any county centroid within the service radius is assumed to be served by the airport. In our analysis, we assume T100D cargo airports are part of the hub-spoke system and are able to connect to other airports. Under this assumption, all T100D cargo airports could be candidate cargo airports for the county within their service radius. Three different scenarios are examined, i.e. 60 -120 miles, 70-120 miles and 90-150 miles. The radius varies proportionally to airport production and attraction demand. In the first scenario, one third of the cargo airports are unable to couple with any county. Therefore the lower bound of the radius is increased and the results are reexamined. By increasing the minima to 90 miles and the maxima 150 miles, all continental cargo airports can be coupled. However, there are counties cannot be coupled any neighborhood airports. In this case, the closest cargo airports are assigned to them. Similarly, if the airport cannot find a nearby county, the closest county is assigned. On average, around 20 counties can be served by one airport, implying a significant overlap in the service area considering a total of 3,000 counties in the continental U.S.

2) REGRESSION ANALYSIS

The outcome of service radius analysis is the association of airport and counties based on demand. Demographics of served counties can be explained as explanatory factors for future airport growth.

Various regression analyses are conducted to explain cargo demand versus socio-economic factors. County demographic data such as total employment and population are the initial trial parameters. As a result of the collinearity test, any pairs of the three parameters cannot be used simultaneously in the regression due to high collinearity. Linear regression produces satisfactory R-square value after

removing several outliers which include large airports attracting and producing enormous cargo or hubs for major air cargo operators such as UPS and FedEx. Therefore, a single variable linear relationship is used. Transportation industry employment and earning are used as independent variables for freight and mail respectively.

3) GROWTH FACTOR

The Woods and Poole Complete Economic and Demographic Data Source provides county demographics data and forecast from 1970 to 2025 for a total of 3,091 counties [28]. Based on county-airport association and W&P demographic prediction, a two-by-two growth factor matrix is developed for each cargo airport by airport type (as origin (production) or destination (attraction) airport) and cargo type (freight or mail).

For outliers such as Memphis and Louisville airport, FAA Terminal Area Forecast (TAF) demand forecast is used. TAF includes forecasts for active airports in the National Plan of Integrated Airport System (NPIAS). It should be noticed that growth factors of these hub airports are smaller than the growth factors of small to medium airports obtained from demographics growth. This can be explained by terminal congestion levels and the potential to use small to medium airports to detour demand of those hub airports. By this procedure, a complete growth factor matrix is built (See Figure 2).

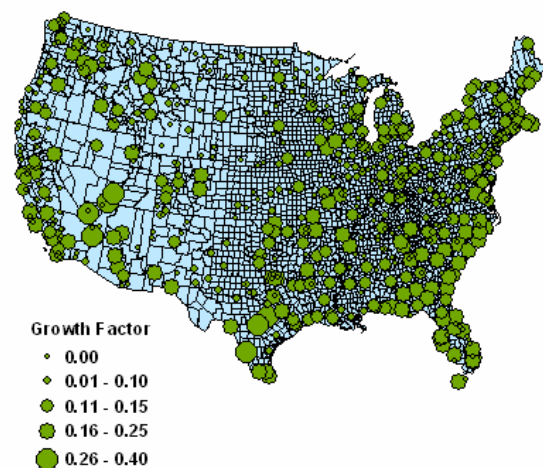


Figure 2: Mail Attraction Growth Factor in Year 2010.

4) FRATAR MODEL

FRATAR model is a growth factor based demand distribution model. Given both origin and destination growth factors, demand distribution prediction can be obtained using following equation.

$$T_{ij} = T_i \frac{t_{ij} G_j}{t_{ix} G_x} \quad \text{Equation 1}$$

Where T_{ij} = number of trips estimated from zone i to zone j
 T_i = future trip generation in zone i
 t_i = present trip generation in zone i
 $G_{j,x}$ = growth factor
 t_{ij} = present trips between zone i and j
 t_{ix} = number of trips between zone i and other zone x

Both of the origin and destination growth factors have been derived from correlated county demographics. FRATAR model is then applied at airport level to predict future cargo origin and destination flow up to year 2025 (

Figure 3). The model converges quickly after a few iterations.



Figure 3: Mail Attraction at County Level in Year 2010.

B. COUNTY DEMAND ASSIGNMENT

Initially, cargo demand at airports is assigned to neighboring counties within service radius based on regression analysis. Initially, cargo demand is distributed to counties proportionally to their share to the total demographics served by the associated airports. Then demand is summed up if one county has multiple service airports.

Two large cargo centers exist in the country, Memphis International Airport (MEM) and Louisville International Airport (SDF), hubs of FedEx and UPS, respectively, function as sink nodes in the initial distribution. These nodes do not represent the true cargo destination and thus warrant a secondary redistribution step. This step treats MEM and SDF as transfer nodes. A simple transfer rate is

assumed: 90% of freight and 85% of mail demand reached MEM is redistributed and at SDF, 90% and 75%. This rule assures a symmetric cargo flow arriving and leaving MEM and SDF.

After the two-step distribution, the county demand OD matrix was collected by rows and columns to obtain county level demand generation.

Figure 4 shows air mail demand attraction at county level.

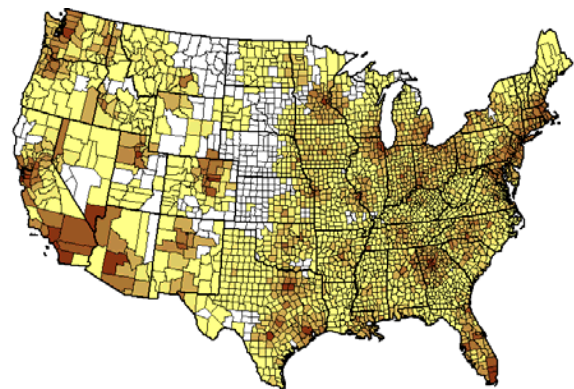


Figure 4: Mail Attraction at County Level.

C. AIR CARGO WEIGHT STUDY

Upon completion of cargo demand generation and distribution, total air cargo demand and OD flows are projected at county level. However, VLJ is not capable of accommodating all the demand due to capacity and weight balancing constraints. Practically VLJs are unable to transport shipments beyond 1,000 lbs. Therefore it is necessary to study air cargo size distribution to estimate VLJ compatible cargo demand.

Air cargo weight distribution can be derived from the Commodity Flow Survey (CFS) from the four regions reported. Study shows that there is no significant difference in the four regions for shipment less than 1,000 lbs. This result justifies a national cargo weight analysis concentrating on shipments less than 1,000 lbs. It shows approximately 40% of current air cargo is compatible with VLJ aircraft.

D. CARGO AIRCRAFT ECONOMICS MODEL

VLJ aircraft economics models are developed using STELLA Research 7.0. Various cost components including hourly variable cost, annual fixed cost, periodic cost, personnel cost, and facility costs are integrated in the model.

The model allows users to modify parameters such as jet fuel cost, load factor, number of pilots, flight hours per year, percent resale value, profit margin, mission stage length, annual pilot salary, and percent repositioning flights. The model provides Life Cycle Cost Metrics, including cost per hour, cost per mile and cost per pound-mile, and a summary of annual cost components.

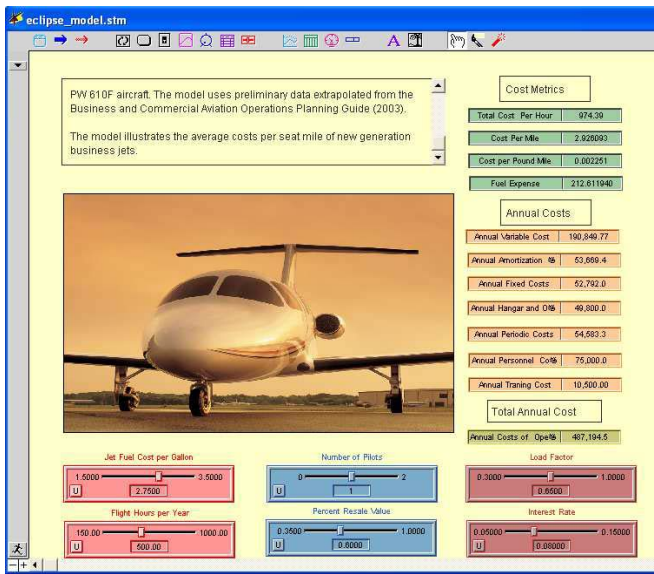


Figure 5: Eclipse 500 Economics Model.

Life-cycle costs of similar small cargo aircraft, Cessna Caravan and Falcon 20F, are also modeled for comparative purposes. The cost estimations are derived from data published by Business and Commercial Aviation and ARG/US in the Year 2004 Operations Planning Guide. Average load factor is obtained from BTS Air Carrier Summary Data T2 as the quotient of revenue ton-miles over available ton-miles. From the historical trend revealed by T2, average load factor is assumed to be 0.65. Life-Cycle cost estimates for the three aircraft are shown in Table 1.

Table 1.

Table 1: Small Cargo Aircraft Economics Model Output.

Aircraft	Total Cost per Hour (Dollars)	Cost per Pound Mile (cents / lb-mile)
Eclipse 500	974	0.25
Cessna Caravan	1,112	0.50
Falcon 20F	2,258	0.63

E. CASE STUDIES

Case studies are conducted to assess the competitiveness of VLJ cargo operations. Commercial cargo

service providers such as UPS and FedEx are modeled as the primary competitor. Both FedEx and UPS offer same day and second day expedite service. In our case studies, UPS Ground and UPS Sonic-Air express service are modeled as the primary commercial service competing against the on-demand cargo service using VLJ.

It is assumed that the VLJ aircraft operates either as a dedicated on-demand service or with a load factor of 0.65. The former means the VLJ aircraft will be used solely for the requested package whereas the latter combines the package with other payloads. Two package sizes, i.e. 100 lbs and 200 lbs, and ten distance categories (100 – 1000 miles) are evaluated. Estimation of VLJ travel time is based on assumption that 3,000+ airports available for VLJ cargo operation [29]. The closest airport to the origin and destination zip code is chosen as origin and destination airport. The result (for 200lbs) indicates that VLJ aircraft require the least transport time thanks to its point-to-point operation concept. In the cost competition, the position of cargo VLJ operation greatly depends on load factor. A load factor of 0.65 is sufficient for the first place in the competition. However, schedule delays may be added for transport time. If the operation is completely dedicated to this package, cargo VLJ will be most expensive one among the three methods.

IV. RESULTS AND CONCLUSIONS

This paper presents demand generation, distribution and partial mode choice of the four-step model to study VLJ cargo transport capability. Air cargo demand is generated and distributed at 900 cargo airports and 3,091 counties for the entire US. A strong linear correlation is found between cargo demand and demographics. Transportation / Communication / Public Service employment and earnings are found to be the best explanatory variables for freight and mail demand respectively.

It is observed that cargo demand concentrates at highly populated area. As a result of cargo airport service area analysis, more than 200 counties cannot locate a cargo airport within 60-120 mile radius. And it is under the assumption that all airports appearing in T100D are reliable enough to offer regular air cargo service. More counties will be isolated if only airports with regular cargo service are to be considered.

Growth factor analysis indicates that the potential growth tends to concentrate at high density where congestion could constrain projected growth in the future. By operating from point-to-point and reduce ground time at under-utilized rural airports, VLJ offers an attractive alternative mode for highly time sensitive shipment. Case studies suggest that VLJ provides substantial travel time savings compared to expedited commercial service currently available. VLJ offers competitive prices if moderate to high load factors are achieved.

V. MODEL LIMITATION AND RECOMMENDATIONS

In the distribution analysis, a static airport network has been applied. Emergence of new cargo OD pairs is not considered in the model. Furthermore, the demand distribution is limited by T100D OD pair network. Intermodal operations that reflect the true origin and destination are approximated by service area. Small to medium hubs' sorting and redistribution function is not considered in this analysis. A dynamic air cargo operator behavior model and the inclusion of small to medium hub redistribution step will enhance our distribution model and in turn improve mode choice. Further validation is needed.

Our primary data source, T100D, does not provide information on small cargo carrier operations such as air taxi. A database collection analysis should be undertaken to understand this market segment.

A complete mode choice module needs to be built to estimate VLJ market. Several parameters need to be addressed, including percentage of highly time sensitive cargo, percent repositioning flights, commodity value of time, commercial cargo transporter expedite service cost structure, and reliability of VLJ operations.

VI. ACKNOWLEDGEMENT

This study has been supported by the National Consortium of Aviation Mobility (NCAM) under a grant with the Virginia SATS Alliance. Thanks are due to Stuart Cooke and Jeff Viken (NASA Langley Research Center) for their constructive criticism and technical oversight. We would like to thank Dr. Hojong Baik and Howard Swingle for their valuable input. The views expressed in this paper are those of the authors and do not reflect the official policy or position of the U.S. government or an airport authority.

References:

- [1] Business and Commercial Aviation, 2004 Operations Planning Guide, the McGraw-Hill Companies, Aug 2004.
- [2] L. Cheng, Developing a Truck/Freight Model for the Los Angeles Metropolitan Area, Report prepared from Metropolitan Transportation Authority from Citilabs, Jun 2004
- [3] Florida DOT Systems Planning Office, Florida Intermodal Statewide Highway Freight Model, Jan 2000
- [4] R.R. Souleyrette, et al., Statewide Transportation Planning Model and Methodology Development Program, Final Report, Nov 1996.
- [5] J. Gliebe, I-75 North South Transportation Initiative Truck Model, <http://transportation.ky.gov/planning/traffic/mug/files/KYMUGmeeting.ppt>, accessed Jan 2008.
- [6] J. Holguin-Veras, et al., An Assessment of Mythological Alternatives for a Regional Freight Model in the NYMTC Region, prepared for New York Metropolitan Transportation Council (NYMTC), May 2001.
- [7] A. Mani and J. Prozzi, State-of-the-Practice in Freight Data: A Review of Available Freight Data in the U.S., Center for Transportation Research at the University of Texas at Austin, Feb 2004.
- [8] Freight Transportation Research Associates, Freight Forecasting Model, <http://www.ftrassociates.net/freightmodel.aspx>, accessed Jan 2008.
- [9] [Global Insight Inc., North American Trade and Transportation Database, <http://www.globalinsight.com/ProductsServices/ProductDetail1024.htm>, accessed Jan 2008.
- [10] R. Schmitt, T. Tang, and R. T. Curlee, Freight Analysis Framework and the Commodity Flow Survey, Commodity Flow Survey Conference, Boston MA, Jun 2005.
- [11] [Federal Highway Administration of the US Department of Transportation, Office of Operations, Freight Analysis Framework (FAF), http://www.ops.fhwa.dot.gov/freight/freight_analysis/faf/index.htm, accessed Jan 2008.
- [12] Federal Highway Administration of the US Department of Transportation, Office of Operations, FAF² Technical Documentation, http://www.ops.fhwa.dot.gov/freight/freight_analysis/faf/faf2_tech_document.htm, accessed Jan 2008.
- [13] Bureau of Transportation Statistics, Freight in America, Jan 2006.
- [14] L. Lundqvist and L.G. Mattsson, National Transport Models – Recent Developments and Prospects, Springer, 2001.
- [15] WSP and John Bates Services, Audit of the Great Britain Freight Model, Final Report – Contract PPAD 9/134/22, Feb 2005.
- [16] G. de Jong, H. Gunn, W. Walker, and J. Widell, Study on Ideas on a New National Freight Model System for Sweden, Prepared for the SAMGODS Group by RAND, 2002.
- [17] SAMPLAN, The Swedish Model System for Goods Transport – SAMGODS. Swedish Institute for Transport and Communications Analysis Report 2001:1. Stockholm, Sweden, 2001.
- [18] I. B. Hovi and A. Vold, An Overview of the National Freight Model for Norway, Conference on National Freight Transport Models, Helsingor, Demark, Sep 2003.
- [19] ME&P – WSP, et al., DfT Integrated Transport and Economic Appraisal, Review of Freight Modeling, Report B2: Review of Models in Continental Europe and Elsewhere, Jan 2002.
- [20] P. Tardieu, and Rijswijk, Validation of European Transport Demand Forecasting Models and Scenarios, TRANSFORUM, Antwerp, Belgium, Feb 2005.
- [21] L.A. Tavasszy, Characteristics and Capabilities of Dutch Freight Transportation System Models, RAND, 1994, ISBN: 0833015494.
- [22] L.A. Tavasszy, et al., Scenario-wise Analysis of Transport and Logistics Systems with a SMILE, TNO Inro. Delft, Netherlands.
- [23] Van de Vooren, F.W.C.J.. A Policy Oriented Model about Economy, Mobility, Infrastructure and Other Regional Features, 8th World Conference on Transport Research, Antwerp, in: World Transport Research, vol. 4, Elsevier Science, pp. 43-56, 1999.
- [24] OAG Cargo Solutions, <http://www2.inforwarding.com/>, accessed Jan 2008.
- [25] Federal Aviation Administration, All-Cargo Reporting, http://www.faa.gov/airports_airtraffic/airports/planning_capacity/passenger_allcargo_stats/reporting/, accessed Jan 2008.
- [26] Bureau of Transportation Statistics, Air Carrier Statistics (Form 41 Traffic) – U.S. Carriers, <http://www.transtats.bts.gov/DataIndex.asp>, access Jan 2004.
- [27] Cambridge Systematic Inc., Commis Corp, and University of Wisconsin-Milwaukee, Quick Response Freight Manual, Final Report, Sep 1996.
- [28] Woods & Poole Economics, Data for All U.S. Counties and Metro Areas, 2005.
- [29] Y. Xu, A.A. Trani and H.J. Baik, Preliminary Assessment of Lower Landing Minima Capabilities in the Small Aircraft Transportation System Program, Transportation Research Record, Volume 1915, 2005.

Track 8

Airlines and ATC Concerns

Untapped potential of on-board advertising

Benedikt Badanik, Hitham Fakh, and Milan Stefanik

Air Transport Department
University of Zilina
Zilina, Slovak Republic
benedikt.badanik@fpedas.uniza.sk
hitham.fakh@fpedas.uniza.sk
milan.stefanik@fpedas.uniza.sk

Abstract— Today advertisements can be seen everywhere, on the seats of grocery carts, on the walls of an airport walkway, on the sides of buses, heard in telephone hold messages and they can even be seen on the fuselage of an aircraft or on-board. On-board advertising recently becomes a significant source of revenues for airlines all over the world. Inspired by ground public transport, air carriers have started to explore a great potential of on-board marketing and advertising. Advertising on-board provides smarter and cost-effective way of communicating with the customer and it is one of the new methods to inform potential customers about products and services and how to obtain and use them. However, compared to other means of transport, the potential of advertising on-board of aircraft seems to be untapped. It is necessary to realise that revenues from on-board advertising can help airlines to keep their competitiveness. Despite the fluctuation of fuel price, airlines will be able to maintain constant fare thanks to revenues from advertising. The main objective of our study is to analyse a real potential of on-board advertising. The study is mainly focused on low-fare airlines as these reach both high fleet utilisation and high load factors. These facts make low-fare carriers being an ideal market for on-board advertising. By means of this study, we would like to answer the question, if Michael O’Leary’s dream of no-fare airline can come true in near future or if it is just utopia.

Keywords; on-board advertising, ancillary revenues, operating costs

I. INTRODUCTION

Public transport is generally considered to be very valuable marketplace for advertising. In big cities, there are thousands of people using public transport services every day. In London, over 3 million passenger journeys are made across the underground network every day and London Tube earns billions of Euros every year from its advertising space [1]

Commercial advertising is an indispensable source of revenue for many coach and railway operators, providing services on the regional, national, and international level. In the field of passenger transportation, advertising becomes a very efficient tool for being competitive.

Compared to other means of transport, air transport seems to be behind in the on-board advertising development. Many airlines have already started to use the advertising space in

aircraft cabins for the promotion of hotel chains or car rental companies but the market is by far untapped. Moreover, the development in the field of on-board services and on-board entertainment will definitely lead to an increase of the on-board commercial advertising potential.

The main aim of this article is to perform an analysis of the on-board advertising potential and find out if further development in this field can provide airlines with sufficient financial resources and thus lead to deployment of the no-fare business model. In our opinion, a good concept of on-board advertising can bring a significant competitive advantage for many airlines regardless of their business model. The advertising revenues can partly compensate growing fuel prices, and traditional airlines do not need to collect fuel surcharges. On-board advertising in combination with latest on-board technologies could also lead to transformation of low-fare airlines to no-fare airlines. Currently, no-fare flights are used as very efficient marketing tool, but in the future the on-board marketing and advertising may be used for reducing or covering the cost of travel.

The possibilities of on-board marketing and advertising are almost unlimited. It does not need to be used only for the promotion of hotel chains or car rental firms. The whole air transportation process provides a great resource of advertising channels and opportunities. The following list presents a fraction of possible advertising and marketing spots which are not exclusively on-board the aircraft:

- Menu card;
- Window application;
- In-flight announcements;
- Headrest covers;
- Product sampling;
- Air sickness bag;
- In-flight magazine;
- Exterior aircraft branding;
- Meal tray table back;
- Overhead bins;

- Cabin crew uniforms;
- Boarding pass;
- Promotional leaflets;
- Happy Snack Bags;
- Drink Tumblers;
- Napkins;
- On-board testing (e.g. wine, perfumes);
- In-flight entertainment (IFE) systems.

II. PSYCHOLOGY OF ON-BOARD ADVERTISING

In general, passengers perceiving advertisements on-board an aircraft are considered to be a good target for marketers. The reason is simple. The aircraft cabin is a closed space restricting the movement of passengers. Remaining seated for several hours at one particular place a passenger would start reading an in-flight magazine, pay attention to headrest covers and read all the booklets. Another reason that marketers are interested in the passenger is free time that he/she has to perceive an advertising message.

Psychological condition of passenger plays an important role. As passenger does not assume huge advertising campaigns as it is seen on TV he will perceive messages from in-flight magazine more quietly. Moreover, there is an opportunity of repeated contact with the advertising message. Having looked through all products offered the passenger may turn once again to the booklet or journal which was liked. According to statistics, 20 % of air passengers take a magazine with them after the flight [16]. Although in-flight magazines can be found in each aircraft (no matter of what kind of airline an aircraft belongs to) the concept of magazine as well as whole on-board marketing and advertising needs to be different for each airline, depending on its business model and taking into account requirements and needs of its passengers, considering that the passengers' needs and requirements can be specific regarding particular city pairs.

It is also necessary to point out that current marketing and advertising opportunities will multiply once the passenger is connected to Internet or is allowed to use his/her Blackberries or phones during the flight.

III. DIFFERENT BUSINESS MODELS

A. Traditional Airlines

The operational model of most traditional airlines is based on hub-and-spoke operations. Their networks usually consist of short-haul and long-haul routes. The flight schedule is set in order to offer passengers the most attractive times and to ensure the connectivity between short-haul and long-haul flights. In other words, during the day there is a time when short-haul flights feed the long-haul flights and there is also a time when long-haul flights feed short-haul connecting flights. Thanks to this feature, the operation of traditional airlines at major airports form typical arrival and departure peaks.

Traditional airlines form alliances (e.g. Star Alliance, One World, and Sky Team). Thanks to membership in alliances, the traditional air carriers are able to take passengers seamlessly from anywhere to everywhere.

Nowadays, the traditional airlines have a monopoly in the long-haul market. Long-haul operation generates an indispensable number of passengers for connecting short-haul flights.

Advertising in the cabin of long-haul aircraft can have disadvantages due to several reasons. Although long-haul aircraft have huge seat capacity, these are able to carry only limited passengers per day. Due to long legs, the aircraft can be operated only on one or two flights per day. It means that only a limited number of people will see the advertisement. Moreover, taking into account a flight lasting 10 – 15 hours the cabin full of posters could be disturbing and could reduce the level of passenger's comfort.

On the other hand, the cabin of long-haul aircraft could be a great place for on-board marketing (e.g. tasting of wine). Wide-body aircraft are also usually equipped with excellent in-flight entertainment systems. Once the in-flight connectivity is introduced, these systems will become a virtual store of unlimited possibilities.

Considering the short-haul flights operated by traditional airlines, these provide more space for on-board advertising. The requirements posed on the level of comfort in the cabins of short-haul aircraft are much less demanding compared to long-haul aircraft. Although daily utilizations of these aircraft are significantly lower, they are able to fly about 6 legs per day. As a result of this, more passengers could be carried during each day.

B. Charter airlines

The charter airline business model is based on selling whole seat capacity of the aircraft to tour operators. Once the contract between airline and tour operator is signed, marketing and consequently the load factor is tour operator's responsibility.

The charter airline operation and business model is currently the most efficient in the market as all flights are profitable for the air carrier. The tour operator bears the risk of financial loss.

Charter airlines usually have very high fleet utilisation. The fleet utilisation can reach as much as 17 hours per day. These carriers operate short- to medium- haul point-to-point flights. Their flights are not scheduled. Time and destination of each flight depends on the requirements of the tour operator.

These airlines usually operate single aisle aircraft, with high density seating configuration. Depending on route lengths, the aircraft can fly 6 to 8 legs per day. Moreover, most of their passengers are holidaymakers. Charter airlines are therefore a huge and very specific market that is considerably easy to address by properly selected advertisement.

C. Low-fare airlines

The low-fare business model is based on reducing operating costs with a view to offer passengers a very competitive fare. The air ticket price usually include only air transportation and all additional services like airport check-in, priority boarding and on-board services are charged extra.

Low-fare airlines operate scheduled, short-haul, point-to-point passenger services serving usually regional or secondary airports. In order to reduce operating costs, these carriers operate single type fleet and cabins of their aircraft have very high density seating configuration. Average fleet utilisation usually reaches more than 12 hours per day and each aircraft flies about 8 legs per day.

As these airlines usually offer the lowest possible fare, they are able to achieve average load factor of more than 80 %. Thanks to these features, the aircraft operated by low-fare airline can carry more than 1,200 passengers per day. It means that low-fare market has very good potential regarding on-board marketing and advertising as it hits relatively high number of potential customers.

It is also possible that growing revenues from advertising will sooner or later lead some low-fare carriers to introducing no-fare business model. Advertising and additional services will become primary source of profit for such airlines and air transportation will be provided to passengers for free. Of course, the transformation from low-fare to no-fare business model will be feasible only for the biggest in the market (e.g. Ryanair). The advertising revenues directly relate to fleet size, aircraft utilisation and load factors. Taking into account Ryanair’s performance in recent years, its aircraft are very attractive and very efficient advertising channels.

For our further analysis, we have selected low-fare business and operational model as it seems that high aircraft utilisation and high load factors make low-fare airlines be a good marketplace for on-board advertising and its further development. Moreover, thanks to low operating costs these airlines are very close to deploy no-fare business model.

IV. ANALYSIS OF LOW-FARE AIRLINES ANCILLARY REVENUES

Revenues from non-ticket sources (ancillary revenues) are of vital importance for many airlines worldwide, especially for those running low-fare business models. These revenues are generated mostly by the services that passengers are to buy before or during their travel experience. Legacy airlines bundle these services into the price of air ticket while low-fare airlines charge them extra. TABLE I shows potential sources of non-ticket revenues for airlines.

TABLE I. : POTENTIAL SOURCE OF ANCILLARY REVENUES FOR AIRLINES

Ancillary Revenues	On-board sales of food and beverages
	Baggage check-in charges
	Excess baggage charges
	Seat assignments charges
	Fee charged for purchases made with credit cards
	Commissions from the sales performed via airline website (e.g. hotels, car rentals, transportation from/to airport)

	Commissions from the sale of travel insurance
	Commissions from the sale of airport lounge access
	On-board advertising

Some legacy carriers use other sources of non-ticket revenues. For example: miles or points sold to banks issuing co-branded credit cards, travel partners such as hotel chains and car rental companies and other partners such as online malls, retailers and communication services. These services refer to frequent flyer ancillary revenues and we will exclude them from our analysis as we are focusing on low-fare airlines’ business model.

To introduce no-fare operational model, low-fare airlines will rely strongly on ancillary revenues and on-board advertising revenues. Even now, many airlines try to attract their passengers to buy as many non-ticket services as possible. Aircraft cabins full of on-board advertisements can be widely seen as well.

This is happening because non-ticket services and on-board advertising revenues are becoming of high value for low-fare airlines. Considering major European low-fare airline Ryanair, its non-ticket services and on-board advertising generated as much as € 8.5 per passenger in 2007 [2]. Other examples of specific ancillary revenue amounts including on-board advertising revenues can be seen on the following figures. Fig. 1 shows ancillary revenue amounts per passenger for three well-known European low-fare airlines in 2006 and 2007. Fig. 2 shows ancillary revenues per aircraft per month for the same carriers.

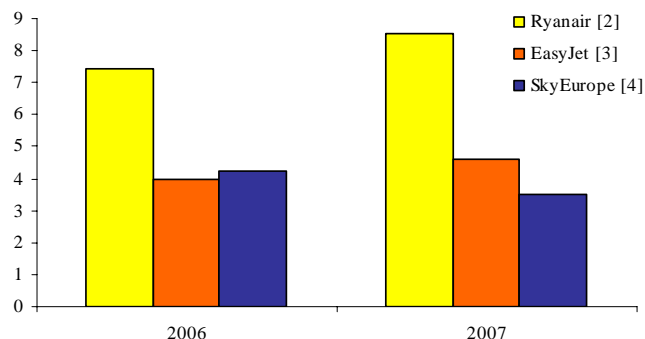


Figure 1. Per passenger ancillary revenues (€)^a

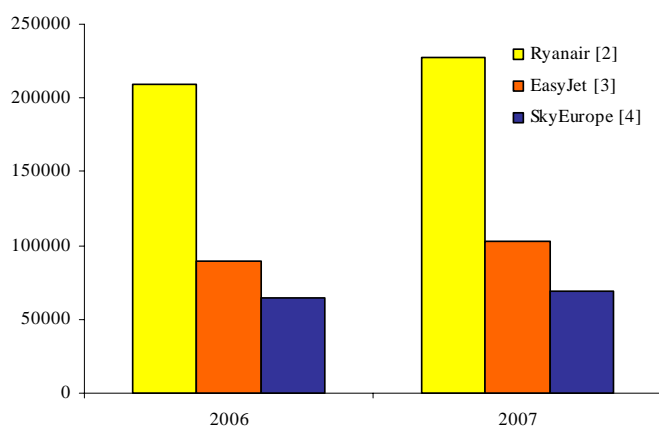


Figure 2. Ancillary revenues per aircraft per month (€)^a

a. Aircraft types and numbers used for the analysis:

<i>Ryanair:</i>	2006	120 B 737-800 (189 seats)
	2007:	133 B 737-800 (189 seats)
<i>EasyJet:</i>	2006:	87 A 319 (156 seats) 32 B 737 - 700 (149 seats) 3 B 737 - 300 (149 seats)
	2007:	107 A 319 (156 seats) 30 B 737 - 700 (149 seats)
<i>SkyEurope:</i>	2006:	14 B 737-700 (149 seats)
	2007:	14 B 737-700 (149 seats)

Ryanair and other carriers offer many of their flights for less than € 10 (when booked in advance). It is interesting that (according to the Fig. 1) ancillary revenues per passenger collected by Ryanair in 2007 are almost the same (€ 8.52) [2]. Average annual growth (2005-2007) of ancillary revenues (15% for Ryanair and more than 20% for EasyJet) shows potential of further increase in number of air tickets given to passengers for free. Although the annual growth in the case of SkyEurope was not so significant, results of our consultations with SkyEurope representatives indicates that this trend will change in near future and SkyEurope will follow its competitors.

V. ON-BOARD ADVERTISING REVENUES AS A SIGNIFICANT SOURCE OF AIRLINES' PROFIT

Some airlines provide exact figures regarding on-board advertising revenues in their annual reports. For example Air Berlin generated revenue of €1,500,000 (€ 0.11 per passenger) from advertising in its in-flight magazine in 2005. This revenue was slightly less than revenues from trip-insurance (€ 0.12 per passenger) and pre-assigned seats (€ 0.12 per passenger) [5].

EasyJet quantified its on-board advertising revenues up to € 55 million last year [3]. The revenue of € 1.43 per passenger from on-board advertising makes this airline 13 times more successful than Air Berlin. These figures give EasyJet a potential for further development and reflect the fact that airline's high revenues from on-board advertising and non-

ticket services are not a utopia. However, there is still the question if Ryanair and other low-fare carriers will be able to introduce no-fare business model in near future.

It is clear that Air Berlin, EasyJet and Ryanair already offer free fares to its customers. One quarter of Ryanair's passengers travel "for free". It is marketing policy that says: "We are offering you the lowest possible fare and you don't need to look for better value for your money."

But Ryanair's chief executive officer Michael O'Leary goes even further and promises making all the flights free by 2010. He might be slightly optimistic because his company generated the biggest ancillary revenues per passenger in the market (as for 2007 [2]) although it currently uses only four on-board advertising channels. As shown in TABLE III, there are some other channels (including prices) that are not used by Ryanair but remain interesting for other carriers.

VI. OPERATING COSTS PER AIRCRAFT PER MONTH

For further analysis we had to estimate monthly operating costs of aircraft operated by low-cost carrier. As our further analysis is focused mainly on SkyEurope's and Ryanair's business models, our estimation of operating costs is based on SkyEurope's and Ryanair's annual costs breakdowns and considers both B737-700 and B737-800 aircraft.

In our calculation, we assume that proportion of particular cost items in single year time horizon (as published in airlines' annual reports) is same as the proportion of particular cost items per block hour. If we know the proportion of particular cost items and value of at least one cost item, we are able to calculate other cost items (see TABLE II). In our calculation, the fuel costs were used as a baseline, as we were able to calculate these for each aircraft type using Eurocontrol BADA's Aircraft Performance Summary Tables [10]. Needless to point out, that both airlines Ryanair and SkyEurope use single type fleets.

TABLE II. ANNUAL COSTS PROPORTIONS [IN THOUSANDS OF €]

Operating costs	Ryanair [2]		SkyEurope [4]	
	Costs	Proportion	Costs	Proportion
Aircraft fuel	693331	39.28%	57892	22.60%
Sales and marketing	23795	1.35%	13231	5.16%
Maintenance, material and repairs	42046	2.38%	19405	7.57%
Staff costs	226580	12.84%	25832	10.08%
Navigation charges	199240	11.29%	23522	9.18%
Airport and handling charges	273613	15.50%	70477	27.51%
Depreciation and amortisation	104859	5.94%	15304	5.97%
Aircraft rental	143503	8.13%	1488	0.58%
Other	58183	3.30%	29029	11.33%
Total	1765150	100.00%	256180	100.00%

TABLE III. : ON-BOARD ADVERTISING REVENUES ACCORDING TO CHANNELS USED BY DIFFERENT LOW-FARE AIRLINES IN 2007 (€)

Channel	RYR [6] ^b	GWI [7] ^b	WZZ [8] ^b	ESK [9] ^b	Generic Airline
Menu card				360	360
Window application				1,500	1,500
In-flight announcements		29,900			29,900
Headrest covers		29,900	7,200		29,900
Product sampling		4,000		570	4,000
Air sickness bag				2,750	2,750
In-flight magazine	Ink Publishing ^c	UHURA ^c	Ink Publishing ^c	Ink Publishing ^c	n/a
Exterior aircraft branding	16,700	12,500		15,000	16,700
Meal tray table back	13,700	12,600	5,000	3,500	13,700
Overhead bins	19,500		8,000	3,500	19,500
Cabin crew uniforms			880	350	880
Boarding pass		24,900			24,900
Promotional leaflets		1,750		430	1,750
Happy Snack Bags		6,250			6,250
Drink Tumblers		6,250			6,250
Napkins		6,300			6,300
E-ticket banner ^d				8,000	8,000
Newsletter banner ^d		1,100		40,000	40,000
Booking engine banner ^d		280			280
Booking confirmation banner ^d		1,080			1,080
Potential advertising revenue per aircraft per month	49,900	136,810	21,080	75,960	191,300

b. Acronyms (according to ICAO): RYR – Ryanair; GWI – Germanwings; WZZ – Wizz Air; ESK – SkyEurope

c. Airlines usually use services of specialised publishers to produce their in-flight magazines. The very competitive business of these specialised publishers (e.g. Ink Publishing, UHURA) is based on production of in-flight magazines for airlines all over the world. Both production costs and profit of the publisher are covered by advertising in these magazines. Each airline has its own in-flight magazine which is adapted to its specific needs and requirements. Magazines are provided to airlines for free and usually do not generate air carriers any direct revenues. The main purpose of in-flight magazines from airlines' point of view is promotion of airlines' services and destinations. The articles in these magazines should motivate passengers to travel with a particular airline. Production of in-flight magazines is based on „Reason to fly“ policy.

The advertising revenues of in-flight magazines relates to number of passengers that can potentially read the magazines. Therefore the in-flight magazines are provided to airlines for free but airlines have to guarantee certain fleet size, aircraft utilisation and load factors. In the case when airlines have operational or financial problems leading to reduction of fleet size or significant decrease of load factor, the publishers can withdraw the contract.

d. These channels are not considered to be on-board advertising channels however they directly relate to flying passengers. Passengers receive their e-tickets and booking confirmations after the online booking reservation (using booking engine) of their flights have been made. Newsletter is also of wide passengers' attention. Electronic newsletter is distributed to more than 3 million e-mail addresses (case of SkyEurope).

According to last Ryanair's annual report [2], average daily utilisation of its aircraft in 2007 was 13.56 block hours. Assuming that an average month has 30 days, the average monthly utilisation of one aircraft is 406.8 block hours.

According to last SkyEurope's annual report [4], average daily utilisation of its aircraft in 2007 was 10.75 block hours. Assuming that an average month has 30 days, the average monthly utilisation of one aircraft is 322.5 block hours.

Using Eurocontrol BADA's Aircraft Performance Summary Tables [10], we have estimated an average fuel consumption of both aircraft types considered. According to our estimations, fuel consumption of B737-700 is 680 US gallons per block hour and fuel consumption of B737-800 is 690 US gallons per block hour. According to IATA Fuel Monitor website [11], the average fuel price in 2007 was € 1.56 per US gallon.

It means that monthly fuel costs per SkyEurope's B737-700 are € 342,108 and monthly fuel costs per Ryanair's B737-800 are € 437,880.

Further calculations are based on operating cost breakdown as stated in the latest Ryanair's [2] and SkyEurope's [4] annual reports. Other operation costs were calculated considering the ratio of fuel costs and particular operating cost items.

Following TABLE IV shows SkyEurope's and Ryanair's monthly operating costs per one aircraft.

TABLE IV. : CALCULATION OF MONTHLY OPERATING COSTS PER AIRCRAFT IN 2007 (€)

Operating costs	Ryanair [2]	SkyEurope [4]
Aircraft fuel	437,880	342,108
Sales and marketing	15,028	78,188
Maintenance, material and repairs	26,555	114,672
Staff costs	143,099	152,652
Navigation charges	125,832	139,001
Airport and handling charges	172,803	416,478
Depreciation and amortisation	90,631	8,793
Aircraft rental	36,746	171,544
Other	66,225	90,438
Total	1,114,797	1,513,875

According to figures from the latest annual reports [2] and [4], Ryanair carried 42,500,000 passengers and SkyEurope carried 3,312,443 passengers in 2007. Considering these figures, one SkyEurope's aircraft carried 19717 passengers in an average per month and one Ryanair's aircraft carried 26629 passengers in an average per month. It means that operating costs per passenger amount € 76.78 in case of SkyEurope and € 41.86 in case of Ryanair.

The following TABLE V. shows potential advertising revenues per passenger calculated according to number of passengers carried by particular airlines in 2007. Analysis in TABLE V. refers to maximum potential of on-board advertising revenues but does not reflect current figures of advertising revenues.

As shown in the table, there are big differences between potential advertising revenues per passenger as they are directly connected to number of passengers carried, fleet size and daily utilisation. Germanwings seems to be the most effective compared to its competitors however it carried almost 6 times fewer passengers than Ryanair. This results from the fact that the number of aircraft used by Ryanair is almost 5 times higher than number of aircraft used by Germanwings as well as from the fact that there is strong imbalance in number of on-board advertising channels used by these airlines.

SkyEurope has started to work on identifying on-board advertising opportunities 6 – 9 months ago. Currently it offers several advertising channels not only on board of its aircraft, but on company's website as well. It is anticipated that in near future the advertising will generate a significant portion of its revenues.

There are still some advertising channels that are not used by particular airlines but remain interesting for others. In order to achieve the maximum potential of on-board advertising; revenue airlines could preferably use all the available advertising channels. This is met in our model of Generic Airline that virtually uses all the above listed channels. We have defined Generic Airline with a view to analyse the maximum potential of all currently known advertising channels taking into account that these would be used by one airline. As channels are considered to be sold at their maximum market

prices (according to the analysed air carriers' price lists), the highest possible advertising revenues can be achieved.

Taking into account the available data, number of passengers per month per aircraft for the Generic Airline was calculated as an average of Ryanair's, SkyEurope's, Wizz Air's and Germanwing's figures.

TABLE V. : ESTIMATION OF POTENTIAL ADVERTISING REVENUES PER PASSENGER (€)

	RYR [2]	GWI [7]	WZZ [8]	ESK [4]	Generic Airline
Number of passengers in 2007	42,500000	7,090000	3,000000	3,312443	n/a
Fleet (number of aircraft) in 2007	133	27	15	14	n/a
Number of passengers per aircraft per month	26,629	21,883	16,667	19,717	21,224
Potential advertising revenue per aircraft per month	49,900	136,810	21,080	75,960	191,300
Potential advertising revenue per passenger	1.87	6.25	1.27	3.85	9.01

VII. CONCLUSIONS AND OBSERVATIONS

A. Generic airline potential

The aim of our analysis was to assess current situation in advertising revenues that are generated by airlines. We also put emphasis on potential advertising revenues estimation.

Based on our calculations (see TABLE II.), operating costs per passenger amount € 76.78 in case of SkyEurope and € 41.86 in case of Ryanair. TABLE IV shows that our Generic Airline (using all advertising channels sold for the maximum market prices) generates slightly more than € 9 potential revenue per passenger. This means that advertising revenues of Generic Airline could possibly cover up to 12 % of current SkyEurope's operating costs per passenger (as referred in TABLE II.) respectively up to 22 % of current Ryanair's operating costs per passenger (as referred in TABLE II.).

Although it seems that our Generic Airline does not have a potential to generate advertising revenues big enough to cover the total operating costs of typical low-fare airline, there are at least two more sources of airlines' revenues. These sources relate to on-board marketing and airlines qualify them as listing and marketing fees for goods sold on-board aircraft (in case of SkyEurope listing fees are € 2,000 per product per year and marketing fees € 5,000 per product per year). In case of SkyEurope, on-board marketing generates inconsiderable revenues. There are also some airlines selling their flight timetables for advertising purposes.

B. Risk analysis

From our point of view we have identified two main risks resulting from airline's business being dependant on revenues

from advertising. Once the model of no-fare airline is introduced airline may possibly face the rapidly changing demographic and social structure of its passengers. This can lead to the fact that price sensitive airline customers will no more be attractive market for advertisers.

There is one more risk that we have identified. Once the airline is dependent on revenues from on-board advertising the airline has to guarantee certain fleet utilisation as well as certain load factors. If there is a crisis in the air transport market (as the one after September 2001), the airline won't be able to fulfill the above mentioned conditions that are usually specified in the contract between carrier and advertisers. This can finally lead to loss of advertising revenues and consequently to worse airline's financial situation.

VIII. FUTURE OF ON-BOARD ADVERTISING

If 12 % of SkyEurope's operating costs per passenger respectively up to 22 % of Ryanair's operating costs per passenger can be covered by existing advertising channels imagine that airlines used more channels of advertising and revenues in near future?

One channel which could help airlines to cover the operating costs from the advertising revenues is the in-flight entertainment (IFE) system. IFE is considered to be the future channel for advertising. IFE with personal LCD screens at every seat has always been a domain of long-haul wide-body aircraft. The passengers on single-aisle aircraft had to be satisfied with drop-down LCD screens or ceiling mounted CRT (Cathode Ray Tube) screens. However, the development of IFE systems goes further and some airlines, such as West Jet and Delta Airlines have already equipped their narrow body aircraft with personal screens on every seat. It is anticipated that development of new technologies like Electronic Paper Display (very light, very thin, flexible paper-like display with ultra-low power consumption) will make personal IFE systems very attractive for single-aisle aircraft [12].

Finally, it can be admitted that transformation of low-fare airlines to the no-fare business model is possible; partly thanks to on-board advertising revenues as well as increased prices of non-ticket services (including e.g. checked baggage fees, check-in fees and priority boarding fees). Such kind of transformation can only be expected in case of having a large fleet airline with sufficient aircraft utilisation and load factor. On the other hand, we are aware that utilisation of all the available ancillary revenues sources would generate significant investment costs as well as would lead to an increase of the operating costs. However, we anticipate that investments in future advertising channels will bring airlines continuous income.

Nevertheless, the introduction of the no-fare business model can lead to a significant change in airlines passenger demographic mix, which can potentially have a very negative impact on airlines attractiveness for the advertising market.

There are certainly many different scenarios how the LCC market strategy will evolve. We don't pretend that on-board advertisement will turn round the market and bring free flying. However, the income from on-board advertisement can be a

contribution to further reduction of cost of travel and can expand the air travel market to new potential passenger who would not fly at all.

As already mentioned till this time the airline companies focused on additional financial sources closely related to the air travel (insurance, car rental, hotel accommodation etc.). The reason why the income from adverts was not used in larger extent yet is that it is completely different business and airlines don't have necessary know-how. With respect to this most airlines can sell the on-board advert capacity to specialised companies. However, to be able to maximise their profit they must be able to specify the advertising potential. Our paper could contribute to this.

IX. FUTURE WORK

Our initial research revealed both, the great potential of on-board advertising and several issues that will be addressed in our future research.

To be able to proceed further we will develop economical model that should allow us to perform demand sensitivity analysis as well as detailed what-if analysis. Our current work is focussing on various simulation scenarios that define various levels of airline's business dependency on on-board advertising revenues. The main aim of our research is to find ideal equilibrium between on-board advertising revenues and other sources of income considering various airline business models.

REFERENCES

- [1] <http://www.tfl.gov.uk/>
- [2] Ryanair Holdings plc Annual Report & Financial Statement 2006, 2007
- [3] EasyJet plc Annual reports and account 2006, 2007
- [4] SkyEurope Holding AG Annual Report 2006, 2007
- [5] Europe's Top 4 Low Cost Carriers Generated 470 Million Euros issued by Idea Works (page 7) October 10, 2006
- [6] Ryanair Rate Card 2008 available at <http://www.ryanairmag.com>
- [7] Germanwings/advertising media 2008 available at www.germanwings.com
- [8] Wizzair on-board advertising/media offer, September 2007 available at www.wizzair.com
- [9] SEA ON-BOARD ADVERTISING winter 2007-2008, presentation
- [10] Aircraft performance summary tables for the base of Aircraft data (BADA), rev. 3.6, issued by Eurocontrol Experimental Centre (pages A 25 – A 26), September 2004
- [11] IATA, Jet Fuel Price Monitor website
- [12] <http://www.eink.com/technology/>
- [13] http://www.iata.org/whatwedo/economics/fuel_monitor/index.htm
- [14] European Central Bank website, average annual exchange rate, 2007
- [15] ICAO Base-line Aircraft Operating Costs, Summer 2000
- [16] http://www.spblp.ru/publicity_eng.shtml
- [17] <http://www.Viacom-Outdoor-wins-London-Underground-and-Victoria-Coach-Station-advertising-contracts.htm>
- [18] Worldwide In-Flight Advertising Revenues, 2004-2006, emarketer, Publication Date: November 15, 2005, Subjects: Advertising Revenues; Airlines, Geographies: Global
- [19] Worldwide In-Flight Advertising Spending, by Region, 2004 & 2006 (% market share), emarketer, Publication Date: November 15, 2005, Subjects: Airlines; Advertising Spending, Geographies: Asia-Pacific; Europe; North America

- [20] Worldwide In-Flight Advertising Spending, by Region, 2006, emarketer
Publication Date: November 15, 2005, Subjects: Airlines; Advertising
Spending, Geographies: Africa; Asia-Pacific; Europe; Middle East;
North America; Latin America
- [21] A radical fix for airlines: Make flying free, Business 2.0 Magazine,
March 31, 2006
- [22] Frequent fliers reach new heights with more than 15 million reward trips
during 2005, Idea Works (Page 7), May 22, 2006
- [23] Ofcom consults on Mobile Communications onboard Aircraft, Thursday
18 October 2007 12:10:37 by John Hunt
- [24] Low-cost carriers now try out new revenue options
11 Apr, 2007, 0240 hrs IST, Vishakha Talreja, TNN
- [25] Inflight advertising to hit US\$1 billion, Ink Publishing reports, 15th
November 2005
- [26] <http://www.interairmedia.com/about.html>
- [27] Ryanair facing possible action over breaches of advertising rules –
watchdog, Thomson Financial News, 04.09.08
- [28] <http://www.mediabuyerplanner.com/2007/08/28/advertising-gets-high-with-ryanair-ads/>, Advertising Gets High with Ryanair Ads
- [29] The Arbitron In-flight media study, exploring frequent flyer engagement
with airline magazines and in-flight TV, 2006 Arbitron Inc., 06-CUS-
113 6/06
- [30] Honomichl, J. J., *Honomichl on Marketing Research*, Lincolnwood, IL:
NTC Business Books, 1986.
- [31] Krugman, H., Memory Without Recall, Exposure Without Perception.
Journal of Advertising Research, July/August, 1977.
- [32] Young, Charles E., *The Advertising Research Handbook*, Ideas in Flight,
Seattle, WA, April 2005, ISBN 0-9765574-0-1
- [33] <http://www.imm-international.com/?gclid=CN2Dm4zozZICFQk-ZwodDm32CQ>
- [34] International service variants: airline passenger expectations and
perceptions of service quality, Author(s): Fareena Sultan, Merlin C.
Simpson, Jr, *Journal of Services Marketing*, Year: 2000, Volume: 14,
Issue: 3, Page: 188 – 216, DOI: 10.1108/08876040010327211,
Publisher: MCB UP Ltd
- [35] STRATEGIC MARKETING OPTIONS IN THE U.S. AIRLINE
INDUSTRY, Author(s): T.K. Das, William D. Reisel, *International
Journal of Commerce and Management*, Year: 1997, Volume: 7, Issue:
2, Page: 84 – 98, DOI: 10.1108/eb047350, Publisher: MCB UP Ltd
- [36] Developments in airline marketing practice, Author(s): John C. Driver,
Journal of Marketing Practice: Applied Marketing Science,
Year: 1999, Volume: 5, Issue: 5, Page: 134 – 150, DOI:
10.1108/EUM00000000004571, Publisher: MCB UP Ltd

Pricing schemes based on air navigation service charges to reduce en-route ATFM delays.

Preliminary results.

Andrea Ranieri, Lorenzo Castelli

Dipartimento di Elettrotecnica, Elettronica e Informatica

Università degli Studi di Trieste

Via A. Valerio 10, 34127 Trieste, Italy

{aranieri, castelli}@units.it

Abstract—In this paper we analyze an incentive scheme based on air navigation service charges modulation that could help in reducing ATFM delays, inducing users to make a better and more uniform use of capacity, especially in those situations in which the distribution of traffic is known to be non homogeneous. A first experiment indicates the feasibility of implementing such a system and further investigations and refinements of the model are going to be performed in the next future.

Keywords: En-route charges, Air navigation service charges; Incentive schemes; ATFM delays.

I. INTRODUCTION

Air transport is a key element for the world's prosperous evolution and represents a major pillar of its economy, as it sustains local national economies, global trade and tourism. In 2005 European Commission stated that "Air transport contributes €220 billion to European gross domestic product and employs 3.1 million people. It is also an important aspect of European cohesion ensuring the rapid and efficient movement of people and goods, but also providing essential access to remote region"[1].

In 2006 the EUROCONTROL Central Flow Management Unit (CFMU) recorded 9.6 million IFR movements in the European skies, which represent an increase of 4.1% with respect to the previous year and this trend is expected to continue in next years, reaching a total number between 15.5 and 18.9 million IFR flight movements in the EUROCONTROL Statistical Reference Area (ESRA) by 2025 [2].

Unfortunately these growth figures are coupled with a worsening in the performances of the whole system. During 2006 overall punctuality deteriorated for the third consecutive year (21.4% of flights arrived late in 2006, while they were 17.2% in 2003).

Part of these delays is due to ATFM measures. In fact when traffic demand is anticipated to exceed the available capacity in en-route control centers or at an airport, ATC units may call for "ATFM regulations" and all the aircrafts subject to ATFM

regulations are held at the departure airport according to "ATFM slots" allocated by the CFMU. The ATFM delay of a given flight is consequence of the most constraining regulation applied either at one airport or over an en-route sector along its route. This imbalance between traffic demand and available capacity and resulting ATFM delays can have various ATM-related (staffing, etc.) and non-ATM related reasons (weather, airport scheduling, etc.).

Total number of minutes of delay caused by en-route ATFM measures has increased by 15% in 2006 and represents the majority of ATFM delays (56%). The average en-route ATFM delay increased from 1.3 minutes to 1.4 minutes per flight in 2006, meaning that the goal of 1 min/flight is still far to be achieved [3].

ATM has currently no mandate and limited scope to help improve punctuality and predictability, hence the opportunity of added value from ATM in air transport is currently underexploited.

In this paper we propose some possible incentive schemes based on en-route charges modulation that could help in reducing ATFM delays, inducing users to make a better and more uniform use of capacity, especially in those situations in which the distribution of traffic is known to be non homogeneous. We maintain consistency with the current charging regulations and do not propose new rules conversely to what proposed by [4]. Congestion pricing is widely viewed by economists as the most efficient mean to alleviate traffic congestion and to influence travelers' choice of route and travel mode. The traditional congestion prices policies are mainly based on a congestion toll imposed on each user which equals the external cost of congestion caused by the user on the system [5], [6], [7]. In economical terms the congestion toll objective is to make the user internalize the external cost caused by its own exploitation of the resource, which can be estimated for instance through the use of empirical data or with queuing theory [8].

Our approach differs from the traditional one, since we try to make the local objective of a selfish profit maximizing user coincide with the global objective of achieving the minimum total delay at system level. To achieve this goal we propose to

provide to those users who contribute to improve the global efficiency (by modifying their demand) an incentive which offsets the increase of their cost.

A first preliminary simulation run on an actual operating environment has shown the potential offered by this system in consistently reducing global delay, with minor changes in flight paths. Further investigations and validation of the model are currently underway.

The remainder of this work is organized as follows. Section 2 gives a brief description of the European charging system. Section 3 describes the principles of re-routing aircraft to enhance capacity, while in Section 4 we propose two possible scenarios of implementation of an incentive scheme. Section 5 summarizes the conclusions.

II. THE EUROPEAN CHARGING SYSTEM

Current ATC service in Europe finances its operations by adequately charging airspace users in accordance with EC Regulation 1794/2006, which came into force on the 1st January 2007, laying down a common charging scheme for air navigation services ('the Regulation') [9]. Air navigation service charges are composed of en-route and terminal charges due to the provision of air navigation services for the en-route and terminal segments of the flight, respectively. In particular, the en-route charge r for a specific flight in a specific en-route charging zone is equal to $r = d \cdot p_r \cdot t$, where d is the distance factor, p_r is the en-route weight factor of the aircraft and t is the en-route unit rate of the en-route charging zone. A "charging zone" is defined as a volume of airspace for which a single cost base and a single unit rate are established. So far charging zones coincide with national boundaries, each under the control of an ANSP establishing its own cost base and unit rate. Nevertheless the Regulation states that a charging zone can be set regardless of national boundaries. The product of the distance and weight factors is referred to as en-route Service Unit (SU), i.e. $SU = d \cdot p_r$. These parameters are calculated as follows:

- d is the distance factor and is obtained by dividing by one hundred the number of kilometers flown in the great circle distance between entry and exit point of the en-route charging zone, according to the latest known flight plan filed by the aircraft operator. The distance to be taken into account is to be reduced by twenty kilometers for each take-off and each landing on the territory of a Member State.
- $p_r = \sqrt{MTOW/50}$, i.e., the en-route weight factor is equal to the square root of the quotient obtained by dividing by fifty the number of metric tons in the maximum certificated take-off weight (MTOW) of the aircraft.
- t is the unit rate for the en-route charging zone.

In all Contracting States except United Kingdom (which adopts a price cap mechanism), the unit rate is based on the Full Cost Recovery (FCR) principle stating that all en-route costs for ANS regulatory and supervisory functions are fully

recovered through en-route charges. When the full cost recovery principle applies, the en-route unit rate is calculated by dividing the forecasted chargeable costs for providing en-route air navigation services by the forecast number of chargeable en-route service units for the relevant year. The balance resulting from under or over recovery of previous years is included in the forecast costs. The amount R of en-route charges due for a flight through states $1, \dots, n$ is equal to

$$R = \sum_{i=1}^n r_i \quad (1)$$

where r_i is the amount due to the i -th ANSP for the en-route air navigation services provided, i.e.,

$$r_i = d_i \times \sqrt{\frac{MTOW}{50}} \times t_i = SU_i \times t_i \quad (2)$$

Unit rates are set for each charging zone on an annual basis and can be modified during the course of the year only if unexpected major changes in traffic or costs occur (Article 13 of the Regulation).

Besides the new definition of charging zone, independent from national boundaries, another major innovation introduced by the Regulation is the first opening to incentive schemes based on en-route charges:

"Member States may establish or approve incentive schemes consisting of financial advantages or disadvantages applied on a non-discriminatory and transparent [...] resulting in a different calculation of charges [...] When a Member State decides to apply an incentive scheme [...] in respect of users of air navigation services it shall, [...], modulate charges incurred by them in order to reflect efforts made by these users to optimize the use of air navigation services, to reduce the overall costs of these services and to increase their efficiency, in particular by decreasing charges according to airborne equipment that increases capacity or to offsetting the inconvenience of choosing less congested routings." [9]

This new distinctive feature provides ANSPs with an operational instrument to exert some demand management which could help in dealing with the congestion problems systematically faced by them.

III. RE-ROUTING TO REDUCE AIRSPACE DELAYS

From an airline perspective en-route charges constitute a direct operating cost associated with the execution of the flight. In fact the direct operating costs of an airline can be divided into flight operations, maintenance and depreciation, where "flight operations" category generally comprehends three main sources of cost [10]:

- flight crew;
- fuel and oil;
- airport, terminal and en-route charges.

The delay also constitutes a cost for Aircraft Operators (AOs) and its unit value can widely vary depending on the type

of aircraft, the type of delay and the total amount of delay experienced by a particular flight. The cost of delay has been deeply investigated by [11]. According to the direct operating cost-breakdown we can assume that an airline that decides to re-route a particular flight performing a longer route than the originally planned one, will experience in general higher fuel costs, while route-charges costs may increase, decrease or remain constant depending on the new entry and exit points of the charging zones overflowed. The flight crew cost, expressed on a time rather than on a distance basis, can be included in the cost of delay, according to the methodology adopted in [11].

Thus a flight delayed by an ATFM regulation will experience a delay cost that may be higher than the cost of re-routing depending on the entity of the delay and the deviation of the re-routed path from the original one. On the other hand, the non homogeneous distribution of traffic in Europe can create as a consequence that different sectors of the same ANSP experience opposite congestion problems: one regulated and the contiguous one with spare capacity, thus implying that a limited rerouting action could make a flight avoiding the regulated sector and respecting its original time schedule. But which is the amount of delay that triggers the convenience point for an AO to re-route and what are the consequences that this re-routed flight generates on the system? We want to obtain some numerical figures that answer to these questions.

The idea is to identify a situation where 2 contiguous traffic volumes (TVs) have different accommodation capabilities: the first is affected by a regulation, meaning that the demand to flight over it exceeds its capacity, and the second without any active regulation, meaning that there is still spare capacity offered to users. We identified this particular situation in the 2 traffic volumes (LIPPNUX and LIPPSUX) under the control of Padova ACC, in Italy (See fig.1). In the period from 6 July to 2 August 2006 (corresponding to AIRAC¹ 284), the traffic volume of LIPPNUX was often affected by a regulation for ATC capacity reasons (see Fig. 2), limiting the maximum number of flights to 42-47 per hour, according to the different regulations. During the same period no regulations were registered on the contiguous traffic volume LIPPSUX. As one can see, during this particular period the call of a regulation on LIPPNUX is exercised almost on a daily basis, while for LIPPSUX there is no regulation in place. This means that the excess of demand over capacity systematically verifies, so this situation could be predicted with reasonable accuracy in the pre-tactical phase.

We reduced the scope of our analysis to a single day, 16 July 2006, when a regulation over LIPPNUX was capping the maximum number of flights to 44 per hour from 7.20AM to 14.30PM. With COOSAC² tool we identified that during the activation of this regulation 321 flights passed through LIPPNUX, 157 of them suffered a delay and the regulation

active on LIPPNUX was the most penalizing one for 100 among them.

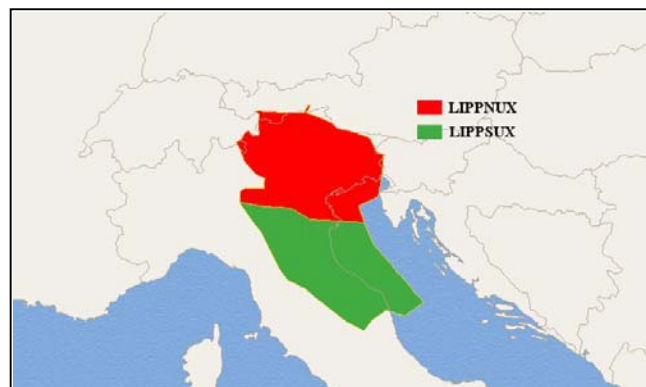


Figure 1. LIPPNUX and LIPPSUX traffic volumes.

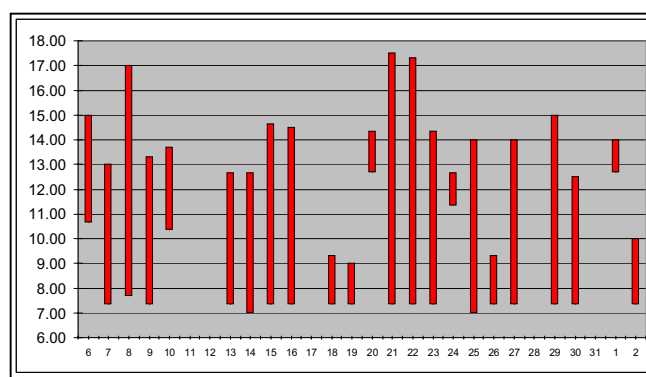


Figure 2. Regulations activated on LIPPNUX during the reference period.

We analyzed the FPLs corresponding to these 100 flights, as recorded by CFMU, and we identified that the option of re-routing through LIPPSUX, in order to avoid LIPPNUX, actually existed for just 2 among them, although this option was not chosen. Hereafter we will refer these 2 flights, respectively as AZA and SWR, according to the code of their operator. For all the other flights in the set, the re-routing option was either unfeasible or leading to an excessive extension of the trajectory, so we considered them as fixed.

Table I reports the direct operating costs associated with both AZA and SWR flights, where the cost of delay has been considered equal to 1 €/min according to [10] and the cost of jet fuel was fixed at 0.53 €/l according to the figures provided by IATA [12]. For each flight the fuel consumption was calculated according to the nominal performances reported by BADA model for the different types of aircraft [13]. Table I shows that both AZA and SWR chose the cheapest option in the original scenario, even if the cost of delay was included in it. Their delay was in fact just 17 and 12 minutes respectively, meaning that the cost associated with it was relatively small and the switch to a conditional route would have represented a less convenient option. Nevertheless the figures in Table II show that benefits accruing from a re-routing are larger if considered from a global traffic volume perspective than for the single flight.

¹ AIRAC (Aeronautical Information Regulation And Control) defines a series of common dates and an associated standard aeronautical information publication procedure for States. Each cycle lasts 28 days, always starting on Thursday.

² COSAAC is a tool initially developed at CENA, used to assess ATFM concepts such as flexible use of airspace and ATFM delays.

TABLE I. DIRECT OPERATING COSTS OF THE SELECTED FLIGHTS

Route	Callsign (first 3 chars)	Dist (NM)	En-route charge Cost	Fuel Cost	Delay Cost	Total Cost
original	AZA	458	281	1409	17	1707
	SWR	746	982	5054	12	6049
rerouted	AZA	467	305	1436	0	1740
	SWR	772	1016	5230	0	6246

All values in €

TABLE II. EFFECTS OF RE-ROUTING

	Actual values on LIPPNUX	Values on LIPPNUX with AZA re-routed	Values on LIPPNUX with SWR re-routed	Values on LIPPNUX with AZA and SWR re-routed
N° Flights	321	320	320	319
N° flights Delayed	157	147	147	141
Total sector delay (mins.)	1795	1683	1696	1619
Delay difference (mins.)	/	-112	-99	-176

Source: COSAAC

A total reduction of 112 minutes is expected on the whole LIPPNUX TV if AZA re-routes through LIPPSUX, 99 minutes if SWR re-routes and 176 if both re-route.

The function that links the delay of the sector to the delay of a single flight is nonlinear and this fact means that the penalties suffered locally by the airline for re-routing its flight are more than compensated globally by the enhanced conditions that it allows in the airspace.

In the case that a reduction in the en-route charge cost would be offered to them, the re-routing could eventually represent the most convenient option and induce them to release their en-route ATFM slots over LIPPNUX, thus leading to a more than proportional benefit for the remaining traffic.

In the following, we estimate the reduction in route charges that has to be offered to these particular users in order to make profitable to them to re-route their flights.

IV. TWO CONGESTION PRICING SCHEMES

A. Each traffic volume as a different charging zone

The first proposed incentive scheme is based on the creation of two different charging zones each corresponding to a different traffic volume. Hence each traffic volume has its own unit rate associated. According to the formula for en-route

charge calculation, dividing a charging zone in several smaller ones represents a general higher cost for users, since the distance factor would be calculated according to the great circle distance between entry and exit points of each charging zone rather than between the entry and exit points of the unique one (see Figure 3).

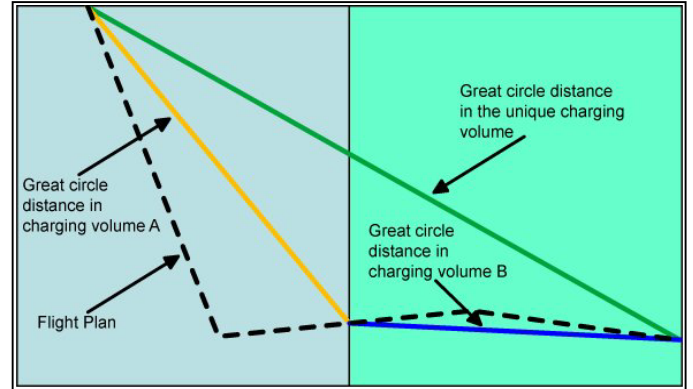


Figure 3. Distance factors under the different scenario

To determine the unit rate threshold values that make it convenient for each flight to re-route, we built the cost function associated to each route. Then we equalized the cost function relative to the original FPL to the one associated with the re-routed one and we obtained a new linear function linking the unit rate on LIPPNUX with the unit rate on LIPPSUX. By assuming that the direct cost of operating a flight is equal to the sum of delay costs, en-route charges and fuel, it follows that:

$$C_{dO} + C_{rO} + C_{fO} = C_{dR} + C_{rR} + C_{fR} \tag{3}$$

where d refers to as the cost of delay, r to as the cost due to en-route charges, f to as the cost of fuel. The subscript O refers to as the original FPL and R to the re-routed one. In accordance with Equations (1) and (2), and with little algebra, it follows that

$$t_{LIPPNUX} = a + b \cdot t_{LIPPSUX} \tag{4}$$

where

$$a = \frac{c_{dR} + \sum_{i \neq LIPPSUX} c_{rR}^i + c_{fR} - c_{dO} - \sum_{i \neq LIPPSUX} c_{rO}^i - c_{fO}}{SU_{LIPPNUX}}$$

and $b = \frac{SU_{LIPPSUX}}{SU_{LIPPNUX}}$.

For each flight we have thus determined the indifference curve, a straight line with slope b and intercept a , formed by that pairs of unit rate values, according to which the 2 options of re-routing and maintaining the original trajectory, are equally preferred by the user as they imply the same direct operating cost (see Fig.4).

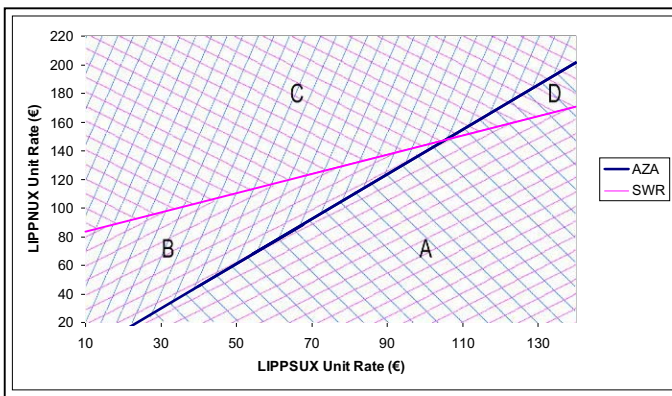


Figure 4. Unit rate values scenario A

For each flight the correspondent function divides the plane into 2 regions, one in which each pair of unit rate values make it cheaper to fly according to the original route and the other containing the pairs of values that make it convenient to fly the re-routed one. We thus defined a total of 4 different regions, identified as A, B, C and D in Figure 4. According to the region in which the pair of values is chosen, the flights will be induced to re-route or to maintain the original flights, in order to comply with capacity constraints also on the other sectors. For instance unit rate values in region A would not promote re-routing by any of the two flights, whereas points in region C would foster re-routing for both flights.

The points on one line correspond to those unit rate pair of values that equal the cost of performing the original route and the cost associated to the modified one.

For example, considering that the Italian unit rate for the sample period was fixed at 67.67 €, the increase of LIPPNUX unit rate to 90€ and the decrease of LIPPSUX unit rate to 40€ triggers the convenience point for AZA to re-route but not for SWR, whereas 155€ and 60€ values would drive both to re-route.

If we decide to maintain the unit rate of LIPPNUX equal to the reference national unit rate at 67.67 €, LIPPSUX unit rate value that makes it cheaper the re-route has to decrease until 54.45 € for the AZA, and to inadmissible negative value for SWR. Otherwise if we would charge a sort of congestion toll on LIPPNUX, leaving the original unit rate on LIPPSUX unchanged, we should rise LIPPNUX unit rate up to 88.50 € before AZA prefers to re-route and up to 122.50 € before both AZA and SWR prefers to re-route.

Even though the impact on total delay is straightforward, these results show that re-routing is achieved by means of really high en-route unit rate values, thus limiting the feasibility of such approach. Previous studies as [14] also claim that “Even if route charges were an important component of airline operating costs they can hardly be used as the sole means of demand management without being associated with any other incentive like punctuality”. In addition, with the implementation of this incentive scheme, the flights which maintain their planned FPL on LIPPNUX have to pay a higher amount of en-route charges due to the

changes in the unit rate values. This situation could be considered as unfair because users would generally pay more for receiving the same ATC services during the execution of their flight. On the other hand this surcharge could be considered as a charge on the externality cost they generate in exploiting a limited resource like capacity. In any case, it is still a challenging issue to mathematically model the impact of such new charges on the route choices of all the TV users.

B. A different unit rate per flight

Another way of implementing the incentive scheme would be to maintain the charging zones as they are defined in the actual situation (e.g. one single charging zone for Italian FIR) and proposing a discount in en-route charges amounts to all the fights that decide to re-route in order to avoid a congested sector (e.g. LIPPNUX). Anyway this would eventually lead to a less controllable behavior of the AOs since a flight that performs most of its flight on Italy would prefer to re-route in order to pay less not only in the contingency sector (i.e. LIPPSUX), but during all the execution of the flight within the national FIR. Under this scenario the link functions between unit rate values modify, and they are reported in Figure 5. With little algebra from Equation (3), for each flight the formula for calculating the unit rate under this incentive scheme B is the following:

$$t_{LI}^I = \frac{1}{SU_{LI}^I} \cdot \left(c_{do} + c_{ro} + c_{fo} - c_{dr} - \sum_{i \neq LI} c_{rR}^i - c_{fR} \right) \quad (5)$$

where LI refers to the Italian airspace under consideration.

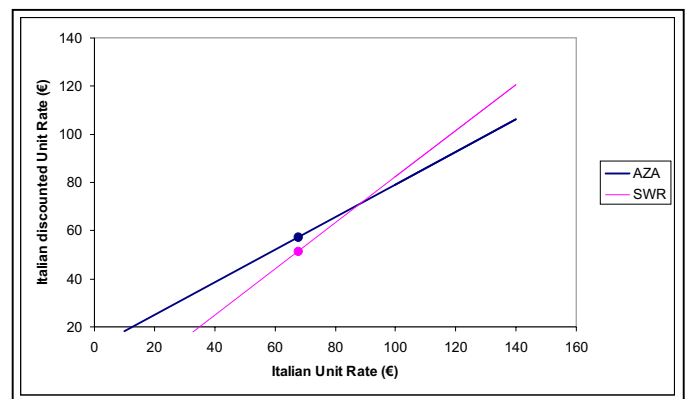


Figure 5. Unit rate values scenario B

Under this second implementation scenario, given the reference national unit rate value at 67.67€, this value should be discounted to 57.07 € and 51.20 € to induce respectively AZA and SWR to re-route in order to avoid the congested airspace. These values are slightly higher than the ones found under scenario A, because now the discount offered to the AOs would be applied along the whole route segment internal to Italy and not only on the LIPPSUX segment.

This scenario could be seen by users as fairer as it does not introduce penalization for flights which does not re-route, but just a benefit to the cooperating AOs, which decide to re-route

their flight to enhance the global situation. On the other hand the ANSP would experience a loss of income since the lower charge collected from the two flights would not be counterbalanced by a higher cost imposed on other users.

To ensure a balance between ANSP experienced cost and revenues, the Full Cost Recovery principle could provide a solution, just as in the present system. Any gap between the actual costs of ANSP and its income would be treated exactly as today when establishing the national unit rate, including it as an under-recovery in the formula, thus sharing the cost of the incentive among all users who can potentially experience the benefits of its implementation.

The ANSP would be responsible for the adoption of such an incentive scheme, establishing in advance its form of implementation and the related unit rate modulation, in order to allow the users to schedule in advance their flight according to the chosen scheme.

V. CONCLUSIONS

We have assessed 2 different potential implementations of an incentive scheme, whose objective is to drive users in making a better usage of capacity resource from a system perspective, by reconciling the single user objective (the minimum cost) with a global one (the minimum total delay). The very first results obtained from a simulation based on a real traffic sample indicate the potential of such a system, which can notably reduce the delay on a specific sector by rerouting a very few number of flights. We assume that the re-route action can not be imposed by a central authority but has to be induced through a modulation of en-route charges amount that each user has to pay for the air traffic services it receives.

En-route charges can be modulated to provide a demand management tool both from a legislative perspective, as this is permitted by the Regulation, and from an operative one, as indicated by the simulation results.

Implementation scenario A represents a finest tool to drive the demand according to the global objective of minimum delay, as the decomposition of a larger charging zone into smaller ones allow to better control the behaviour of single users and to closely bind the unit rate of a TV with its level of traffic. Nevertheless it would require some major changes with respect to the present charging system which, though compliant with the Regulation, could give rise to some objections from users about equity and cost effectiveness. On the other hand the implementation scenario B requires only minor changes to the present system but could eventually lead to less deterministic responses of users' behaviour as well as less cost reflectivity.

Other implementation scenarios and their impact on stakeholders deserve further study, especially the assessment of the consequences of the scheme on the Full Cost Recovery principle and the estimation of the non-linear function that links the presence of an aircraft with the total ATFM-delay of the sector.

ACKNOWLEDGMENT

The authors would like to thank Mr. Bernard Kerstenne from the EUROCONTROL Experimental Centre for its technical support during the simulations with COSAAC tool.

REFERENCES

- [1] European Commission, Communication to the Council on the project to develop the new generation European air traffic management system (SESAR) and the establishment of the SESAR Joint Undertaking, 25 November 2005.
- [2] EUROCONTROL Air Traffic Statistics and Forecast Service, "Long-Term Forecast: IFR Flight Movements 2006-2025", December 2006.
- [3] EUROCONTROL Performance Review Commission, "An Assessment of Air Traffic Management in Europe during the Calendar Year 2006" May 2007.
- [4] M. Raffarin, "Congestion on European Airspace. A pricing solution?" *Journal of Transport Economics and Policy*, vol. 38, pp. 109-126, January 2004.
- [5] W.S. Vickrey, "Pricing in urban and suburban transport", *American Economic Review*, vol. 53, pp. 452-465, May 1963.
- [6] R. Lindsey and E. Verhoef, "Traffic Congestion and Congestion Pricing", Tinbergen Institute Discussion Paper, November 2000.
- [7] P. Bergendorff, D.W. Hearn and M.V. Ramana, "Congestion toll pricing of traffic networks", in *Network Optimization, Lecture Notes in Economics and Mathematical Systems*, P.M. Pardalos et al., Ed. Springer-Verlag, Berlin, Germany, vol 450, pp.51-74, 1997.
- [8] A. Odoni, "Congestion Pricing for Airports and for En-route Airspace", in *New Concepts and Methods in Air Traffic Management*, L.Bianco et al., Ed. Springer-Verlag, Berlin, Germany, pp.31-44, 2001.
- [9] COMMISSION REGULATION (EC) No 1794/2006, laying down a common charging scheme for air navigation services, 6 December 2006.
- [10] R. Doganis, *Flying Off Course: The Economics of International Airlines*, Routledge 3rd edition, pp 90-91, 2002.
- [11] A. Cook, G. Tanner and S. Anderson, "Evaluating the True Cost to Airlines of One Minute of Airborne or Ground Delay: final report", Project Report. EUROCONTROL, Brussels, Belgium. May 2004.
- [12] IATA Jet Fuel Price Development (www.iata.org).
- [13] EUROCONTROL Experimental Centre, "User Manual for the Base of Aircraft Data (BADA)", EEC Note No. 10/04, July 2004.
- [14] L. Brunetta, L. Righi and G. Andreatta "Analysis of alternative routes for selected city pairs". Final Report of WorkPackage 6, part II, of the EUROCONTROL CARE Innovative project 'Innovative Route Charging Schemes', 2004.

Estimation of Aviation Infrastructure Condition from a Biased Sample

Gautam Gupta, Jasenka Rakas, Mark Hansen
Civil and Environmental Engineering Department
University of California at Berkeley
Berkeley, California, USA
ggupta@berkeley.edu

Abstract - In this paper, we address the problem of making inferences about a population of infrastructure facilities from a subset that is a biased sample. We consider the case in which the sample is biased towards facilities in worse condition or requiring more expensive repair. Two methods are developed that incorporate a model of the process through which the sample is selected. One of the methods is based on well-known truncated distributions, whereas the other assumes that the bias operates continuously. The methods are applied to a class of facilities under the FAA's jurisdiction known as "un-staffed facilities." These consist of structures housing radars, navigation aids, radio beacons, and other ground-based equipment, and no previous system-wide evaluation has been attempted for these facilities. We present and discuss the estimates obtained from both the methods, and examine their goodness-of-fit with the sample. Given the premise that bias exists, the continuous bias model proved more suitable. However, the continuous bias model did not surpass the truncation models in terms of goodness-of-fit.

I. INTRODUCTION

Infrastructure maintenance and repair decisions, along with supporting budgets, are based on data about facility condition, cost factors, and budgetary constraints. In some infrastructure systems, comprehensive condition surveys are done periodically. In other systems, condition surveys over the entire population of facilities are not done. Reasons for this can include excessive cost, accessibility constraints, and a reactive "fix it when it breaks" approach to infrastructure management. But even in such cases, there may be partial data (of a subset of the population) on condition, replacement or repair needs, and associated costs. With adequate knowledge about the procedure used to gather such data, reasonable extrapolations can be made about the entire population.

In this paper, we address such a problem of making inferences about a population of infrastructure facilities from data for a sample of them. The unique aspect of the problem is that the sample is biased in a particular fashion. Specifically, we consider the case in which the sample of facilities is biased toward facilities in worse condition or requiring more expensive repair (or even replacement). Such a bias may exist for a variety of reasons. For example, if the infrastructure manager is accustomed

to budgets that are insufficient to bring all facilities to "like-new" condition, it will reasonably focus its condition-monitoring resources on the more "urgent" and "expensive" facilities.

We develop a method to address such a bias in the sample. The method incorporates a model of the process by which the sample is selected. We then demonstrate this method for a case study, which is a set of Air Traffic Control (ATC) facilities operated by the Federal Aviation Administration (FAA). The contributions of this paper are two-fold. First, we develop a method for utilizing biased sample data to derive information about the entire population. The kind of bias treated here is generic and may be encountered in a wide variety of situations. Second, we apply this method to a class of facilities under the FAA's jurisdiction, known as "un-staffed facilities" and consisting of structures housing radars, navigation aids, radio beacons, and other ground-based equipment, for which no previous system-wide evaluation has been attempted. FAA has identified such a comprehensive un-staffed facility evaluation as critical to the ongoing re-structuring of its infrastructure assets [1].

The rest of this paper is organized as follows. We first motivate and state the problem. We then identify a possible method for addressing the problem derived from previous literature, and discuss its shortcomings. We next propose more innovative and appropriate methods, followed by a description of the case study. We then apply the alternative methods to several different classes of FAA un-staffed facilities, compare their results, and demonstrate the advantages of our proposed method.

II. PROBLEM STATEMENT AND MOTIVATION

Consider a system of diverse infrastructure facilities spread over a wide region, managed by a government or private agency. The periodic allocation of maintenance and replacement funds to facilities is based, at least in part, on information about facility condition. The specific information provided is the cost of bringing a subset of facilities to "like new" state. Only a subset of facilities is included because it is not feasible (or even desirable) to bring all facilities to such a state, and the cost of developing the information of any given facility is non-negligible. Finally assume that agency policy is to prioritize maintenance and repair

activities on facilities in the worst condition, and thus with the highest restoration costs, there being a strong correlation between poor condition and high restoration cost.

Given these circumstances, it is reasonable to suppose that the subset of facilities for which cost information is provided will not be representative of the entire set, but rather be skewed towards those in the poorest condition. These cost estimates provided for a given class of facility can be treated as a sample of the population, but it is a biased sample in which facilities in a poor condition (and thus with higher restoration costs) are more likely to be included.

Suppose now that the agency wishes to use restoration cost data from these biased samples to estimate properties of the populations from which they are drawn. These properties might include the cost of restoring all facilities to "like-new" condition, or the probability distribution of these costs for individual facility types. There may be various motivations for this, including internal "budget drills" or the wish to publicly document the extent to which maintenance budget is "under-funded." Whatever the reason, and irrespective of its validity, the technical problem is to use available data for a purpose it was not originally intended for, and hence is imperfectly suited. Our aim is to investigate how to do this.

Let us now formalize the above problem. Let X be a random variable that is the cost of bringing a given type of facility to "like-new" condition. Suppose there are N facilities of this type, and that we have cost information for n of these facilities. Let p_i ($i = 1, 2 \dots N$) be the probability that facility i is included in the sample, and suppose that p_i depends on x_i the value of X for facility i : $p_i = p(x_i)$. Further, assume that the function $p(\cdot)$ is positive monotonic; more specific assumptions about the function will be discussed below. Given our sample data $x_1, x_2 \dots x_n$, our objective is to estimate the probability density function (PDF) for X , $f_X(x)$.

III. ALTERNATIVE APPROACHES

We now present three approaches to this estimation problem. The main difference between them is the specific assumptions we make about the function $p(\cdot)$.

A. Truncation Models

The most basic approach for modeling this kind of data is to assume that the sample is a truncated sample, with truncation point being a . Thus, we assume in this case that the cost data is drawn from the set of facilities whose restoration cost is above a . Given this assumption, a truncated distribution function can be constructed from the probability density function (PDF) and cumulative distribution function (CDF). Thus, if $f_X(x)$ is the PDF of the un-truncated distribution and $F_X(x)$ is the cumulative distribution function (CDF), the density of the truncated random variable can be written as [2]:

$$f_X(x|x > a) = \frac{f_X(x)}{\text{Prob}(x \geq a)} = \frac{f_X(x)}{1 - F_X(a)} \quad (1)$$

Along with the parameters of $f_X(x)$, the truncation point a will also be an unknown parameter to be estimated. The estimation of a is subject to the constraint that $a \leq x_1$, where x_1 is the lowest value in the sample. The likelihood function for this approach can be written as

$$L = \prod_{i=1,2,\dots,n} \frac{f_X(x_i)}{1 - F_X(a)} \quad (2)$$

For estimation of this model, it can be shown that the constraint $a \leq x_1$ is binding (proof is included in appendix 1). Thus, the truncation point would be the lowest value in the sample, and estimation involves determining the parameters of $f_X(x)$ only.

The above approach employs the cost data only. In most cases, the total number of facilities, including those without cost data, is also known. This information can be used in the estimation process. In order to do so, we must make an assumption about the process that determines whether or not a given facility appears in the sample. Two assumptions may be considered. First, we can assume that all of the facilities whose cost is above a are included. In this case, we know that if a facility is excluded, its cost must be below a . This yields the likelihood function:

$$L = \prod_{i \in P \setminus S} F_X(a) \prod_{i \in S} f_X(x_i) \quad (3)$$

where S be the set of facilities in the sample, and P is the entire set, with $S \subseteq P$. Again, the estimation of a is subject to the constraint that $a \leq x_1$, where x_1 is the lowest value in the sample. Using arguments similar to appendix 1, it can be easily seen that this constraint is binding. We call this approach truncation with complete sampling (TWCS).

Alternatively, we can assume that the data represents only a fraction of the facilities whose restoration cost is greater than the truncation value. This sampling fraction thus becomes an additional parameter to be estimated, along with the truncation point and the parameters of $f_X(x)$. In effect, we assume that the facilities are initially screened to eliminate those whose restoration cost is less than a , and that a fraction p of the remaining facilities are included in the sample. The process could also begin by choosing a sample from all the facilities, and then eliminating from that sample those with a cost below a . The likelihood function is the same regardless of the sequence, but is most intuitively expressed if it is assumed that the initial sample includes all the facilities. It is given by:

$$L = \prod_{i \in P \setminus S} [pF_X(a) + (1 - p)] \prod_{i \in S} pf_X(x_i) \quad (4)$$

The first product is for facilities not in the sample, either because they were initially sampled and had a cost less than a , or because they were not initially chosen for the sample. The second product corresponds to facilities in the sample. Note that when $p = 1$, this model reduces to the previous one. As before, estimation of a is subject to the constraint that $a \leq x_1$, and using arguments similar to appendix 1, it can be seen that this constraint is binding. We call this approach truncation with incomplete sampling (TWIS).

Both of these models have an important limitation. The bias toward facilities with higher restoration costs takes the form of rule that simply excludes facilities with costs below a certain value. The data is treated as an unbiased sample of those facilities that pass the cost test. A more plausible assumption is that the bias operates continuously: the higher the cost, the more likely the facility will be included. This is the basis for the next set of models.

B. Continuous Bias Models

These models assume that as facility repair cost increases, the probability of the facility being included in the data increases in a continuous fashion. These may be no absolute minimum cost, but facilities with low repair costs are very unlikely to be sampled. In the previous truncation based models, we modeled this selection as a constant probability, p , for all facilities with repair cost above the minimum value. Now selection is modeled as a monotonically increasing function of the repair cost.

Perhaps the simplest such model is that the selection probability is a linear function of the repair cost, with the probability being zero for the lowest repair cost in the sample, and 1 for highest. Formally, if x_1 and x_n be the lowest and highest values in the sample respectively, the probability of inclusion in the sample is:

$$p(x_i) = \begin{cases} 0, & x_i \leq x_1 \\ \frac{x_i - x_1}{x_n - x_1}, & x_1 \leq x_i \leq x_n \\ 1, & x_n \leq x_i \end{cases} \quad (5)$$

In this case, the likelihood function can be written as:

$$L = \prod_{i \in P \setminus S} \left(1 - \int_0^\infty p(y) f_X(y) dy \right) \prod_{i \in S} p(x_i) f_X(x_i) \quad (6)$$

The first product term covers the facilities that are not included in the sample, and the second term gives the contribution of the sample in the likelihood function. The integral $\int_0^\infty p(y) f_X(y) dy$ is the probability that a facility is selected. The likelihood maximization should yield a result such that $\int_0^\infty p(y) f_X(y) dy$ nearly equal to the ratio of sample to population.

The advantage of the above method of linearly increasing probability is the ease of estimation. Equation (5) gives a pre-determined probability for each data point in the sample, and

there are no added parameters to be estimated besides the underlying distribution. This stems from the linear relationship and assigning probability values to the largest and smallest data-points. However, these assumptions themselves are a shortcoming of the above approach. This is because firstly, the relationship need not be linear, and secondly, even if the relationship is linear, determining the parameters of the linear relationship by assigning probabilities to smallest and largest values is restrictive. A more refined approach is to estimate the parameters of the sample selection model along with those of the cost distribution.

As stated earlier, the functional form used to model sample selection should be positive monotonic, since the selection is skewed towards higher values. Further, for the ease of estimation, the function should preferably be smooth for $x > 0$, where x is the repair cost. Equation (5) shows that estimation involves evaluating a definite integral that varies with the parameters being estimated, and a non-differentiable selection function would add to the complexity of an involved estimation. This criterion questions the logical extension to the functional form in (5), where x_1 and x_n would be parameters to be estimated rather than lowest and highest values in the sample. The function $p(x_i)$ in (5) is not differentiable at x_1 and x_n , and this would lead to a complicated estimation procedure.

Due to the above reasons, we adopt a binary logit model for sample selection. The probability of a facility with a particular repair cost, x , being present in the sample is:

$$p(x_i) = \frac{e^{\beta + \alpha f(x_i)}}{1 + e^{\beta + \alpha f(x_i)}} \quad (7)$$

where α and β are parameters to be estimated and $f(\cdot)$ is a positive monotonic function. For $\alpha > 0$, $p(x_i)$ approaches 1 as x_i becomes very large. Moreover, in this model sample selection can be represented as a utility maximization process, in which the expression $\beta + \alpha f(x_i)$ can be interpreted as the deterministic utility of including facility i in the sample as compared to the alternative of not including it, whose deterministic utility is assumed to be 0. The likelihood function remains the same as in (6). Estimation, however, is considerably more difficult as compared to the earlier methods because the definite integral includes unknown parameters of both the probability function and the repair cost distribution. We call this approach the continuous bias approach (CB).

IV. CASE STUDY: UN-STAFFED FACILITIES IN THE NATIONAL AIRSPACE SYSTEM

A. Introduction and Objective

The Federal Aviation Administration, as a part of its ongoing re-structuring of its infrastructure assets, plans to comprehensively evaluate the *un-staffed* facilities in the National Airspace System (NAS). Un-staffed facilities are structures that house communication, navigation, and surveillance equipment. The number of physical assets in the NAS is significant. The

NAS contains about 5,000 unstaffed facilities and 9,000 structural towers. Although the number and importance of such facilities is significant, it appears that the state of the art in assessing facility conditions and its performance is not very advanced [1,3,4].

NAS facilities are very diverse in terms of their type, construction and size, geographic location, environment and the traffic area they serve [9]. This diversity represents one of the major challenges in assessing facility condition and performance at an aggregate level. The entire NAS has been divided into nine different regions, based on different climactic and local conditions [9]. Because of the large number of diverse unstaffed facilities distributed across the entire NAS it would be extremely expensive to systematically assess and evaluate each facility and to establish a comprehensive data-base within a short time-frame. Instead, it is proposed to assess the condition of a representative sample of facilities from different regions.

The ultimate objective of our work is to develop a sampling methodology for this assessment. In order to do this, it is desirable to develop preliminary estimates of the mean and variance of repair cost for different classes of facilities so that accuracy of estimates yielded by different sample sizes can be predicted. These preliminary estimates are based on samples of facilities for which cost data are available. As noted above, these existing samples are skewed toward facilities with high repair costs. The preliminary estimates can then be used to design a sampling methodology based on stratified sampling [5].

B. Data and Methodology

As stated before, the entire NAS is divided into 9 regions, with each region having its own sub-divisions. There are 12 different types, and a table with abbreviations for different facilities is reproduced from [9] in appendix 2. For each facility, the respective sub-division assigns a subjective measure (Facility Condition Index, FCI) of the condition of the facility, with the assignment being updated at least once every year. This FCI is on a scale of 1 to 5, with 1 denoting a new facility, and 5 denoting need for replacement.

The data provided included the following:

- The population of a particular type of facility in any one of the 9 regions.
- For a certain subset of facilities, the *deferred maintenance cost* (DMC) was provided. DMC represents the cost of the repair that has been deferred for that financial year. This estimate is based on subjective judgment and periodic cursory inspection of facilities suspected to be in need of repair. Further, most of the data is for facilities which are judged to have an FCI value of 4 or 5.

In [9], the authors give the sizes of the total population and the sample for which deferred maintenance cost is available for each facility type. These data are not the result of a thorough inspection but based on quick appraisals, to be used for budget allocation. Nevertheless, some FAA managers state that these

estimates are a good representation of the real cost. In the following sections, we assume these data are accurate and use them to estimate, for individual facility types, the underlying distributions of deferred maintenance cost. We do this for all types except TDWR's and ASDE's, whose sample sizes of 8 and 6 are too small to yield meaningful results.

C. Analysis

In this analysis, we assume that the repair costs are lognormal, or the log of the costs is normally distributed. This is clearly more plausible than the normal, because cost must be non-negative, but includes only two parameters, making estimation tractable. We also conducted experiments with other distributions including the folded normal and the exponential distribution (see table 4), but found the lognormal to have the best results. More complicated distributions, such as the Gamma might also be tried, but they would make estimation more difficult. Moreover, as discussed below, goodness-of-fit tests performed after estimation yielded acceptable results for most of the facility types. The functional forms for the lognormal distribution are given below in (8).

$$f_Y(y) = \frac{1}{By\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln y - A}{B}\right)^2}$$

$$F_Y(y) = \Phi\left(\frac{\ln y - A}{B}\right)$$
(8)

1) *Truncation:* We first present results from maximizing the likelihood function in (2), which considers only the sample cost values, not the fraction of the population sampled. As mentioned above, the maximum likelihood estimate for the truncation point is the lowest value in the sample, leaving just the two log-normal distribution parameters for numerical estimation. The results of this estimation are given in table 1. The parameters *A* and *B* are the same as defined in (8).

The Kolmogorov-Smirnov (KS) test was performed on the truncated probability function defined in (2). As shown in table 1, the estimated distribution passes the KS test for the .05 significance level in almost all cases.

TABLE 1. ESTIMATION RESULTS FROM SIMPLE TRUNCATION

Facility Type	A	B	KS Test Value
ALS	8.768 (0.428)	1.945 (0.327)	0.083*
ARSR	<u>10.483</u> (0.162)	1.476 (0.105)	0.063*
ASR	9.47 (0.209)	1.718 (0.183)	0.057*
AWOS / ASOS	9.27 (0.181)	<u>1.035</u> (0.123)	0.236
GS	<u>9.295</u> (0.115)	1.471 (0.078)	0.111
LOC	9.3 (0.134)	1.717 (0.098)	0.118
MALS / SSALS	<u>9.422</u> (0.855)	<u>2.221</u> (0.87)	0.073*
RCAG	9.6 (0.101)	1.45 (0.055)	0.076*
RTR	<u>9.673</u> (0.124)	<u>1.529</u> (0.097)	0.079*
VOR	9.588 (0.062)	1.255 (0.038)	0.049*

(Values in brackets are standard errors for estimates)
Underlined and italicized estimates are significant at 0.05 level
 * Estimated distribution passes KS test at 0.05 level

As discussed before, the likelihood function used in the above estimation ignores potentially important information concerning the proportion of the population included in the sample. This information is particularly relevant if we assume that the sample includes all facilities whose DMC is greater or equal to the truncation value. In this case the likelihood function is given in (3). Estimation results, which appear in table 2, are much different than those in table 1, with lower A values and higher B values. Moreover, in virtually all cases, the fitted distribution fails the KS test. This suggests that the assumption that the data is a complete sample of values exceeding the truncation value is wrong.

TABLE 2. ESTIMATION RESULTS FROM MODIFIED TRUNCATION WITH COMPLETE SAMPLING (TWCS)

Facility Type	A	B	KS Test Value	Sample Size	Population
ALS	<u>4.874</u> (0.572)	<u>3.826</u> (0.702)	0.177 ^d	42	126
ARSR	<u>2.769</u> (0.409)	<u>3.89</u> (0.603)	0.361	85	136
ASR	<u>3.824</u> (0.594)	<u>4.804</u> (0.839)	0.234 ^d	77	249
AWOS / ASOS	-0.851 (2.73)	<u>5.02</u> (1.721)	0.370	36	600
GS	0.445 (0.839)	<u>6.136</u> (0.952)	0.316	187	914
LOC	-0.909 (0.96)	<u>6.652</u> (0.941)	0.273	194	1150
MALS / SSALS	-7.494 (5.383)	<u>2.382</u> (2.661)	0.156 ^d	17	711
RCAG	<u>0.983</u> (0.599)	<u>2.442</u> (1.349)	0.412	217	633
RTR	-0.172 (0.997)	<u>6.451</u> (1.007)	0.293	179	1030
VOR	<u>5.9</u> (0.148)	<u>4.202</u> (0.314)	0.335	487	967

(Values in brackets are standard errors for estimates)
Underlined and italicized estimates are significant at 0.05 level
^d Estimated distribution passes KS test at 0.05 level

The third truncation model, described by (4), relaxes the assumption that the entire population above the cutoff point is included in the sample, but still exploits information about the proportion of the facilities included in the sample. The parameters for estimation are the parameters of the lognormal distribution (A and B) and the sampling fraction (p). Further, it should be noted that as a consequence of the estimation, the expression $1 - p(1 - F_X(a))$ should be equal to the ratio of the sample to the population. The results of the estimation, along with the values of this expression and the ratio of sample to population are given below in table 3.

As expected, the sample-to-population ratio predicted matches to observed ratio. Moreover, the estimated values for (p) are very close to this ratio as well. This means that virtually all the exclusions of facilities excluded from the sample are the result of incomplete sampling rather than truncation. This also explains why the estimated values for A and B in Table 3 are so close to the estimates for the Simple Truncation model (Table 1). If exclusions from the sample are nearly always a consequence of random sampling rather than truncation, then accounting for the excluded facilities in the log likelihood function has very little effect.

2) *Continuous Bias Models:* Estimation results from the truncation models reveal that, if the samples in our data are indeed biased toward more facilities with higher DMC values,

then this bias is not well represented using models based on truncation. It appears from those results that if a bias in fact exists, it results not from categorically excluding facilities whose DMC is below a certain value, but from a tendency to include more facilities with high DMCs. This suggests the use of a continuous bias model. Since we are using the lognormal distribution, we used natural logarithm for the function $f(\cdot)$ in (7).

TABLE 3. ESTIMATION RESULTS FROM MODIFIED TRUNCATION WITH INCOMPLETE SAMPLING (TWIS)

Facility Type	A	B	p	$\frac{1-p}{1-pF_X(a)}$	Sample Fraction	Sample Size	KS Test Value
ALS	<u>8.764</u> (0.427)	<u>1.947</u> (0.327)	<u>0.375</u> (0.056)	0.333	0.333	42	0.083 ^d
ARSR	<u>10.478</u> (0.162)	1.476 (0.105)	<u>0.625</u> (0.042)	0.624	0.625	85	0.062 ^d
ASR	<u>9.468</u> (0.208)	<u>1.716</u> (0.182)	<u>0.316</u> (0.03)	0.309	0.309	77	0.058 ^d
AWOS / ASOS	<u>9.266</u> (0.181)	<u>1.035</u> (0.123)	<u>0.061</u> (0.01)	0.060	0.060	36	0.234
GS	<u>9.292</u> (0.114)	<u>1.47</u> (0.078)	<u>0.205</u> (0.013)	0.205	0.205	187	0.112
LOC	<u>9.297</u> (0.134)	<u>1.718</u> (0.099)	<u>0.17</u> (0.011)	0.169	0.169	194	0.119
MALS / SSALS	<u>9.419</u> (0.857)	<u>2.222</u> (0.871)	<u>0.028</u> (0.009)	0.024	0.024	17	0.073 ^d
RCAG	<u>9.597</u> (0.101)	<u>1.449</u> (0.055)	<u>0.342</u> (0.019)	0.342	0.343	217	0.077 ^d
RTR	<u>9.671</u> (0.121)	<u>1.529</u> (0.097)	<u>0.174</u> (0.012)	0.174	0.174	179	0.080 ^d
VOR	<u>9.584</u> (0.062)	<u>1.356</u> (0.038)	<u>0.504</u> (0.016)	0.503	0.504	487	0.050 ^d

(Values in brackets are standard errors for estimates)
Underlined and italicized estimates are significant at 0.05 level
^d Estimated distribution passes KS test at 0.05 level

As stated before, this approach involves evaluating a large definite integral that varies with the parameters to be estimated. One way to do this is to use maximum simulated likelihood, where simulated probabilities are used instead of actual probabilities [6, 7]. However, our model definition involves only a one-dimensional definite integral, and hence we used the Newton-Cote's quadrature rules (trapezoidal rule) to approximate the integral [8]. Newton-Cote's formulas work by using interpolating functions to evaluate the integral, and in our case, we use the linear interpolation. The results from the estimation are given in table 4.

TABLE 4. ESTIMATION RESULTS FROM THE CONTINUOUS BIAS APPROACH

Facility Type	A	B	α	β	$\int_0^{\infty} \frac{p(y)}{f_X(y)} dy$	Sample Fraction	Sample Size	KS Test Value
ALS	<u>7.025</u> (1.088)	<u>2.33</u> (0.668)	<u>1.206</u> (0.659)	<u>-9.891</u> (1.262)	0.334	0.333	42	0.091 ^d
ARSR	<u>9.415</u> (0.269)	<u>1.973</u> (0.232)	<u>1.846</u> (0.691)	<u>-16.136</u> (2.612)	0.620	0.625	85	0.055 ^d
ASR	<u>8.94</u> (1.521)	1.676 (0.313)	0.339 (0.82)	-3.898 (3.322)	0.309	0.309	77	0.066 ^d
AWOS / ASOS	<u>9.221</u> (0.757)	<u>0.998</u> (0.118)	0.086 (0.789)	-3.544 (3.641)	0.060	0.060	36	0.242 [†]
GS	<u>7.86</u> (0.41)	<u>1.682</u> (0.166)	<u>0.84</u> (0.201)	<u>-8.443</u> (0.628)	0.203	0.205	187	0.100 [†]
LOC	<u>7.474</u> (0.456)	<u>1.95</u> (0.184)	<u>0.782</u> (0.155)	<u>-8.073</u> (0.451)	0.167	0.169	194	0.100 [†]
MALS / SSALS	<u>7.779*</u> (1.971)	<u>1.767</u> (0.482)	<u>0.66*</u> (0.438)	<u>-9.566</u> (1.35)	0.022	0.024	17	0.104 ^d
RCAG	<u>8.291</u> (0.381)	<u>1.764</u> (0.194)	<u>0.941</u> (0.261)	<u>-8.783</u> (0.849)	0.340	0.343	217	0.067 ^d
RTR	<u>8.196</u> (0.566)	<u>1.71</u> (0.197)	<u>0.774</u> (0.249)	<u>-8.39</u> (0.842)	0.173	0.174	179	0.068 ^d
VOR	<u>8.247</u> (0.17)	<u>1.908</u> (0.132)	<u>1.81</u> (0.331)	<u>-14.909</u> (0.948)	0.502	0.504	487	0.077

(Values in brackets are standard errors for estimates)
Underlined and italicized estimates are significant at 0.05 level
^{*} Estimates significant at 0.1 level
^d Estimated distribution passes KS test at 0.05 level
[†] Estimated distribution passes KS test at 0.01 level, fails at 0.05 level

The parameter that captures bias in this model is α . Estimates for this parameter are positive in every case, implying that, as expected, higher cost facilities are more likely to be included in the sample. Moreover, based on a one-tailed test, α is significant at the .10 level in eight cases and at the .05 level in seven. Thus in most cases we can be fairly sure that a bias toward higher DMC facilities exists. Furthermore, based on the earlier estimation results, the truncation models are essentially equivalent to the continuous model with $\alpha = 0$. Thus rejection of this hypothesis implies that the continuous bias model is the most valid of those considered.

Table 4 also summarizes the KS test results comparing the predicted distribution for the observations in the sample to the observed data. In six of the 10 cases, the null hypothesis that the data came from the fitted distribution cannot be rejected at the .05 level. Of the remaining four cases, in three the null hypothesis cannot be rejected at the .01 level, while in one it is rejected at both .05 and .01 levels.

3) *Model Comparison:* We now compare results of the various models, in particular the CB and the TWIS, which we found to be the most satisfactory of the truncation models. With regard to goodness-of-fit, KS test results from the continuous bias and TWIS models closely resemble one another. The three facilities with the poorest distribution fits are the same in both models, and the magnitudes of the KS statistics in these cases—and most others—are quite similar. Thus, while estimation results, as well as the perception of FAA facility managers, comport better with the continuous bias model, the goodness-of-fit results do not support this conclusion.

To further explore the differences between the CB and TWIS models, we plotted the fitted PDFs and sample selection probabilities, and compared predicted CDFs and observed data. The plots for three facility types are given in figure 1. The PDF derived from the TWIS model is almost always shifted to the right of the one obtained from the CB model. The shift is greater when the probability of selection derived from the CB model changes more (in a proportional sense) over the central region of the PDF—this is the situation in which the bias will have the greatest effect. An instructive counterexample is the AWOS/ASOS case, where the PDFs are nearly identical and the selection probability is nearly constant in the central region.

While the PDFs and sample selection probabilities generally look very different for the two models, the resulting CDFs for sampled observations are strikingly similar. As the KS tests also revealed, for most facilities modeled CDFs also fit the empirical distributions quite well. Of the three cases with the poorest fits, in two the modeled distributions appear to have thicker right tails than the observed data, while in the other (which has just 17 observations), it appears that the central part of the distribution is more complicated than suggested by either model.

The ultimate aim of these models is to estimate the average DMC for each type of facility. Estimates from the four models appear in Table 5. Estimates from the CB model are less than

those from any of the others. Comparing CB and TWIS estimates, the difference ranges from around 9% for ARSR's, to over 300% for localizers. It is also notable that only the CB model yields estimates of the population mean that are consistently below those of the sample mean. This again demonstrates that the CB model was the only one that actually demonstrated the bias believed to exist by FAA subject matter experts.

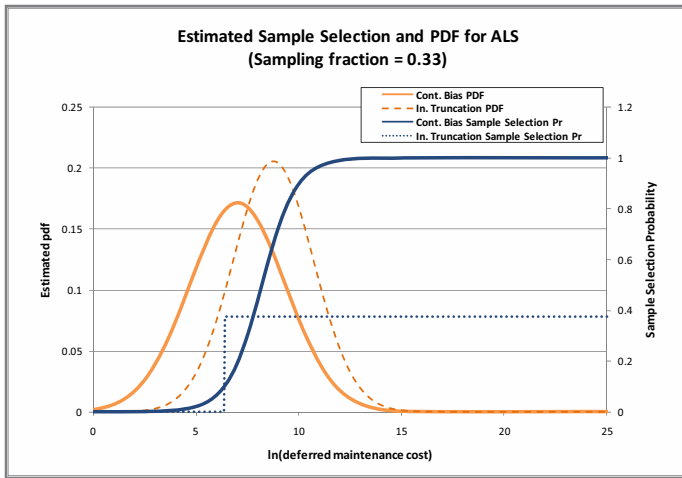
TABLE 5. COMPARING SAMPLE MEAN TO ESTIMATED POPULATION MEAN

Facility Type	Mean of Data or Sample Mean ($\times 10^3$)	Estimated Population Mean ($\times 10^3$)			Continuous Bias
		Simple Truncation	Truncation with complete sampling	Truncation with incomplete sampling	
ALS	42	43	198	43	17
ARSR	95	106	4,571	106	86
ASR	46	57	4,699	57	31
AWOS / ASOS	17	18	127	18	17
GS	30	32	233,800	32	11
LOC	44	48	1,636,000	48	12
MALS / SSALS	86	146	378,900	146	11
RCAG	40	42	2,948,000,000	42	19
RTR	40	51	916,400	51	16
VOR	42	37	2,501	37	24

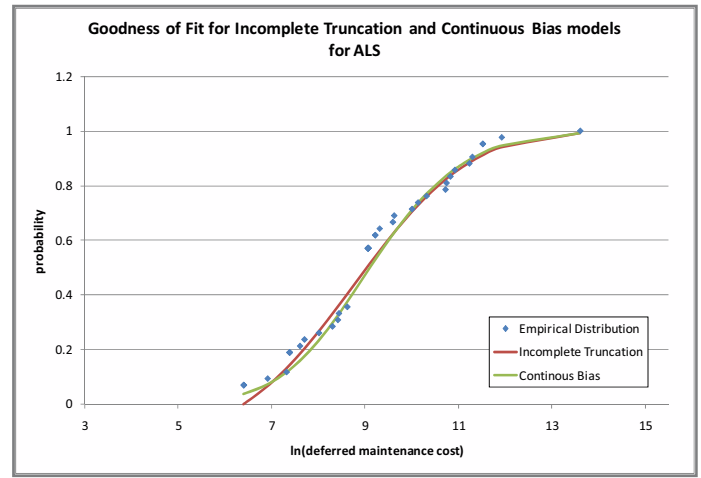
V. CONCLUSION

We have investigated the problem of estimating the parameters of a distribution from a sample of data in the face of known, or assumed, biases in the sampling process. In our application, the data are costs of restoring unstaffed facilities maintained by the FAA to support flight operations. Cost estimates are available for some facilities, but the samples are believed to be biased toward high cost instances. We have sought practical methods of inferring the cost distribution for the entire population that take this bias into account. There are many other settings, in infrastructure management and beyond, in which such a situation may exist, and to which our methods may also apply.

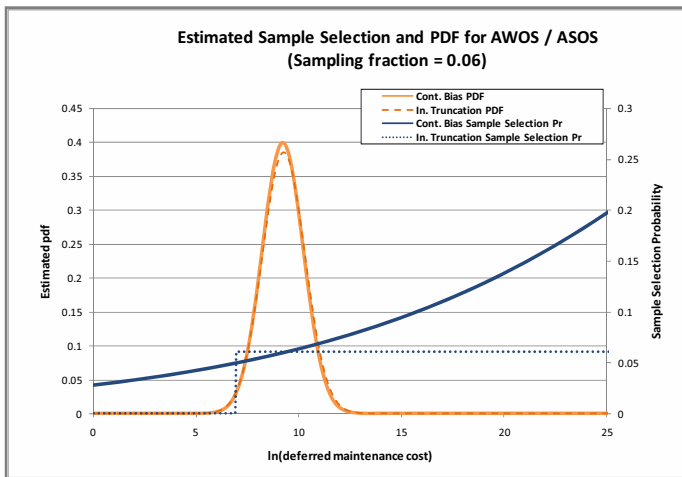
We have experimented with two ways of modeling the bias: (1) truncation, which assumes that facilities are systematically excluded if their restoration cost falls below a certain value, and (2) continuous bias, which allows the sampling probability to increase gradually as cost increases. Estimation results for the truncation models suggest that truncation is not a major source of sample bias, while results from the continuous model do suggest bias. Thus, if we accept the premise that such bias exists, then the continuous bias model proved more suitable. On the other hand, the latter model did not surpass the truncation models in terms of goodness-of-fit. In other words, the data do not, in and of themselves, favor the continuous bias model.



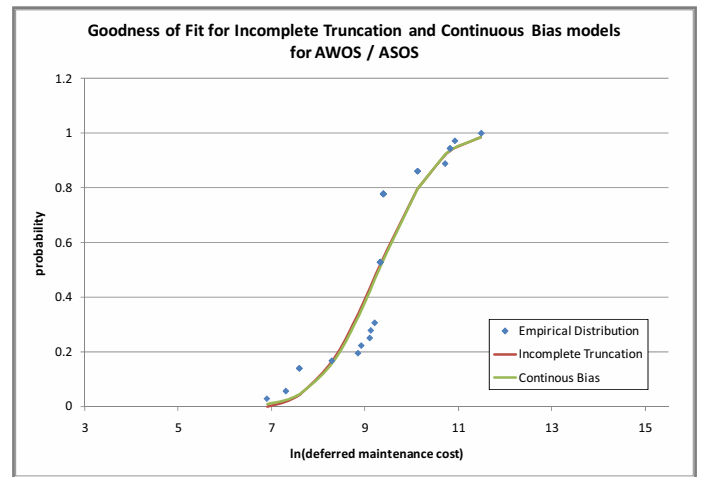
(a)



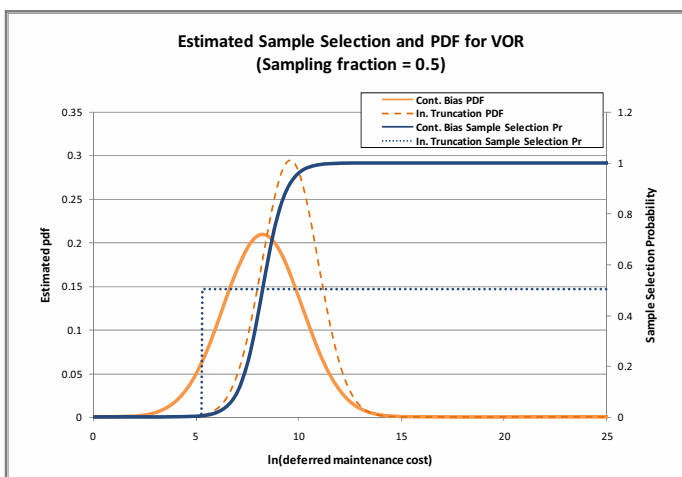
(b)



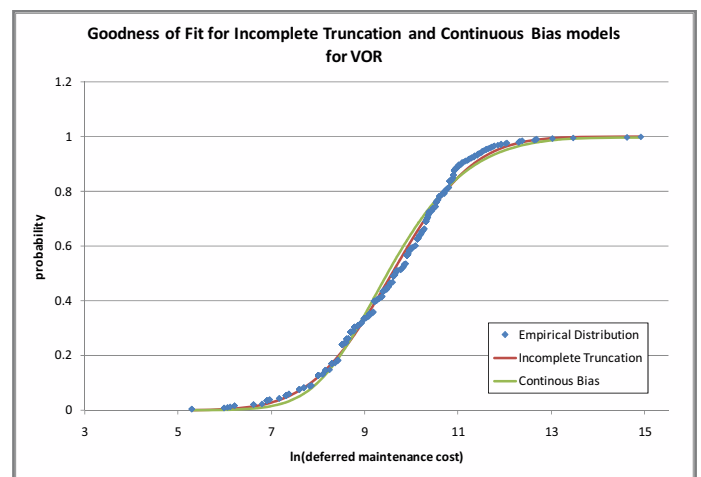
(c)



(d)



(e)



(f)

Fig. 1. Estimated Sample Selection and PDF for ALS, AWOS / ASOS, and VOR. Goodness of Fit for Incomplete Truncation and Continuous Bias models for ALS, AWOS / ASOS, and VOR.

Throughout our analysis, we have maintained the assumption that the restoration cost has a log normal distribution. This was chosen as the most plausible two-parameter distribution for our setting, and results show that it works quite well for most facilities. It would certainly be desirable to extend these methods to more general distributions, but this may prove difficult, particularly for the continuous bias model. Estimation proved challenging even with a two-parameter distribution. As the number of parameters increases, bias effects and properties of the actual distribution will become harder to disentangle. Ultimately, the better approach is almost certainly to collect an unbiased sample. The methods presented here are best used as means of developing preliminary estimates from which efficient sampling strategies can be devised.

Appendix 1: Binding Nature of Constraint on Truncation Point for Simple Truncation

Let $f_X(x)$ be the probability distribution function (pdf) of the underlying population distribution that we are trying to estimate, with μ and σ being the parameters of the distribution. Let a be the truncation point for the estimation, with a being a parameter to be estimated too. Let $x_1, x_2 \dots x_n$ be the sample values, sorted in increasing order. Thus, x_1 is the smallest value in the sample, and the estimation of the truncation point a is done subject to the constraint that $a \leq x_1$. Now, the truncated distribution that we are trying to estimate can be written as

$$f'_X(x) = \frac{f_X(x)}{1 - F_X(a)} \tag{9}$$

where

$$F_X(a) = \int_0^a f_X(x) dx \tag{10}$$

Thus, the likelihood function can be written as

$$\mathcal{L} = \prod_{i=1}^n \frac{f_X(x_i)}{1 - F_X(a)} \tag{11}$$

And the log-likelihood function becomes

$$\log \mathcal{L} = \sum_{i=1}^n \log f_X(x_i) - n \log(1 - F_X(a)) \tag{12}$$

If μ and σ are the parameters of $f_X(x)$, then both the numerator and denominator of the likelihood function are dependent on μ and σ . However, only the denominator $(1 - F_X(a))^n$ depends on the truncation point a . Consider the cumulative distribution function $F_X(a)$. For any parameters μ and σ , the value of $F_X(a)$ increases monotonically with a , since it is the area under the pdf for $x \leq a$. Thus, the function $\frac{1}{(1 - F_X(a))^n}$ also increases monotonically with a for a given μ and σ . Hence, the constraint $a \leq x_1$ becomes binding

while maximizing the log-likelihood function in terms of μ , σ and a .

Appendix 2: List of abbreviations related to different types of facilities

Abbreviation	Facility Type
TDWR	Terminal Doppler Weather Radar
ASR	Airport Surveillance Radar
ASDE	Airport Surface Detection Equipment
ARSR	Airport Route Surveillance Radar
RTR	Remote Transmitter Receiver
RCL	Radio Communication Link
RML	Remote Microwave Link
TML	Television Microwave Link
VOR	VHF Omni-directional Range
VORTAC	VOR collected with TACAN
TACAN	Tactical Aircraft Control and Navigation
LOC	Localizer
ALS	Approach Light System
MALS	Medium Intensity Approach Lighting System
SSALS	Simplified Short Approach Lighting System
AWOS	Automated Weather Observation System
ASOS	Automatic Surface Observing System
NEXRAD	Next Generation Weather Radar
LLWAS	Low Level Wind Shear Alert System
RCAG	Remote Communication Air / Ground
GS	Glide Slope

REFERENCES

- [1] FAA, National Airspace System Capital Investment Plan 2003-2007, <http://www.faa.gov>, 2002
- [2] R. J. Larsen and M.L. Marx, An introduction to mathematical statistics and its applications. 3rd ed. 2001, Upper Saddle River, NJ: Prentice Hall. x, 790.
- [3] T. Brantley, "FAA's Aging ATC Facilities: Investigating the Need to Improve Facilities and Worker Conditions". Statement before the House Committee on Transportation and Infrastructure - Subcommittee on Aviation, July 24, 2007, 2007
- [4] P. Forrey and P. Gilbert, "FAA's Aging ATC Facilities: Investigating the Need to Improve Facilities and Worker Conditions". Testimony before the House Committee on Transportation and Infrastructure - Subcommittee on Aviation, July 24, 2007. <http://www.natca.org/legislationcenter>
- [5] P.S.R.S. Rao, Sampling methodologies with applications. Texts in statistical science. 2000, Boca Raton, Fla.: Chapman & Hall/CRC. 311.
- [6] C. Arias and T. L. Cox, "Maximum simulated likelihood : a brief introduction for practitioners". Agricultural & applied economics staff paper series, 1999(no 421): p. 15p.
- [7] K. Train, Discrete choice methods with simulation. 2003, New York: Cambridge University Press. vii, 334.
- [8] C. W. Ueberhuber, Numerical computation : methods, software, and analysis. 1997, Berlin ; New York: Springer.
- [9] G. Gupta, J. Rakas, and M. Hansen, "Cost-effective sampling to evaluate condition of un-staffed facilities in National Airspace System". in Collection of Technical Papers - AIAA 5th ATIO and the AIAA 16th Lighter-than-Air Systems Technology Conference and Balloon Systems Conference. 2005. Arlington, VA, United States: American Inst. Aeronautics and Astronautics Inc., Reston, VA 20191-4344, United States.

Track 9

Environment, Weather & Economy

On the Use of Near Field Computational Fluid Dynamics for Improving Airport Related Dispersion Models

Syoginus S. Aloysius and Luiz C. Wrobel

School of Engineering and Design

Brunel University

Uxbridge, United Kingdom

Syoginus.Aloysius@brunel.ac.uk, Luiz.Wrobel@brunel.ac.uk

Abstract— This paper discusses one of the major problems concerning current dispersion modelling techniques used around airports; the source dynamics characterization. Due to the lack of information and non-availability of experimental data, common dispersion models rely on very simple source approximations. Through a staged process, the paper shows a more accurate representation of the plume dynamics of an aircraft during the take-off phase. Using Computational Fluid Dynamics, useful data can be collected to represent and understand the fluid mechanics associated with the dispersion process. The results can help dispersion modelers with better source dynamics representation and benefit the management of aircraft time separation delays in the take-off and landing phases.

Keywords— *Source Dynamics; Airports; Take-off; CFD; LIDAR; LES; Buoyant Jets; Ground Effects; Dispersion Models; Air Quality.*

I. INTRODUCTION

There is a growing concern on the pollution resulting from airport operations because of the expansion of air traffic over the years. It is forecasted that future air traffic movements will increase at a mean annual rate of 5 to 7% [1]. According to Schafer et al., 15,000 aircraft are already populating the sky and an expected 2,200 billion of passenger kilometers are flown each year [2].

It is estimated that 3.5% of the global warming from human activity comes from air transportation, and it is predicted that this figure will rise to 15% by 2050 if no measures are taken to control air traffic, according to the intergovernmental panel on climate change [3].

Global effects such as depletion of the ozone layer and global warming are a direct consequence of local activities. To assess these impacts and quantify the amount of pollution resulting from airport operations, airport operators and regulators rely on different techniques to estimate existing situations and predict future scenarios. On-site monitoring is one technique that has been used but it is rather costly and does not isolate airport-related sources. Another option is the use of dispersion modelling techniques to approximate emission dispersion within a virtual domain. There are three modelling

techniques commonly used in the aerospace industry, namely Gaussian, Lagrangian and Eulerian models. Each technique has advantages and drawbacks associated with the way they treat the problem. This paper examines one problem that has been known to both Gaussian and Lagrangian models, the source dynamics characterization.

It is known that the main source contributors at an airport are the emissions from aircraft engines and emissions from traffic inside and around the airport [4]. These are all moving sources, thus there is a real need of properly representing them in a simulation. This paper will first discuss why there is a need for improvement and how source dynamics are actually implemented in Gaussian and Lagrangian models. A discussion will then follow on different ways to characterize a moving source, before focusing on the real aim of the paper: the use of near-field Eulerian simulation to help improve Gaussian and Lagrangian models. To achieve this goal, several reports prepared as part of the ALAQS project (Airport Local Air Quality Studies), managed by EUROCONTROL, will be recalled and a staged process analysis will be done to finally analyze the dispersion process of a complete aircraft during take-off.

II. THE NEED FOR IMPROVING EXISTING AIRPORT DISPERSION MODELS

There are three main types of dispersion modelling techniques, namely Gaussian, Lagrangian and Eulerian. The Gaussian model relies on a simple formula that calculates the concentration field emitted by a source “under stationary meteorological and emission conditions” [5]. In the Lagrangian particle models, the concentration is calculated through integration across the entire computational domain [5]. A gridding system is used to discretize the domain and to calculate localized concentrations within the grid. The Eulerian technique, on which Computational Fluid Dynamics (CFD) simulators are based, solves the governing equations of fluid flow with numerical methods. The basic idea of CFD is to discretize the control volume into small sub-domains, creating a grid system similar to the Lagrangian technique. The fundamental equations of fluid flow are solved either explicitly in the case of Direct Numerical Simulation (DNS) or are sub-

modelled in the case of Large Eddy Simulation (LES) or Reynolds Averaged Navier-Stokes (RANS).

Fleuti [4] provided a simple comparison of these models in an airport context using the commercial software packages ADMS-Urban, LASPORT and EPISODE for Gaussian, Lagrangian and Eulerian models, respectively. The report shows that Lagrangian and Eulerian methods offer advantages over Gaussian models in both spatial and temporal resolutions [4].

Farias & ApSimon [6] studied the contribution of NO_x from traffic and aircraft emissions around Heathrow airport. The Gaussian ADMS-Urban software was used to compare results obtained at different monitoring stations located in populated areas. They found some discrepancies between the two; the ADMS-Urban model overestimated aircraft contribution and underestimated the traffic contribution in comparison with monitoring data.

This over-prediction of airfield-related traffic was also observed by Fleuti & Hoffmann [7] in their study of Zurich airport with the Lagrangian model LASAT, in which simulation results were compared with monitoring sites located at different positions across the airfield. They concluded that the landing and take-off phase emission factors were lower in real condition than the simulated ones.

Fleuti et al. [8] also carried out a sensitivity analysis of Zurich airport, and pointed out that the emissions from the aircraft main engine dominate all other sources of pollutants. This was later confirmed by Celikel et al. [9] in their emission inventory at Zurich airport. Using the ALAQS-AV tool set, they showed that NO_x emissions from aircraft sources around Zurich airport in 2003 were found to be approximately 80.7% of the total emissions [9].

III. SOURCE DYNAMICS TREATMENT BY GAUSSIAN AND LAGRANGIAN MODELS

Both ADMS and LASAT have integrated functions to take into account the dynamics of an airplane, but they are in a way very simplistic. The Gaussian ADMS model treats this problem by assigning an accelerating jet source to represent the effects of buoyancy and momentum of an aircraft engine [10]. Farias and ApSimon [6] criticized this approach because it is lacking features that “may contribute to thermal buoyancy” for the traffic emissions. In the case of aircraft emissions, the pollutant release occurs at different heights with fast moving sources during a short period of time [6]. They later added that ADMS, although highly suitable for dispersion from stacks, needs adjustments for ground traffic use because there are differences in “mechanics governing dispersion from these emissions” [6].

LASAT, on the other hand, rely on parameters calibrated by means of Dusseldorf DOAS measurements, which are also available for APU, GPU and vehicle source dynamics [11]. To characterize the take-off phase, LASPORT used an initial uniform box of 50m wide and 25m high having parameters (velocity and turbulence) of the exhaust plume decaying in a time scale of a minute or two [12]. Fleuti et al. [8] pointed out a major drawback concerning this method, because these parameters are not fixed and rely on constant changes to get the

adequate results. For instance, they needed to increase the vertical source extent for vehicle exhaust they studied from 2m to 8m, and the Auxiliary Power Unit (APU) was also changed using more recent information from the Frankfurt airport air quality studies [8].

The ALAQS-AV tool set has the capability to impose a smooth and shift approach on these different modelling techniques [11]. This concept is based on the principle that emissions from engine exhausts can be shifted spatially downwind and smoothed to replicate the real plume dynamics of the jet engine. The smooth and shift approach used by ALAQS-AV employs LASPORT default parameterization values for different types of moving sources present at the airport [13].

Because of its time and length scales, the emission is done through an hourly passive grid source. A major drawback of this technique is its non-accountability of the meteorological conditions; the effects of the wind direction and magnitude are not properly treated, and this has a consequence in the directional exit of the plume and its rise [13]. These parameters are set in the model without any differentiation of the type of aircraft which was operational within the hour. This “one type set of data fits all” approach has also been criticized by Farias and ApSimon [6] because ADMS uses only one set of buoyancy and momentum parameters for all aircraft.

Yu et al. [14] showed that the major disadvantage of using air quality models is the lack of knowledge on emissions rates of pollutants from different types of engines, and these vary greatly during different phases of an aircraft. This argument was also raised by Schafer et al. [2], who pointed out that the only source of data available up to now is the ICAO database, but unfortunately this does not give any information whatsoever about the source dynamics. Fleuti et al. [8] also expressed the need for information concerning the plume characteristics in their sensitivity analysis; this parameter was found to have a major influence in the dispersion process. Like Schafer et al. [2], they argued that the knowledge of how much emission is being released is important but the dynamics associated with it is also of great relevance.

IV. INADEQUACY OF LIDAR MEASUREMENTS FOR SOURCE DYNAMICS CHARACTERISATION

One possible answer is to use Light Detection and Ranging (Lidar) equipment. Lidar systems consist of the transmission of a light signal pulse through a laser sheet; particles emitted from an engine backscatter the light and the laser receives back the signal, showing the presence of a certain particle at a location along the sheet [15]. Depending on the quality of the equipment, it is possible to analyze the results up to every second, with a resolution of about 10 meters, operating at the ultraviolet wavelength of 355nm [16], but this comes at a high price as a complete mobile facility has to be built. Examples are the Rapid Scanning Lidar Facility (RASCAL) introduced by the Manchester Metropolitan University or the Ozone Profiling Atmospheric Lidar (OPAL) by the National Oceanic and Atmospheric Administration. RASCAL is a self-contained mobile unit with the necessary equipment to work autonomously with on-board meteorological facility. Apart

from the equipment cost, there are several disadvantages associated with the use of Lidar to characterize source dynamics.

Starting with the meteorological conditions, it is likely that the wind speed and direction will not be constant and the changes that may occur will alter the fluid dynamics, hence dispersion. This makes it harder to compare one observation to another and correlate the results in order to have a proper understanding of the dynamics.

To be able to capture a pollutant, the particle has to reflect. In other words, the size of the particle is important. Usually aerosol, in the order of "100 nanometers with an average of 30 nanometers", is measured to give an approximate description of the plume on the laser sheet [15]. Wayson et al. [15] also found that these particles can only be seen using their OPAL system operating in the very low ultra-violet wavelength (355nm). This means that this device has to be used to its limit to capture the parameters of interest.

This search for small particles causes some problems, especially around airports where several other sources can pollute the Lidar results. Eberhard et al. [16] reported some of the problems concerning the low contrast between the backscattering and the signal from the ambient air. They found that some aircraft presented very low or no plume relative to the ambient air, and explained that this was due to backscatter particles having the same size as the ambient air. What is more intriguing is the fact that they found conditions where the plume has less backscatter than the ambient air, due to "a combination of volatilization of the ambient particles passing through the engine and low particle emissions" [16]. Another possible explanation is that the background concentration and interaction with other sources present around the airport are greater than the emission released by the engines.

According to Angus Graham from Manchester Metropolitan University (private communication), Lidar data have to be processed in order to be readable, and this involves a considerable amount of processing and manpower, at a cost that cannot be neglected. Eberhard et al. [16] also pointed out that only 40% of their measurements were retained after this "processing and quality control". This finally gave the plume characteristics of only 21 types of aircraft.

To conclude this discussion on Lidar, only a sheet can be analyzed making the analysis two-dimensional. It is often interesting to know the dispersion process not only on the vertical plane but also in the lateral direction, because some concentration may be trapped into wing-tip vortices and travel around the airport, as will be shown later in this paper. Another problem with the analysis is that it is only qualitative. Moreover, there are disagreements concerning the take-off phase. Wayson et al. [15] found compact ground-based plumes for all types of airplanes, whereas Graham et al. [17] in the Project for the Sustainable Development of Heathrow (PSDH) report found a non-compact ground plume for the B747-400. Yamartino et al. [12] tried to directly compare LASPORT's initial box with the plume size dimensions reported by Wayson et al. [15], and found it rather impossible to do so because of the time scale used by both analyses. Ignoring this, it was revealed that LASPORT produced values twice as large as the

ones reported by Wayson et al. The explanation for this difference is that the measurement values taken by Wayson et al. are for the plume at a very early stage, before even the dissipation of the plume's internal and thermal energy had taken place [12]. Yamartino et al. [12] concluded that issues of plume entrainment and rise characterization have to be resolved before Gaussian or Lagrangian models can perform accurately in airport-related dispersion studies.

V. EULERIAN MODEL AS A WAY TO UNDERSTAND AND CHARACTERIZE SOURCE DYNAMICS

As discussed previously, Fleuti [4] compared the results of the Eulerian commercial package EPISODE to ADMS-Urban and LASPORT. This is a unique comparison between the three techniques, but unfortunately EPISODE is not a 100% Eulerian model. The calculation of the large scale dispersion process is done through an Eulerian grid where an averaged form of the Navier-Stokes equation is computed via the K-theory [4]. The small scales, on the other hand, are calculated through a subgrid scale Lagrangian or Gaussian model [18]. These near source treatments are usually done by point or line source dispersion [18].

Within the ALAQS project initiated in 2003 by the Eurocontrol Experimental Centre, Aloysius et al. [19] provided a comparison between CFD and LASAT simulations for airfield emission dispersion. It is believed this report was the first attempt to compare full Eulerian and Lagrangian models for pollutant dispersion around an airport. The CFD studies were based on the Large Eddy Simulation (LES) method to predict the dispersion of NO_x, assumed to be a non-reactive pollutant, from the airport runways. A number of other simplifications were introduced to the model to reduce computing time, such as the terrain and the runways which were respectively flat and modelled as area sources.

Although the simulation of the air traffic at Zurich airport was done for a one-day period only, the results for the fourth hour of the day were interesting because of the changes in wind direction and magnitude, and emissions values, that happen during that period of time. Those changes were found to happen four times during that particular day, mainly due to airport operations or severe meteorological conditions.

The predicted dispersion patterns and the magnitudes of the NO_x concentrations showed good agreement between the Eulerian and Lagrangian models. The notable differences were that LASAT predicts higher concentrations at ground level than CFD. At higher altitudes, LASAT predicts an intuitive dispersion throughout the control volume whereas CFD presents some recirculation. Although the magnitude of this recirculation is in the order of less than 1 ppb in this case, this can be higher for larger airports. In addition, it was found with the CFD simulation that:

- Vortices play an important role at high wind speeds, making the flow recirculate around the control volume.
- Buoyancy is dominant when emissions are high and wind speed relatively low.

- Large vortices contain high concentration of pollutant and do most of the transport.
- The influences of small eddies are small, and their number increases when large vortices break down due to the surrounding flow and the wind magnitude.
- Changes in wind direction alter the flow rotation creating vortices in different directions.

It has also become apparent from this study that the application of CFD to the simulation of a full airfield is currently unreasonable. The CFD model required more than 8 days of processing time to reach its transient solution for a single hour simulation using 8 dual core processors optimized for parallel processing. In comparison, LASAT took about 60 seconds to compute the solution. But it was concluded from this report that CFD allows for a much in-depth assessment of how emissions are transported, and the wealth of information it provides means that it would be extremely beneficial if applied to more localized airfield studies such as source dynamics.

The strategy adopted in this study to validate the CFD results was to gradually increase, through a staged process, the complexity of the simulation towards representing the near-field effects of aircraft exhaust plumes under realistic conditions. This paper will go through this process from a free jet engine to a complete aircraft on the runway.

The LES model was initially used to investigate the differences between turbulent buoyant and non-buoyant jets in a free atmosphere condition, highlighting the mechanism of dispersion behind the exhaust. The non-buoyant free jet was compared with existing experiments and analytical results [20]. The non-buoyant simulation results were found to agree with classical results, replicating the characteristic self-preserving behavior of a free jet after the flow development region (Fig.1). The buoyant jet, on the other hand, breaks the symmetrical pattern of the flow showing a rise of the axial velocity above the centerline axis, as illustrated in Fig. 2.

The disappearance of the potential core was found to be the

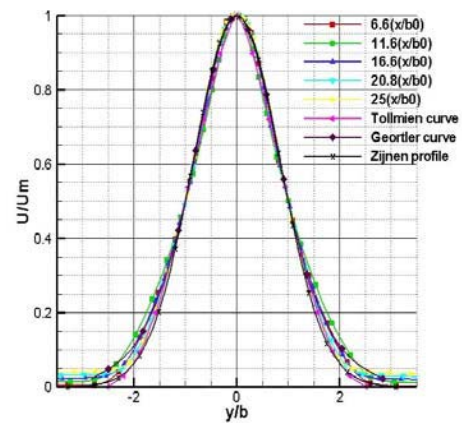


Figure 1. Vertical velocity comparison between CFD and theoretical methods for non-buoyant free jet

result of turbulence penetration, leading to the fully developed region. In the potential core, the jet entrains the ambient fluid surrounding it, triggering an enhancement of turbulence. Its intensity begins to act and increases further along the axis, resulting in a decay of the axial velocity. Compared to the buoyant jet, the non-buoyant jet has a longer potential core because it is not restrained by the buoyancy effects acting on the flow.

The spreading of the jet was found to be linear in the case of non-buoyant release and followed very closely the theoretical predictions. The buoyant jet, on the other hand, can be divided into three different linear regions:

- The jet region, closest to the exhaust, where the linear curve matches closely the non-buoyant line.
- The plume region, furthest away from the exhaust, where a high spreading of the jet can be observed because the buoyancy effect is dominant.
- The intermediate region, where the flow undergoes a transition from pure jet-like to plume-

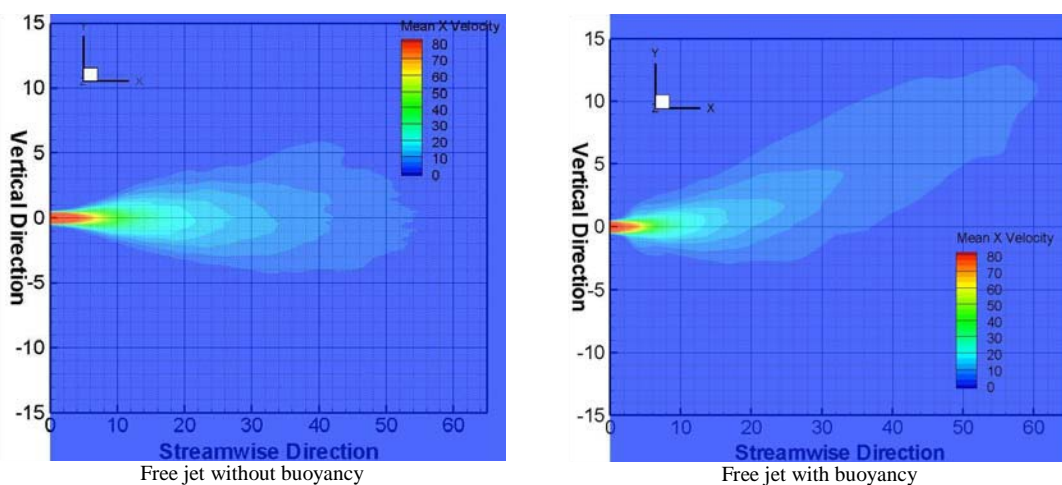


Figure 2. Mean velocity profile comparison between buoyant and non-buoyant free jet

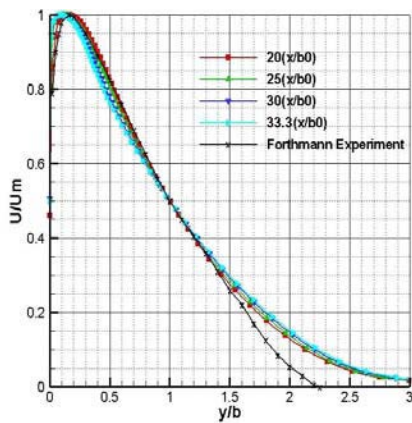


Figure 3. CFD and experimental results for the non buoyant wall jet vertical velocity profile

like behavior.

Vortices around the spanwise direction play an important role in the jet dispersion, as they regulate the potential core length when the counter-rotating vortices collide in the centerline axis and help the buoyancy jet to rise higher than the non-buoyant jet.

A second study within the ALAQS project aimed to provide an understanding of the impact of the presence of the ground on the fluid dynamics of the jet [21]. Before carrying out such comparison, a validation similar to the one presented in [20] was done on this wall jet against existing experimental and analytical results for a non-buoyant condition. Only then the buoyant case was analyzed and used to assess the impact of the solid boundary on the fluid mechanics of the jet flow. Once again, the results of the CFD simulation of the non-buoyant

wall jet agreed very well with the classical results. An illustration of this is presented in Fig. 3, where it successfully replicated the boundary layer profile created by the wall and the free shear layer profile generated by the ambient fluid.

The comparison between the buoyant free and wall jets (Fig. 4) revealed several differences. First, the potential core was found to be much longer for the wall jet than for the free jet. This has an effect on the flow penetration through the control volume; the wall jet offers a deeper penetration than the free jet. The maximum velocity decays much faster for the free jet than for the wall jet. This leads to a correlation between this parameter and the penetration properties previously discussed, as the penetration involves higher velocity pushing into the control volume.

There is also an interconnection between all the parameters discussed previously and the streamwise vortical structure of the buoyant wall jet. As in the case of the buoyant free jet, counter-rotating vortices are created on one side by the surrounding fluid and on the other by the solid boundary. What is different from the free jet situation is the presence of the wall generating continuous vortices, whereas the influence of the vortices created by the surrounding fluid gradually decreases.

The first point of merging of the counter-rotating vortices occurs in the potential core. As the flow progresses, the intensity of the vortices generated by the wall is much stronger than the ones created by the surrounding fluid, causing a clinging of the flow, also known as the Coanda effect. This pushing-down phenomenon restricts the growth of the jet vertically, hence creating a lower rate of spread than for the buoyant free jet. As the velocity further away from the jet exhaust decreases, the vortices created by the wall decrease and buoyancy takes over, with positive streamwise vortices lifting up the flow from the ground.

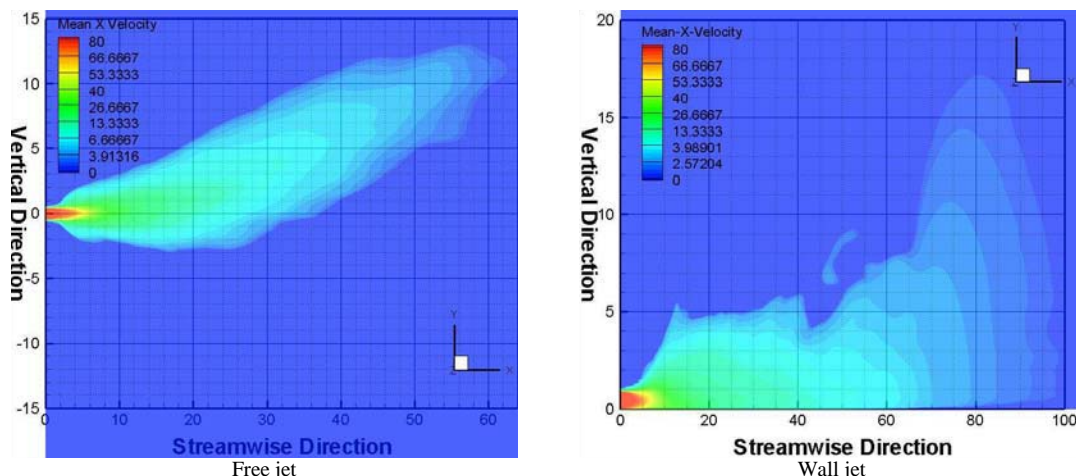


Figure 4. Comparison of mean velocity profile of buoyant free and wall jet after 10s

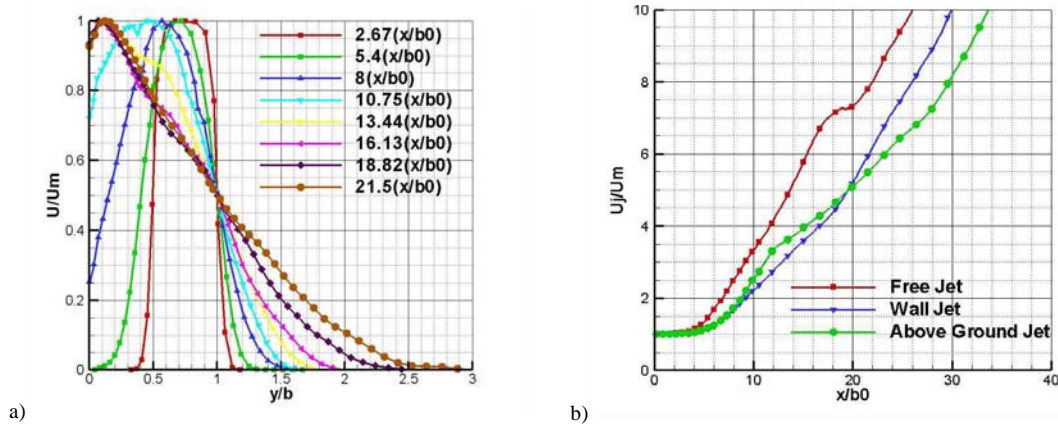


Figure 5. Vertical velocity profile (a) and velocity decay comparison between different jets (b)

Raising the jet above the ground showed a combination of both free and wall jet velocity profiles. Near the engine exhaust, the velocity profile resembles that of the free jet whereas further downwind the characteristic self-similar profile of the wall jet can be found (Fig. 5a). The rate of decay of the maximum velocity of different jets is shown in Fig. 5b, with the above ground jet having the lowest rate. This is in line with the results reported by Davis and Winarto [22] and has a consequence in terms of penetration through the control volume. The above ground jet configuration is expected to have a deeper penetration of the exhaust gases through the control volume.

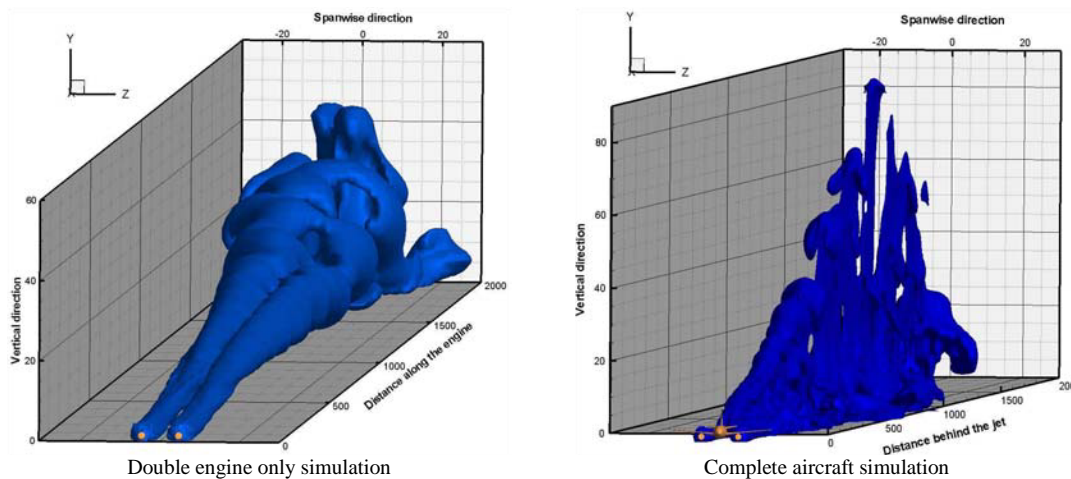
A further study within the ALAQS project aimed at characterizing the effects of engine geometry and acceleration in the source dynamics by incorporating the complete geometry of an engine (CFM56-3C-1) [23]. Two engine configurations (single and double engine) were simulated under two types of wind (a headwind of 2.5m/s and a crosswind of 4.5m/s at an angle of 40 degrees from the engine centre). The results highlighted the importance of several parameters on plume dispersion.

The first parameter is the wind configuration; headwinds directly applied to the engine will increase the plume

penetration downwind. Crosswinds, on the other hand, increase the lateral spread of the plume. The second parameter is related to the vortical structure configuration. After the jet regime, divergence of the jet core from its centre occurs when the plume starts to weaken. This is the result of counter-rotating vortices wrapping around the “self-induced” rotation of the jet. A sinusoidal instability pattern was reported due to the divergence from the centre, which increases in amplitude and period as the flow progresses through the control volume, leading to the breakup of the instability.

In the case of a double engine, the wind configuration and vortical structures parameters still affect the plume dispersion process. In fact, it was found that the presence of two engines enhanced it because of the interaction between the two jets’ plumes favoring a “straining” mechanism, resulting in the development of the instabilities (increase in amplitude), thus facilitating the breakup.

The final step was to include the full geometry of the airplane; a Boeing 737 was chosen because it uses the CFM56-3C-1 engines and is the most popular at medium size airports such as Zurich. Presently, results are only available with a double engine on the runway with a headwind condition. The results show some differences when the body, the wings and



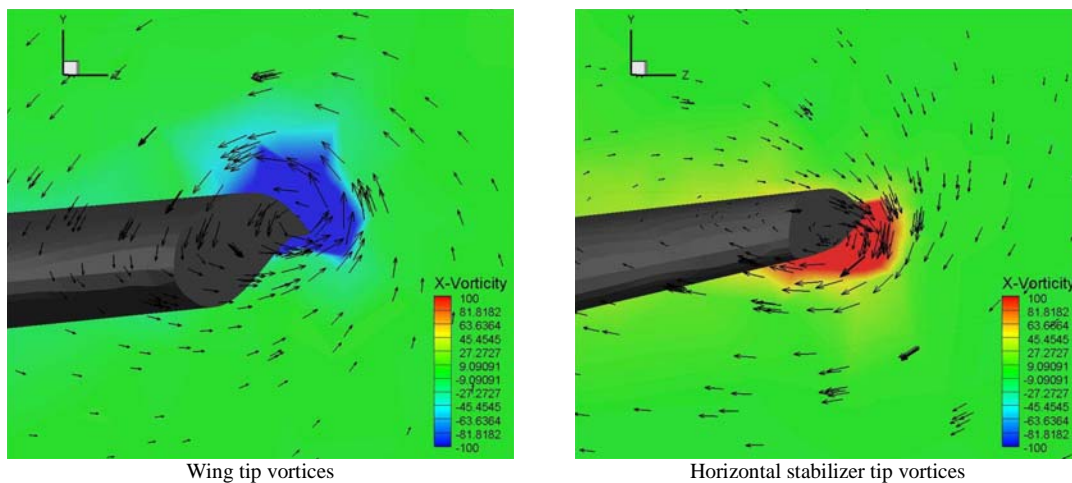


Figure 7. Wing and horizontal stabilizer tip vortices comparison

the empennage of an aircraft are added to the simulation.

Fig. 6 represents an iso-surface of 1 ppm NOx concentration. The whole aircraft restricts the flow progress downwind through the control volume. The penetration length is about 1,860m behind the exhaust, whereas for the engines alone it is extended by 40m. Without any geometrical disturbances, the flow was allowed to travel deeper.

On the other hand, some geometrical disturbances were found to enhance the vertical and lateral dispersion. These are due to the vortices created by the wing and the horizontal stabilizer. Fig. 7 shows a comparison of their tip vortices; the arrows represent the flow pattern in the Y-Z plane and the magnitude of the vortices is shown by the contour plot scale. The differences concern the rotation of the flow; the air stream over the wing creates a negative rotation at the tip around the X-direction, whereas the horizontal stabilizer's tip shows a positive rotation. This is expected as the horizontal stabilizer is in fact an inverted wing, but what is least expected is the impact they have on the plume dispersion.

Because the horizontal stabilizer is located just after the jet exhaust and almost on its way, it will be the first to affect the

flow. Its disturbance is characterized by an enhancement of the vertical dispersion as can be seen in Fig 8. Part of the engine exhaust is attracted by the rotation of the fluid created by the horizontal stabilizer's tip and spreads upward.

The influence of the wing-tip can only be observed further away from the engines' exhausts when the jet plume has spread sufficiently laterally to meet the wing-tip vortices core. Its disturbance of the flow pattern is characterized by an enhancement of the lateral dispersion. Fig. 8 clearly shows this; the concentration of NOx rolls around the negative x-vortices core and spreads laterally, at low distances from the ground.

Other parameters of interest, such as velocity, temperature and turbulence, can be obtained from the CFD simulation. These are the major parameters to characterize the source dynamics in any problem. In addition, the simulation could serve several other purposes, e.g. related to air traffic operations. Results can shed light on the strength of the wake turbulence when different types of aircraft take-off and land. As a consequence, a proper time delay separation can be obtained for different classes of aircraft, thus increasing the airport traffic capacity.

The need for a better characterization of source dynamics in an airport environment is important not only for moving aircraft during the take-off and landing phases, but also necessary when they are immobile and using their APU. In this study, a non-reactive pollutant was introduced to simplify the problem but chemistry models exist for CFD simulations that allow the prediction of chemical transformations associated with plume dispersion.

VI. CONCLUSIONS

The work reported in the present paper is related to the need from airport dispersion modelers for better source dynamics characterization. Gaussian and Lagrangian techniques are well suited for complete airport dispersion calculations, but only include very simplified models to represent moving sources during take-off and landing. Unfortunately, these are not appropriate for adequate characterization of the near-field fluid dynamics. Conversely,

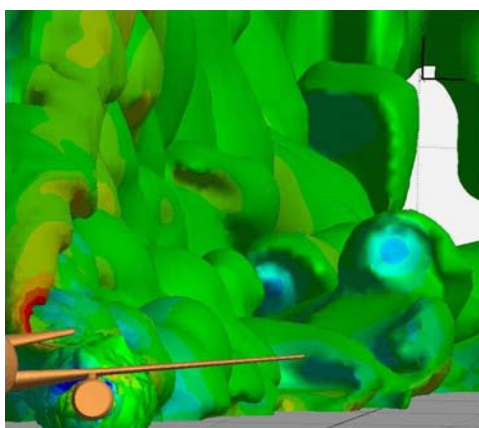


Figure 8. Iso-surface of 1ppm of NOx plotted with x-vorticity

CFD techniques are presently not appropriate for the simulation of a full scale airport, but it can help to improve the understanding of some of the flow characteristics, including source dynamics.

This paper summarizes our CFD work related to airport local air quality studies. The strategy adopted was to use a staged approach, starting from a single engine free in the atmosphere and increasing the simulation complexity step by step. This allowed an initial validation with known theoretical and experimental results. After investigating free jets in buoyant and non-buoyant conditions, a wall jet simulation was analyzed and validated before the jet was raised to a proper distance above the ground. Important results of the simulations include the influence of the maximum velocity decay in the plume penetration through the control volume, and the role of counter-rotating vortices in the dispersion process.

The incorporation of the engine and the aircraft body, including the wings and empennage, revealed the effects of several other parameters such as the horizontal stabilizer's and wing's tip vortices. These were found to play an important role in the vertical and lateral dispersion, whereas the aircraft was found to slow down the plume penetration through the control volume.

This paper has demonstrated the applicability of CFD methods to airport-related flow and dispersion problems, and their capability to characterize source dynamics. The results can help dispersion modelers with better source dynamics representation and benefit the air traffic management of aircraft time separation delays in the take-off and landing phases.

ACKNOWLEDGMENT

The work reported in this paper has been undertaken as part of the Airport Local Air Quality Studies (ALAQS) project commissioned and sponsored by EUROCONTROL. The authors would like to thank all the members of the ALAQS consortium, especially Dr Ian Fuller from EUROCONTROL.

REFERENCES

- [1] A. Graham and D. Raper. "Air Quality in Airport Approaches: Impact of Emissions entrained by Vortices in Aircraft Wakes". <http://www.cate.mmu.ac.uk/documents/Publications/Woct03.pdf>. (2003).
- [2] K. Schäfer, C. Jahn, P. Sturm, B. Lechner and M. Bacher. "Aircraft Emission Measurements by Remote Sensing Methodologies at Airports". *Atmospheric Environment*, Vol. 37, pp. 5261-5271. (2003).
- [3] J.E. Penner, D.H. Lister, D.J. Griggs, D.J. Dokken and M. McFarland. "Aviation and the Global Atmosphere". <http://www.grida.no/climate/ipcc/index.htm>. Intergovernmental Panel on Climate Change special report.
- [4] E. Fleuti. "Airport Local Air Quality Assessment", Eurocontrol report C20126E/DM/02. Eurocontrol Experimental Centre. (2002).
- [5] P. Zannetti. "Air Pollution Modeling: Theories, Computational Methods and Available Software", Avon, UK, Computational Mechanics Publications, 1990.
- [6] F. Farias and H. ApSimon. "Relative Contribution from Traffic and Aircraft NOx Emissions to Exposure in West London", *Environmental Modelling and Software*, Vol.21, pp. 477-485. Feb. 2005.
- [7] E. Fleuti and P. Hofmann. "Airport Local Air Quality Studies: Case Study Zurich Airport 2004", Eurocontrol report. Eurocontrol Experimental Centre. (2005).
- [8] E. Fleuti, P. Hofmann and C. Talerico. "Airport Local Air Quality Studies, Sensitivity Analysis Zurich Airport 2004", Eurocontrol report C21197/05. Eurocontrol Experimental Centre. (2006).
- [9] A. Celikel, N. Duchene, I. Fuller, M. Silue, E. Fleuti, P. Hofmann and T. Moore. "Airport Local Air Quality Studies, Case Study: Emission Inventory for Zurich Airport with Different Methodologies. Eurocontrol report EEC/SEE/2004/010. Eurocontrol Experimental Centre. (2004).
- [10] D. Carruthers, C. McHugh, M. Jackson and K. Johnson. "Developments in ADMS-Airport to Take Account of Near Field Dispersion and Applications to Heathrow Airport". Proceedings of the 11th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes. (2007).
- [11] U. Janicke and I. Fuller. "Derivation of Smooth and Shift Parameters to Account for Source Dynamics in ALAQS-AV Emissions Grids". Eurocontrol report C21197/05. Eurocontrol Experimental Centre. (2005).
- [12] R.J. Yamartino, P. Builtjes and R. Stern. "Status of the Current Level of Development and Understanding in the Field of Modeling Pollutant Dispersion at Airports". UFOPLAN-Ref. No. 20341253/01. (2004).
- [13] N. Duchene, I. Fuller and U. Janicke. "Verification of ALAQS Hourly 3D Grid Source Approach with Smooth and Shift Parameters to Account for Plume Dynamics". Proceedings of the 11th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes. (2007).
- [14] K.N. Yu, Y.P. Cheung and R.C. Henry. "Identifying the Impact of Large Urban Airports on Local Air Quality by Non Parametric Regression". *Atmospheric Environment*, Vol. 38, pp. 4501-4507. (2004).
- [15] R.L. Wayson, G.G. Fleming, B. Kim, W.L. Eberhard and W.A. Brewer. "Final Report: The Use of LIDAR to Characterize Aircraft Initial Plume Characteristics". FAA-AEE-04-01. (2004).
- [16] W.L. Eberhard, W.A. Brewer and R.L. Wayson. "LIDAR Observation of Jet Engine Exhaust for Air Quality" 2nd Symposium on LIDAR Atmospheric Applications. (2005). <http://ams.confex.com/ams/pdfpapers/83405.pdf>.
- [17] Project for the Sustainable Development of Heathrow. Report for the Department for Transport. <http://www.dft.gov.uk/pgr/aviation/environmentalissues/heathrow/>. (2006).
- [18] R.S. Sokhi, N. Kitwiroom and L. Luhana. "FUMAPEX-Datasets of Urban Air Pollution Models and Meteorological Pre-processors". Assessment of Different Existing Approaches to Forecast UAP Episodes". EVKL-CT-2002-00097. (2002).
- [19] S. Aloysius, D. Pearce, L.C. Wrobel and M. Silue. "ALAQS- Comparison of CFD with Lagrangian based Simulations for Airfield Emissions Dispersion". Eurocontrol report. Eurocontrol Experimental Centre. (2006).
- [20] S. Aloysius and L.C. Wrobel. "ALAQS- CFD Comparison of Buoyant and Non-Buoyant Turbulent Jets". Eurocontrol report. Eurocontrol Experimental Centre. (2007).
- [21] S. Aloysius and L.C. Wrobel. "ALAQS- CFD Comparison of Buoyant Free and Wall Turbulent Jets". Eurocontrol report. Eurocontrol Experimental Centre. (2007).
- [22] M.R. Davis and H. Winarto. "Jet Diffusion from a Circular Nozzle above a Solid Plane". *Journal of Fluid Mechanics*. Vol. 101, part 1, pp.201-222. (1980).
- [23] S. Aloysius and L.C. Wrobel. "CFD Simulation of Emissions Dispersion from an Aircraft Engine during Take-off". Eurocontrol report. Eurocontrol Experimental Centre. (2006).

Optimal departure aircraft trajectories minimising population annoyance

Xavier Prats*, Vicenç Puig, Joseba Quevedo, Fatiha Nejari

Technical University of Catalonia (UPC)

Av. Canal Olímpic 15

08086 Castelldefels. Spain

(*) Corresponding author: xavier.prats@upc.edu

Abstract—This paper presents a strategy for designing noise abatement procedures aimed at reducing the global annoyance perceived for the population living around the airports. By using fuzzy logic techniques it is shown how annoyance can be modelled in function of the maximum perceived noise level at a specific noise sensitive location and the time of the day when the departure takes place. Thus, the annoyance is computed for different kinds of sensibility areas, such as residential zones, industrial zones, schools or hospitals and an annoyance figure is obtained for each possible trajectory. Then, a non-linear multi-objective optimal control problem is presented in order to obtain the minimum annoyance trajectory for all sensitive locations. Lexicographic optimisation is used to cope with the difficulties that arise when several criteria appear in the optimisation process. Finally, a practical example is given for an hypothetical scenario where different optimal trajectories are obtained at different day periods.

I. INTRODUCTION

The noise produced by aircrafts during take-off and landing operations around airports is a very serious ecological and social problem. Aircraft noise can be very annoying for people living in the vicinity of the airports. Therefore, the design of noise abatement procedures aimed at reducing the noise exposure of the population around airports is one of the main issues that airport authorities and national navigation services providers have to address. Noise is generally defined as an unwanted sound and its effects can be appreciated physiologically but also psychologically [1]. Annoyance is a concept that is hard to quantify because there is no underlying physically measurable scale. However, it is usually qualitatively assessed with social surveys. It is clear that fuzzy techniques can help to make more accurate predictions by incorporating the vagueness and uncertainty into the modelling and reasoning process. Recently, few research papers based on fuzzy logic in noise pollution area have been reported [2], [3], [4]. In [3], annoyance is considered as a function of noise level, its duration of occurrence, and the socioeconomic status of a person and the results were applicable to the urban areas of India. In [4], a fuzzy model has been developed, on the basis of field surveys conducted by various researchers and reports of World Health Organisation, for predicting the effects of sleep disturbance by noise on humans as a function of noise level, age and duration of its occurrence. Fuzzy set theory is a generalisation of traditional set theory and provides a means for the representation of imprecision and vagueness. Zadeh [5]

further developed the corresponding fuzzy logic to manipulate fuzzy sets.

The International Civil Aviation Organisation (ICAO) publishes two different Noise Abatement Departure Procedures (NADP), defined in [6]. NADP are generic procedures and are far from being the optimum ones regarding noise minimisation. This is due to several factors, such as the impossibility to define a general procedure satisfying the specific problems that may affect each particular airport, air traffic management and airport capacity constraints or even the the limitations of nowadays on-board technology. Nevertheless, some research in theoretical optimal trajectories minimising the noise impact in departure or approaching procedures is also found in the literature. For instance, in [7], [8] and [9] is presented a tool combining a noise computation model, a Geographical Information System (GIS) and a dynamic trajectory optimisation algorithm, aimed at obtaining optimal noise procedures. A similar methodology is proposed in [10], and an adaptative algorithm for noise abatement can be found in [11]. On the other hand, in [12] and [13] it can be found a dynamic programming technique for minimising noise in runway-independent aircraft operations. All the results and conclusions arisen from these works are encouraging and will set the basis for new noise abatement procedures, specially regarding the forthcoming new navigation concepts, such as area navigation (RNAV) or Performance Based Navigation (PBN). These concepts will allow for air navigation procedures to be designed with a higher level of flexibility than conventional radionavigation ones [14].

This paper is organised as follows: in Section 2 the optimisation criteria are presented introducing how annoyance can be modelled by using fuzzy logic. Section 3 is devoted to the optimisation strategy that is proposed to solve this multi-criteria objective problem. Finally, section 4 shows the results obtained for a hypothetical airport scenario containing two residential zones, a school, a hospital and an industrial zone.

II. OPTIMISATION CRITERIA

This section presents two kinds of optimisation criteria. First one deals with the noise annoyance produced when the trajectory is flown. Second criterion takes into account airliner costs, such as time or fuel consumption.

TABLE I
RULE BASE TABLE FOR THE ANNOYANCE IN FUNCTION OF THE PERCEIVED NOISE AND THE HOUR OF THE DAY

	Residential zone			School			Hospital			Industrial zone		
	Mor.	Aft.	Night	Mor.	Aft.	Night	Mor.	Aft.	Night	Mor.	Aft.	Night
No noise	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Very low noise	NA	NA	SA	SA	SA	NA	SA	SA	MA	NA	NA	NA
Low noise	NA	SA	MA	MA	MA	NA	MA	MA	HA	NA	NA	NA
Medium noise	SA	MA	HA	HA	HA	NA	HA	HA	EA	NA	SA	SA
High noise	MA	HA	EA	HA	HA	NA	EA	EA	EA	SA	MA	MA
Very high noise	HA	EA	EA	EA	EA	NA	EA	EA	EA	MA	HA	HA

A. Noise annoyance

The annoyance or perception of the acoustic noise describes the relation between a given acoustic situation and a given individual or set of persons affected by the noise and how cognitively or emotionally they evaluate this situation. The acoustic annoyance of the aircraft flights around an urban airport depends logically of the acoustic behaviour produced in the sensitive locations, using for example, the L_{max} or Sound Exposure Level (SEL) metrics, but it is not a sufficient measurement to define completely the annoyance behaviour of a noise. An additional list of non acoustic elements to take into account to define the annoyance behaviour could be:

- Types of affected zones (rural zone, residential zone, industrial zone, hospitals, schools, markets,...)
- Time interval during the noise event (day, evening, night)
- Period of time between two consecutive flights
- Personal elements (emotional, apprehension to the noise, personal healthy, age,...)
- Cultural aspects (young or aged people habits, activities, holiday,...)

In conclusion, the annoyance is a subjective and a complex concept which can be studied as a qualitative form using fuzzy logic sets, as previous similar works in this area have been done (see for instance [2], [3], [4] and [15]). In this paper, the annoyance generated by the aircraft trajectories will be represented by fuzzy logic sets from the fuzzification of the maximum sound level (L_{max}) and from the hour of the day where the trajectory is supposed to be flown regarding four typical zones around an urban airport: a residential zone, a hospital, a school and an industrial zone.

1) *Noise model*: The maximum perceived sound level at location i is defined as:

$$L_i(\vec{z}) = \max_t Noise_i(\vec{z}(t)) \quad (1)$$

where $Noise(\vec{z}(t))$ is the perceived noise level at location i for a given trajectory $\vec{z}(t)$ (being t the time variable).

In this work, the same methodology employed by the Integrated Noise Model (INM) program [16] is implemented when computing noise functions. INM is developed by the Federal Aviation Administration¹ (FAA) and has been adopted as the standard package for noise studies and assessments

¹<http://www.faa.org>

in many countries. INM deals with several noise metrics and, in particular, noise levels are computed at a given point by selecting and interpolating appropriate noise values from a noise-thrust-distance (NTD) table, which is derived from empirical measurements.

2) *Annoyance model*: In [17] the authors presented a basic methodology for modelling aircraft noise annoyance by using fuzzy logic. Essentially, two membership set functions are defined. The first set introduces five linguistic terms to describe the magnitude of the maximum sound level (L_{max}):

- Very high noise
- High noise
- Medium noise
- Low noise
- Very low noise

A second set is related with the hour of the day introducing the following linguistic terms:

- Morning
- Afternoon
- Night

Afterwards a rule base is established to represent the annoyance of an event defined by the two fuzzy logic sets for each of the 4 zones considered. The annoyance concept has been represented by the following linguistic terms:

- Extreme Annoyance (EA)
- High Annoyance (HA)
- Moderated Annoyance (MA)
- Small Annoyance (SA)
- Null Annoyance (NA)

Table I shows the rule base of the annoyance at all sensitive locations. For each couple of sound level and time of day linguistic terms a rule is established giving a specific annoyance term. By using this kind of fuzzy rule base it would be easy, for example, to model the output of a population survey, asking for the annoyance produced by the airport.

Finally, the fuzzy set of the annoyance is defined as a crisp set to obtain a normalised degree of annoyance. Extreme annoyance corresponds to a normalised value of 1, high annoyance takes 0.75 value, medium annoyance takes 0.5, small annoyance takes 0.25 and finally null annoyance corresponds to 0. Figures 1, 2, 3 and 4 show in a plot this normalised annoyance in function of the two input variables (L_{max} and the hour of the day) for each noise sensitive location.

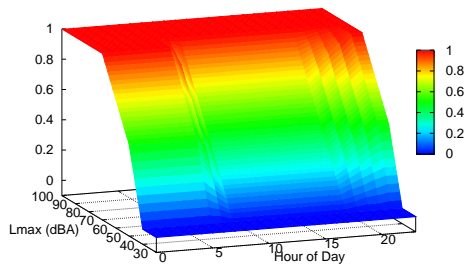


Fig. 1. Normalised Annoyance. **Hospital**

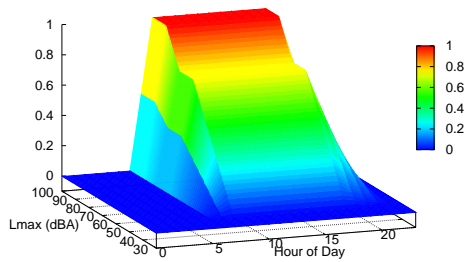


Fig. 2. Normalised Annoyance. **School**

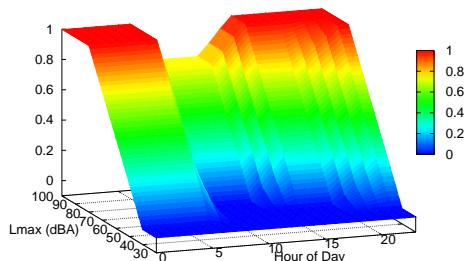


Fig. 3. Normalised Annoyance. **Residential Zone**

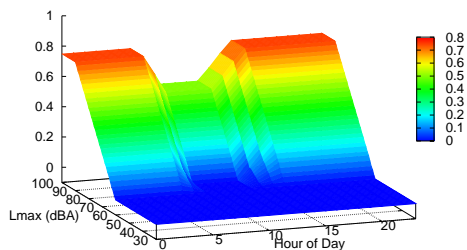


Fig. 4. Normalised Annoyance. **Industrial Zone**

B. Airliner costs

Airliner cost and Air Traffic Management (ATM) efficiency should also be taken into account when designing aircraft

trajectories. In this context, *Fuel* and/or *Time* spent during the trajectory may be considered as optimisation objectives too.

Being t_0 and t_f the initial and final time of a given trajectory, fuel cost C_f associated to this trajectory can be computed as:

$$C_f = \pi_c \cdot Fuel = \pi_c \int_{t_0}^{t_f} FF(t) dt \quad (2)$$

where π_c is the fuel price and $FF(t)$ is the total fuel flow, which in turn can be expressed in function of the current thrust setting.

On the other hand, time cost represents the different constant rate costs associated with aircraft operations (insurances, traffic control fees, crew salaries, etc). This can be easily written as:

$$C_t = \pi_t \cdot Time = \pi_t(t_f - t_0) \quad (3)$$

where π_t is the cost attached to one unit of time of delay.

Current Flight Management and Guidance Systems (FMGS) equipping a wide number of aircraft deal with a compound cost function which involves fuel and time consumption during the flight. A cost index parameter (*CI*) relates the cost of time delay to the price of the fuel and its value is carefully chosen by the operator prior to each flight. Cost index (*CI*) is defined as:

$$CI = \frac{\pi_t}{\pi_c} \quad (4)$$

Fuel saving flights are associated with low values of the cost index while more direct and faster flights are associated with high values of this index. As mentioned above, this strategy is currently used in civil aircraft operations giving optimal flight levels and speed settings for all phases of flight.

In addition, for this study it would be incomplete to consider only these magnitudes regardless of the altitude achieved at the end of the procedure. Reaching a low final altitude $h(t_f)$ would lead to small time or fuel consumption figures during the departure but the consumption would increase in the following phase, when trying to gain the altitude required to reach the optimal cruise flight level. Therefore, the final altitude must be also taken into account as an optimisation criterion to be maximised. Following the same philosophy, an *Height Index (HI)* is proposed in this work. Finally, the airliner cost compound function is defined as:

$$C_a = Fuel + CI \cdot Time - HI \cdot h(t_f) \quad (5)$$

where, by definition, $CI > 0$ and $HI > 0$.

III. THE OPTIMISATION STRATEGY

In [18], the authors presented a framework to optimise departing or approaching trajectories which can be summarised in figure 5. The involved airport, with its surrounding cartography, geography and meteorological data, will define a *scenario* which will be used to compute a given *noise nuisance* in

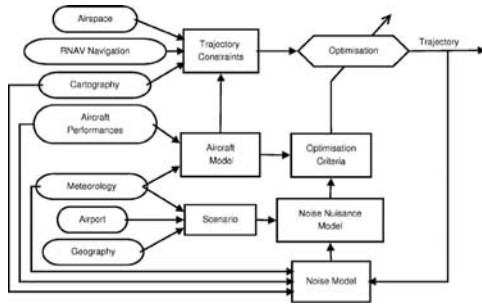


Fig. 5. Framework for the noise abatement optimisation strategy

function of the emitted aircraft noise along its trajectory. This value, together with some airliner economic considerations, will define one or several *optimisation criteria*. Then, an *optimisation algorithm* will compute the best departing or approaching trajectory minimising these criteria and satisfying a set of *trajectory constraints* which, in turn, will depend on the dynamics of the aircraft, navigation constraints and specific airspace configurations.

A. Statement of the problem

This optimisation process can be formally written as a *constrained multi-objective optimal control problem* in a given time interval $[t_0, t_f]$. In this case, the value of t_f is let *free* during the optimisation, meaning that this value is a decision variable itself and will be fixed by the optimisation algorithm. Let $\vec{x}(t) \in \mathbb{R}^{n_x}$ be the state vector describing the trajectory of the aircraft over the time t , $\vec{u}(t) \in \mathbb{R}^{n_u}$ the control vector that leads to a specific trajectory and $\vec{p} \in \mathbb{R}^{n_p}$ a set of control parameters not dependent on t . The goal is to find the best trajectory that minimises a given set of optimisation objectives (or criteria) $\vec{J} \in \mathbb{R}^{n_j}$. Namely:

$$\min_{\vec{z} \in \mathcal{Z}} \vec{J}(\vec{z}) = \min_{\vec{z} \in \mathcal{Z}} [J_1(\vec{z}), J_2(\vec{z}), \dots, J_{n_j}(\vec{z})] \quad (6)$$

where $\mathcal{Z} \subseteq \mathbb{R}^{n_x+n_u+n_p+1}$, is the admissible set of decision variables $\vec{z} = [\vec{x}(t), \vec{u}(t), \vec{p}, t_f]^T$, and $J_i(\vec{z})$ are scalar valued functions representing each individual criterion or objective.

In order to guarantee a feasible and acceptable trajectory as a result of the optimisation process presented above, several constraints must be taken into account and are summarised as:

- *dynamic constraints* describing the trajectory of the aircraft:

$$\dot{\vec{x}}(t) = \mathbf{f}(\vec{x}(t), \mathbf{u}(t)) \quad (7)$$

- *end point or event constraints* fixing the initial and final boundary conditions:

$$e_L \leq \mathbf{e}(\vec{x}(t_0), \vec{x}(t_f), t_0, t_f) \leq e_U \quad (8)$$

- *mixed state-control path constraints* allowing to restrict the behaviour of some variables:

$$\mathbf{h}_L \leq \mathbf{h}(\vec{x}(t), \mathbf{u}(t), t) \leq \mathbf{h}_U \quad (9)$$

- *box constraints on the state and control variables* allowing to bound them:

$$\begin{aligned} \mathbf{x}_L &\leq \mathbf{x}(t) \leq \mathbf{x}_U \\ \mathbf{u}_L &\leq \mathbf{u}(t) \leq \mathbf{u}_U \end{aligned} \quad (10)$$

Function f is a non linear function that contains the dynamical model of the aircraft trajectory. Vectorial functions \mathbf{e} and \mathbf{h} define the *event* and *path* constraints respectively and vectors $e_L, e_U, h_L, h_U, x_L, x_U, u_L$ and u_U are respectively the *Lower* and *Upper* values which bound all constraints. For a detailed description of these functions and vectors, please refer to [18].

B. Numerical solution of the optimisation problem

The optimal control problem described in section III, which contains differential and algebraic constraints, is transformed in two steps into a non linear programming (NLP) problem with only algebraic constraints. First, differential equations (7) are written in its equivalent integral form:

$$\vec{x}(t) = \vec{x}(t_0) + \int_{t_0}^t \vec{f}(\vec{x}(\tau), \vec{u}(\tau), \vec{p}) d\tau \quad (11)$$

Then, equation (11) is discretised using a sampling time $\Delta t = t_{n+1} - t_n$ where t_{n+1} and t_n are two consecutive time instants using an explicit numerical integration rule to approximate the above integral, as Euler or Runge-Kutta. For example, in case of using the Euler rule, the following equivalent discrete-time form is obtained:

$$\vec{x}(k+1) = \vec{x}(k) + \Delta t \cdot \vec{f}(\vec{x}(k), \vec{u}(k), \vec{p}) \quad (12)$$

Once the problem is formulated as a NLP, it can be solved using a commercial optimisation software. In this paper, the General Algebraic Modelling System (GAMS)² is the optimisation package used to code and solve the NLP problem. The numerical optimisation method used to solve the problem is a generalised reduced gradient search [19], implemented in the NLP solver CONOPT³ available in the GAMS optimisation package, which can cater for the nonlinearities of the performance index and constraints.

The CONOPT optimisation algorithm starts by finding a feasible solution; then, an iterative procedure follows, which consists of:

- finding a search direction, through the use of the Jacobian of the constraints, the selection of a set of basic variables and the computation of the reduced gradient.
- performing a search in this direction, through a pseudo-Newton process until a convergence criterion is met.

A detailed description of the CONOPT algorithm and its implementation may be found in [20] and in the manuals available at the GAMS web page.

²<http://www.gams.com>

³www.aimms.com/aimms/product/solvers/conopt.html

C. Lexicographic algorithm

A solution \bar{z}^* of the multi-objective optimisation problem, presented in equation (6), is said to be *Pareto optimal* iff there does not exist another $\bar{z} \in \mathcal{Z}$ such that $J_i(\bar{z}) \leq J_i(\bar{z}^*)$ for all $i = 1, \dots, n_j$ and $J_j(\bar{z}) < J_j(\bar{z}^*)$ for at least one index j . In other words, a solution is Pareto optimal if and only if an objective $J_i(\bar{z})$ can be reduced only at the expense of increasing at least one the other objectives. In general, there may be many Pareto optimal solutions to an optimisation problem.

Lexicographic optimisation establishes a hierarchical order among all the optimisation objectives. If such a priority exists, a unique solution exist on the Pareto hyper-surface (see [21] and the references therein).

Let the objective functions be arranged according to the *lexicographic order* from the most important J_1 to the least important J_{n_j} . A given $\bar{z}^* \in \mathcal{Z}$ is a *lexicographic minimiser* of equation (6) iff there does not exist a $\bar{z} \in \mathcal{Z}$ and a j satisfying $J_j(\bar{z}) < J_j(\bar{z}^*)$ and $J_i(\bar{z}) = J_i(\bar{z}^*)$ for all $i = 1, \dots, j-1$. An interpretation of this definition is that a solution is a lexicographic minimum iff an objective J_i can be reduced only at the expense of increasing at least one of the higher-prioritised objectives $\{J_1, \dots, J_{(i-1)}\}$. Hence, a lexicographic solution is a special type of Pareto-optimal solution that takes into account the order of the objectives. This hierarchy defines an order on the objective function establishing that a more important objective is infinitely more important than a less important objective.

A standard method for finding a lexicographic solution is to solve a sequential order of single objective constrained optimisation problems. After ordering, the most important objective function is minimised, subject to the original constraints. If this problem has a unique solution, it is the solution of the whole multi-objective optimisation problem. Otherwise, the second most important objective function is minimised. Now, in addition to the original constraints, a new constraint is added to guarantee that the most important objective function preserves its optimal value. If this problem has a unique solution, it is the solution of the original problem. Otherwise, the process goes on iteratively. More formally, the lexicographic minimum of equation (6), $\text{lex min}_{\bar{z} \in \mathcal{Z}} \bar{J}(\bar{z})$, can be found by using the following algorithm:

- 1: $J_1^* = \min_{\bar{z} \in \mathcal{Z}} [J_1(\bar{z})]$
- 2: **for** $i = 2$ to n_j **do**
- 3: $J_i^* = \min_{\bar{z} \in \mathcal{Z}} [J_i(\bar{z}) | J_j(\bar{z}) \leq J_j^*, j = 1, \dots, i-1]$
- 4: **end for**
- 5: Determine the lexicographic minimiser set as:
 $\bar{z}^* = \arg(J_{n_j}^*)$

Lexicographic optimisation permits to sort *a priori* the different optimisation criteria according to its relative importance. This method has shown several benefits in front of the classical weighting methodology [22], [23] and has been started to be widely used in control engineering applications (see, for instance [24], [21] and [25]).

TABLE II
HYPOTHETICAL SCENARIO DATA

Departing runway heading	70°
Minimum climb gradient	3.3%
Initial point coordinates	[0, 0] km
Final point coordinates	[10, 20] km
Minimum height at final point	4000 ft
Maximum height at final point	10000 ft
Cost Index (CI)	CI = 1
Height Index (HI)	HI = 0.1

We can assume that the procedure designer in charge of publishing such a departure trajectory (i.e. the decision maker of this optimisation process) has a clear idea of what prioritisations should give to each location, maybe influenced by some *political reasons*. In that case, previous algorithm leads to the best trajectory according to the desired hierarchy. In the case where this prioritisation is not clear, or when a more accurate scenario study is necessary, it is possible to run all optimisations by using all possibilities in the prioritisation order. The number of different prioritisations is $n_P = n_j!$, where n_j is the total number of noise sensitive locations. Then a *performance index* is defined aimed at choosing the best trajectory among all the possibilities.

Let J_i^* be the minimum annoyance that can be achieved at sensitive location i (i.e. when location i is in the first priority). Let J_i^P be the annoyance at location i reached with the optimal trajectory corresponding to priority P . For each priority P a performance factor Δ^P can be defined as:

$$\Delta^P = \max_i (J_i^P - J_i^*) \quad (13)$$

Then, the best trajectory, \bar{z}^* corresponds to the priority minimising this performance factor Δ^P :

$$\bar{z}^* = \arg(\min_P (\Delta^P)) \quad (14)$$

IV. APPLICATION EXAMPLE

This section presents a practical example concerning an hypothetical scenario where a departure route should be optimised.

A. Scenario description

Table II summarises the different data that define this scenario. In a departure trajectory it is enforced that speed and altitude may not decrease during all the procedure. In addition, being all trajectories below 10000 ft maximum air-speed becomes $v_{max} = 250$ Kt [6]. The chosen aircraft model corresponds to the Airbus A340-600 equipped with Trent 556 engines and operating at its Maximum Take-off Weight, ($m = 368000$ kg). Take-off is supposed to be performed with CONF3 flaps/slats configuration. The initial take-off phase going from ground level to a height 400 ft will not be considered in the optimisation process since the standard operational regulations almost restrict all degrees of freedom

TABLE III
NOISE SENSITIVE LOCATIONS

Sensitive location	Acronym	East coord.	North coord.
School	S	2000 m	15000 m
Industrial Zone	I	6000 m	25000 m
Residential Zone 1	R1	4000 m	5000 m
Hospital	H	7000 m	8000 m
Residential Zone 2	R2	6000 m	13000 m

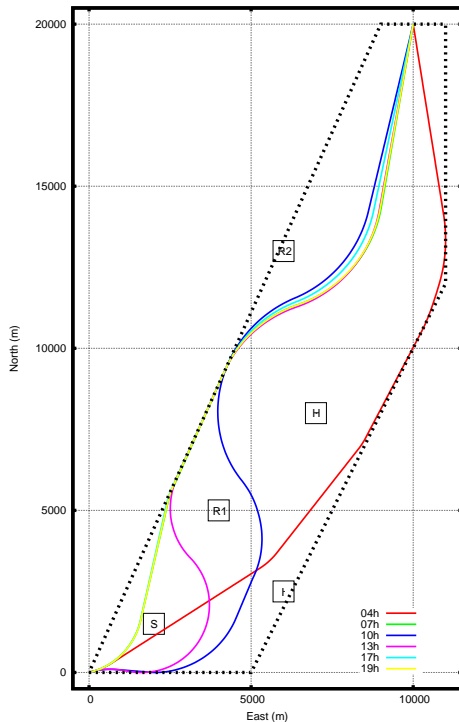


Fig. 6. Optimal trajectories at different hours of the day. **Horizontal tracks**

during this particular phase [6], [26]. In this initial phase the aircraft follows a straight trajectory, following the departing runway heading, at a constant speed (usually v_2), which depends on the aerodynamics and the actual weight of the aircraft. For this problem, and for the sake of simplicity, initial horizontal coordinates are set to zero at the point where the aircraft reaches a height of 400 ft above the runway. Moreover, during a normal take-off, the landing gear has been completely retracted when passing 400 ft so it is not considered in the simulations. Finally, five different noise sensitive locations have been located in the vicinity of the departing runway (see table III).

B. Optimal trajectories

Table IV contains the minimum annoyance values corresponding to the trajectories that minimise only one noise sensitive criterion in function of the hour of the day. In other words, these values are the best annoyance figures that can be achieved with independent single objective optimisations at each sensitive location for this particular scenario. The

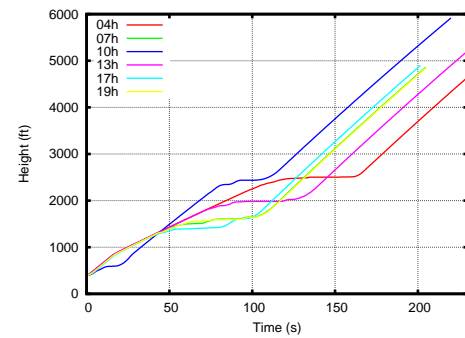


Fig. 7. Optimal trajectories at different hours of the day. **Vertical paths**

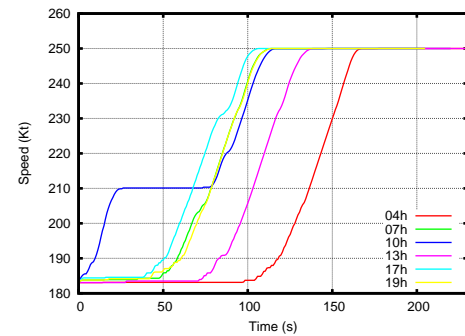


Fig. 8. Optimal trajectories at different hours of the day. **Speed profiles**

corresponding L_{max} values that produce such annoyance values are also given in the table. As it was commented in section III, in a multi-objective optimisation problem there exist multiple *Pareto-optimal* solutions regarding all objectives. Lexicographic optimisation presented in the same section allows to obtain a solution of the *Pareto* front for a given order in the optimisation objectives. Finally, by using equation (14) the “best” *Pareto-optimal* solution is chosen, according to the performance index stated in equation (13). Tables V and VI show, for different hours of the day, the prioritisation P giving the best performance index. These tables contain the annoyance values at each noise sensitive location as well as the corresponding L_{max} values. Finally, it is shown the time and fuel used in the optimal trajectory and the height reached at the end of the procedure.

Figure 6 shows the corresponding 6 optimal trajectories (flown at 04h, 07h, 10h, 13h, 17h and 19h). As it can be seen, optimal night trajectories (04h) start with a straight segment following runway heading and almost over-flying the school location (S). However annoyance in this location is zero since in night periods schools are not annoyed (see table V). This initial path allows to keep a trade-off distance to residential zone 1 (R1) and the industrial zone (I), producing a relatively low value of annoyance (0.38 and 0.21, respectively). The hospital (H) is passed following the east airspace restriction (dotted line in figure 6) producing a relatively high amount of annoyance (0.8) due to the high sensibility of this location during night periods. Finally the annoyance produced in the

TABLE IV
MINIMUM ANNOYANCE VALUES FOR EACH SINGLE OBJECTIVE TRAJECTORY. CORRESPONDING L_{max} VALUES ARE GIVEN IN PARENTHESES

Hour of the day:	04h	07h	10h	13h	17h	19h
$J_S^* = \min_{z \in Z} J_S$	0.00 (87.4 dB)	0.26 (61.1 dB)	0.53 (61.4 dB)	0.53 (61.4 dB)	0.52 (61.1 dB)	0.39 (61.1 dB)
$J_I^* = \min_{z \in Z} J_I$	0.00 (42.6 dB)	0.00 (42.6 dB)	0.00 (42.6 dB)	0.00 (42.6 dB)	0.00 (42.6 dB)	0.00 (42.6 dB)
$J_{R1}^* = \min_{z \in Z} J_{R1}$	0.13 (45.7 dB)	0.06 (45.7 dB)	0.00 (54.9 dB)	0.00 (48.6 dB)	0.00 (45.5 dB)	0.04 (45.6 dB)
$J_H^* = \min_{z \in Z} J_H$	0.32 (47.1 dB)	0.24 (47.1 dB)	0.16 (47.1 dB)	0.16 (47.1 dB)	0.16 (47.1 dB)	0.20 (47.1 dB)
$J_{R2}^* = \min_{z \in Z} J_{R2}$	0.02 (39.1 dB)	0.01 (39.1 dB)	0.00 (54.9 dB)	0.00 (48.6 dB)	0.00 (45.1 dB)	0.00 (41.5 dB)

TABLE V
OPTIMAL ANNOYANCE VALUES FOR THE BEST TRAJECTORY. CORRESPONDING L_{max} VALUES ARE GIVEN IN PARENTHESES

Hour of the day:	04h	07h	10h
Best prioritisation P :	$J_S, J_{R2}, J_H, J_{R1}, J_I, C_a$	$J_I, J_H, J_{R2}, J_S, J_{R1}, C_a$	$J_S, J_H, J_{R1}, J_I, J_{R2}, C_a$
$J_S \text{lex min}_{z \in Z} [\bar{J}^P(z)]$	0.00 (81.7 dB)	0.42 (79.7 dB)	0.53 (61.4 dB)
$J_I \text{lex min}_{z \in Z} [\bar{J}^P(z)]$	0.21 (69.0 dB)	0.01 (41.8 dB)	0.03 (68.1 dB)
$J_{R1} \text{lex min}_{z \in Z} [\bar{J}^P(z)]$	0.38 (55.3 dB)	0.30 (60.9 dB)	0.16 (66.7 dB)
$J_H \text{lex min}_{z \in Z} [\bar{J}^P(z)]$	0.80 (60.9 dB)	0.24 (47.1 dB)	0.21 (49.1 dB)
$J_{R2} \text{lex min}_{z \in Z} [\bar{J}^P(z)]$	0.02 (38.9 dB)	0.28 (59.9 dB)	0.07 (61.4 dB)
$t_f \text{lex min}_{z \in Z} [\bar{J}^P(z)]$	230 s	237 s	241 s
$h_f \text{lex min}_{z \in Z} [\bar{J}^P(z)]$	4645 ft	4787 ft	6417 ft
$Fuel \text{lex min}_{z \in Z} [\bar{J}^P(z)]$	1576 kg	1619 kg	1659 kg

TABLE VI
OPTIMAL ANNOYANCE VALUES FOR THE BEST TRAJECTORY. CORRESPONDING L_{max} VALUES ARE GIVEN IN PARENTHESES

Hour of the day:	13h	17h	19h
Best prioritisation P :	$J_I, J_S, J_{R1}, J_H, J_{R2}, C_a$	$J_I, J_H, J_{R1}, J_S, J_{R2}, C_a$	$J_I, J_H, J_{R2}, J_S, J_{R2}, C_a$
$J_S \text{lex min}_{z \in Z} [\bar{J}^P(z)]$	0.55 (62.2 dB)	0.83 (79.6 dB)	0.62 (79.7 dB)
$J_I \text{lex min}_{z \in Z} [\bar{J}^P(z)]$	0.00 (53.8 dB)	0.00 (41.6 dB)	0.00 (41.7 dB)
$J_{R1} \text{lex min}_{z \in Z} [\bar{J}^P(z)]$	0.16 (62.8 dB)	0.24 (60.2 dB)	0.33 (61.0 dB)
$J_H^* \text{lex min}_{z \in Z} [\bar{J}^P(z)]$	0.18 (47.9 dB)	0.16 (47.1 dB)	0.20 (47.1 dB)
$J_{R2} \text{lex min}_{z \in Z} [\bar{J}^P(z)]$	0.10 (59.5 dB)	0.26 (61.0 dB)	0.30 (59.8 dB)
$t_f \text{lex min}_{z \in Z} [\bar{J}^P(z)]$	201 s	212 s	212 s
$H_f \text{lex min}_{z \in Z} [\bar{J}^P(z)]$	5791 ft	5370 ft	5481 ft
$Fuel \text{lex min}_{z \in Z} [\bar{J}^P(z)]$	1402 kg	1459 kg	1458 kg

residential zone 2 (R2) is almost null (0.02) due to the high distance kept from this location. In addition, in this trajectory, the initial segment of the trajectory is also used to climb as much as possible (see figures 7 and 8 where the vertical paths and speed profiles are plotted in function of the time). This climb allows to reduce the annoyance in R1, I and H.

Best trajectory for 07h is significantly different. At this time, the school area starts to be annoyed by the over-flying aircraft

so the optimal trajectory for this hour of day starts with an immediate left turn when the aircraft reaches 400 ft above runway threshold. This left turn allows to keep the maximum distance to the residential zone 1 and the hospital locations as long as the west airspace restriction permits, producing a medium amount of annoyance (0.30 for the residential zone and 0.24 for the hospital). When the influence zone of the hospital is passed the aircraft performs a right turn improving

the annoyance at the second residential zone (0.28). At 10h the annoyance produced when over-flying a residential zone is relatively low, being just the opposite when over-flying the school. Therefore, the optimal trajectory at 10h starts with an initial right turn in order to avoid as much as possible the school location (producing a value of 0.53 of annoyance). Then, the aircraft turns left passing in between the industrial zone, the residential zone 1, the hospital and the second residential zone. It should be noted that the initial part of this trajectory is used to accelerate instead of climbing. This acceleration improves further climbing and maximises the aircraft height when approaching all the remaining locations. The best trajectory at 13h is very similar to the previous one but, at this time, the aircraft passes in between the school and the first residential zone because in the afternoon residential zones start to be more annoyed. This influence is noticed in optimal trajectories corresponding to 17 and 19. These trajectories are essentially the same as the optimal trajectory corresponding to 07h.

V. CONCLUSIONS

A technique for designing noise abatement departure procedures is presented in this paper. Noise annoyance produced by over-flying aircraft is modelled by using fuzzy logic in function of the received noise level during the trajectory, the sensibility of the areas being over-flown and the time of the day when the aircraft departure takes place. A non-linear multi-objective optimal control problem is formally written specifying the different objective functions considered. This problem is transformed to a Non Linear Programming (NLP) problem after a suitable discretisation and it is solved by using a lexicographic multi-objective optimisation technique. Finally, an application example is shown considering an hypothetical scenario with an hospital, a school, two residential zones and an industrial zone. Results show how this strategy is valid for solving this kind of multi-criteria optimisation problem, obtaining optimal trajectories that minimise nuisances in the population at different hours of the day. Work is underway extending this study including different aircraft types and different departures (different final points) corresponding to a realistic scenario of an existing airport with its surrounding features. In addition, further work will deal with the model of actual residential or industrial areas, treating them as a surface in the optimisation process and not only as a single point as it is presented in this work.

ACKNOWLEDGEMENT

The authors wish to thank the support received by the Research Commission of the Generalitat of Catalunya (group SAC, ref. 2005SGR00537).

REFERENCES

- [1] A. Muzet, "Environmental noise, sleep and health," *Sleep Medicine reviews*, vol. 11, pp. 135–142, 2007.
- [2] D. Botteldooren and A. Verkeyn, "A fuzzy rule based framework for noise annoyance modeling," *Journal of Acoustical Society*, vol. 114, no. 3, pp. 1487–1498, 2003.
- [3] Zaheeruddin, J. K. Vinod, and S. V. Guru, "A fuzzy model for noise-induced annoyance," *IEE Trans. Syst. Man. And Cyber-Part A: Systems an Humans*, vol. 36, no. 4, pp. 697–705, 2006.
- [4] Zaheeruddin and J. K. Vinod, "A fuzzy expert system for noise-induced sllep disturbance," *Expert systems with applications*, vol. 30, pp. 761–771, 2006.
- [5] L. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, pp. 338–353, 1965.
- [6] ICAO, *Procedures for Air Navigation Services - Aircraft Operations (PANS-OPS) - Volume I, Flight Procedures*, 4th ed., International Civil Aviation Organisation (ICAO), Montreal (Canada), 1993, doc. 8168-OPS/611.
- [7] H. Visser, "Generic and site specific criteria in the optimization of noise abatement procedures," *Transportation Research Part D: Transportation and Environment*, vol. 10, pp. 405–419, Sep. 2005.
- [8] H. Visser and R. Wijnen, "Optimization of noise abatement departure trajectories," *Journal of Aircraft*, vol. 38, no. 4, pp. 620–627, Jul. 2001.
- [9] R. Wijnen and H. Visser, "Optimal departure trajectories with respect to sleep disturbance," *Aerospace Science and Technology*, vol. 7, pp. 81–91, 2003.
- [10] J.-P. Clarke and R. J. Hansman, "A systems analysis methodology for developing single event noise abatement procedures," MIT Aeronautical Systems Laboratory, Cambridge, Massachusetts (USA), Tech. Rep., 1997, report No ASL-97-1.
- [11] K. Feng Zou and J.-P. Clarke, "Adaptative real-time optimization algorithm for noise abatement approach procedures," in *AIAA's 3rd Annual Aviation Technology, Integration, and Operations (ATIO) Technology Conference*, vol. 1, Nov. 2003, AIAA paper No 2003-6771.
- [12] E. M. Atkins and M. Xue, "Noise-sensitive final approach trajectory optimization for runway-independent aircraft," *Journal of Aerospace Computing, Information and Communication*, vol. 1, pp. 269–287, Jul. 2004.
- [13] M. Xue and E. M. Atkins, "Noise-minimum runway-independent aircraft approach design for baltimore-washington international airport," *Journal of Aircraft*, vol. 43, no. 1, pp. 39–51, Feb. 2006.
- [14] EUROCONTROL, "Navigation strategy for ECAC," European air traffic control harmonisation and integration programme, Tech. Rep., Mar. 1999, doc: NAV.ET1.ST16-001.
- [15] A. Verkeyn, "Fuzzy modeling of noise annoyance," Ph.D. dissertation, Gent University, Gent (The Netherlands), Feb. 2004.
- [16] FAA, *Integrated Noise Model (INM) Version 6.0 Technical Manual*, Office of Environment and Energy, Washington, DC. USA, Jan. 2002, fAA-AEE-02-01.
- [17] X. Prats, J. Quevedo, V. Puig, and F. Nejari, "Planning of aircraft departure trajectories by using fuzzy logic and lexicographic optimization," in *The 2007 Congress and Exposition on Noise Control Engineering*, Istanbul (Turkey), Aug. 2007.
- [18] X. Prats, F. Nejari, V. Puig, J. Quevedo, and F. Mora-Camino, "A framework for RNAV trajectory generation minimizing noise nuisances," in *2nd International Congress on Research in Air Transportation (ICRAT)*, Belgrade (Serbia), Jun. 2006.
- [19] A. Drud, "A GRG code for large sparse dynamic nonlinear optimization problems," *Mathematical Programming*, vol. 31, pp. 153–191, 1985.
- [20] —, "CONOPT: A large-scale GRG code," *ORSA Journal on Computing*, vol. 6, pp. 207–216, 1992.
- [21] E. Kerrigan and J. Maciejowski, "Designing model predictive controllers with prioritised constraints and objectives," *Proceedings of IEEE International Symposium on Computer Aided Control System Design*, vol. 1, pp. 33–38, 2002.
- [22] R. Marler and J. Arora, "Survey of multi-objective optimization methods for engineering," *Structural Multidisciplinary Optimization*, vol. 26, pp. 369–395, 2004.
- [23] B. Srinivasan, P. Myszkowski, and D. Bonvin, "A multi-criteria approach to dynamic optimization," in *Proceedings of the American Control Conference*, Seattle, Washington (USA), 1996.
- [24] E. Aggelogiannaki and A. Sarimveis, "Multiobjective constrained mpc with simultaneous closed-loop identification," *International Journal of Adaptive Control and Signal Processing*, 2008, (in press).
- [25] C. Ocampo, A. Ingimundarson, V. Puig, and J. Quevedo, "Objective prioritization using lexicographic minimizers for mpc of sewer networks," *IEEE Transactions on Control Systems Technology*, 2008, (in press).
- [26] JAA, *Joint Aviation Requirements. JAR-OPS: Airplane Operations*, Joint Aviation Authorities (JAA), Hoofddorp (The Netherlands), 2003.

An Artificial Intelligence Approach to Operational Aviation Turbulence Forecasting

Jennifer Abernethy

National Center for Atmospheric
Research
University of Colorado
Boulder, CO
aberneth@ucar.edu

Robert Sharman

National Center for Atmospheric
Research
Boulder, CO
sharman@ucar.edu

Elizabeth Bradley

University of Colorado
Boulder, CO
lizb@cs.colorado.edu

Abstract— Turbulence is a major aviation hazard for both commercial and private aircraft. Currently, the clear-air turbulence forecasting tool Graphical Turbulence Guidance (GTG) is used by airline meteorologists and dispatchers for flight planning, and in part to determine operational Airman's Meteorological Information (AIRMET) turbulence advisories; however, GTG has much higher resolution and intensity discrimination than do AIRMETs, providing more pinpointed locations of moderate or greater turbulence. Because numerical weather prediction (NWP) models cannot explicitly predict aircraft-scale turbulence, we use artificial intelligence (AI) algorithms to capture the relationships between large-scale atmospheric conditions and turbulence. This paper provides an overview of GTG and details beginning work for development of the next release of GTG using in-situ turbulence observation data. We apply two AI techniques, support vector machines and logistic regression, to clear-air turbulence prediction. We show improved forecast accuracy over the current product performance, and begin specializing forecasts by geographic region and altitude. We show the algorithms' feasibility as part of a real-time operational turbulence forecasting system.

I. INTRODUCTION

Pilots' ability to avoid turbulence during flight affects the safety of the millions of people who fly commercial airlines and other aircraft every year. Turbulence is a rare event in terms of percentage of the atmosphere that is turbulent at any given time [10], however, of all weather-related commercial aircraft incidents, 65% can be attributed to turbulence encounters, and major carriers estimate that they receive hundreds of injury claims and pay out "tens of millions" per year [26]. Turbulence can occur in and around thunderstorms, over mountains, near the ground, in clouds, and even in clear air. At upper levels, clear-air turbulence, or CAT, is particularly hard to avoid because it is invisible to traditional remote sensing techniques. The dynamical scales in which CAT appears, however, are far finer than those of any current weather prediction model. And observations of the state of the system – reports of 'light' or 'moderate/severe' radioed in by pilots who encounter turbulence – are sparse and subjective.

In 1998, the Federal Aviation Administration (FAA) Aviation Weather Research Program (AWRP) funded the National Center for Atmospheric Research Research Applications Lab (NCAR/RAL) to develop a graphical decision support tool, now called Graphical Turbulence Guidance (GTG), which provides clear-air turbulence (CAT)

forecasts over the continental U.S. (CONUS). GTG became operational in 2003. Meteorologists and dispatchers at the major airlines have access to GTG forecasts through the National Weather Service (NWS) Aviation Weather Center's (AWC) Aviation Digital Data Service (ADDS) website to use in planning and altering flight routes. AWC forecasters consider GTG forecasts when producing Airman's Meteorological Information (AIRMET) turbulence advisories, also available on ADDS. Future development plans include merging CAT forecasting with other forecasting products and weather information in the Joint Product Development Office (JPDO) Next Generation Air Transportation System (NextGen), a comprehensive four-dimensional weather information source for aviation decision support (information available online at <http://jpdo.gov>)

The turbulence forecasting difficulty is due to two main factors: (1) turbulent eddies at the scales that affect aircraft (~100m) are a microscale phenomenon and operational numerical weather prediction (NWP) models cannot resolve that scale (NWP models that are run operationally by the National Weather Service, for instance, which produce hourly and daily weather predictions, only capture what's happening every 10 to 20 km) and (2) lack of objective observational turbulence data. The prior factor has been addressed during the past 50 years, by assuming that most of the energy associated with turbulent eddies at aircraft scales cascades down from larger scales of atmospheric motion [9,20,28]. The turbulence forecast problem then becomes one of linking large-scale features resolvable by NWP models to the formation of aircraft-scale eddies. Numerous 'rules of thumb' empirical linkages, termed turbulence *diagnostics*, were developed by the National Weather Service, airline meteorologists and academic researchers. The forecast skills of these diagnostics depend on the forecaster (for manual forecasts) and diminish with lead time. The diagnostics' skills reflect in part researchers' imperfect understanding of the atmospheric processes involved.

The second problem is being addressed by a new, better source of turbulence observations, termed *in-situ data*. In-situ data are sensor data from aircraft: measures of atmospheric eddy dissipation rate (Cornman et al., 2004). While the study of CAT is necessarily limited to that directly experienced by aircraft since it cannot be seen, in-situ data is so much more

plentiful than PIREP observations that researchers now have enough data to explore additional AI techniques for forecasting.

This paper details the atmospheric and observational data used in turbulence forecasting, the current operational algorithm, and how new AI approaches, used both over the entire domain and regionally, in combination with new, more plentiful in-situ turbulence observations to improve operational CAT forecasting.

II. BACKGROUND

A. Turbulence Diagnostics

Through the years when forecasts were done manually, forecasters developed “rules of thumb” about what atmospheric conditions typically indicated turbulence. These rules of thumb were an attempt to link the available large-scale meteorological information to the micro-scale CAT that was the subject of the forecast [13]. Forecasters later quantified these rules, creating CAT diagnostics. A CAT diagnostic is a simple turbulence model (equation) calculated from NWP model data. For instance, a major cause of CAT is the Kelvin-Helmholtz instability: when gravity waves become steep and unstable, they may break into a chaotic motion [9]. This typically happens in areas of strong vertical shear (the difference in velocity between horizontal layers) and low local Richardson number (Ri , the ratio of static stability and wind shear), so many CAT diagnostics involve shear and Ri . There are many different diagnostics, each linking a large-scale condition to small-scale turbulence. Their predictive powers vary, depending upon the large-scale condition that each represents and how directly it is linked to turbulence. A full explanation of the forty CAT diagnostic equations can be found in [26].

Forecasters use these diagnostics by mapping their values to different turbulence severity levels. As an example, low Ri indicates high turbulence. Early on, forecasters determined some unofficial thresholds to quantify the severity of turbulence that corresponded to a given diagnostic value--- “ $Ri < 0.25 =$ moderate or greater turbulence,” for example [9]. In this way forecasters were able to transform their qualitative knowledge to a quantitative form (diagnostics) which could be used in automated systems. GTG developers used several years’ worth of PIREPs to develop threshold values for each diagnostic that map to different levels of PIREP turbulence severity. This allows the diagnostics to work neatly with the qualitative PIREP observations. Since there are many problems with PIREP accuracy, and optimal thresholds may change depending on the day or season, we hope to avoid this step in the next forecasting system. We expect that AI techniques will do this thresholding step within their algorithm and thus can either find the best thresholds themselves, and/or respond to the dynamic relationship between large-scale and small-scale atmospheric processes by adjusting these thresholds during training.

PIREP and In-Situ Data

(United 737/757)
Oct-Mar 2006/07

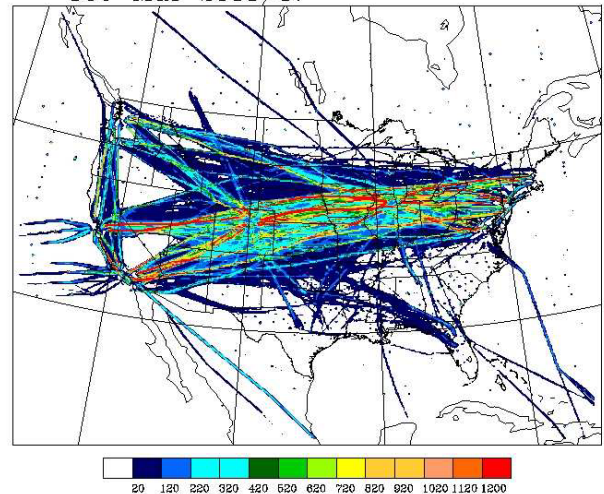


Figure 1. The PIREP and in-situ data used in this study, showing the geographic distribution of in-situ data. PIREP data are all but invisible under dense in-situ data along United flight paths; some can be seen as points in the southeast and surrounding the U.S. Color indicates frequency of observations.

B. In-Situ Data

In-situ turbulence measurements are sensor data that are recorded by special software on commercial aircraft every minute during flight. Detailed coverage of in-situ data methods can be found in [6] and [7]. Specifically, in-situ measurements are an estimate of atmospheric turbulence intensity called the eddy dissipation rate (EDR) around an aircraft. Eddies are irregular currents of air, and the rate at which eddies break down is recognized as a good measure of atmospheric turbulence intensity [23]. Compared to PIREPs, in-situ data are more objective, more accurate, more plentiful, and more representative of the actual distribution of turbulence in the atmosphere [8, 26]. At any one time, over 99% of the atmosphere is expected to be free of turbulence [10].

Currently, in-situ measurements of EDR are being gathered from 197 United Airlines aircraft. Several other airlines will deploy the same system in the coming year. The in-situ data used in this study, from October-March 2006/7, is shown in Figure 1. Each in-situ data report is a location triple (latitude, longitude, altitude) and a median and peak (95th percentile) EDR reading from measurements taken over the corresponding minute. Each of the two EDR fields is binned and the two binned values are transmitted to the ground. The binning turns otherwise continuous quantitative observation data into a set of eight discrete values that are cognate to the eight PIREP intensity levels. Currently, we consider bin 4 to correspond to a ‘moderate’ PIREP, although study to better qualitatively understand in-situ data is ongoing [2].

C. Performance Metrics

It is not trivial to assess the accuracy of a forecast because we do not know the ‘truth’; we must use available observation

data, however flawed or irregular. We follow the verification practices of [27], which include the Receiver Operating Characteristic (ROC) curve and area under the ROC curve (AUC) [22], and True Skill Score (TSS). A ROC curve measures how well an algorithm discriminates between Moderate or Greater (MOG) and less than moderate (LTM) turbulence. To construct the curve, we vary the threshold that separates these two classes over a range of 0 to 1 and measure the discrimination accuracy at each threshold. Two numbers are used to capture this: PODY, “probability of detecting a yes” (forecast made a correct positive (MOG) prediction), and PODN, which corresponds to a correct negative (LTM) prediction. Higher the PODY/PODN combinations over the range of thresholds implies greater classification accuracy, so the area under the curve (AUC) is a useful single-number metric for forecast accuracy. The TSS considers PODY and PODN at one threshold: $TSS = PODY + PODN - 1$. As mentioned in subsection B, our threshold for this study is in-situ bin 4 and ‘moderate’ PIREPs: bin 4 and above constitute the MOG category, and below bin 4 constitutes the LTM category.

D. Graphical Turbulence Guidance System

The GTG forecasting product produces a graphical display of turbulence severity for each flight level, FL100 to FL450, over the CONUS, for zero, six, nine and 12 hour forecasts, updated every hour. Displays of the operational product (GTG1) are available in real-time on the ADDS website, <http://adds.aviationweather.noaa.gov>. The newer version, GTG2, is available on the experimental ADDS website, <http://weather.aero>. An example is shown in Figure 2, with the AIRMET forecast for comparison of forecast specificity.

Every hour, the algorithm calculates ten diagnostics from the National Center for Environmental Prediction’s (NCEP) Rapid Update Cycle (RUC) model at 20km resolution for that hour [3]. These diagnostics are paired with incoming PIREPs from a time window around the RUC time, usually 1.5 hours. A fuzzy logic technique scores each diagnostic based on its agreement with the observation data, deriving a set of weights such that the weighted sum of the diagnostic values is between zero and one. The scoring function, per diagnostic over the CONUS, incorporates TSS and percentage of the CONUS forecast as MOG turbulence (f_{MOG}):

$$\phi_n = \left(\frac{TSS + 1.1}{1 + C f_{MOG}} \right) \tag{1}$$

Currently, C=1. From the n scores (one for each diagnostic over the entire forecast area), weights are formed:

$$W_n = \phi_n^2 \tag{2}$$

$\sum_{m=1}^n W_m = 1$. The diagnostics are combined into a weighed sum to form the GTG combination, an estimated turbulence intensity for every grid point:

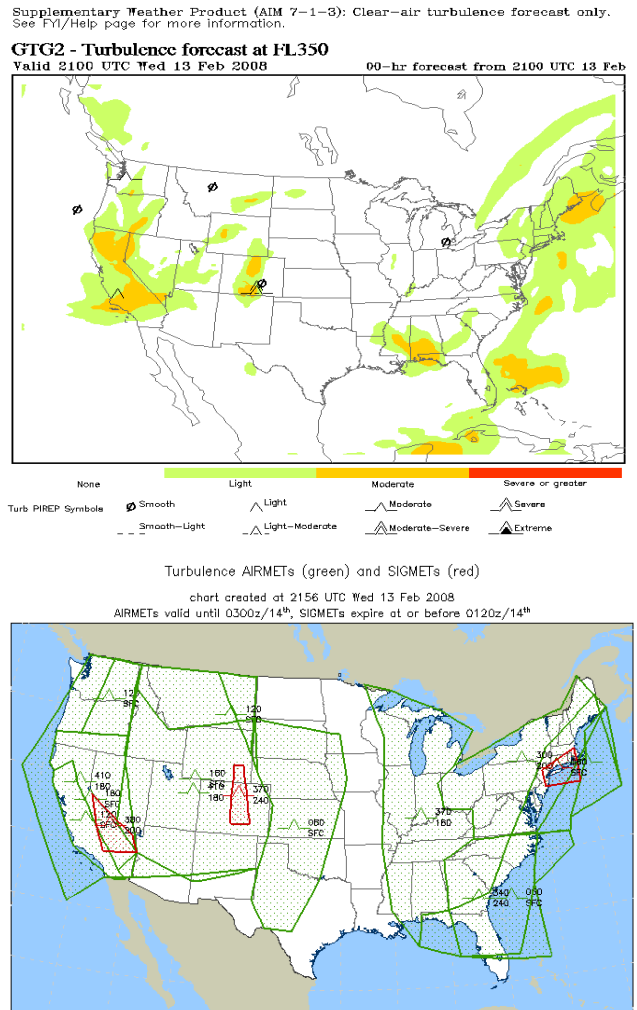


Figure 2. An example of a GTG2 forecast display (above) which designates light, moderate, severe and extreme CAT per hour for each flight level, and AIRMET (below), in green, which show much larger polygon areas of moderate turbulence for a corresponding 6-hour period over all flight levels. Both displays are from <http://weather.aero>.

$$GTG(i, j, k) = \sum_{m=1}^n W_m D_{m, i, j, k} \tag{3}$$

The set of weights is applied to the RUC diagnostics calculated from the 6, 9 and 12 hour RUC forecast. This dynamic weighting allows GTG to respond to changing conditions every hour. While the fuzzy logic handles sparse PIREP observation data well, the new, plentiful in-situ data allows for more choices in prediction algorithms.

E. Artificial Intelligence Techniques

Generally, a classifier is an algorithm that predicts a data classification given (presumably) relevant data features. The Support Vector Machine (SVM) is a popular machine learning technique for classification. The SVM produces a model that predicts the class label by setting parameter values of an optimization problem based on its input data [15]. Here, class

labels are MOG turbulence, and LTM turbulence (see subsection C).

In order to learn the relationships (parameter values) between these data features and the class label, we first train a classifier by giving it many known feature/class pairs. Each pair in the training set is known as a data instance. A data instance k consists of a set of features $x_{i,k}$ $i=1\dots n$ and a target class label y .

During training, each feature vector X_k is mapped into a higher dimensional space. The SVM finds a linearly separating hyperplane with the maximal margin between class means in this higher dimensional space. To classify an example, the SVM calculates the distance of that example to each class mean through a series of dot products, and classifies it in whatever class has the closest mean [5]. This series of dot products is at the heart of the model and is a measure of vector similarity called a kernel function:

$$K(x_i, x_j) = \phi(x_i^T) \phi(x_j) \quad (4)$$

For implementation of the SVM, we use the LibSVM library [4]. LibSVM provides four basic kernels and an optional program, "grid.py", which selects the model (i.e., does a parameter search). Previous studies [1,2] show good performance for the radial basis function kernel:

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \gamma > 0 \quad (5)$$

The radial basis function kernel only has two parameters: γ and C , a penalty parameter for the SVM error term. A model can output probabilities for class membership, also.

A logistic regression equation is solved iteratively over the training set, determining a prediction equation with a parameter, or weight, for each feature. The response variable, between zero and one, is the log odds that the class label is one. We can interpret the log odds as a probability, and decide that $P(y) \geq 0.5$ should be classified as MOG. Background on logistic regression can be found in [14].

III. METHODOLOGY

Our initial application of AI techniques to operational turbulence prediction consisted of testing Support Vector Machine (SVM) and logistic regression algorithm performance over our entire prediction domain, the CONUS. For each algorithm, for both zero-hour and six-hour forecasts, we used a subset selection search to pick a subset of CAT diagnostics which together had the highest forecast accuracy. We then tested the performance of each model in a simulated operational real-time system using either a static model for each hour's forecast or dynamic training of the model using the previously-chosen subset. We've taken the first steps to make specific forecasts, with specific sets of diagnostics, for different regions of the CONUS. The following subsections summarize our data and methods.

A. Data

Data used in the current development stage consist of weather model and observation data – both PIREPs and in-situ data – from October through March 2005/6 and 2006/7, shown in Figure 1. The weather model is the RUC NWP model at 13km resolution, run operationally and disseminated every hour by the National Center for Environmental Prediction (NCEP) [3]. RUC model data was used to calculate forty CAT diagnostics for each RUC model grid point and observation data was matched by time and location to the forty diagnostics for a grid point. Since we are primarily interested in predicting CAT, an upper-level phenomenon, only matches above 20000ft were used.

The distribution of the data used during the training process is a very important factor in the ability of a classifier to discriminate between the two classes [17]. The winter 2006/7 data set, as an example, contained nearly nine million observation/diagnostics matches. Over 98% were LTM turbulence. SVMs, for instance, aim for the lowest overall error rate, and could simply classify everything as LTM and have a less than 2% error. This is well-supported in the literature [16, 29, 30]. To work well, the training data set must have a large number of examples from each class; we rebalanced the training data such that 40% of the data were of MOG, and 60% were LTM. We did this by keeping all the MOG observations and choosing LTM observations randomly to be 60% of the set. This proportion of MOG/LTM resulted in the best SVM classification rate in an earlier study of SVMs with CAT diagnostics and in-situ data [1]. We found 20% MOG and 80% LTM to be a good distribution for logistic regression training data.

Analysis of the data reveals that PIREPs dominate the MOG category (>92%) – partly due to inadequate special coverage of in-situ data at this time – and in-situ data dominates the LTM category (>98%). Thus, PODY is effectively a measure of the algorithm's ability to predict PIREPs and PODN is a measure of in-situ prediction capability. Since in-situ data is more objective, we know using only in-situ data to train the algorithm improves performance [2]. However, our forecasting product will be verified using PIREPs, at least in part, and they cover more geographical area than do in-situ data, thus we cannot abandon them yet.

B. Subset Selection Search

Turbulence forecasting, in its current state, is essentially the task of classifying atmospheric indicators of turbulence: the forecast reflects the number of diagnostics which indicate turbulence in an area. While it might seem obvious to simply use the individually best-performing diagnostics for forecasting, as was done with GTG, that approach allows one to possibly miss a different set of diagnostics that might perform better, as a group, than the set of the individually top-ranked diagnostics [12, 19, 20]. Our search for the best subset of diagnostics is essentially the task of feature subset selection [12]. We are faced with the choice between 40 diagnostics, knowing that some may not improve our current forecasting accuracy. In addition, it is infeasible to calculate and use all 40 in a real-time operational system. The wrapper method in feature subset selection executes a state space search for a good

feature subset, estimating prediction accuracy using an induction algorithm – here, we used SVMs or logistic regression [19], with TSS as a final scoring metric. We used a simple hillclimbing search. Each state is a subset of diagnostics, and the search operator is “add a diagnostic”. The search chooses the best addition to the current subset based on the classification performance of the induction algorithm using the current subset plus an additional diagnostic. This approach to the search is called forward selection. Thus, we start with an empty subset and added diagnostics stepwise; our stopping condition was no further classification performance improvement (measured by no change in TSS). Searches were performed for SVM and logistic models for both zero and six-hour forecasts using training, testing and holdout data sets from 18Z over winter 2006/7.

C. Simulated Real-Time System

We have created a simulated real-time forecasting system capable of using either SVMs or logistic regression to create a turbulence forecast every hour for the CONUS (like GTG). The system is capable of training a model for every forecast hour or using a pre-trained model so that we may test performance differences between dynamic and static weighting, respectively. For both, we use the sets of diagnostics found in the searches explained in subsection B. Results are from trials over the fifteen day period of 2/1/2007 to 2/15/2007. Thus far, we have concentrated on zero-hour forecasts in this step.

We had to take several steps to make the SVM algorithm feasible for a real-time system. Since LibSVM uses ASCII files, 13km-resolution gridded RUC data caused each forecast to take over an hour. To streamline processing, we built a NetCDF file format interface onto the library. We also replaced the exponential function with an approximation. Both changes cut the forecast time down to a more operationally-appropriate five minutes.

D. Regionalization

Thus far, we have conducted regionalization studies using SVMs only, on winter 2005-2006 data. We employed subset searches for each of these regions: west of 100W meridian, east of 100W, above and below 30000ft (20000 to 30000ft), and by both geography and altitude (e.g., east of 100W and below 30000ft: low east). We plan to further refine and divide regions

TABLE I. AREA UNDER THE CURVE, TRUE SKILL SCORE, AND SUBSET SIZE RESULTS FOR FEATURE SELECTION SEARCHES AND 0-HR 15-DAY REAL-TIME SIMULATION RUNS. GTG SKILLS FOR THE SAME DATA ARE IN ITALICS. HIGHER TSS AND AUC INDICATE GREATER SKILL.

	AUC	TSS	Subset size
<i>GTG 0hr fcsts</i>	<i>0.795</i>	<i>0.390</i>	<i>10</i>
Log search 0hr	0.801	0.478	13
SVM search 0h	0.7825	0.471	8
<i>GTG 6hr fcsts</i>	<i>0.78</i>	<i>0.366</i>	<i>10</i>
Log search 6hr	0.79	0.467	6
SVM search 6h	0.78	0.4643	12
<i>GTG 0-hr 15days</i>	<i>0.799</i>	<i>0.350</i>	<i>10</i>
Log static 0-hr	0.823	0.466	13
SVM static 0-hr	0.796	0.459	8
Log dyn. 0-hr	0.786	0.45	13
SVM dyn. 0-hr	0.775	0.464	8

TABLE II. INITIAL REGIONALIZATION RESULTS USING SVM FOR WINTER 2005/6. THE GTG TSS FOR THIS PERIOD WAS 0.453; EVEN ARBITRARY REGIONALIZATION SHOWS IMPROVEMENT WITH SPECIALIZED SETS OF DIAGNOSTICS.

Region	TSS	Set of Diagnostics
West	0.465	6
East	0.562	4
>=30000ft	0.447	5
<30000ft	0.607	5
High west	0.441	4
High east	0.516	4
Low west	0.614	5
Low east	0.519	2

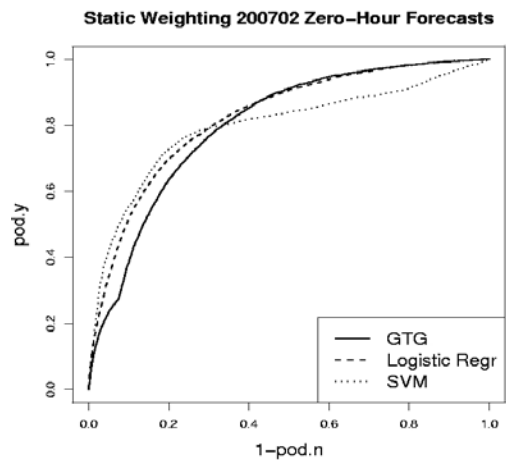


Figure 3. Receiver operating characteristic (ROC) curves comparing performance for 15-day real-time simulation of 0-hr forecasts using static weighting. The solid line is the current GTG performance for the same 15-day period. Lines closer to top left corner indicate better forecasting performance. See Table 1 for areas under the curves.

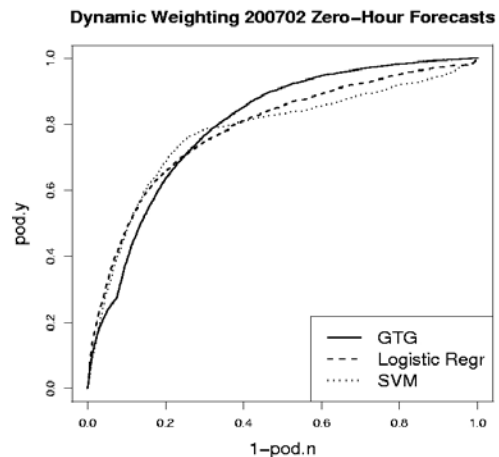


Figure 4. Like Figure 3, for dynamically-trained models.

in the near future, but for this study, we have simply isolated the mountainous terrain in the west region from the non-mountainous terrain in the east.

IV. RESULTS

Results of our forward selection subset searches and real-time simulations are shown in Table 1. While we do not list the exact diagnostics chosen by each search for the sake of simplicity, we did find that there was significant — though not complete — overlap in the diagnostics chosen by each search, indicating high predictive capability for a core subset of four or five diagnostics. Logistic regression shows a small improvement in AUC over the overall performance of the current GTG algorithm for both 0 and 6-hr forecasts (about a 0.01 difference), however, the true-skill scores (TSSs) for both algorithms are significantly improved over GTG (0.09 – 0.1 improvement). This is most likely due to the fact that our search used TSS as the heuristic to choose the sets of diagnostics.

Figures 3 and 4 show the ROC curves for our static- (model trained in the search step is applied to data from each hour) and dynamic-weighting (new model is trained every hour) 15-day real-time simulations. It should be noted that GTG has been tuned using years of PIREPs, thus its PODY scores are highest (since PIREPs dominate the PODY category). Logistic regression using pre-determined (static) weights improves significantly upon the current GTG product, increasing the AUC from 0.799 to 0.823 and the TSS from 0.350 to 0.466. While the static-weighting SVM and both dynamically-weighted models had similar improvements in TSS over GTG, we saw no improvement in AUC (mainly due to reduced prediction skill for the MOG category). TSS is discrimination skill at the MOG threshold, 0.375; AUC measures classification skill at many thresholds. Thus, we have improved forecasting performance at the operational MOG threshold, although the ROC curves show us that there is still need for improvement in the algorithms overall.

Our initial regionalization results are shown in Table 2. For winter 2005/2006, the baseline TSS for the GTG algorithm was 0.453; thus, most regions showed improvement in forecasting accuracy when using chosen subsets of diagnostics. In addition, the fact that different diagnostics were chosen in the different regions indicate that diagnostics can perform differently in different areas of the country, reflecting the geographic differences in the large-scale atmospheric processes they represent and good potential for the regionalization approach.

V. CONCLUSIONS

Forecasting clear-air turbulence is critical to aviation safety. AI techniques can be very useful in meeting the challenges inherent in this process because they smoothly handle sparse, noisy data sets, significant levels of uncertainty, and gaps in the understanding of the underlying physical mechanisms, all of which are characteristics of the turbulence prediction domain.

This paper has detailed the first steps in applying the artificial intelligence techniques of support vector machines and logistic regression to clear-air turbulence forecasting. While the current GTG product uses fuzzy logic, the algorithmic choices were limited by the sparse PIREP observation data; now, the more objective and plentiful in-situ data vastly widens the choices for prediction algorithms. We've shown not only improvement in forecasting

performance both globally and regionally, but also the feasibility of implementing these AI algorithms in a real-time operational product setting. Future work includes continued study of these algorithms for regionally-specific forecasting and probabilistic forecasting.

ACKNOWLEDGMENT

This research is in response to requirements and funding by the FAA. The views expressed are those of the authors and do not necessarily represent the official policy or position of the FAA.

REFERENCES

- [1] Abernethy, J., 2005: Domain Analysis Approach to Clear-Air Turbulence Forecasting Using In-situ Data. Dissertation Proposal, Department of Computer Science, University of Colorado.
- [2] Abernethy, J., Bradley, E.; and Sharman, R. 2006: Qualitative reasoning about small-scale turbulence in an operational setting. In Proceedings of the Qualitative Reasoning Workshop. Hanover, NH.
- [3] Benjamin, S. G., G. A. Grell, J. M. Brown, T. G. Smirnova, and R. Bleck, 2004: Mesoscale weather prediction with the RUC hybrid isentropic-terrain-following coordinate model. *Mon. Wea. Rev.*, **132**, 473-494.
- [4] Chang, C. and C. Lin. LIBSVM – a library for support vector machines. <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [5] Chen.P., C. Lin and B. Scholkopf, 2003: A tutorial on v-support vector machines. <http://kernel-machines.org>.
- [6] Cornman, L. B., C. S. Morse, and G. Cuning, 1995: Real-time estimation of atmospheric turbulence severity from in-situ aircraft measurements. *J. Aircraft*, **32**, 171-177.
- [7] Cornman, L., G. Meymaris, and M. Limber, 2004: An update on the FAA Aviation Weather Research Program's in-situ turbulence measurement and reporting system. Preprints, *Eleventh Conf. on Aviation, Range, and Aerospace Meteorology*, Hyannis, MA, Amer. Meteor. Soc., P4.3.
- [8] Dutton, J., 1980: Probability forecasts of clear-air turbulence based on numerical output. *Meteor. Mag.*, **109**, 293-310.
- [9] Dutton, J., and H. A. Panofsky, 1970: Clear Air Turbulence: A mystery may be unfolding. *Science*, **167**, 937-944.
- [10] Frehlich, R., and R. Sharman, 2004a: Estimates of turbulence from numerical weather prediction model output with applications to turbulence diagnosis and data assimilation. *Mon. Wea. Rev.*, **132**, 2308-2324.
- [11] Frehlich, R., and R. Sharman, 2004b: Estimates of upper level turbulence based on second order structure functions derived from numerical weather prediction model output. Preprints, *Eleventh Conf. on Aviation, Range and Aerospace Meteorology*, Hyannis, MA, Amer. Meteor. Soc., P4.13.
- [12] Guyon, I. and A. Elisseeff, 2003: An introduction to variable and feature selection. *J. Machine Learning Research*, **3**, 1157-1182.
- [13] Hopkins, R. H., 1977: Forecasting techniques of clear-air turbulence including that associated with mountain waves. WMO Technical Note No. 155, 31 pp.
- [14] Hosmer, D. and S. Lemeshow. 1989. *Applied Logistic Regression*. John Wiley and Sons, Inc.
- [15] Hsu, C., C. Chang and C. Lin, 2003: A practical guide to support vector classification. Published online with Libsvm documentation at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [16] Japkowicz, N., 2000: Learning from imbalanced data sets: a comparison of various strategies. *AAAI Workshop on Learning from Imbalanced Data Sets*, Menlo Park, CA.

- [17] Kay, M., J. Henderson, S. Krieger, J. Mahoney, L. Holland and B. Brown, 2006: Quality assessment report: Graphical turbulence guidance (GTG) version 2.3.
- [18] Kohavi, R., and D. Sommerfield, 1995: Feature subset selection using the wrapper method: overfitting and dynamic search space topology. *First International Conference on Knowledge Discovery in Data Mining (KDD-95)*.
- [19] Kohavi, R. and G. John, 1997: Wrappers for Feature Subset Selection. *J. Artificial Intelligence*, **97**, no1-2, 273-324.
- [20] Koshyk, J. N., and K. Hamilton, 2001: The horizontal energy spectrum and spectral budget simulated by a high-resolution troposphere-stratosphere-mesosphere GCM. *J. Atmos. Sci.*, **58**, 329-348.
- [21] Kronebach, G. W., 1964: An automated procedure for forecasting clear-air turbulence. *J. Appl. Met.*, **3**, 119-125.
- [22] Marzban, C. 2004: The ROC Curve and the Area Under It as a Performance Measure. *Weather and Forecasting*, Vol. 19, No. 6, 1106-1114.
- [23] Panofsky, H and J. Dutton, 1983: Atmospheric turbulence: models and methods for engineering applications. John Wiley & Sons.
- [24] Schwartz, B., 1996: The quantitative use of PIREPs in developing aviation weather guidance products. *Wea. Forecasting*, **11**, 372-384.
- [25] Sharman, R., G. Wiener and B. Brown, 2000: Description and verification of the NCAR integrated turbulence forecasting algorithm. *Proceedings of the 38th Aerospace Sciences Meeting and Exhibit, Reno, NV*.
- [26] Sharman, R., C. Tebaldi, G. Wiener and J. Wolff, 2006: An Integrated Approach to Mid- and Upper-Level Turbulence Forecasting. *Weather and Forecasting*.
- [27] Takacs, A., L. Holland, R. Hueftle, B. Brown and A. Holmes, 2005: Using in-situ eddy dissipation rate (edr) observations for turbulence forecast verification.
- [28] Tung, K. K., and W. W. Orlando, 2003: The k^{-3} and $k^{-5/3}$ energy spectrum of atmospheric turbulence: Quasigeostrophic two-level model simulation. *J. Atmos. Sci.*, **60**, 824-835.
- [29] Weiss, G. and F. Provost, 2001: The effects of class distribution on classifier learning: an empirical study. *Technical Report ML-TR-44, Department of Computer Science, Rutgers University*.
- [30] Wu, G and E. Chang, 2005: KBA: Kernel boundary alignment considering imbalanced data distribution. *IEEE Transactions on Knowledge and Data Engineering*.

An Environmental Airport ATM Modelling Support Tool and its use in Stacking and CDA Scenarios

Alexander Goman
Helios Technology
Farnborough, UK
alex.goman@askhelios.com

David Atkins
Systems Engineering Innovation Centre
Loughborough University
Loughborough, UK

Abstract— Air Traffic Managers increasingly need to consider environmental impacts when planning future operations or reviewing current procedures, particularly in relation to noise and emissions. In response to this need an environmentally aware Air Traffic Management (ATM) modelling tool has been designed and implemented in the context of the Environmentally Friendly Airport ATM Systems (EFAS) Project. This paper focuses specifically on the support environment provided by the ATM modelling tool and how it was used to inform the decision making process in an example case study examining the impact of various amounts of stacking on the environmental efficiency of Continuous Descent Approaches (CDAs).

It is found in a pessimistic scenario, (where no delay is absorbed ‘up-stream’), traffic arriving at a medium sized UK airport subjected to increasing traffic levels, (from 2004 out to 2030), experience exponentially growing stacks. In a 2015 timetable scenario, for example, stacks are found to generate approximately 11% more CO₂ and 5% more NO_x than top of descent CDAs alone. This finding underlines, from an environmental perspective, the need for the use of advanced ATM techniques such as Airborne Separation Assistance Systems (ASAS), Arrival Management tools (AMAN) and Collaborative Decision Making (CDM), to produce a set of efficient, de-conflicted, flight movements.

Keywords - *Modelling, Stacking, Continuous Descent Approaches, Air Traffic Management, Environment, On demand delay.*

I. INTRODUCTION

A. Background

NATS, the UK’s Air Navigation Service Provider, (ANSP), has forecasted that by 2015 its air traffic will have increased by 45%. Environmental impacts, if not addressed, will lead to unsustainable growth and constrain the development of airports. Consequently airport stakeholders need to make an informed exploration of the procedural and technological ATM options at their disposal to help the airport meet increasingly stringent environmental targets [1].

B. The challenge of developing Environmentally Friendly Airport ATM

It is the convergence of the air traffic network at its airports where some of the most intricate logic in ATM comes into play. This is true of how the traffic is managed and what metrics are used to measure its performance (principally the capacity it provides versus the emissions and noise it produces) [2].

One can develop hypotheses for reducing environmental impact but the testing and validation of these concepts is not always practical due to the cost, time and equipment issues involved. Therefore, when examining the impact of radical operational changes at a system wide level, (both for present and future scenarios), other methods are required to explore and rationalise the trade-offs involved. Computational models, (sometimes referred to as Synthetic Environments), are ideal for testing such hypotheses as they can realise the consequences of solutions without the cost of actual implementation. This helps to evaluate the properties or behavior of a potential solution, thereby allowing scenarios to be optimised to the requirements of the project. This process not only informs the planning and decision making process; it can also help to evolve the solution space itself.

II. THE EFAS PROJECT

The Environmentally Friendly Airport ATM Systems (EFAS) Project (see acknowledgments for details) identified a number of candidate ATM technologies and systems that were expected to reduce the environmental impact of the growth in air traffic. To assess the proposed operational improvements of these ‘solutions’ a modelling environment was developed.

The EFAS Model is a purpose built simulation environment designed to capture the environmental impact of an operational ATM system. It includes detailed information about aircraft movements, the noise they generate and the chemical emissions they produce. This environment is an integrated implementation of a bespoke trajectory model that utilises Eurocontrol’s BADA (Base of Aircraft Data) model 3.6 [3], the

FAA’s INM (Integrated Noise Model) 6.0 [4] and the QinetiQ Emission Model (the latter uses Boeing Fuel Flow Methodology and was previously used in [5]).

Each ‘solution’ is assessed by modelling one or more scenarios. Each scenario represents a day of operations at a medium sized UK airport; it considers aircraft taxi, landing and take off (LTO) operations up to 20,000ft (although emissions calculations were capped at 10,000ft due to processing constraints). A scenario is defined by a number of configuration files, each describing an aspect of an ATM system. One such file is the arrivals timetable; one was created for each of the scenario years simulated (a busy day in 2004, 2010 and then out to 2030 in 5 year steps) by Cranfield University (see Fig. 1). Future aircraft types are mapped to existing types in BADA that had emissions and noise characteristics judged to be nearest. Linear reductions are applied to future aircraft type’s fuel burn and emissions to account for future aircraft type efficiencies.

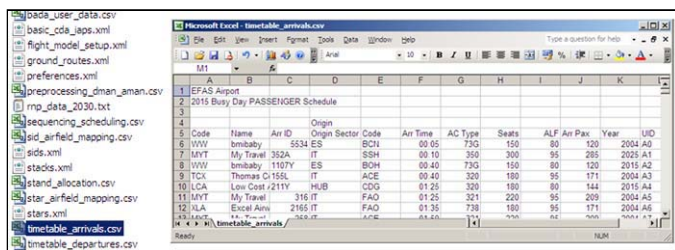


Figure 1. A number of scenario configuration files, with a preview of the arrival timetable configuration file.

Due to the processing power required to model a large number of scenarios, the EFAS Model is hosted at a QinetiQ data processing facility. As the facility is a shared resource, there were only a limited number of batch execution runs of the Model available to the EFAS project. Prior to a batch execution, each partner was invited to submit a collection of zip files to QinetiQ, with each file describing a model configuration designed to emulate a particular scenario.

Constructing scenarios is a complex and technical process, prone to error as it requires a high level of planning and forethought. As many of the configuration files are complex and dense, syntactical and semantic errors can be easily made. Without a tool to support the construction of scenarios any errors in configuration files will only be identified when the scenario is processed by the EFAS Model as they would result in execution failure. As use of the model is limited, such an execution failure is an undesirable waste of a limited resource.

Furthermore, as a large number of scenarios were planned for each batch execution (up to 100) transferring scenario zips between the EFAS consortium partners and the QinetiQ facility presented a technological and logistical challenge. This was further complicated due to the fact that once results were added to scenario zips the files became too large to email.

Therefore to overcome the problems discussed a scenario builder tool was developed to support the construction, submission and subsequent analysis of scenarios.

A. The EFAS Scenario Builder Tool

The Scenario Builder Tool enables the use of the EFAS Model using an access anywhere web-based interface. As the tool is web-based, it is platform independent, the user is not required to install any software and it could be securely accessed anywhere¹.

1) Scenario Construction

The user can create an unlimited number of scenarios using the tool. When first accessing the tool, the user is presented with a list of all existing scenarios and their current state, such as ‘Under construction’, ‘Ready for processing’ and ‘Results available’ (see Fig. 2).

Scenarios created by group 'HELIOS'	Status	Updated	Created By
HELIOS_EE_FF_B_3_1 (5, 3)			
HELIOS_EE_FF_2004_B_3_1	Results Available	22:45 18/10/07	HELIOS
HELIOS_EE_FF_2015_B_3_1	Results Available	22:34 18/10/07	HELIOS
HELIOS_EE_FF_2020_B_3_1	Ready for processing	14:20 12/02/08	HELIOS
HELIOS_EE_FF_2025_B_3_1	Under Construction	14:20 12/02/08	HELIOS
HELIOS_EE_FF_2030_B_3_1	Results Available	22:00 18/10/07	HELIOS

Scenarios with results available	Status	Updated	Created By
HELIOS_EE_FF_B_3_1 (3, 3)			
HELIOS_EE_FF_B_3_2 (3, 3)			
HELIOS_EE_FF_B_3_3 (3, 3)			

Figure 2. Scenario list filtered for batch run 3.

The tool allows a scenario to inherit all or part of its configuration from other scenarios built using the tool, encouraging reuse of configuration files between partners and ensuring maximum comparability between different scenario results. A number of default or ‘baseline’ scenarios were provided for each modelled day.

Configuration of a scenario is split into a number of functional sections, such as Airspace Design, Timetable & Aircraft, Airport & Weather, etc. For each individual configuration file the user can upload a custom version of the file, preview the file’s contents directly in the browser, use the default version of the file or inherit a version of the file from an existing scenario.

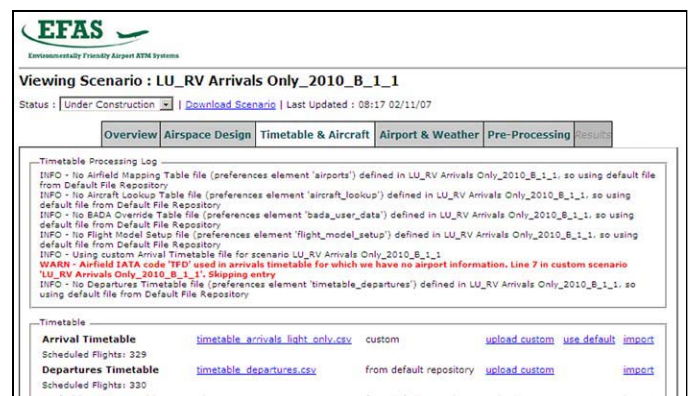


Figure 3. Configuring a scenario.

¹ Access to the tool required a username and password and all transmissions to and from the server were SSL encrypted.

Automated error checking validates a modelling scenario's configuration for syntactic and semantic errors, ensuring that errors are caught at the development stage - before the modelling scenario is processed by the EFAS Model. Any errors are provided to the user in an easy to understand in-context manner, (see the red text in Fig. 3).

2) Job Submission

Once the construction of the scenario has been completed, the user can set the modelling scenario as 'Ready to process'. The tool provides a facility to allow bulk download of all scenarios marked as 'Ready to process'.

3) Result Delivery

When modelling results are available for one or more modelling scenarios, they can be uploaded to the Scenario Builder Tool. Results are then accessible through each modelling scenario's 'results' tab (see the far right tab in Fig 3).

B. Scenario Assessment Lifecycle

The Scenario Assessment Lifecycle (Fig. 4) shows how the Scenario Builder Tool is used to analyse, assess and iteratively develop a solution in EFAS.

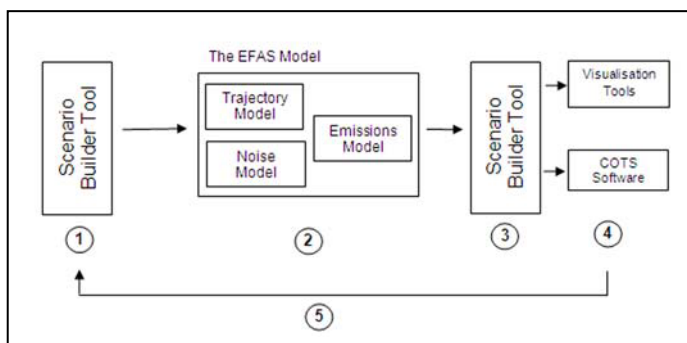


Figure 4. The Scenario Assessment Lifecycle.

1. The Web Based Scenario Builder Tool is used to describe a scenario. When a scenario description is complete, the user can change the scenario's state from 'Under Construction' to 'Ready to Process'.
2. As use of the facility that hosts the EFAS Modelling Environment is limited, scenarios are processed in batches. On a number of fixed dates, all scenarios with a state of 'Ready to Process' are downloaded from the Scenario Builder Tool and processed by the EFAS Modelling Environment. The results of the modelling are then uploaded to the Scenario Builder Tool.
3. The user can now access the modelling results for each scenario using the same interface used to describe the scenario.
4. Commercial off the shelf and bespoke external tools can now be used to analyse the modelling results.
5. Using the findings from scenario analysis, a scenario can be refined and re-processed.

An example case study examining the impact of holding on Continuous Descent Approaches (CDA) that made full use of the Scenario Builder Tool is now illustrated.

III. EXAMPLE CASE STUDY: THE IMPACT OF AIRBORNE HOLDING ON CDA

A. The use of holding and path-stretching in ATM

Efficient airport ATM operations rely on the timely deployment of optimal aircraft operating principles during approach, landing, taxi, take off and departure. An efficient approach to landing is one that yields an effective loss of the aircraft's potential energy as it descends without needing additional energy input, (e.g. to account for level segments or holding periods). This is achieved by managing its momentum from Top Of Descent (TOD), En-route or stack altitude such that the vertical and lateral speed of the aircraft as it is accepted onto the glide slope are optimised for the aircraft type to ensure a safe and efficient landing. In advancing aircraft operations towards this envisaged 'optimal trajectory' such concepts as Low Power/Low Drag (LPLD) and Continuous Descent Approaches have been conceived.

Application of LPLD principles seeks to minimise noise at the aircraft source by specifying the maintenance of a low thrust and low drag profile of the aircraft. The aircraft is maintained in the most aerodynamically 'clean' configuration for as long as operationally possible in the prevailing air traffic and meteorological environment. Ideally the aircraft minimises its air brake use and delays the deployment of both the flaps and landing gear until they are absolutely necessary for the safe energy management of the approach. By comparison a CDA minimises the periods of level flight while maintaining as close to thrust idle as possible. Unlike an LPLD approach a CDA specifies the nature of the vertical approach profile to be flown.

Due to safety considerations it is generally accepted that the aircraft must be established on the glide-slope in its final configuration for landing by 2000 ft Above Ground Level (AGL). After this point the aircraft is flown solely to capture the correct speed at touch down.

However, busy airports operating at close to maximum capacity need to preserve flexibility in aircraft arrival times to accommodate wake vortex safety considerations and the permanently evolving nature of the Terminal Maneuvering Area (TMA) (e.g. meteorology, traffic etc). As such aircraft are not allowed to fly freely down their optimum descent path because the controller has to repeatedly intervene to appropriately space the traffic (e.g. by using 'path-stretching' techniques). Various projects around the world are attempting to find the correct balance between these two conflicting needs by utilising state-of-the-art technology [6][7][8].

The application of simple queuing theory to this problem dictates the necessity for airborne holding (currently stacks are used) from which the approach controller can select and vector into land in the optimum sequence. This method ensures that any slack in the system is immediately accounted for, however its use leads to inefficiency and environmental damage [9][10][11].

This case study aims to determine the impact of stacking upon the environmental efficiency of CDAs by exploring current and future scenarios where stacking is used; (a) to excess, (b) in moderation and (c) eliminated entirely in the EFAS model.

B. Scenario Building and Analysis

The scenario builder tool provided access to all of the parameters used as inputs into the simulation – as such it was used to modify the BADA configuration files to emulate the use of LPLD procedures as part of CDAs. Thus the scenario builder enabled maximum flexibility when constructing scenarios while still ensuring that the files submitted were fit for processing and offered maximum comparability with each other and their parent scenario.

Further, the scenario builder allowed the amount of delay introduced into the scenario to be varied by allowing the user to specify the maximum duration of delay that could be applied to any given flight. From this input the pre-processing feature of the tool produced a timetable that was modified to include this ‘on-demand’ delay. The revised timetable could then be reviewed online before submitting the scenario to the EFAS model. Users effectively controlled the amount of stacking within the scenarios by jointly manipulating this delay parameter with another controlling the threshold that determined when the delay held by a given aircraft required it to stack. The traffic handling logic dictated that the time spent by a given aircraft in the hold remained constant; the duration for which an aircraft was held for was exponentially related to the delay it possessed. In addition users effectively were able to control the amount of path-stretching activity within a scenario – it too being a function of the delay introduced.

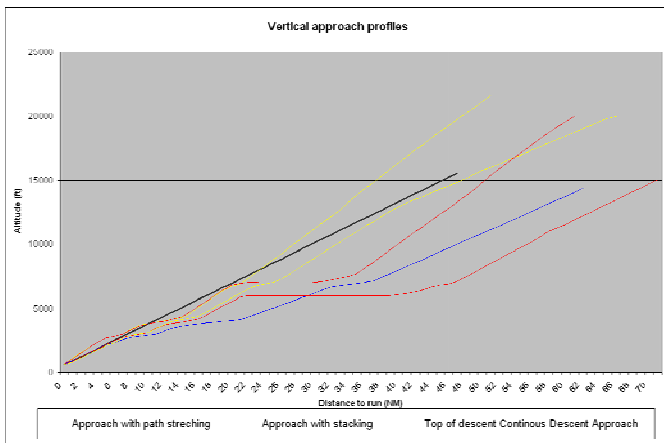


Figure 5. Vertical approach profiles (black reference line indicates the angle of a 3 degree approach).

1) Decision making

Initially the tool was used to build scenarios that compared the noise and emissions associated with stacking, path-stretching and TOD CDAs using LPLD settings (corresponding vertical profiles are shown in Fig. 5). Pre-processor parameters were manipulated firstly to reduce the total stacking duration in the baseline scenario (generated using relatively large airborne separation requirement assumptions) by approximately a half

(a 54% reduction in duration was actually achieved) and secondly to eliminate it entirely.

2) Results

Reducing the amount of airborne delay in the EFAS model meant fewer aircraft flew the longer path-stretch routes on final approach - thereby concentrating the arrival noise contours over a smaller area, as shown in Fig. 6. If these path-stretch routes are to be used in combination with CDAs Standard Terminal Arrival Routes (STARs) with dynamic TOD points will need to be defined and used as part of routine operational procedures. The suitability of this mitigation technique will depend upon the geographical distribution of noise sensitive areas surrounding a given airport.

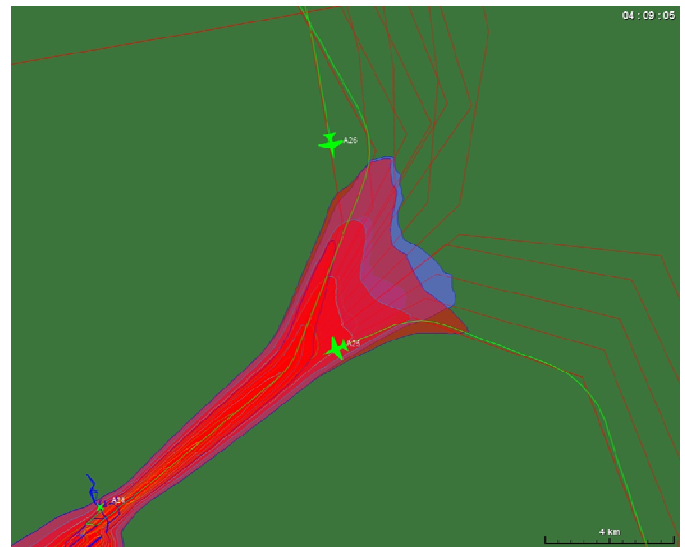


Figure 6. The final approach 57 Leaq noise contour in the 2030 scenario year. Blue contour is that generated with delay included in the model, red without. Screenshot taken from the ‘Route Builder’ visualisation tool built by David Atkins to support step 4 of the scenario assessment lifecycle.

As traffic levels increase in future years the time spent in the stack will grow exponentially (see Fig. 7). In reality it is unlikely that the two runway EFAS airport would operate with a total stacking duration longer than that used at major airports [12] as capacity is likely to have become saturated by this point. Consequentially 2.0×10^5 seconds could be reasonably assumed as a maximum total stacking duration. Therefore the scenarios where the approach traffic stacks for a longer duration than this (i.e. 2025 & 2030 of the 100% stacking scenario and 2030 of the 54% stacking scenario, as shown in Fig. 7) represent a hypothetical case where;

- no delay is propagated upstream (e.g. through Eurocontrol Central Flow Management Unit imposed ground delays); and;
- traffic demand continues to grow as forecast beyond the capacity of the airport (in reality excessive stacking would reduce demand for slots at the airport).

Because random perturbations are not simulated in the EFAS model it is thought that we currently underestimate the amount of stacking likely to be already occurring in the earlier scenario years.

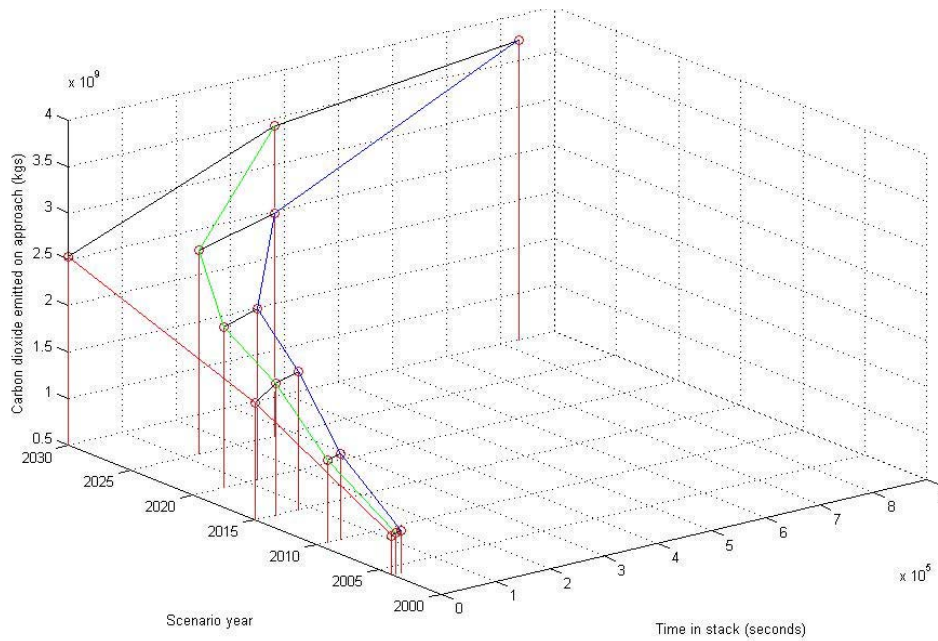


Figure 7. The relationship between total CO₂ approach emissions, stacking duration and scenario year (blue shows 100% stacking, green shows 54% reduction in stacking duration and red shows the approach emissions produced by no stacking at all).

Further, these results show that the amount of carbon dioxide generated during the approach is directly proportional to;

- the scenario year – reflecting the trade-off between increased traffic levels and the ‘cleaner’ aircraft types of future years (the rate of CO₂ emission production in the stack decreases by 91% from 54800 kg s⁻¹ in 2004 to 4460 kg s⁻¹ in 2030);
- the stacking duration (i.e. the length of time in mode); and;
- the number of aircraft entering the stack (i.e. the number of power-ups required to support level flight).

TABLE I. INCREASE IN EMISSIONS DUE TO STACKING.

Year	Full Stacking	54% reduction in stacking duration	No Stacking
<i>% Increase in Carbon Dioxide Emissions</i>			
2004	4.3	2.7	0.0
2015	11.1	7.2	0.0
2030	30.9	20.9	0.0
<i>% Increase in Nitrous Oxide Emissions</i>			
2004	2.2	1.4	0.0
2015	5.6	3.7	0.0
2030	20.6	14.5	0.0

Figures indicate the increase in emissions as a % of those generated solely by TOD CDAs.

Increased traffic levels will generate more emissions in the stack, (Ref. Table 1), without the introduction of new technologies and procedures. These will need to be capable of reducing excessive traffic demand at a particular airport (i.e. by managing peaks in traffic through flow management and strategic timetabling) and improving the logic used to handle the approaching traffic.

3) *Implications for the ATM System*

The results describe the need for an “environmentally friendly” ATM system to implement TOD CDA techniques at a system wide level. To meet this requirement advanced ATM is called for at both the strategic and tactical levels to create a set of efficient, de-conflicted, flight movements. Various ATM techniques that may be involved in the creation of such a set of movements are discussed in this section, primarily in the context of the airport and TMA environments. Further detailed investigation, modelling and validation of these concepts are required to establish their environmental performance.

A suite of tactical control tools will help to provide more certainty in the Required Time of Arrival (RTA) for the aircraft on approach, relative to each other. This could be deduced in a high traffic environment by the approach Air Traffic Control Officer (ATCO) utilising an Arrivals Manager (AMAN) software toolset. AMAN attempts to optimise the sequence of arrivals primarily by grouping aircraft of a similar weight together to minimise wake vortex separations. The longer the range of AMAN the more optimised the sequence becomes – principally through better planning and a more effective tailoring of the upper airspace arrival time. The end result of using the AMAN tool is the creation of a sequence (i.e. the attachment of a number to each inbound aircraft depending on their position in the queue to land). Once sequenced the controller then issues a set of clearances based on this information.

Secondly the short term assurance of adequate spacing between arriving aircraft can be maintained by tactical

Airborne Separation Assistance Systems (ASAS). This capability enables aircraft to maintain a relative spacing between each other in order to safely reduce the distances between arriving flights. It works by providing airborne surveillance assistance to the flight crew (thereby helping them to maintain the separation of their aircraft from others) and improving the situational awareness of the ATCO during the approach. This reduces the workloads of both Pilot and Controller while possibly allowing overall runway capacity to be improved. It is worth noting that various categories of ASAS application exist with varying degrees of responsibility placed on the flight crew (rather than the controller) to provide adequate spacing between their aircraft and the surrounding traffic. These concepts range from straightforward surveillance to operating high-density airspace without ground controllers. Most applications of airborne spacing in the descent are based on maintaining a fixed relative spacing in *time*, rather than in *distance*. Since the groundspeed reduces on the approach a fixed time spacing means that aircraft close up in distance as they descend.

ASAS is sufficiently flexible to allow its users to fly preferred pre-defined routes - such as those STARS that could facilitate CDAs or LPLD approaches. This particular functionality has been proven in the design of the CDA into UPS's Hub in Louisville, Kentucky [13]. This places a requirement upon the aircraft's ability to navigate along the STAR to within given navigational accuracies - such as those specified by Precision Radio Navigation (P-RNAV)². A P-RNAV route is pre-programmed into and controlled by, the Flight Management System (FMS) - the flight crew select the route upon receiving an appropriate clearance from Air Traffic Control. Strategic AMAN and tactical ASAS, underpinned by P-RNAV, provide different facets for the planning and accommodation of the arrivals sequence - their joint use provides optimal performance. The flexibility in the system is derived from the dynamic nature of the ASAS and AMAN systems which continually respond to the developing TMA situation in real time.

From a more strategic point of view stacking is essentially a manifestation of the inefficiency associated with the runway capacity bottleneck in the airport system. Effective ATM to overcome this challenge can be realised through flow management, ideally that made in a Collaborative Decision Making (CDM) - Network Operations Planning (NOP) environment supported by a System Wide Information Management (SWIM) infrastructure (as envisaged by the SESAR Concept of Operations [2]). Such a situation would allow detailed planning of flight operations both when drawing up timetables and imposing strategic control tactics, such as ground or En-route delays.

² The 'P-RNAV' standard gives a lateral track keeping performance to within 1 NM (Nautical Mile) for 95% of the time. The Vertical Navigation 'VNAV' capability is optional for P-RNAV - the vertical profile may be flown either by pilots or automatic systems (the latter using approach specific information held in an on-board navigation database).

IV. CONCLUSIONS

This paper has shown how a web-based modelling support tool has supported the EFAS Project, reducing the risk of potential delays and ensuring that project partners utilise the EFAS ATM modelling tool to its maximum potential. In particular a case study has outlined how the tool was used to introduce different amounts of delay into scenarios exploring the impact of Stacking on TOD CDAs. The result of this modelling effort is the illustration of a function that describes how total stacking duration will grow exponentially with the increasing traffic levels of future years. The amount of CO₂ generated in the stack is directly proportional to the stacking duration, the number of aircraft entering the stack and the scenario year. These findings highlight, from an environmental perspective, the importance of the technologies and procedures that will assist in minimising the amount of stacking in the air transport system of the future.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the support of the EFAS Partners and the UK Government Technology Strategy Board in the production of this paper. The EFAS Project was launched in October 2005 and is due to run until early 2008, co-funded by the Technology Strategy Board (TSB) - a non-departmental public body sponsored by the UK Department for Innovation, Universities and Skills (DIUS). The Thales ATM-led consortium includes Helios Technology, Selex Sistemi Integrati, QinetiQ, NATS and BAE Systems from the private sector; Manchester Metropolitan University Centre for Air Transport and the Environment (CATE); and Loughborough University Systems Engineering Innovation Centre.

Particular thanks go to Professor Roy Kalawsky of Loughborough University for his support on the EFAS Project. Special mention must also go to Tomas Dyjas and Ben Stanley both of Helios for preparing Fig. 7 and for providing information on the role of ASAS & AMAN respectively.

REFERENCES

- [1] "The Future of Air Transport", UK Government white paper, Department for Transport, July 2004, <http://www.dft.gov.uk/>
- [2] "SESAR Definition Phase; Milestone 3", SESAR Consortium for the SESAR Definition Phase Project co-funded by the European Commission and Eurocontrol, Doc. Ref. No. DLM-0612-001-02-00, Section 10.2.4 "Sustainability Assessment", September 2007
- [3] Eurocontrol's "Base of Aircraft Data", Aircraft performance model, <http://www.eurocontrol.int/eec/>
- [4] FAA's Integrated Noise Model, <http://www.faa.gov/>
- [5] "Project for the Sustainable Development of Heathrow", Department For Transport, UK Government, 2006 <http://www.dft.gov.uk/pgr/aviation/environmentalissues/heathrow/>
- [6] "The tailored arrivals concept", Boeing/NASA, October 2006, <http://www.boeing.com/phantom/news/>
- [7] "LFV's Green Approaches", October 2007, <http://www.enviro.aero/>
- [8] "The Atlantic Interoperability Initiative to Reduce Emissions (AIRE)", FAA/Eurocontrol, <http://www.faa.gov>
- [9] S. Carlier, I. de Lépinay, J. C. Hustach, F. Jelinek, "Environmental Impact of Air Traffic Flow Management Delays", Eurocontrol Experimental Centre, ATM R&D Seminar, Barcelona, Spain, July 2007

- [10] S. Carlier and J. Hustache, "Project GAES – Environmental impact of delay", Project for Eurocontrol Experimental Centre, (EEC/SEE/2006/006), 2006
- [11] Dr A. Cook "Evaluating the true cost to airlines of one minute of airborne or ground delay", Performance Review Commission, Report prepared by the University of Westminster, Transport Studies Group, University of Westminster, London, February 2004
- [12] "Improving the air passenger experience; an analysis of end to end journeys with a focus on Heathrow", Department for Transport, UK Government, November 2007, <http://www.dft.gov.uk/pgr/aviation/airports/improveairpassenger.pdf>
- [13] John-Paul Clark, Dannie Bennett, Kevin Elmer, Jeffery Firth, Robert Hilb, Nhut Ho, Sarah Johnson, Stuart Lau, Liling Ren, David Senechal, Natalia Sizov, Robert Slattery, Kwock-On Tong, James Walton, Andrew Willgruber, David Williams, "Development, design and flight test evaluation of CDA procedure for night-time operations at Louisville International Airport", Partnership for Air Transportation Noise and Emissions Reduction (PARTNER), January 9th 2006 <http://web.mit.edu/aeroastro/partner/projects/project4.html>

ANALYSIS OF EMISSIONS INVENTORY FOR “SINGLE-ENGINE TAXI-OUT” OPERATIONS

Vivek Kumar

Center for Air Transportation
Systems Research
George Mason University
Fairfax, USA
vkumar3@gmu.edu

Lance Sherry

Center for Air Transportation
Systems Research
George Mason University
Fairfax, USA
lsherry@gmu.edu

Terrence Thompson

Metron Aviation Inc.
Herndon, USA
thompson@metronaviation.com

Abstract—Stringent federal and state programs along with technology innovation have resulted in declining emissions from static sources (e.g. power plants) and are projected to meet national quality standards by 2025. The same cannot be said for mobile sources of emissions from flight operations at airports. In the absence changes in airport operations, the forecast rates of growth in flight operations will jeopardize State’s abilities to lower emissions to meet Federal standards.

Recent studies indicate that 96% of flights in the U.S. accrue their delays at the airports and directly impact local non-attainment through emissions. This paper examines the sensitivity of emission factors (number of engines, engine efficiency and fleet mix, taxi-out time) through a case-study of departure operations at Orlando (MCO) and New York-LaGuardia. Under the assumptions of a representative fleet mix, departure schedule, runway assignment, and taxi flows, “feasible single engine” taxi-out procedures reduced emissions (CO/NO_x/SO_x/HC) by 27% at MCO and 45% at LGA. To achieve the same level of emissions reduction requires a 25% decrease in taxi-out time at MCO, and 44% decrease at LGA. The implications of these results on optimization of surface operations to minimize emissions are discussed.

Keywords-component; emissions, noise, surface optimization, single engine, taxi, fuel, pollution, environment.

I. INTRODUCTION

Flight operations result in the emission of a host of air pollutants that adversely affect public health and the environment. Nitrogen Oxides (NO_x) and Hydrocarbons (HC) are precursor emissions of ground-level ozone, which causes lung irritation and aggravates diseases such as asthma, chronic bronchitis, and emphysema. Particulates have adverse cardiopulmonary effects and contribute to regional environmental problems such as haze and acid rain. Toxics such as benzene and formaldehyde are known or probable human carcinogens.

Stringent Federal and State regulatory programs, along with innovations in technology, have resulted in significant reductions in projected emissions from static sources (e.g. power plants) by 2025 to meet Federal standards (Cooper et al, 2003, page I-6). Despite improvements in aircraft engine technologies, the forecast growth in flight operations will yield large increases in airport emissions. These emissions are

projected to jeopardize the ability of States in non-attainment of criteria pollutant National Ambient Air Quality Standards (NAAQS) from meeting Federal ambient levels of these pollutants.

Recent studies indicate that 96% of flights in the U.S. accrue delays at the airports as opposed to airborne delays (Xu, 2007; Chappell, 2004). These delays are the result of carrier delays (i.e. gate push-back), delays that are the result of departures scheduled in excess of the departure capacity of the airport (i.e. departure congestion), delays that are the result of predicted airborne delay (e.g. Miles-In-Trail, Airspace Flow Program), and destination airport delays (Ground Delay Program).

These delays, contribute to local emissions in excess of the ambient emissions from non-delayed operations at the airport. This paper examines the sensitivity of emission factors (number of engines, engine efficiency and fleet mix, and taxi-out time) through a case-study of departure operations at Orlando (MCO) and New York-LaGuardia. Under the assumptions of representative fleet mix, departure schedule, runway assignment, and taxi flows, the following results were identified:

- “Single Engine” taxi-out procedures have the potential to reduce emissions (CO/NO_x/SO_x/HC) by 27% at MCO.
- “Single Engine” taxi-out procedures have the potential to reduce emissions (CO/NO_x/SO_x/HC) by 45% at LGA
- To achieve the same level of CO/NO_x/SO_x/HC reductions as “single engine” operations, would require an average of 27% (27, 26, 22, 24)% decrease in taxi-out time at MCO
- To achieve the same level of CO/NO_x/SO_x/HC reductions as “single engine” operations, would require an average of 44% (46, 45, 43, 44)% decrease in taxi-out time at LGA.

The paper is organized as follows: Section 2 provides an overview of aircraft emissions. Section 3 describes the method of analysis. Section 4 describes the results of the analysis. Section 5 the implications of these results on optimization of

surface operations to minimize emissions and on the design of an Airport Environmental Dashboard are discussed.

II. AIRCRAFT EMISSIONS AND MITIGATION STRATEGIES

The total emissions burden associated with airport operations is the result of emissions from aircraft, Ground Support Equipment (GSE), Ground Access Vehicles (GAV), stationary sources, and private vehicles (Cooper, Ulbrich 2003). Figure 1 illustrates the contribution of each of these sources to NOx emissions at Logan airport in 1999.

A. Chemistry of Emissions:

Oxidation of fuel can be represented by the following chemical reaction

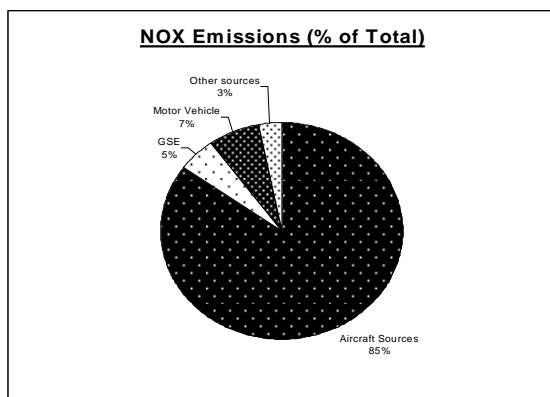
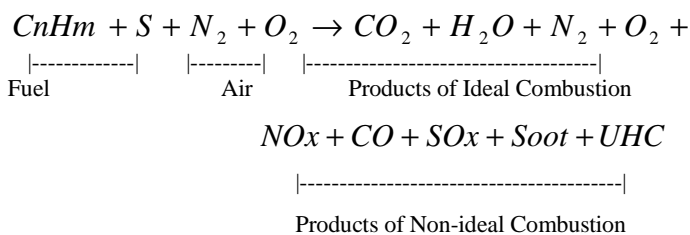


Figure 1 – Contribution of sources of emissions to total NOx emissions at Logan International in 1999.

National Ambient Air Quality Standards

Pollutant	Primary Stds.	Averaging Times	Secondary S
Carbon Monoxide	9 ppm (10 mg/m ³)	8-hour	None
	35 ppm (40 mg/m ³)	1-hour	None
Nitrogen Dioxide	0.053 ppm (100 µg/m ³)	Annual (Arithmetic Mean)	Same as Pri
Sulfur Oxides	0.03 ppm	Annual (Arith. Mean)	-----
	0.14 ppm	24-hour	-----
	-----	3-hour	0.5 ppm (1300 µg/m ³)

Table 1- NAAQS Source- <http://www.epa.gov/air/criteria.html>

Aircraft fuel is a mixture of hydrocarbons (essentially kerosene) and sulphur. Its combustion in the presence of oxygen and atmospheric nitrogen (inert), produces several products. Some of the products, i.e. carbon dioxide, water, left over nitrogen and oxygen are products of ideal combustion and non-harmful to the environment. However, not all of the fuel undergoes ideal combustion, resulting in harmful by-products like nitrous oxides, sulphur oxides, carbon monoxide, and unburned hydrocarbons.

B. Permissible safe level of the pollutants

The Environmental Protection Agency (EPA), under the influence of Clean Air Act, has set National Ambient Air Quality Standards for pollutants considered harmful for public health and the environment. National Ambient Air Quality Standards for six principal pollutants, known as ‘criteria’ pollutants, are listed in Table 1. The units of measure for the standards are parts per million (ppm) by volume, milligrams per cubic meter of air (mg/m³), and micrograms per cubic meter of air (µg/m³).

Although most US airports are well within the National Ambient Air Quality Standards (NAAQS) set by EPA (Environmental Protection Agency), some of the airports fall in the non-attainment zone for ozone for non-compliance with NAAQS: For example, the state of Texas has four areas that are out of compliance with the NAAQS for ozone, two of which -- Houston-Galveston-Brazoria and Dallas/Forth Worth -- are also home to major airport facilities. The Houston-Galveston-Brazoria non-attainment area alone hosts three major commercial airports: (1) George Bush Intercontinental, (2) William P. Hobby, and (3) Ellington Field. The Dallas/Forth Worth non-attainment area is also host to four commercial airports: (1) Dallas/Fort Worth International (2) Fort Worth Alliance; (3) Meacham; and (4) Dallas Love Field. Several airports, including Sacramento International Airport (SMF), are located in the Sacramento area of California, which is classified as being in severe non-attainment of the ozone NAAQS. (see Cooper, Ulbrich 2003, pg IV -16 for details)

C. Mitigating Aircraft Emissions

There are two approaches to mitigating aircraft emissions for airport operations (Figure 2). The first approach involves improvements in the technology (aircraft engines, fuel, and aircraft design). The second approach is related to modifying airport operations. The following discussion focuses on airport operations.

Emission mitigation strategies that do not involve changes to current engine technology or aircraft design include: (1) Decreasing Taxi-out time (TOT), (2) Lower Emission per passenger through Load Factor and Upgauge, (3) Reducing Power Output during Taxi (see Cooper, Ulbrich 2003, pg III 7-10 for details)

(1) Decreasing Taxi-out time (TOT): Aircraft emissions of HC and CO tend to be particularly high during taxiing operations. Operational changes that reduce aircraft idling and taxi time

directly reduce pollutant emissions. A possible option is

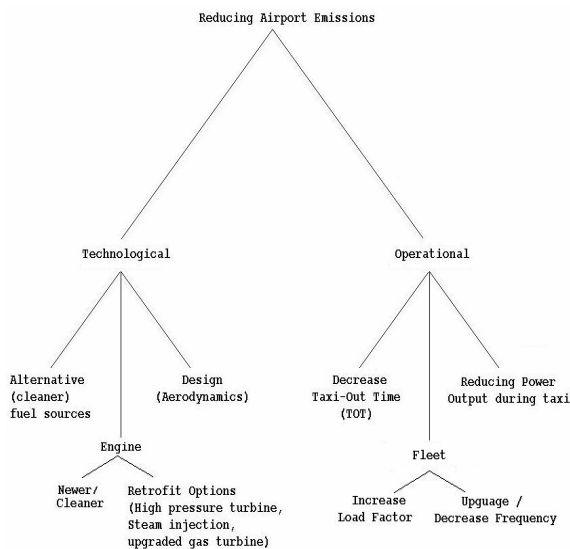


Figure 2: Mitigating Aircraft Emissions

“dispatch towing”. High-speed tugs can be used to move aircrafts between the terminal gate and runway more efficiently and with fewer frequent stops than with standard practices. Such tugs have already been tested successfully by Virgin Atlantic at LAX and SFO. Taxi-times can also be reduced by planning the airport more efficiently. For example, using a decentralized gate designs wherein passengers are brought to and from the aircraft by other transport vehicles. Also, if better dynamic real time estimates of the taxi-out time were available to the airline and the ground controller, one could possibly hold the aircrafts at the gate (if possible) or ask the pilot to operate on fewer engines. There has been some promising work in this area recently.(Ganesan, Poornima 2008).

(2) **Fleet:** Airlines can improve their operational efficiency by maximizing the number of passengers on each flight, thereby minimizing emission per passenger. Airlines already have enough profit incentive to increase their “load factors”. Upguaging is another approach: A single flight serving more passengers on a larger plane may reduce emissions – and airline costs – compared to multiple flights using smaller airplanes to serve the same route. However, other considerations often apply, such as the desire to provide the customers with more frequent flight options. Also, depending on how landing fees may be structured it might be more expensive to land one large airplane (weight based landing) compared to two smaller airplanes.

(3) **Reducing Power Output during Taxi:** Another way of reducing aircraft emission is more judicious use of aircraft engines in taxi mode. Most of the modern day aircraft are equipped with two to four engines, one or more of which can be shut down during taxiing. This not only reduces emissions, but allows other engines to operate more efficiently (i.e. at higher RPMs) resulting in lower consumption of fuel and less HC and CO emissions per pound of fuel consumed. Heathrow airport in the United Kingdom already encourages this practice. It should be noted that execution of “single engine ops” is at

pilot discretion, as engines have their own warm up and cool down time when they achieve thermal stability.

III. METHOD FOR COMPUTATION OF ENVIRONMENTAL FACTORS FOR DEPARTURE

Estimates of emissions from aircraft can be computed using models developed by ICAO or FAA. The FAA model is known as the Emission and Dispersion Modeling System (EDMS) EDMS was developed by the FAA in cooperation with the U.S. Air Force in the mid 1980’s as a complex source microcomputer model to assess the air quality impact of proposed airport development projects. EDMS is designed to include the contribution of airport emission sources, particularly aviation sources, which consists of aircrafts, auxiliary power units (APUs) and ground support equipments.

Emission is calculated using the following formula:

$$Emissions(grams) = TOT(\quad) * EI(grams / kg) * Fuel.Flow(kg / min) * \#.Of.Engines$$

Each aircraft engine type has its own emission characteristic. Depending on the engine type levels of CO, NOx, SOx and HC (Hydro Carbon) emissions may vary. This information is comprehensively captured in what is called an Emission Index (EI). The Emission Index is a function of engine type, phase of flight (only taxi mode for our study) and pollutant (hence there is an EI corresponding to each category of pollutants NOX, SOX etc.). The emission indices are based on information provided by the engine manufacturers and documented by the FAA and ICAO (EPA, 1985) (Lang and Chin, 1998).

Row # 8 labeled NOX_EI_G_KG means that for every kilogram (1kg = 2.2 lbs) of fuel that B767 consumes, each of its 2 engine emits about 3.4 grams of NOx into the atmosphere. Other rows can be interpreted similarly.

Note: We used EDMS as a source of the aircraft Emission Indices. Even though EDMS is a FAA progeny and is widely accepted in the industry, there are some issues which are dealt better in an alternate model called NESCAUM (Northeast States for Coordinated Air Use Management). EDMS simplifies the air fleet mix. Only one (the most common) engine is assigned to each aircraft. However we know that each aircraft can be outfit with more than one engine types. For example: Boeing 757-200 can be fitted with 4 different types of engine. NESCAUM improves on EDMS by taking a weighted average of the different engine types used on each airline’s fleet of aircraft. The same weighted average methodology is followed while calculating the emission indices for the APUs. Hence NESCAUM provides a finer and more accurate detail level than EDMS (Cooper, Ulbrich 2003).

IV. CASE STUDIES: LGA AND MCO

Case studies of Orlando (MCO) and New York –LaGuardia (LGA) were conducted to understand the factors that contribute to emissions. Analysis was done for the month of July 2007 for our case study.

Taxi-out times and ETMS equipment number were extracted from the ASPM database (Aviation Safety

ETMS_Name	B767
ETMS_Descr	Boeing Company B767/CF6-80A
Num_Engine	2
Fuel_Flow	0.150000006
Taxi_Fuel	0.150000006
CO_EI_G_KG	28.20000076
HC_EI_G_KG	6.28000021
NOX_EI_G_K	3.400000095
SOX_EI_G_K	1

Table 2 A sample row of EDMS/ICAO database showing Emission Index of B767s

Performance Metrics database maintained by the FAA) for individual take-off operations. This study only considered departure operations for our study.

The ETMS equipment number extracted from the ASPM database was then mapped to its corresponding Emission Index (EI) number in the EDMS (Emission and Dispersion Modeling System) database.

Some assumptions: (1) the study considered only ground operations. Future work will take into account the full Landing Take-off Cycle (LTO). (2) only considered aircraft pollution not Ground Service Vehicles etc. See Figure 1. (3) All aircrafts with the ETMS Equipment number were assumed to have the same engine type. This is the underlying principle behind ETMS emission estimating methodology.

A. Daily Emissions Profile

Emissions were estimated using the equations described in Section 3. Figure 3 shows the daily profile of emissions for MCO and LGA. The X axis is the time of day in 15 minute epochs. The Y axis is the pollutant emissions (i.e. CO, HC, NOx, SOx) weight in grams. The dotted line in each curve represents the cumulative emissions over the course of the day. Peak periods result in a steeper rise in the cumulative curve.

B. Monthly Emissions Profile

Figure 4 summarizes the monthly emission profile at MCO for July 2007. The histograms indicate the number of days with varying levels of cumulative emissions. The bars on the right indicate the days for which the emission level of a particular pollutant was higher than the rest. The ‘bad’ days are a result of higher taxi out times on those particular days (since the fleet mix operating at MCO is pretty much constant throughout).

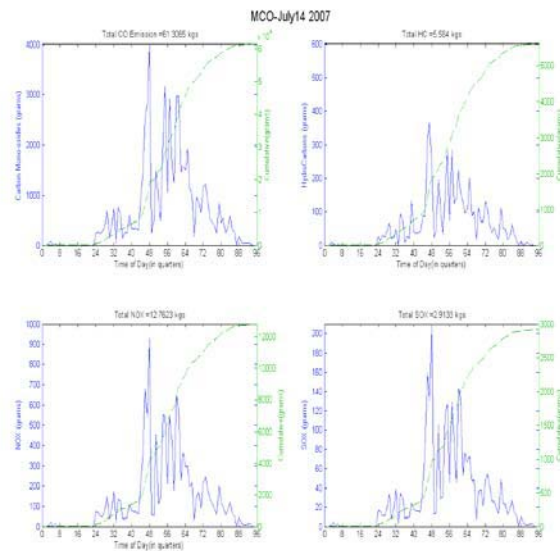


Figure 3(a)– Emissions on a typical day at MCO (July 14th, 2007)

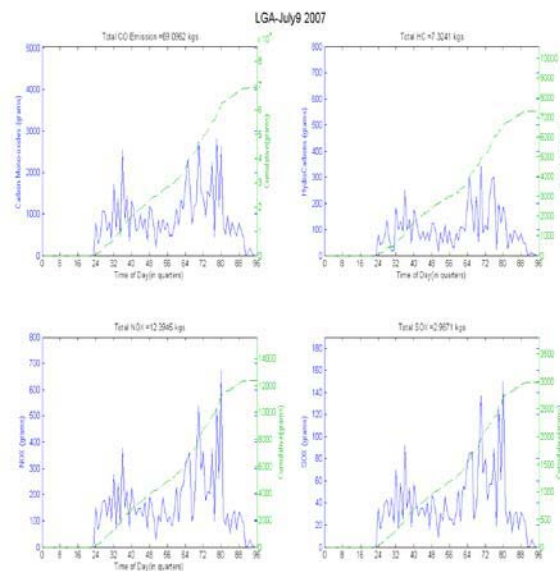


Figure 3(b) – A typical day at LGA (July 9th, 2007)

C. Comparison between MCO and LGA

Figure 5 provides the total emissions per day over the course of the month for each pollutant. The LGA emissions are greater than the MCO emissions except on Saturdays. Comparison between LGA and MCO- Operational and Fleet mix interplay

Figure 6 shows a comparison of the rate of emissions for each of the pollutants between MCO and LGA. The rate of emissions was derived from only weekday fleet mix and schedule for July 2007. The X-axis is the total taxi-out time.

The Y-axis is the total pollutant level. The slope of the graph gives the rate of pollution metric tonnes/minute of taxi out time. A higher slope represents a higher polluting fleet independent of taxi time. The slope is a product of number of engines, fuel flow and emission index (for the particular category of pollutant).

The MCO fleet exhibits slightly higher average emissions for the three of the four pollutants than the LGA fleet. The total

daily emissions can be determined by reading up from the X-axis total daily taxi-out time to the rate-line and then reading across to the Y-axis to the total emissions.

D. Single Engine Taxi Operations

Taxi on reduced engines, such as a single engine for double engine aircraft, has the potential to reduce emissions by half. This practice is performed at some airports where emissions are a concern (e.g. London Heathrow Airport (LHR)). In practice, due to engine start-up and shut-down procedures, the single engine operations can only be used when taxi delays are anticipated to be in excess of 15 minutes.

Figure 7 illustrates the impact of single-engine taxi for taxi-out delays in excess of 15 minutes at MCO and LGA. Since taxi-out times are shorter at MCO, single-engine taxi-out resulted in an average reduction in emissions of 27%. CO/NOx/SOx/HC reductions of 27%, 26%, 22%, and 24%

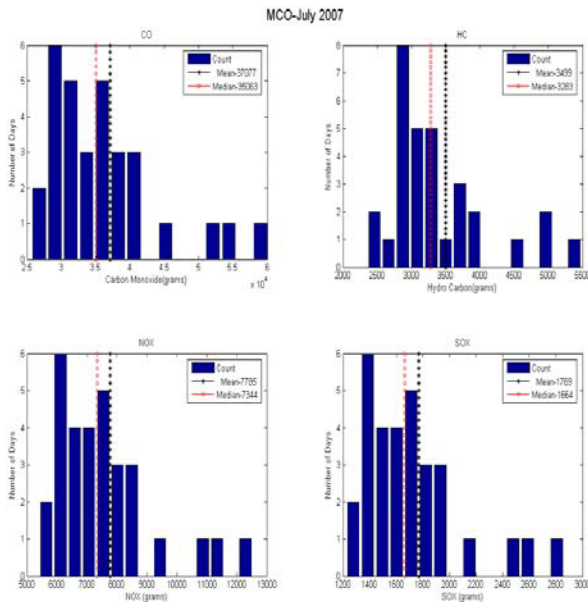


Figure 4(a) – Histogram for MCO showing CO, NOx, SOx and HC level for the month of July 2007.

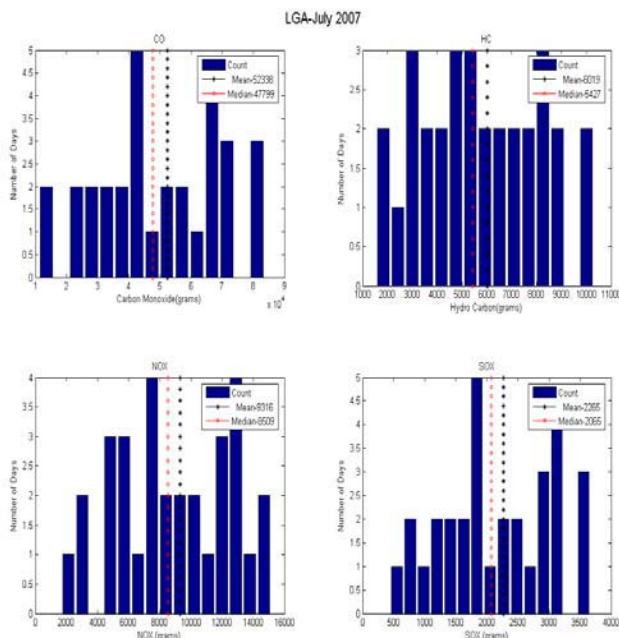


Figure 4(b) – Histogram for LGA showing CO, NOx, SOx and HC level for the month of July 2007.

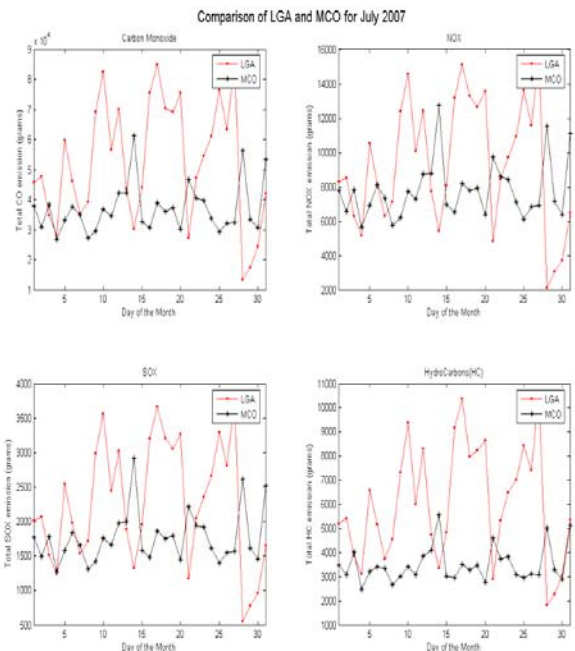


Figure 5 – Comparison of emissions at MCO and LGA for July 2007 respectively.

Single-engine taxi for taxi-out delays in excess of 15 minutes at LGA achieved an average reduction of 44%. CO/NOx/SOx/HC reductions of 46%, 45%, 43%, 44% respectively.

One of the main roadblocks with implementing a single engine taxi policy more stringently is the fact that engines typically require about 4-5 minutes of warm up time prior to take off to achieve thermal stability. It is counter productive to taxi on single engine and then wait at the runway trying to warm up the engines. In order to avoid such a situation one could use taxi out time prediction in issuing ‘single engine taxi’ advisories. A robust algorithm for Taxi out time prediction, which incorporates and captures the dynamic, stochastic nature of airport surface operations could help the Air traffic

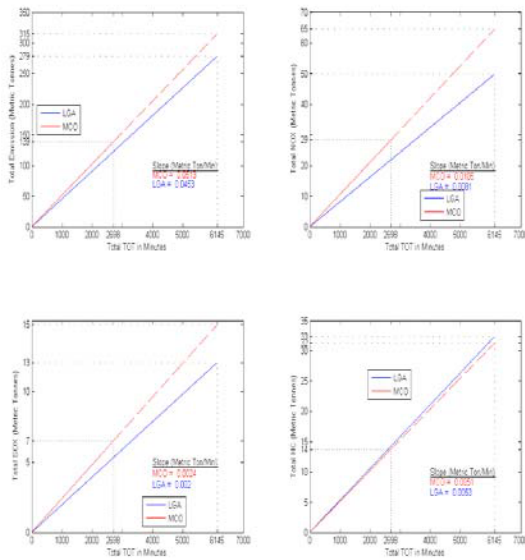


Figure 6 – Comparison of rate of emissions at MCO and LGA for July 2007

control and the airlines to make better decisions about whether or not they should issue a ‘single engine taxi’ advisory to the pilots. For example, if the Taxi out time is greater than certain threshold (say 15 min), one could advise the pilot to taxi out on a single engine for the first 8-10 minutes (depending on the confidence of prediction in Taxi Out time), and thus save on fuel and emissions.

V. CONCLUSIONS & FUTURE WORK

The paper examined the sensitivity of the factors contributing to emissions in taxi-out operations. The results indicate that under the constraint of a fixed-fleet, schedule, runway assignment procedures, and taxi-out operations, single-engine taxi provides the potential to reduce emissions up to approximately 44%. Under the constraints of a practical single-engine taxi procedure, that requires a minimum of 15 minutes taxi, the procedure provides the biggest advantage at airports with periods of delays in excess of 15 minutes.

Future Work:

- (1) The ICAO/EDMS database that we referenced for obtaining engine emission characteristic is conservative in a sense that it assumes that all aircrafts with a given ETMS equipment number are fitted with the same engine type (the most common type). However it is not uncommon to have more than one engine types fit on the same aircraft type. NESCAUM improves on EDMS by taking a weighted average of different engine types used on each airline’s fleet. Using this model in the future has a potential benefit of improving the accuracy of our emission computation
- (2) Understanding and modeling the change in fuel flow while shifting to single engine taxi for different aircraft engine types would be helpful in estimating the benefits of a single engine taxi more accurately.

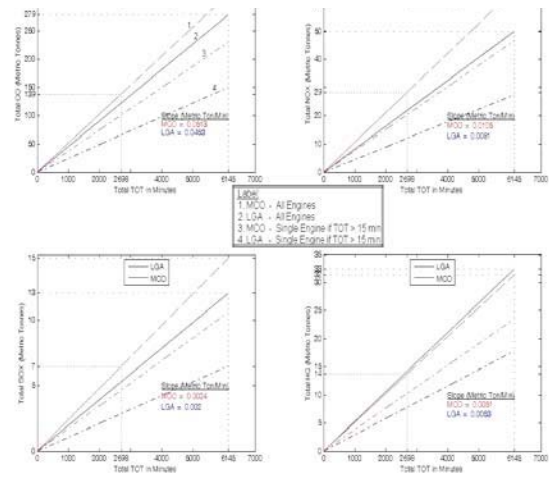


Figure 7: Effect of Single Engine Taxi

This model could then be used to obtain dollars gained (as a result of lesser fuel burnt) to incentivize the airlines.

- (3) It is also required to understand the physical/operational constraints that prevent airlines/controller from adopting single engine taxi. For example:
 - Engine warm up time to achieve thermal stability (3-5 minutes)
 - No tight turns allowed on a single engine
 - Abnormal weather conditions, i.e. Snow/ice
 - Uneven ground surface at the airport.

ACKNOWLEDGMENT

This Research has been funded by NASA Ames under project MEFISTO (Modeling Environment Factors in Surface Traffic Optimization) and thanks are due to Gilena Monroe, Sandy Lozito, Len Tobias, John-Paul Clarke and Ed Knoesel(NYNJPA).

REFERENCES

- [1] Coralie Cooper, Dave Park, Ingrid Ulbrich et al, “Controlling Airport-Related Air Pollution”, June 2003
- [2] Emission estimates from “1999 Environmental Status and Planning Report,” Massport 2002..
- [3] Emissions and Dispersion Modeling System Policy for Airport Air Quality Analysis; Interim Guidance to FAA Orders 1050.1D and 5050.4A
- [4] Lang and Chin, CNS/ATM Enhancements to Reduce Aircraft Emissions, 1998
- [5] Voluntary Airport Low Emission Program. Technical Report, Version 4. Office of Airports, Airport Planning and Programming. September 2007.
- [6] David Daggett, Robert C Hendricks, Rainer Walther, Edwin Corporan, “Alternative Fuels for use in Commercial Aircraft”.
- [7] Rajesh Ganesan, Poornima Balakrishna, “Taxi-out time prediction using approximate dynamic programming”, 2007
- [8] Bruno Miller, Kenneth Minogue, John-Paul Clarke, “Constraints in aviation infrastructure and surface aircraft emissions”, 2001

On the Use of Visualization Tools to Present Complex Simulated Environmental Data for Policy Making

Daniel Pearce^{1,2}, Luiz Wrobel¹, Ian Fuller²

¹ School of Engineering and Design
Brunel University
Middlesex, UB8 3PH, UK

daniel.pearce@brunel.ac.uk, luiz.wrobel@brunel.ac.uk

² EUROCONTROL Experimental Centre
Centre de Bois des Bordes, BP 15
F-91222 Brétigny sur Orge
CEDEX, FRANCE
ian.fuller@eurocontrol.int

Abstract — The future of aviation is the subject of considerable debate and policy discussion. There is also increasing emphasis on the inclusion of public consultation and participation within the planning and decision making system. Yet, presenting the findings of complex, multi-dimensional research in a style that is accessible by a potentially lay audience is no simple challenge. This is especially true for subjects whose findings are controversial, such as airport expansion plans and possible health implications of activities.

Text heavy documents laden with equations, graphs and tables will in most cases act only to alienate a non-expert. However, a method that has found favor in presenting to the non-expert is the use of visualizations, particularly visualizations that allow for a degree of interactivity. This paper investigates a number of the barriers that exist between science and policy making, and then proposes the virtualization application Google Earth as ideally suited for presentation of aviation-related subjects. The paper includes examples of Google Earth models which make use of the 3D capabilities of the tool combined with streamed geographical content to provide stakeholders with a novel, intuitive and interactive presentation that requires no expert understanding to convey its findings.

Keywords-Environmental policy, aviation, visualization, science-policy divide, lay audience, Google Earth

I. INTRODUCTION

While the emissions of most industries are anticipated to decrease over the coming decade, air passenger traffic and consequently the impact of aviation on the environment is projected to rise [1]. This has generated a significant increase in media and public attention on the growing needs to establish policy that better mitigates aviation's environmental impact. Subsequently, environmental controls have been highlighted as a threat to future capacity [2,3]. However, the creation of mitigation strategies is no easy task.

In order for science to contribute to policy making decisions, research findings must be communicated to stakeholders. Engel-Cox [4] summarized that scientific data intended for use in the creation of policy needs to meet 5 essential criteria if it is to be useful: *relevance, timeliness, integrity, clarity, and visualization*. While relevance and

timeliness are tied into the project definition, integrity, clarity and visualizations are the responsibility of the researcher. Furthermore, increased awareness has placed a greater motivation to include public consultation in the decision-making process. This increases the number of potential lay recipients of scientific data and means the ability to present scientific conclusions in a manner that is accessible to non-experts is even more important. Still, the utilization of visualization techniques is often ignored and it remains the norm for researchers to publish their findings in long technical documents or journal papers that are difficult to understand by all but a few specialists. This is specifically true of the multi-dimensional data generated by complex mathematical simulations. This generates an understandable divide between the scientific and lay communities which hampers the exchange of information from one camp to another.

There is an increasing body of literature that seeks to expose the underlying causes of this barrier between science and successful policy making. This includes growing support for the implementation of information transfer models that better support the general public as a stakeholder. Here, we seek to present an overview of this barrier between science and policy and how it applies to the transfer of aviation research into information that proves useful in the policy making process. We then look at the ways in which visualizations are being used to contribute to the policy making process. The paper concludes with the presentation of a novel method of displaying complex 4D aviation data to non-experts using the well known and intuitive geographical visualization tool, Google Earth.

II. RESEARCH CULTURE

Political opinions that lack a credible supporting argument will not influence policy. Likewise, environmental forecasts based solely on scientific objectives have little chance of contributing to the formation of new regulation [5]. Therefore, there must exist an inextricable link between science and policy; with governmental agencies relying heavily on scientific findings to guide decision making, and research institutes increasingly dependent on government funding. However, there exists a historical divide between scientists and

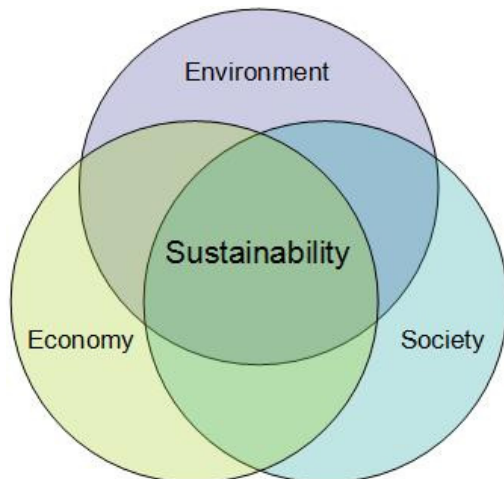


Figure 1. The concept of sustainable development requires a balance between social, economic and environmental factors

policy making communities that hampers the free flow of information from one side to another.

Alcamo [6] pointed out that scientists and decision makers “speak different languages”. Price [7] hypothesized that the problem stems from differences in the motivations of the parties involved. On one side of the divide are scientists, motivated by the pursuit of knowledge through thorough investigation. On the other side we find politicians, motivated by electoral support and whose decisions are based on debate, compromise and response to shifts in public opinions. This disparity generates an inevitable conflict. These opinions are echoed by Lee [8], who notes in addition that the differences in cultures, standards and practices of science and politics can lead to distrust. Fischer [9] highlights the “politicization of science”, whereby politicians have in the past misinterpreted or misused scientific data to support their own agendas. Because the general public will often take publicized information at face value, this can be interpreted as if the scientists are for or against a particular argument and subsequently affect the reputation and credibility of those involved.

Furthermore, science has historically adhered to the paradigm that research is objective and value free. However, this is rarely the case in reality. Subjective decisions to define a research problem, establish methodologies, choose which models to use, which assumptions to apply, and how results are presented are made at all stages of the research process. Cullen [10] argued that this means science is unable to deliver non-arguable conclusions because personal interpretations are made by scientists all the time. Sundqvist [11] stated that the barrier is perhaps not between science and policy, but established by social constraints where all parties “act strategically by drawing boundaries which suit their own interest”. Engel-Cox [4] echoed this thought, stating that the key to using scientific data in policy is the relationship between the scientists and the policy makers who are directly involved in the exchange of information.

Norse [12] notes that the process of policy development itself can create a barrier between scientists and policy makers. He observed that policy development is rarely linear or logical,

can take an extremely long time and the end result is not guaranteed to have any scientific rigor. Engel-Cox [4] explains that regardless of the clarity of the presentation, science is never able to overrule political and social considerations. A recent example is the introduction of international agreements on emissions caps. Despite worldwide agreement from the scientific community that human activity is a “very likely” cause of accelerated climate change [13], the United States and China continue to oppose the introduction of emissions caps. Such is the frequency of these occurrences, the acronym ‘NIMTO’ (not in my term of office) has come to being solely to describe policy decisions that go against the grain of logic and/or science in favor of increased electoral support. This additional bureaucracy combined with the risk of negative exposure often represents too great a danger of negative repercussions and prevents active scientific engagement.

Despite this, the need for interdisciplinary collaboration has never been greater. Sustainability and the pursuit of sustainable development are touted with ever-increasing regularity and achieving a sustainable future requires a balance between social, economic and environmental factors (figure 1). If research is to adapt to better transition science-policy divide, it is necessary to understand the needs of each party and how data evolves as it moves through the decision making system.

Norse [12] explained that information that is useful by policy makers will represent a considerable reduction in both size and complexity from the original scientific data. This reduction in volume has the potential to greatly increase subjectivity and increase the likelihood of misinterpretation. This is equally true of the transfer of information to the general public because there will inevitably be components of the original data that are beyond their technical understanding.

A solution would be to ensure that scientific guidance is present throughout the transfer process. Anderson [14] highlights the importance of including a ‘broker’ to ensure the smooth translation of scientific findings into policy. However, performing such a role requires the ability to understand both the science and the policy making process. Individuals who are both capable and prepared to perform such a role at the highest level are few. Lee [8] refers to them as ‘philosopher kings’; a term lifted from Plato’s writings on philosophy and political theory where it is stated that, in an ideal society, those in positions of control must have a firm understanding of philosophy; the science of the time. “...philosophers [must] become kings...or those now called kings [must]...genuinely and adequately philosophize” [15]. Fischer [9] applied the less elaborate term “movement scientists” to describe those capable of generating meaningful policy from scientific findings.

Environmental research often results in controversial recommendations that invite strong challenges. Bickerstaff [16] speaks of a ‘halo effect’ whereby the general public refuses to acknowledge the influence of their own activities as they perceive them to be insignificant. They instead place blame on factors they feel they have no control over. Environmental research relating to aviation is particularly susceptible to challenge because it is known that its environmental impact is projected to increase. Aviation is therefore subject to contest from Governments, NGOs, local councils, local residents, the

general public as well as a host of other industries looking to divert attention.

The problem with providing scientific guidance at all stages therefore stems not from the scarcity of “movement scientists”, but that it is not feasible to provide continual expert guidance in most scenarios, especially those that include the public as a stakeholder. Consequently, there is the need for methods of presentation that anticipate the effects of the policy process and are capable of bridging this divide between experts and the lay audience without expert intervention.

There is an increasing number of advocates for the benefits of a “civic science” for environmental research [17,18,19,20]. Civic science is the use of science for policy making through the involvement and understanding of society as a whole. Shannon [20] summarized: “*civic science involves scientists as citizens and citizens as lay scientists in a process in which knowledge production is integrated with and therefore cannot be separated from [...] the moral effects of political deliberation and choice*”. Social issues are not treated as separate issues but encapsulated in the research at conception. This holistic approach complements the concept of sustainability but means that it becomes essential that information is presented in a format that supports the lay participant.

III. USING VISUALIZATIONS IN POLICY

Visualization techniques are already used in a number of scientific fields to convey complicated data to the non-expert. The best format in which to present research findings will depend entirely on the anticipated audience. Engel-Cox [4] lists five basic formats of visualization: graphical, symbolic, metaphorical, photographic and quantitative.

Graphical visualization refers to the basic graphics one would find in standard reporting. These are obviously essential in a technical context but large, multidimensional data sets pose a number of problems, and their visualizations (in the form of charts, tables, graphs etc...) can end up complex, confusing, and difficult to read.

Symbolic visualizations refer to those that carry particular meanings. This could be as simple as a slogan or logo (e.g. the Fair Trade logo) or even an event (e.g. the 2005 Live 8 event intended as a visual spectacle to encourage support for the Make Poverty History campaign).

Metaphorical visualizations are those that explain one concept in terms of another. This is commonly used by news channels to quantify something a viewer would not normally understand in terms of something that they can relate to (e.g. explaining the number of calculations a computer performs in one second in terms of the number of man hours required to achieve the same result). Metaphorical visualizations also encompass virtualization tools like Google Earth. This tool allows many ‘layers’ of separate information (satellite images, terrain maps, street and road networks etc...) to be combined into a cohesion that represents the world we live in.

Photographic visualization is the use of real world images and has been used to great effect in cases where the audience may not know anything about a subject or feel that they have

no reason to associate with it. Some of the most striking examples of how photographic visualizations have been used to try and influence policy decisions come from NGOs, who commonly use images or short films to make the lay public aware of the importance and severity of their causes which in turn generates public support. If successful, this increasing support creates a motivation for decision makers to take action.

Quantitative visualization describes images that have additional characteristics that represent numerical data, such as contour maps for local air quality or a world map in which the area of a country is linked to oil consumption. Quantitative visualizations allow potentially complicated numerical data to be displayed in context in a format that requires very little understanding of the underlying science.

Perhaps the most obvious field in which visualization has been exploited for several decades is meteorology, where computer graphics have now been routinely used to display weather forecasts for several decades. In this time, graphics intended for a diverse audience have evolved from simple 2D quantitative maps displaying isobars and cloud patterns into complex 3D and animated virtual earth models which combine all 5 forms of visualization into a single, easily understood presentation of a complex, 4D dataset [21].

Environmental fields are also increasingly turning to augmented visualizations to present their findings. One recent application has been in attempting to present the anticipated impacts of global warming. Because the effects of climate change materialize so slowly, most people are unaware that they are happening around them. Furthermore, most are aware that the dramatic effects of climate change will not occur during their lifetime. This can make it difficult to inspire people into taking action. Dockerty [22] proposed a method influencing agricultural policy by presenting photorealistic images of the local landscape that have been modified to show the likely future effects of climate change (e.g. change in vegetation and land use) for various policy scenarios. Visualizations of this type are intended to illustrate the potential implications of different policy decisions on the local environment in an attempt to motivate stakeholders into making decisions that better consider the wider environmental implications. Sheppard [23] commented that there were perhaps several scenarios that warranted the use of visualizations that were specifically extreme enough so as to sway stakeholders into taking action.

However, Lowe [24] has highlighted the potential negative consequences of the exaggeration of negative effects. He found that films like “The Day After Tomorrow”, which take scientific theory and exaggerate the consequences for dramatic effect, have the effect of reducing people’s belief that such extreme consequences are possible results of climate change. It is therefore essential to ensure that, however simplified visualizations maybe, they retain a firm scientific grounding.

IV. INTERACTIVE VISUALIZATION

So far, we have touched only on methods of presenting findings directly to an audience. Yet, dramatic increases in the graphical capabilities of personal computers have resulted in the increasing use of tools that allow interactive data



Figure 2. Airport terminal buildings can be rendered in 3 dimensions and color coded according to function

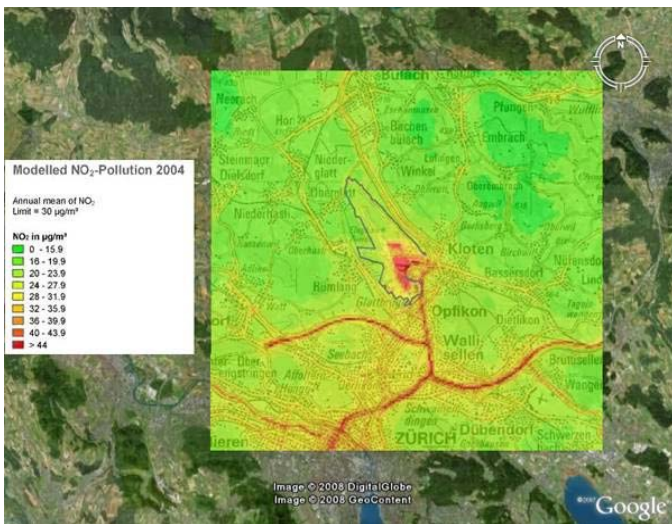


Figure 3. Image overlays can be viewed alongside streamed content. This example shows high NO_x concentrations that align with roadways



Figure 4. Additional information can be embedded into the visualisation. Here, a point marking the location of an air quality sensor contains additional information when clicked

presentation. Haase [21] noted that complex data can be understood much faster if the recipient has control of all spatial dimensions during visualization. There is now a vast range of software packages devoted to the display of 3D and 4D data. They are particularly useful in cases where the data to be visualized is simulated and therefore already stored electronically. A problem does exist though, in that there are many simulation tools and subsequently a large number of bespoke data formats. This means that a lot of data sets can only be visualized using bespoke applications or are tied to expensive, third party package (e.g. ArcGIS). In such cases, the costs associated with acquiring licenses for these packages and the time required to learning how to use them can be significant. This means interactive visualization of such data sets is often limited to experts.

Furthermore, if visualizations are tied to third party applications, the project is bound to the vendor’s development schedule. This could introduce unexpected compatibility problems at any time. Similarly, research projects who design and maintain their own 3D and 4D virtualization applications must deploy significant time and specialization. It is therefore not surprising that a number of bespoke applications often lack the functionality of third party alternatives or are simply so constrained by the availability of resources that they fail to keep up with changes in technology.

A possible solution is to integrate software that is either Open Source or freeware. Open Source tools, such as graphics engines, are community maintained and free to use (according to one of several general public licenses). The benefits of using an open source graphics engine is that they offer the developer immediate access to the latest graphical techniques with no personal development required, leaving them the flexibility to focus on integration of the features they need for their own applications. Examples of popular open source graphics engines include ORGE 3D, Genisis3D and jmonkeyengine to name but a few.

Freeware refers to applications that are developed and distributed free of charge. Although using freeware still means a developer has limited control over future functionality, the fact that they are free means that a data set can be viewed by anyone with a sufficiently capable computer.

Which freeware application is appropriate could be a factor of license, application, functionality, support and ease of use; and there are many to choose from. Some of these are developed and maintained by some of the largest organisations in their field. For example, for visualizing data on a global, regional or local scale, one could choose from World Wind (developed and maintained by NASA), ArcGIS Explorer (developed and maintained by ESRI), Virtual Earth (designed and maintained by Microsoft) and Google Earth (developed and maintained by Google), which represent a tiny sample of those available.

Many applications make use of the fact that most people are now web enabled. The internet has long been seen as an ideal environment to publish data where interested parties are able to view data interactively and at their leisure [25]. Many of these visualization applications run within a web browser (e.g. Google Maps and FlashEarth) while large numbers of other



Figure 5. Files can be made to show data on local, regional and global scales. Here we see an example of trans-Atlantic flight paths from Brussels to Toronto

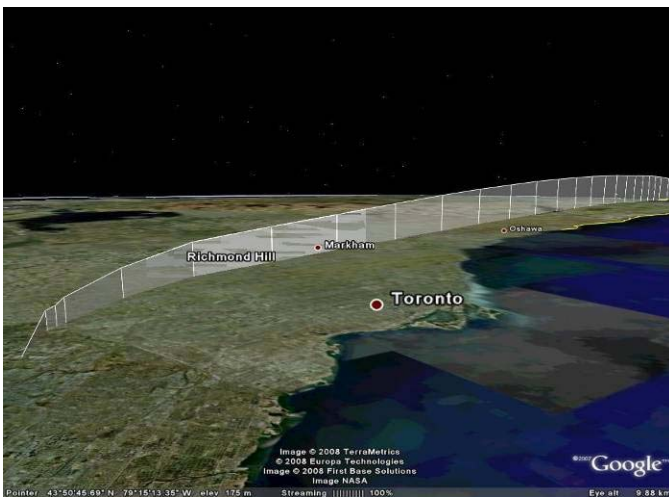


Figure 6. Zooming in and tilting the viewing access allows the profiles to be viewed with altitude and great circle path clearly displayed

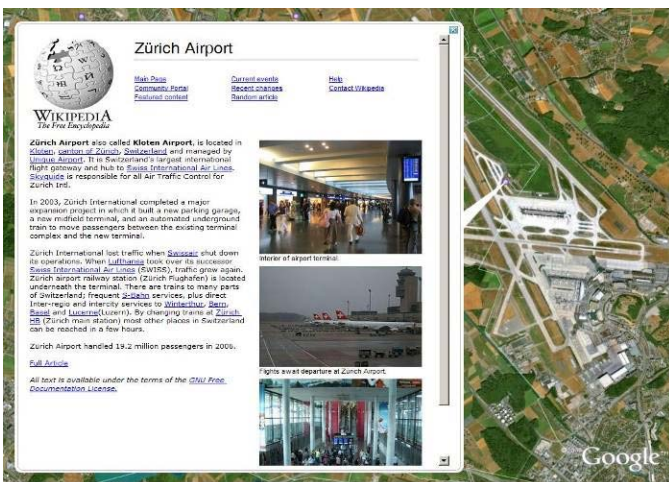


Figure 7. Content can be streamed into the virtual environment from any external server using basic HTML of Network Links. This image shows a viewer calling on additional background information on Zurich Airport from the popular free encyclopaedia 'Wikipedia'

applications establish network links to stream the content being displayed directly from remote servers. There are many benefits from streaming content to a viewer. Initially, the dataset remains under the control of its creator. That is to say, when the dataset is updated on the server, all viewers will immediately be able to stream the most up-to-date information. A second benefit comes from the fact that streaming content means that the viewing applications themselves can be made into extremely small downloads. This makes the application more accessible as only the most enthusiastic are prepared to wait for a very large application to download.

V. A VISUALIZATION EXAMPLE USING GOOGLE EARTH

Among the many virtualization tools currently available, we consider Google Earth to be particularly suitable for the visualization of aviation data. Presented below is an illustration of how Google Earth could be used to present complex, 4D, aviation research data to a lay audience. The reason for choosing this application over its rivals is that the application is already familiar to millions of users and is used by several prominent media companies (including the BBC and CNN) to display geographical information to a non-expert audience. Google Earth is also being used in many other fields to simplify the visualization process.

A. 4D Content

One of the fundamental problems with presenting paper or electronic reports is that display is constrained to a static 2D presentation. This can make it extremely difficult to present findings that are three-dimensional (e.g. aircraft flight paths) or that are time evolving (e.g. local air quality or noise contours). Virtualization models do not suffer from this limitation.

Google Earth allows user-generated content to be displayed within the virtual environment alongside the huge amount of continuously improving geographical information it streams directly from the Google servers (e.g. satellite imagery, road maps, place names, landmarks). User-generated content is written in a very simple, plain text XML format. Features such as points, lines, areas, volumes, image overlays, pop ups and dynamic, streamed content (e.g. embedded web pages, videos or real time data) are all supported.

A viewer is given full control over the height, direction, viewing angle and perhaps most importantly the time they spend viewing each piece of data. This allows the viewer to move through and around the data set at a speed at which they are comfortable. By adding point markers, lines, areas and volumes to the virtualization, features like sensor locations, roadways, runways, flight paths, buildings, areas of interest etc... can all be added to the data set to achieve a much higher level of understanding of their spatial relevance.

Figure 2 shows a mock-up with a number of main terminal buildings and hangers of Zurich Airport. In this instance, the buildings have been colored so as to differentiate between functions (yellow for terminal buildings, blue for aircraft hangers). The viewer is able to toggle all user-generated feature on and off if they care to reduce the number of obstructions from the view.

Figure 3 shows how image overlays can be integrated into the view. In this case, a quantitative map indicating simulated NO_x concentrations is presented above Zurich Airport. This is supported by a graphic that is fixed relative to the viewing window which provides a scale for interpretation.

The Google Earth environment is also able to streaming and displaying complex terrain information from the internet. At present, this data is not accessible for use outside of the application (e.g. as shape files for use in emission dispersion models), however it does add an additional component to the model that enhances the user's ability to relate the data presented with the real world.

B. Augmented Visualizations

When presenting data in a paper report, there is only so much information that can be presented within a single graph, image or table before clarity will be lost. In cases where there is a large amount of data to present, this may require several pages to explain a single scenario in sufficient detail.

Using an interactive visualization tool like Google Earth, the same effect can be achieved by augmenting a single basic data set with additional information. The viewer can recall if and when they choose to view this additional information. Figure 4 shows how additional information can be added to the basic, user defined, and geometric additions - in this case a point marker indicating the location of an air quality sensor. Clicking on the marker brings up additional information about the type of receptor, including pictures of the sight and links to published concentration values. Having this information available but displayed only on request allows the virtualization to retain clarity.

Figures 5 and 6 show a number of trans-Atlantic flight paths from Brussels to Toronto. In a paper report, one would be able to do little better than the 2D image shown in figure 5. However, the virtualization allows these flight paths to be viewed on a tilted axis, exposing the additional component of height. Furthermore, each flight path can be extended to ground, showing the grand circle path of each route.

Google Earth allows all basic markers to be augmented with formatted text, images, external web links and even embedded videos, either from within the 3D environment or via links added to the side bar. Subsequently, a great deal of information can be contained within a single file that a viewer can explore intuitively in an uncluttered fashion.

C. Network Links

A feature that is of great use in virtual environments (like Google Earth) is the ability to stream content dynamically from remote servers. This allows data to be published and maintained from a single central location, and have that newest data sets automatically feature within the viewers virtual world.

An area where a centralized dataset could provide an immediate contemporary benefit would be in storing airport layouts. Currently, defining the locations of runways, taxiways, gates, service roads and buildings must be done for every new simulation or model. A centralized dataset that is continually improved and/or extended by the community would mean that everyone would have access to the latest airport definitions in

XML format. These definitions could also be parsed and used by other applications, such as flight path data.

Perhaps of more interest for many aviation applications is the ability to present information in real time (4D data). Google Earth already offers support for GPS tracking information to be integrated in real time within the virtual environment, and a number of commercial services are already offering support for real time aircraft tracking (e.g. fbweb.com and aereoseek.com). If real time data is not available, users can link their data sets to a timeline that features on the viewers' machine. Using video-like start, stop and pause commands or a slide bar, data sets and the corresponding visualizations are streamed from your remote server, presenting a 4D viewing environment.

A final example of how using the internet connectivity of virtualization tools can be used to improve visualizations is shown in figure 7. The Google Earth environment already provides links to a wide range of additional data sources that will automatically provide the viewer with a host of supplementary data, supporting and background reading, images, 3D buildings and videos. Presenting your data alongside other information sources means that the viewer has the option to support his or her understanding without the bias one would associate with a presentation authored by a single person or organisation. This adds an additional degree of integrity to the findings. The example shows the viewer pulling additional information on Zurich Airport from the popular community encyclopedia 'Wikipedia', although the application itself has an inbuilt web browser allowing access to any online site.

VI. CONCLUSION

This paper has investigated a number of problems that exist in generating and presenting research data capable of providing information that can be useful during the policy generation process. It has been found that there remains a number of barriers between the science and policy communities, and that the complexity of these barriers is continually changing as the general public are increasingly encouraged to participate in the decision-making process.

Having noted some of the causes of these barriers and the effects this can have on interactions between stakeholders, we summarised a few of the problems that face the researcher when presenting complex, multi-dimensional data to non-expert audiences. With an appreciation of how data can change as it moves from one party to another, we further explored the nature of visualizations and how they can be used to present complex data in an intuitive manner to non-experts. It is seen that no single method suits all audiences. However, recent developments in virtualization software allow for several formats to be combined into a single, interactive environment.

Although many virtualization tools exist, Google Earth has been shown to be particularly suitable for the visualization of aviation data because of the simplicity of the data format required, its simple and intuitive interface, its ability to display all data formats (including video), because data can be viewed from local, regional and global scales and the fact that data can be streamed to the viewer from any remote server.

Although Google Earth is not a replacement for scientific reporting, the paper has presented a number of scenarios within the Google Earth environment that would be extremely difficult and/or time consuming to present with sufficient understanding in a standard document.

In closing, we conclude that changes in the policy-making process facilitate a need for simplified data presentation and that all but the biggest research projects, who can afford to develop bespoke viewing software, would be unwise to overlook the potential of using visualization tools like Google Earth for presenting to a lay audience.

ACKNOWLEDGMENT

This work is supported by EPSRC and EUROCONTROL and has been completed as part of the Brunel University Engineering Doctorate in Environmental Technology.

REFERENCES

- [1] IPCC, "Aviation and the global atmosphere." J. E. Penner, D. H. Lister, D. J. Griggs, D. J. Dokken, and M. McFarland, eds., Cambridge University Press, 1999.
- [2] Upham, P., Thomas, C., Gillingwater, D., and Raper, D., "Environmental capacity and airport operations: current issues and future prospects." *Journal of Air Transport Management*, 9, 2003, pp.145-151.
- [3] Upham, P., Raper, D., Thomas, C., McLellan, M., Lever, M., and Lieuwen, A., "Environmental capacity and European air transport: stakeholder opinion and implications for modelling." *Journal of Air Transport Management*, 10, 2004 pp.199-205.
- [4] Engel-Cox, J. A., and Hoff, R. M., "Science-policy data compact: use of environmental monitoring data for air quality policy." *Environmental Science & Policy*, 8, 2005, pp.115-131.
- [5] Sarewitz, D., Pielke, R. A., and Byerly, R., "Prediction in Science and Policy", Island Press, 2000, Washington, DC.
- [6] Alcamo, J., Shaw, R., Hordijk, L., "The RAINS Model of Acidification: Science and Strategies for Europe". Kluwer Academic Publishers, 1990, Dordrecht.
- [7] Price, D. K., ed. "The Scientific Estate", The Belknap Press of Harvard University Press, Cambridge, MA, 1965.
- [8] Lee, K., "Compass and Gyroscope: Integrating Science and Politics for the Environment", Island Press, 1993, Washington, DC.
- [9] Fischer, F. Citizens, "Experts and the Environment", Duke University Press, 2000, Durham.
- [10] Cullen, P., "The turbulent boundary between water science and water management." *Freshwater Biology*, 24, 1990, pp. 201-209.
- [11] Sundqvist, G., Letell, M., Lidskog, R., "Science and policy in air pollution abatement strategies", *Environmental Science & Policy*, 5, 2002, pp. 147-156.
- [12] Norse, D., and Tschirley, J. B., "Links between science and policy." *Agriculture Ecosystems & Environment*, 82, 2000, pp.15-26.
- [13] IPCC, "Climate change 2007: Synthesis report – a summary for policy makers" 2007.
- [14] Anderson, J. E., "Public Policy Making", Houghton Mifflin, 2002.
- [15] Plato., "The Republic", Penguin Classics; 3rd edition (31 May 2007), 360BC.
- [16] Bickerstaff, K., "Public understanding of air pollution: the 'localisation' of environmental risk." *Global Environment Change*, 11, 2001, pp.133-145.
- [17] Bagby, K., and Kusel, J., "Civic science partnerships in community forestry: building capacity for participation among underserved communities." Pacific West Community Forestry Center, 2003.
- [18] Clark, J. S., Carpenter, S. R., Barber, M., Collins, S., Dobson, A., Foley, J. A., Lodge, D. M., Pascual, M., Pielke, R., Pizer, W., Pringle, C., Reid, W. V., Rose, K. A., Sala, O., Schlesinger, W. H., Wall, D. H., and Wear, D., "Ecological Forecasts: An Emerging Imperative." In: *Science*, 2001. pp. 657-660.
- [19] Park, P., "People, knowledge, and change in participatory research." *Management Learning*, 30(2), 1999, pp.141-157.
- [20] Shannon, M. A., and Antypas, A. R., "Civic science is democracy in action." *Northwest Science*, 70(1), 1996, pp.66-69.
- [21] Haase, H., Bock, M., Hergenröther, E., Knöpfle, C., Koppert, H. J., Schröder, F., Trembilski, A., and Weidenhausen, J., "Meteorology meets computer graphics - a look at a wide range of weather visualisations for diverse audiences." *Computers & Graphics*, 24, 2000, pp.391-397.
- [22] Dockerty, T., Lovett, A., Appleton, K., Bone, A., and Sunnenburg, G., "Developing scenarios and visualisations to illustrate potential policy and climate influences on future agricultural landscapes." *Agriculture Ecosystems & Environment*, 114, 2006, pp.03-120.
- [23] Sheppard, S. R. J., "Landscape visualisation and climate change: The potential for influencing perceptions and behaviour." *Environmental Science and Policy*, 8, 2005, pp.637-654.
- [24] Lowe, T., Dessai, S., Doria, M., Haynes, K., and Vincent, K., "Does tomorrow ever come? Disaster narrative and public perception of climate change." *Public Understanding of science*, 15, 2006, 435-457.
- [25] Lesjak, M., Boznar, M., and Mlakar, P., "Internet applications as a link between environmental information systems and public." In: *Seventh International Conference on Air Pollution, Computational Mechanics Publications*, 1999, California, USA.

Peak oil, fuel costs and the future of aviation

Richard Klophaus

Centre for Aviation Law and Business
Trier University of Applied Sciences
Trier, Germany
klophaus@umwelt-campus.de

Abstract—This research investigates the impact of global oil production peak transmitted via soaring fuel prices on future air traffic. The paper analyzes the short-term impact of higher fuel prices on airline operating costs, passenger fares and demand for short-haul and long-haul services. Results indicate that the rate of air traffic growth constrained by scarcity of kerosene is much lower - and may even be negative - than unconstrained air traffic growth. Services offered by low-cost carriers and long-haul services are most adversely affected. It is also contended that a strong increase in fuel prices outweighs the potential impact of proposed emission trading systems for the aviation industry. Looking beyond the peak in oil production the paper provides a brief discussion of potential substitutes for petroleum kerosene as jet fuel.

Keywords-environment; peak oil; airline operating costs; air fares; traffic growth

I. INTRODUCTION

The aviation industry is probably the economic sector most depending on fossil fuels besides the petrochemical industry itself. Today, commercial aviation is characterized by growing passenger numbers and cargo volumes as well as expanding airport and airline capacities. According to a recent Airports Council International (ACI) forecast the current number of 4.5 billion passengers worldwide is expected to reach 9 billion by 2025[1]. Similar forecasts of air traffic growth are issued by manufacturers of commercial jetliners like Airbus and Boeing and proliferated by other sources including public agencies and academia. Based on these forecasts of air traffic growth the demand for kerosene is bound to grow.

The prospering world economy leads to a soaring demand for crude oil aside from aviation. OPEC projects a growth of oil demand from 84 million barrels per day (mb/d) in 2005 to 118 mb/d in 2030 [2]. In 2007 crude oil prices have been boosted by growing worldwide demand. By the time peak oil is reached and half of the global oil resources are exploited, costs for oil extraction will rise and keeping up the production level will become increasingly difficult. The depletion of the world's oil reserves results in an upward price trend for crude oil and also for its refinery products such as kerosene. To our knowledge the potential impact of peak oil on commercial aviation has only recently been addressed [3]. This paper investigates the short-term economic impact of higher crude oil prices on fuel costs, air fares and air passenger demand. The analysis uses the methodology developed for quantifying the impact of emission trading on aircraft operators [4]. With regard to the relationship between kerosene prices and airlines'

fuel costs different fuel hedging scenarios are considered. The paper indicates that the rate of air traffic growth constrained by scarcity of kerosene is much lower - and may even be negative - than unconstrained air traffic growth, especially leading to a strong reduction of demand for leisure traffic and long-haul services.

The paper is structured as follows: Section 2 contains considerations on peak oil including an overview of predictions on the time of global oil production peak and forecasts for the future price of crude oil. Section 3 examines the short-term economic impact, i.e. assuming one-year horizon, of higher fuel prices on airline costs, ticket prices and passenger demand for short-haul and long-haul services. Short-haul is further differentiated into routes operated by full service network carriers (FSNCs) and low-cost carriers (LCCs). In the long run the aviation industry has to look beyond the fuel-efficient '3 liter aircraft' and search for new groundbreaking ways to become less dependent on fossil fuels. Hence, Section 4 gives an overview of current research directions in the fields of future aircraft technology and evaluates potential alternative fuels to kerosene. The closing Section 5 summarizes the paper's results and concludes that peak oil has the potential to stop and even reverse long-term air traffic growth.

II. PEAK OIL AND FUTURE FUEL PRICES

End of Oct. 2007 the spot price of Brent-Europe crude oil reached for the first time \$90 a barrel. This reflects that world oil demand has continued to grow much faster than oil supply but also ongoing geopolitical risks, OECD inventory tightness, worldwide refining bottlenecks and speculative trading. \$90 a barrel is about 50 percent more than Oct. 2006. In real terms, adjusted for inflation, oil is at its highest price since the early 1980s when it hit its peak following the Iranian Revolution and the beginning of the Iran-Iraq war (Fig. 1).

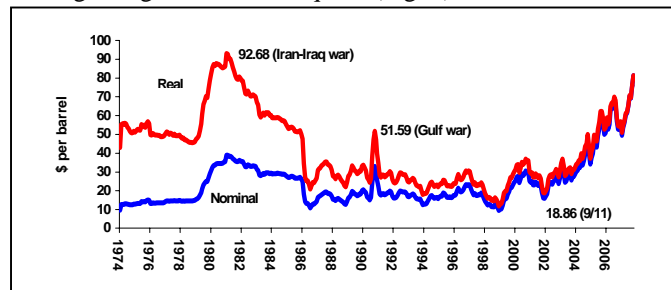


Figure 1. Real and nominal crude oil prices, 1974-2007 (real prices in 2007 dollars).

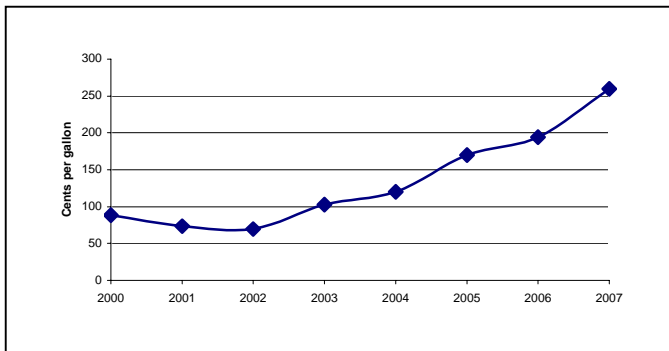


Figure 2. Kerosene-type jet fuel prices in Rotterdam, 2000-2007 (2000-2006: annual average price, 2007: spot price in Oct.).

Kerosene is produced by distilling crude oil. Hence, the product price of kerosene is closely linked to crude oil prices. End of Oct. 2007 the spot price for kerosene-type jet fuel in Rotterdam was about 260 cents per gallon (Fig. 2). This translates into close to \$110 per barrel. This spread of approx. \$20 on the spot price of Brent-Europe crude oil reflects the gross refining margin. According to the International Air Transport Association (IATA) fuel outranked labor as largest single cost item in the global airline industry in 2006 [5]. Fuel accounted for 25.5% of total operating costs in 2006 compared to 13.6% in 2001. The rise in fuel costs reflects a sharp increase in the price of crude oil but also a widening in the refining margin.

Counteracting soaring fuel costs airlines intensified their efforts to improve fuel efficiency and to obtain cost savings in non-fuel cost items. In particular, labor productivity has improved resulting in a falling labor share of airline operating costs to 23.3% in 2006. The 25.5% fuel share of total operating costs calculated by IATA rests upon an average price of jet fuel of approx. \$82 per barrel. With a kerosene price of \$110 per barrel at the end of Oct. 2007 the share of fuel costs further increases in the airline industry even if fuel hedging contracts lock a percentage of the fuel purchases at lower prices. Airlines react by increasing average prices for passenger tickets and rates for air cargo. For example, European airlines like Air France-KLM (AF-KLM) or Lufthansa raised their fuel surcharges for passenger tickets several times in 2007 (see Section 3).

Crude oil prices react to the balance of demand and supply. Hence, the current spiking of fuel prices creates concerns about a global shortage of future oil supplies. If actors in the oil market expect a shortage of oil supplies, oil prices increase before a shortage actually occurs. This is reflected in contracts for future deliveries of crude oil, called futures. In Oct. 2007 the prices of crude oil futures soared to all-time highs after Energy Information Administration (EIA) indicated a drop in commercial US crude inventories to the lowest level in two years. EIA providing the official energy statistics from the US

government publishes an International Energy Outlook [6]. In the so-called reference case of its most recent outlook, EIA projects a growth of world consumption of petroleum products by more than 40% from 84 mb/d in 2005 to 118 mb/d in 2030, an average annual growth rate of 1.4%. The demand of China grows much stronger with a forecasted rate of 3.5%. Strong growth is also projected for the other non-OECD economies with the exception of Russia. In addition to the reference case EIA also analyzes high and low oil price cases. Despite considerable differences between oil prices the demand projections for 2030 do not vary substantially indicating that long-term demand is relatively inelastic to oil price changes. It is a question whether the suggested lack of demand elasticity remains a valid proposition once production of crude oil falls short of demand due to finite oil reserves. If global crude oil production cannot be increased even with mounting oil prices there has to be a demand adjustment.

The US Government Accountability Office (GAO) examined more than twenty studies on the timing of the peak in oil production conducted by government authorities, oil companies and oil experts (Fig. 3) [7]. According to this meta-analysis most studies estimate peak oil sometime between now and 2040. The range of estimates on the timing of peak oil is wide due to multiple and uncertain factors including (1) the amount of oil still in the ground, (2) technological, cost and environmental challenges to produce that oil, (3) political and investment conditions in countries where oil is located and (4) the future global demand for oil. Some of the studies cited by GAO consider only the peak in conventional oil, while other studies include non-conventional sources of oil – oil sands, heavy and extra-heavy oil deposits and oil shale. The production process of oil from non-conventional sources is more costly, uses larger amounts of energy and presents environmental challenges.

According to the recently published energy outlook by the International Energy Agency (IEA) the oil production in most countries outside the Middle East has already peaked or will do in the near future. Approximately 70% of the estimated remaining global oil reserves are located in politically insecure regions respectively are kept under OPEC control [8]. OPEC statements concerning strategic oil reserves may be questioned. Oil production represents a major sector of economy in OPEC countries, and the admission of declining oil reserves harms their financial standing and political importance. The number of discovered oil fields decreases year by year. About 42,000 oil fields have been discovered until today, the 400 largest represent about 75% of global oil reserves. The annual worldwide crude oil consumption exceeds the amount of discovered reserves since 1981. The predominant part of extracted crude oil nowadays derives from oil fields discovered in the 1970s [9].

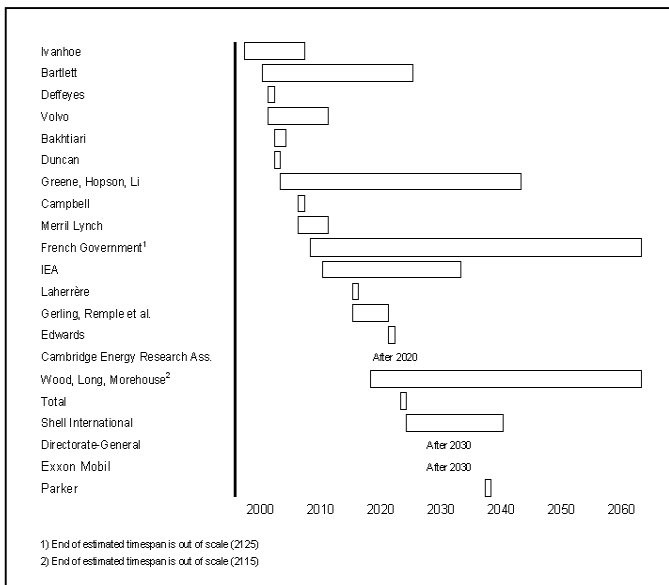


Figure 3. Estimates of the timing of peak oil (Source: GAO [7])



Figure 4. Direct economic impact of higher fuel prices

The paper assumes no disruptions in normal economic activity and that the overall political and economic setting for commercial aviation remains intact. The time horizon is only one year allowing to differentiate a scenario with high level of airline fuel hedging and a scenario with no fuel hedging a year ahead. This short-term approach justifies not to account for fuel efficiency measures and also to use current operational and cost data.

The volatility of kerosene prices is an important issue for the airline industry. In 2006, the fuel consumption of Lufthansa (LH) amounted to 6,940,587 tons equivalent to 54,564,363 barrels (1 barrel = 159 liter, 1 liter kerosene = 0.8 kg) [9]. Without fuel hedging a fuel price rise by \$1 a barrel increases LH's operating costs by more than \$50 millions. Fuel hedging is often touted as the solution to this problem.

Fuel hedging means stabilizing fuel costs by locking in the costs of future fuel purchases to protect against sudden cost increases from rising fuel prices. However, fuel hedging also prevents savings from decreasing fuel prices and might even lack a theoretical justification [11]. In practice, fuel hedging strategies vary significantly between airlines, some opting to hedge their entire fuel needs, while others leave themselves exposed to fluctuations in fuel costs. The lack of fuel hedging might not be strategy-driven but simply the result of insufficient cash or credit. According to Morrell and Swan, most airlines typically hedge between one- and two-thirds of fuel costs and look forward six months in their hedging, with few hedges more than a year ahead. Hence, the fuel hedging policy of AF-KLM with regard to the time period covered seems to be rather exceptional (Table I).

Finite oil resources and global economic growth lead to an upward trend for crude oil prices. However, due to multiple and uncertain factors concerning near-term and long-term oil production and the future development of global oil demand it is not surprising that forecasts on future prices show a wide range. EIA differentiates three world oil price cases [6]. In the high world oil price case, world oil prices climb from \$43/barrel to \$100/barrel in 2030 (all values in 2005 real dollars). In the low price case, oil prices moderate fairly quickly to \$49/barrel in 2010 and then further to \$34/barrel in 2015 and remain at that level through 2030. The reference case oil prices rise steadily after 2015 to \$59/barrel in 2030.

For the purpose of this paper the 50% increase of world oil prices in the one-year period Oct. 2006 – Oct. 2007 from around \$60/barrel to \$90/barrel is simply prolonged for another year. This results in \$135/barrel in Oct. 2008. This figure is not to be considered as another oil price forecast but only as starting point for the analysis of the possible short-term impact of higher kerosene prices on commercial aviation. \$135/barrel is beyond the oil price ranges of most recently published short-term forecasts. In November 2007, EIA expects the average West Texas Intermediate (WTI) crude oil price in 2008 at nearly \$80/barrel [10]. However, most short-term oil price forecasts published in recent years underestimated the actual oil price development. For example, in January 2007 EIA also projected a WTI crude oil spot price of \$65 for the 3rd and 4th quarter of 2007.

III. ECONOMIC IMPACT OF SOARING KEROSENE PRICES

This section considers the short-term response of passenger demand for air travel resulting from the cause-and-effect chain depicted in Fig. 4. This cause-and-effect chain is referred to as direct impact of higher fuel prices. In addition, there are potential indirect impacts of soaring fuel prices, for example, a reduction in air travel demand resulting from lower disposable income of households or an increase in airlines' operating costs for other cost items than fuel due to inflation.

TABLE I. AF-KLM FUEL HEDGING POLICY

Year	2007-08	2008-09	2009-10	2010-11
Forecasted spot price (Brent, \$/barrel)	79	83	79	78
Hedged consumption (%)	77	67	51	31
Average hedged price (Brent, \$/barrel)	62	61	68	69
Final average price (Brent, \$/barrel)	66	68	73	75

Source: www.airfranceklm-finance.com, visited 26th October 2007.

AF-KLM forecasts a spot price of \$79/barrel in 2007-08. The average fuel price for AF-KLM is locked with fuel-hedging contracts that secure 77% of the airline's fuel requirements in 2007-08. Even for 2010-11 31% of the fuel needs are hedged. The average hedged price in 2007-08 with \$62/barrel is only two thirds the spot price of crude oil at end of October 2007 of more than \$90/barrel. By using the futures markets AF-KLM managed to soften the affect of higher fuel prices but still increased its fuel surcharges on air fares several times in 2007.

From the information publicly available, AF-KLM hedges only the crude oil price. This leaves the price difference between crude oil and kerosene uncovered. The kerosene price is driven by crude oil price developments but is also influenced by other specifics, especially refinery capacities and price switches between diverse oil products. End of October, the spot price for kerosene-type jet fuel in Rotterdam was about \$110/barrel, i.e. approx. \$20 higher than the spot price of Brent-Europe crude oil.

For the purpose of this paper the 50% increase of world oil prices in the one-year period Oct. 2006 – Oct. 2007 from a level of \$60/barrel to \$90/barrel is simply prolonged for another year. This results in \$135/barrel in Oct. 2008. As already stated in Section 2 this figure is not another oil price forecast but only serves as starting point for the analysis of the short-term impact of higher kerosene prices on commercial aviation. Additionally, a constant refining margin of \$20/barrel is assumed. Hence, all further calculations are based on a spot price for kerosene-type jet fuel of \$155/barrel.

All airlines are confronted with volatility of fuel prices. Besides the structure of fleet and network different fuel hedging policies lead to a varying effect on fuel costs among airlines. In Europe, network carriers like AF-KLM and LH pass along higher fuel costs to passengers through higher ticket prices with changing fuel surcharges added to air fares. Ryanair (FR) and other European LCCs do not add fuel surcharges but increase the average fare level. The following calculations estimate the increase in ticket prices and the resulting changes in passenger demand due to rising fuel prices. As reference value to measure fuel price increases the final average price of \$68/barrel forecasted by AF-KLM for 2008-09 by end of Oct. 2007 is used (Table I). This fuel price is further on referred to as AF-KLM base. As no similar information about spot price forecasts, hedged and average fuel prices for LH and FR has been available to the authors, it is assume that the AF-KLM base is also valid for LH and FR.

The airline's fuel costs (\$/liter kerosene) at a given future time (e.g. Oct. 2008) based on the future spot crude oil price, the airline's hedged consumption, the average hedged price and the refiner margin can be calculated as follows:

$$C = ([\alpha \cdot p_h + (1-\alpha) \cdot p_s] + c_r) / 159 \quad (1)$$

with

C Future fuel costs (\$/liter kerosene),

α Share of fuel consumption hedged (%),

p_h Average hedged crude oil price (\$/barrel),

p_s Future spot crude oil price (\$/barrel),

c_r Gross refining margin (\$/barrel).

Assuming a spot price $p_h = \$135$ for crude oil in Oct. 2008, the AF-KLM fuel hedging policy ($\alpha=0.67$; $p_h=\$61$) and a refiner margin $c_r=\$20$, the fuels costs (\$/liter kerosene) for AF-KLM in Oct. 2008 resulting from (1) amount to $C=0.66$. In comparison, the AF-KLM forecast of a spot price $p_h = \$83$ for crude oil in Oct. 2008 (see Table I) leads to $C=0.56$ other things being equal. Hence, with a high level of fuel hedging fuel costs would be 18% higher than forecasted. In a scenario with fuel hedging contracts running out or no airline fuel hedging a year ahead ($\alpha =0$) future fuel costs even rise to $C=0.97$. To account for soaring fuel costs, airlines like AF-KLM increase fuel surcharges on passenger tickets.

The economic impact of higher fuel costs passed on to passengers via higher ticket prices is investigated for exemplary routes (Table II). Based on the operational data provided by [12] following routes are analyzed:

- Frankfurt (FRA) – London-Heathrow (LHR) served by LH.
- Hahn (HHN) – London-Stansted (STN) served by FR.
- LH-operated intercontinental route FRA – Singapore (SIN).

For each route the analysis differentiates the impact of fuel price increases with fuel hedging ($\alpha=0.67$) and without fuel hedging ($\alpha=0$). Results are compared with AF-KLM base of \$68/barrel. Lacking airline-specific information it is assumed that any increase in fuel costs in excess of the AF-KLM base is fully passed on to passengers via higher ticket prices, i.e. fuel cost increase equals ticket price increase. FR has a no fuel surcharge policy and accommodates higher fuel prices by increasing average ticket prices. LH increases its fuel surcharge on air fares. In Oct. 2007 LH's fuel surcharge on long-haul tickets amounted to €67 per sector and for short-haul tickets €14 per sector [13]. This fuel surcharge has been increased several times by LH and other European carriers in 2007. Table II shows fuel consumption and average passenger number per flight for the three selected exemplary routes. Based on this data, the route-specific future fuel costs per passenger can be calculated for AF-KLM base as reference value ($C=0.56$) and two fuel hedging scenarios ($C=0.66$ and $C=0.97$) as follows:

$$c_{pax} = C \cdot \frac{k}{n} \quad (2)$$

with

c_{pax} Future fuel costs per passenger (\$/PAX),

C Future fuel costs (\$/liter kerosene),

k Fuel consumption per flight (liter kerosene),

n Average passenger number per flight (PAX).

TABLE II. INCREASES IN FUEL COSTS AND TICKET PRICES DUE TO HIGHER KEROSENE PRICES

Route (carrier)		HHN-STN (FR)	FRA-LHR (LH)	FRA-SIN (LH)
Operational data	Distance flown	572	695	10,603
	Aircraft type	B 737-800	A 321-100	A 340-300
	No. of seats	189	182	247
	Avg. seat load factor	76.1%	66.9%	80.5%
	Avg. no. of passengers	144	122	199
	Fuel consumption (liter kerosene)	3,250	4,125	107,500
Fuel costs (C_{pax})	AF-KLM base	12.6	18.9	302.6
	$\alpha = 0.67$	14.9	22.3	356.5
	$\alpha = 0$	21.9	32.8	524.0
Fuel costs (€/PAX)	AF-KLM base	8.7	13.0	208.7
	$\alpha = 0.67$	10.3	15.4	245.9
	$\alpha = 0$	15.1	22.6	361.4
Avg. ticket price (per sector, €/PAX)		44	136	602
Abs. price increase (€/PAX)	$\alpha = 0.67$	1.6	2.4	37.2
	$\alpha = 0$	6.4	9.6	152.7
Rel. price increase	$\alpha = 0.67$	3.6%	1.8%	6.2%
	$\alpha = 0$	14.5%	7.1%	25.4%

As FR and LH denominate their ticket prices in Euro, fuel costs need to be converted into Euro as well. Table II uses the Dollar/Euro exchange rate €1 = \$1.45 valid end of Oct. 2007.

The short-term impact of soaring kerosene prices on fuel costs and ticket prices depicted in Table II remains relatively moderate as long as fuel hedging by airlines mitigates fuel price increases. In absolute terms the price increase for the two short-haul routes is €1.6 (HHN-STN) and €2.4 (FRA-LHR) corresponding to a relative price increase per passenger and sector of 3.6% and 1.8% respectively. The higher price increase in per cent for FR results from its significantly lower average ticket price per sector and passenger compared to LH which cannot be compensated by FR's shorter flight distance and the higher average number of passengers per flight. For the long-haul route FRA-SIN the impact is already more pronounced, with an absolute price increase of €37.2 corresponding to 6.2% in relative terms. It should be noted that even in a scenario with fuel hedging ($\alpha=0.67$) the impact of increasing fuel prices is higher than the financial burden due to the introduction of the emission trading scheme as proposed by the European Commission [12].

For the scenario with fuel hedging contracts running out or no airline fuel hedging ($\alpha=0$) the impact is much stronger. In relative terms, LH's short-haul operation is less affected than the operation of FR. The average ticket price sold by LH on FRA-LHR rises by 7.1% (€9.6) and FR's price on HHN-STN by 14.5% (€6.4). The impact on surcharges on long-haul traffic largely exceeds the impact on short-haul traffic. For FRA-SIN operated by LH an additional fuel surcharge of €152.7 would occur. Based on an average fare per passenger and sector of €602, the additional fuel surcharge represents a relative fare increase of 25.4%.

€152.7 is the additional fuel surcharge calculated for LH resulting from assuming spot prices for crude oil to rise by 50% compared to current 90\$/barrel and no softening of spot prices by fuel hedging. In principle, this paper equates short-term with a one-year horizon. In its hedging practice, LH hedges up to 90% of its planned fuel requirement on a revolving basis over a period of 24 months. In April 2007 even 70% is hedged one year ahead [13]. LH reduces its hedging ratio from this share each month by 5% leading to a growing exposure to fluctuations in fuel prices after one year. Hence, our results seem to overestimate the short-term impact of rising fuel costs on LH's fuel surcharge. However, FR has recently complained to EU over abusive increases in fuel surcharges based on spot prices on the global crude oil markets rather than hedged prices. According to FR, carriers like LH do not only increase ticket prices for passengers in lockstep with their higher fuel costs but even beyond their additional fuel costs.

As a result of shifting costs to passengers via higher ticket prices, demand for flights is expected to decrease. Table III shows how passenger demand reacts to higher ticket prices. The average price elasticity for short-haul leisure and business demand as well as for long-haul leisure and business demand is taken from a synoptic study [14], the shares of business travelers are adopted from [12] and the relative increases in ticket prices from Table II.

In the fuel hedging scenario ($\alpha=0.67$), the estimated change in passenger demand for HHN-STN - a typical short-haul flight operated by FR - is -4.7%, while for LH's short-haul FRA-LHR and long-haul FRA-SIN it amounts to -2.0% and -4.0% respectively. Passenger demand for LCCs like FR will be more negatively affected by soaring fuel prices than demand for full service network carriers like LH. The higher demand reduction for FR results from a higher relative fare increase compared to LH as well as a higher share of more price-sensitive leisure travelers. Compared to short-haul routes like FRA-LHR demand for long-haul routes such as FRA-SIN will be more affected due to the relative strong increase in ticket prices and despite lower price elasticities for long-haul travel.

Without fuel hedging ($\alpha=0$), the short-term impact on passenger demand is even stronger. The estimated change in passenger demand due to higher ticket prices for HHN-STN is -19.1%, -7.9% for FRA-LHR and -16.6% for FRA-SIN.

TABLE III. DEMAND REDUCTION DUE TO HIGHER TICKET PRICES

Route (carrier)		HHN-STN (FR)	FRA-LHR (LH)	FRA-SIN (LH)
Avg. price elasticity	Business	-0.7	-0.7	-0.265
	Leisure	-1.52	-1.52	-1.04
Share of business travelers		25%	50%	50%
Rel. price increase	$\alpha = 0.67$	3.6%	1.8%	6.2%
	$\alpha = 0$	14.5%	7.1%	25.4%
Change in demand	$\alpha = 0.67$	-4.7%	-2.0%	-4.0%
	$\alpha = 0$	-19.1%	-7.9%	-16.6%

Table III does not differentiate the relative price increases with regard to leisure and business market segments. As the average ticket price per passenger is higher for business travelers compared to leisure travelers, the reduction in leisure demand is even stronger and the reduction in business demand lower than shown in Table III.

This section only estimated the isolated effect of soaring fuel prices transmitted via higher ticket prices on passenger demand. This cause-and-effect chain corresponds to the direct impact of higher fuel prices. There will be indirect impacts of soaring fuel prices, for example, a reduction in air travel demand resulting from lower disposable incomes. Looking at the calculated short-term reduction in passenger demand of more than 15% for typical short-haul services operated by LCC as well as for long-haul services in the no fuel hedging scenario, higher ticket prices due to soaring fuel prices strongly influence commercial aviation. However, this direct impact may be compensated by other factors influencing travel demand. Fuel surcharges levied by airlines in recent years did not keep aviation from growing more than 5% annually [12]. In addition, the estimated changes in passenger demand for services offered by FR and LH have to be set in due proportion to the future growth trend in commercial aviation, especially the currently expected growth rates for European LCCs that go well beyond 5% p.a. The reduction in air travel demand caused by soaring fuel prices may only confine the overall demand increase.

IV. ALTERNATIVES TO KEROSENE AS JET FUEL

The previous results show that the rate of air traffic growth constrained by scarcity of kerosene will be much lower - and may even be negative - than unconstrained air traffic growth, especially leading to a strong reduction of demand for leisure traffic and long-haul services. Hence, the entire aviation industry has to look beyond the fuel-efficient '3 liter aircraft' and search for new groundbreaking ways to become less dependent on fossil fuels. This section provides a brief overview how to save fuel or even replace kerosene as jet fuel.

At present, aircraft and engine manufacturers improve aircraft design (e.g. blended wing aircraft) and fuel-efficiency of engines in order to reduce fuel consumption. Fuel saving strategies by airlines include shorter air routes, carrying less minimum fuel, increased fuel blending, shorter sector lengths, modern fleet, increased load factors and more efficient ground operations (e.g. reduction of ground delays). All these efforts contribute to fuel conservation by commercial aviation but do not provide a substitute to conventional petroleum kerosene.

Kerosene is considered the ideal jet fuel. First reason is its high energy content. The energy content of fuel is measured as specific energy which is the energy content per unit mass (joules/kg) and as energy density which is the energy per volume (joules/liter). The high energy content of kerosene positively affects the total size and weight of the aircraft. Operationally, the heavier the aircraft is at takeoff, the more fuel is required to lift it into the air. With regard to safety criteria, the Jet A-1 kerosene used in commercial aviation has a high flash point of not lower than 40° Celsius reducing explosion hazards and a low freezing point Kerosene also does

not contain or absorb water which means that in cold temperatures no ice crystals form that block fuel filters and ultimately lead to fuel starvation. These safety over a wide temperature range is an important selection criterion for jet fuels.

Below alternative fuels (synthetic kerosene, bio-fuels and liquid hydrogen) and a new aircraft propulsion technology (fuel cells) are presented and briefly evaluated with reference to [15], [16], [17] and [19] along following criteria: high energy content, safety, environmental impact, availability and price. Ethanol and methanol are not considered because of their unfavorable properties at jet fuel.

A. Synthetic kerosene

This is a carbon-based fuel synthesized by using a Fischer-Tropsch conversion process. According to the raw material used three types of synthetic kerosene are differentiated:

- Biomass to liquid (BTL).
- Gas to liquid (GTL).
- Coal to liquid (CTL).

Today, synthetic kerosene is only approved in commercial aviation as a blend with petroleum kerosene despite of having basically the same energy content and safety qualities. Semi-synthetic fuels (50 percent normal fuel and 50 percent synthetic fuel) for the aviation industry have been produced in South Africa since 1999 [16].

BTL is more environmentally clean than GTL and CTL as the combustion process of BTL releases carbon dioxide (CO₂) in the same quantity as the plants have absorbed from the atmosphere during their growth process. However, the CO₂ benefits of BTL must be assessed by life cycle analyses considering emissions generated by cultivation, processing and transport. BTL can be produced from almost any type of plants and offers new perspectives for farmers but implies the risk of competition with food production.

B. Bio-fuels

Bio-fuels refer to fuels derived from feedstock such as rapeseed, soybeans or algae without a Fischer-Tropsch synthesis as in the case of BTL. Bio-fuels have a somewhat lower energy content than kerosene [16]. The primary concern with the use of bio-fuels are their low temperature properties with freezing points near 0° Celsius, much higher than the maximum freezing point of petroleum kerosene (-40° Celsius). With additives the low temperature operability at cruising altitudes of bio-fuels can be improved. There are doubts that bio-fuels can be mass-produced affordably because of limited farmland [17]. For these reasons, bio-fuels are currently not considered as alternative jet fuels on their own but more suitable for blending with kerosene.

C. Hydrogen

Liquid hydrogen is the liquid state of the element hydrogen. It is probably the most commonly discussed long-term alternative to kerosene. Hydrogen provides 2.5 times the

energy per kg than kerosene but is also about four times more voluminous. Liquid hydrogen is non-corrosive. A major potential advantage of hydrogen compared to kerosene is the significant reduction of harmful emissions. The primary combustion product of hydrogen is water. A negative byproduct of its combustion is water vapor as greenhouse gas. Depending on how hydrogen is produced there are significant CO₂ emissions generated during its life cycle.

Today, hydrogen is expensive to produce and difficult to store. Due to the large volume and the requirement to cool down hydrogen to the liquid state (-253° Celsius), the cryogenic storage of hydrogen constitutes a major challenge for aircraft manufacturers. A hydrogen powered aircraft will look very different from today's kerosene aircraft. Hydrogen will not be stored in conventional wings because of pressurization and insulation requirements. The positioning of fuel tanks in the fuselage results in an enlarged fuselage or less passenger capacity. Ensuring explosion safety of cryogenic aircraft is a challenge. Hydrogen will also require a radical change in engine design. Yet the Russian aircraft manufacturer Tupolev managed these technical challenges with the cryogenic fuel aircraft TU-155 performing its maiden flight already in 1988 [18].

Hydrogen aircraft also pose a major challenge for airport infrastructure which at present is only designed for kerosene aircraft. A prerequisite for a change from kerosene to hydrogen already in the transition stage is the global availability of two parallel fueling systems at airports. Hence, a transition to hydrogen-powered aviation may take decades, especially considering the long life-span of aircrafts currently in operation.

D. Fuel cells

Fuel cells have been used in spacecrafts since the 1960's to power auxiliary engines. Experimental aircraft powered only by a fuel cell supported by lightweight batteries during takeoff and climb is on its way. A fuel cell is an electrochemical device that converts hydrogen directly into electricity and heat without combustion. Fuel cells are emission-free and quieter than hydrocarbon fuel-powered engines. The main challenge is to develop compact and lightweight electric propulsion systems with more power. Today, using fuel cell technology as primary power for a passenger airplane leads to a propulsion system several times heavier than conventional aircraft engines and still far from their efficiency. However, chilled superconducting magnets carrying electricity without resistance have been proposed that may allow for lightweight and powerful electric jet engines in the long run [19].

Table IV summarizes the pros and cons of jet fuel alternatives relative to petroleum kerosene along selected criteria. "+" indicates that the potential substitute performs better with regard to the respective criterion, "o" suggests equal and "-" worse properties compared to conventional kerosene.

TABLE IV. ASSESSMENT OF ALTERNATIVE JET FUELS AND FUEL CELLS RELATIVE TO PETROLEUM KEROSENE

Criterion		Energy Content	Safety	Environmental impact	Availability and price
Synthetic kerosene	BTL	o	o	+	-
	CTL	o	o	-	-
	GTL	o	o	-	-
Bio-fuels		-	-	+	?
Hydrogen		+	?	+	-
Fuel Cells		-	?	+	-

The assessment does not account for ground-breaking technology developments and, hence, has to be regarded as preliminary. Evaluating alternatives to petroleum kerosene in the near future, synthetic kerosene holds the greatest promise as it basically can be used in existing aircraft either alone or blended with petroleum kerosene. The main problem for synthetic kerosene with the exception of BTL is the large amount of CO₂ generated during production. In the long run, hydrogen seems to be a promising candidate to replace kerosene if safety standards of civil aviation can be secured but asks for a fundamental change in aircraft design and new ground infrastructure at airports.

V. CONCLUSIONS

Conventional wisdom in commercial aviation is that global air traffic will continue to grow in the coming decades. This implicitly assumes no constraint in traffic growth due to finite oil resources. This is in stark contrast to studies that estimate peak oil sometime between now and 2040.

This paper analyzed the short-term economic impact of soaring fuel prices on commercial aviation. The time horizon was only one year from now allowing fuel hedging by airlines to balance increasing spot prices. The analysis was restricted to the direct effect of higher kerosene prices on operating costs, fare levels and passenger demand. Indirect effects on passenger demand resulting from a reduction of purchasing power, an increase in unemployment and higher costs for other input factors besides kerosene were not considered. The analysis also ignored possible political crisis and economic shocks for oil importing countries forced to spend significantly more on their energy purchases. Hence, the scope of this paper has been somehow limited.

However, the limited approach already shows that the rate of air traffic growth constrained by scarcity of kerosene will be much lower - and may even be negative - than unconstrained air traffic growth, especially with regard to price-sensitive leisure demand. Services offered by low-cost carriers and long-haul services will be most adversely affected by higher fuel prices. Further, the impact of soaring fuel prices largely exceeds the impact of the proposed EU emission trading system (ETS) for the aviation industry. This leads to the question whether ETS is actually needed in view of finite

supplies of fossil fuels that may restrict or even terminate air traffic growth. In addition, high fuel prices are a strong incentive to use more fuel-efficient engines, to optimize minimum fuel policies, to improve air routes and ground operations, etc., in the same direction as intended by ETS.

The fuel price development will also influence the typical air service pattern, for example, there may be a renaissance of technical stops for re-fueling on intercontinental routes or more point-to-point traffic in order to avoid fuel burning detours via hubs. To avoid high fuel costs, regional carriers have already replaced regional jets on some routes by turboprops. The in-depth analysis of the relative economic benefit of competing services patterns and the use of turboprops instead of regional jets in times of high fuel prices is an interesting issue for further research.

Peak oil will happen, the open question is when. It is a problem that may soon replace the global warming debate in commercial aviation as jets are not as fuel-flexible as ground vehicles. Aviation industry and politicians better face the long-term implications of finite oil resources. Airline and airport managers should no longer exculpate themselves by referring to future air frame designs to be developed by aircraft manufacturers or increased blending of other fuels with kerosene by the petrochemical industry. More research than today should be devoted to the economic evaluation of kerosene substitutes in combination with the associated future requirements for airline fleets and airport infrastructure.

REFERENCES

- [1] Airports Council International (ACI), Global Traffic Forecast 2006 – 2025, Executive Summary, Geneva, 2007.
- [2] Organization of the Petroleum Exporting Countries (OPEC), World Oil Outlook, Paris, 2007.
- [3] A. Kuhlman, “Peak oil – impacts on commercial aviation,” *Airlines*, e-zine edition, Issue 38, 2007, pp. 1-5.
- [4] J. Scheelhaase and W. Grimme, “Emissions trading for international aviation - an estimation of the economic impact on selected European airlines,” *Journal of Air Transport Management*, vol. 13, 2007, pp. 253-263.
- [5] International Air Transport Association (IATA), IATA Economic Briefing, Montreal, June 2007.
- [6] Energy Information Association (EIA), International Energy Outlook, Washington, 2007.
- [7] United States Government Accountability Office (GAO), “Uncertainty about future oil supply makes it important to develop a strategy for addressing a peak and decline in oil production,” report to congressional requesters, Washington, 2007.
- [8] International Energy Agency (IEA), International Energy Outlook 2007 – Petroleum and Other Liquid Fuels, Paris, 2007.
- [9] Deutsche Lufthansa AG, Balance 2007, Cologne, 2007
- [10] Energy Information Association (EIA), Short-term Energy Outlook, November 6, 2007.
- [11] P. Morrell and W. Swan, “Airline jet fuel hedging: theory and practice,” *Transport Reviews*, vol. 26, pp. 713-730, 2006.
- [12] J. Scheelhaase, W. Grimme, W. and M. Schaefer, “European Commission plans emissions trading for aviation industry,” *Airlines*, e-zine edition, Issue 36, 2007, pp. 1-5.
- [13] www.lufthansa-financials.de/servlet/PB/menu/1023437_12/index.html (visited Oct. 31, 2007)
- [14] D. Gillen, W. Morrison and C. Stewart, “Air travel demand elasticities - concepts, issues and measurement,” study commissioned by the Canadian Department of Finance, Ottawa, 2004.
- [15] C. Smith, “Aviation and oil depletion,” presentation at Energy Institute, London, November 2006.
- [16] Chevron, “Alternative jet fuels,” supplement to Chevron’s Aviation Fuels Technical Review, Houston, 2006.
- [17] Boeing, Alternate Fuels for use in Commercial Aircraft, Seattle, 2007.
- [18] www.tupolev.ru/English/Show.asp?SectionID=82 (visited Nov. 20, 2007).
- [19] P.J. Masson and C. A. Luongo, “High power density superconducting motor for all-electric aircraft propulsion,” *IEEE Transactions on Applied Superconductivity*, vol. 15, 2005, pp. 2226-2229.

Analysis of Air Transportation for the New York Metroplex: Summer 2007

Liya Wang; George Donohue;

Karla Hoffman; Lance Sherry

Center of Air Transportation and Systems Research
George Mason University
4400 University Dr.
Fairfax, VA 22030

Rosa Oseguera-Lohr

NASA Langley Research Center
Aeronautics Systems Analysis Branch M. S. 442
Hampton, VA 23681

Abstract—The New York metroplex airports (JFK, LGA, EWR) provide air transportation service to this critical international economic hub. In the summer of 2007 the flights servicing the NYC metroplex airports experienced excessive delays and cancellations that added significant costs to doing business in New York. These delays can be attributed to changes in daily airport capacity (due to weather) and to airline practices, in accordance with regulations, of scheduling in excess of airport capacity. Previous research has demonstrated that maintaining airline seat capacity by increasing aircraft size and reducing frequency is an economically efficient and feasible solution. This paper analyzes the characteristics of the air transportation service to the New York metroplex airports. The metroplex has service to 104 domestic airports. 36.5% of airports serve all three New York airports, while 35.6% serve two of the airports. For all the routes to NYC, the average number of flights per day is 6 with a maximum of 32. These routes have an average aircraft seat size ranging from 19 to 238 with an average of 94 seats per flight. These routes had passenger load factors ranging from 0.26 to 0.95 with an average of 0.78. This yields an average of 281 unused seats per day on these routes. Additional statistics and discussion of these results on the implications for consolidation of service with larger aircraft and reduced frequency is discussed.

Keywords- JFK; LGA; EWR; metroplex; air transportation

I. INTRODUCTION

In 2007, domestic U.S. airline travelers experienced the lowest on-time performance on record. Approximately 30 percent of all flights were either cancelled or delayed more than 15 minutes [1].

Airline service to the international economic hub of New York City (NYC) was particularly hard hit. The on-time percentage for the three New York airports (EWR, LGA, and JFK) was 71.5% departures, 62% arrivals, and 3.46% cancellations [2]. The national average was 76.5%, 73.4% and 2.16% respectively [1]. Also, airline service experienced the worst cancellation rate in the nation. Since approximately a third of the nation's air traffic passes through NY airports, delays in NYC ripple through the system causing delays at other airports [3].

Analysis of these delays identified that two functional causes of delays; (i) changes in daily airport capacity (due to weather) as high as 20% reductions from good weather

capacity, and (ii) airline practices, in accordance with regulations, of scheduling in excess of airport capacity.

Previous research [4] demonstrated that maintaining airline seat capacity by increasing aircraft size and reducing frequency is an economically efficient and feasible solution. Airlines flying larger aircraft, with higher load factors increase revenue. Air Traffic Control has reduced operations leading to marginal delays. The airport increases passenger throughput and provides reliable service to its customers.

This paper describes the results of an analysis of the air transportation characteristics of the NYC metroplex airports:

- The metroplex has service to 104 domestic airports.
- 36.5% of airports serve all three New York airports, while 35.6% serve two of the airports.
- For all the routes to NYC, the average number of flights per day is 6 with a maximum of 32.
- These routes have an average aircraft seat size ranging from 19 to 238 with an average of 94 seats per flight.
- These routes had passenger load factors ranging from 0.26 to 0.95 with an average of 0.78.
- This yields an average of 26,197 unused seats over 119,004 provided seats on all the routes each day.

Additional statistics are also provided: number of airports served, redundant service at NYC airports, flight number per day of NYC airports, number of competing airlines of NYC airports, load factor of NYC airports, aircraft seat size of NYC airports, seat size vs. load factor for NYC Airports, unit revenue (\$/mile) vs. load factor for NYC Airports, flight frequency vs. seat size classified by load factor for NYC Airports, and flight frequency vs. seat size classified by unit revenue for NYC Airports.

The paper is organized as follows: Section 2 provides an overview of the scheduled/actual flights and available capacity at NYC airports. Section 3 describes the methodology and algorithms for analysis of the air transportation data. Section 4 describes the results of the analysis. The conclusions and future work are discussed in Section 5.

II. BACKGROUND: DEMAND VS. CAPACITY

Competition coupled with high demand force the airlines to schedule multiple flights during peak hours. To cut costs, flights are served by smaller planes thereby allowing each airline to provide frequency during the most popular times while maintaining reasonable costs. The result for the airport, is that more flights are scheduled than the runway can handle. Figure 1 through 3 show that 2007 summer (06/01-08/30) departure capacity of mean ADR (Airport Departure Rate), average number of scheduled departures, and wheels-off delays in 15 minutes bin from 6:00 am to 12:00 am for NYC metroplex three airports. Obviously, during peak hours of 8:00am-9:00 am and 4:00pm-6:00pm, the airports are over-scheduling. Figure 1 to 3 also illustrate how over-scheduling in one period without sufficient under-scheduling to allow the queues to dissipate, forces delays to become longer as the day progresses. Examining Figure 3, one sees that although there is only a small amount of over-scheduling in any given period, the delays continues to get larger and larger as the day progresses. In addition, JFK has the worst wheels-off delays from volume standpoint. Table 1 summarized the over-scheduling time percentage from 6:00am-10:00pm in 2007 summer for scheduled and actual demands respectively. Table 1 shows that departures are more over-scheduled than arrivals. From time standpoint, LGA is the most over-scheduled airport.

In addition, the study also shows the airport capacity is not fully, efficiently, and properly used. Figure 4 and 5 show the actual and scheduled demands distribution for 2007 summer respectively for JFK. In each cell, the probability of demands is calculated, and then colors are used to distinguish them. Figure 4 shows some quarterly scheduled demands are exceeding the airport capacity very much. For example, (23, 7) cell shows a period of time where there are 23 scheduled arrival demand when the airport is capable of handling only 14. At the same time, Figure 5 also shows that system is under used at most of the time in that high probability of demands is distributed far away from the capacity line.

Therefore, to reduce the delays at NYC metroplex airports, first the transportation characteristics should be achieved. Based on that, then the possible improvement space can be identified. In the next part, algorithms to processing data for achieving air transportation characteristics are in detail presented.

Table 1: NYC airports quarterly overscheduled percentage (6:00am-10:00pm)

Airport	Actual Arr.(%)	Actual Dep.(%)	Scheduled Arr. (%)	Scheduled Dept. (%)
EWR	1.8%	4.8%	6.5%	6.6%
JFK	2.9%	8.9%	5.4%	13.8%
LGA	5.1%	10.3%	9.3%	12.2%

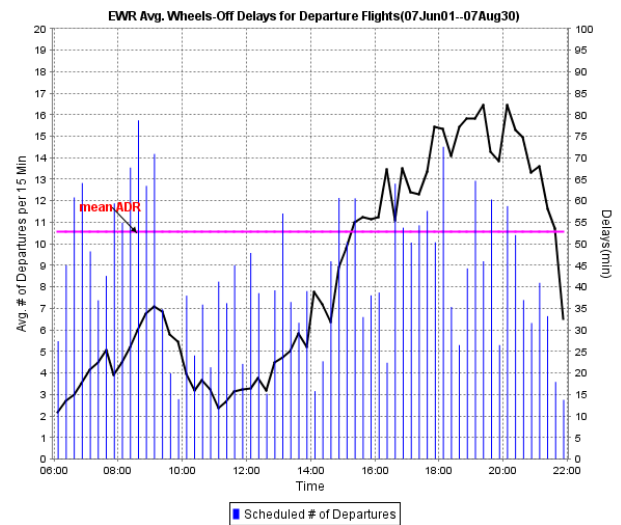


Figure 1. 2007 summer EWR mean capacity, departures and wheels-off delays per 15 min

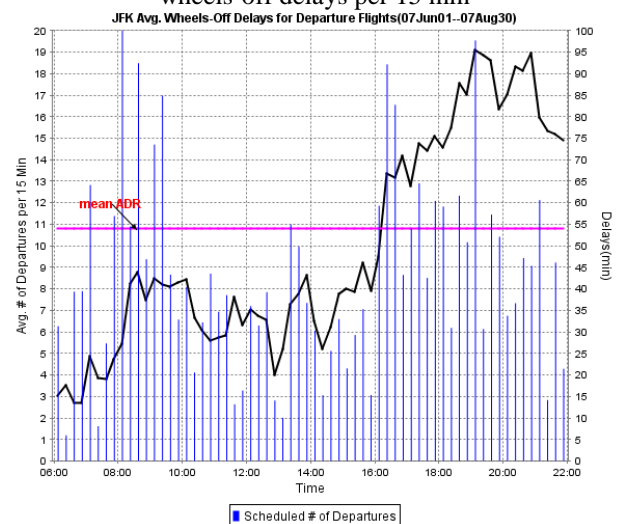


Figure 2. 2007 summer JFK mean capacity, departures and wheels-off delays per 15 min

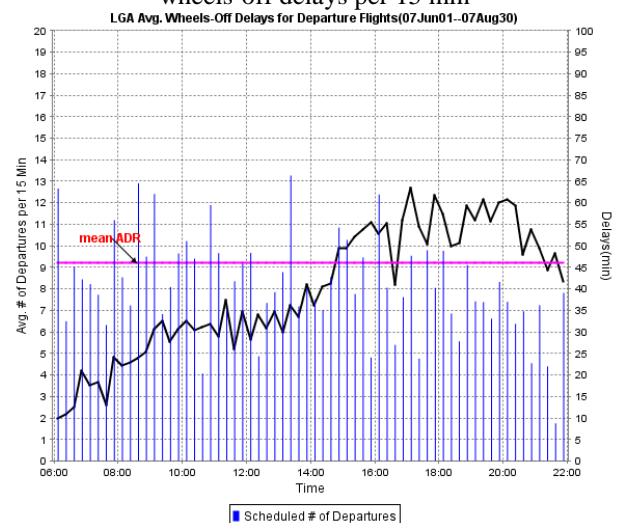


Figure 3. 2007 summer LGA mean capacity, departures and wheels-off delays per 15 min

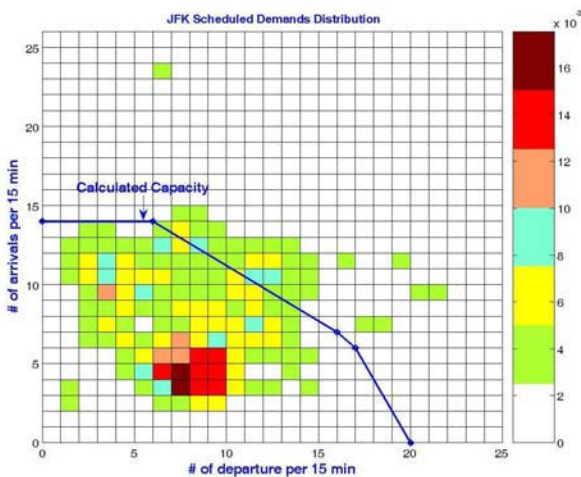


Figure 4. 2007 summer JFK scheduled demands distribution

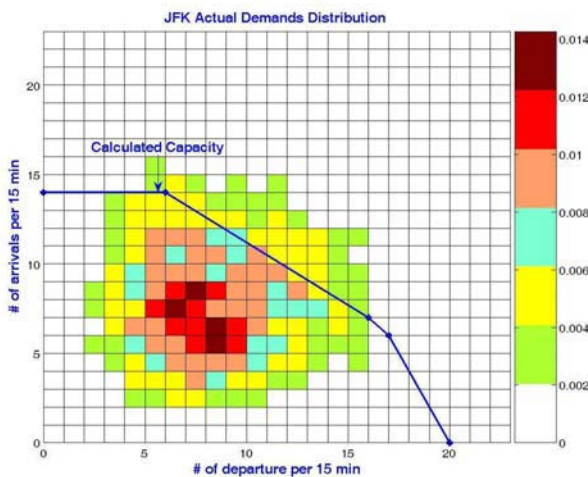


Figure 5. 2007 summer JFK actual demands distribution

III. METHODOLOGY

To get NYC metroplex airports aviation transportation characteristics, a number of aviation databases from Bureau of Transportation Statistics (BTS) [5] and Federal Aviation Association (FAA) Aviation System Performance Metrics (ASPM) [6] databases are used for the study. This section mainly describes the data processing needed to perform our analysis. Figure 6 shows the data processing flow chart. Next, the details of data processing algorithms will be discussed.

A. Data Processing Algorithms

1) Determining seat size, load factor, frequency using T-100 Domestic Segment (U.S. Carriers) Dataset

T-100 Domestic Segment (U.S. Carriers) database contains domestic non-stop segment data reported by U.S. air carriers, including carrier, origin, destination, aircraft type and service class for transported passengers, freight and mail, available capacity, scheduled departures, departures performed, aircraft hours, and load factor when both origin and destination airports are located within the boundaries of the United States and its territories.

Therefore, from the T100 database, we can extract market information for each airport such as the number of carriers, average plane seat size, load factor, and average flight number per day. Average airplane seat size can be calculated in Eq. (1). The calculation of load factor and average flight number per day is expressed in Eq. (2) and (3) respectively.

$$\text{Seat Size} = \frac{\text{total \# of seats}}{\text{total \# of departures}} \quad (1)$$

$$\text{Load factor} = \frac{\text{total \# of passengers}}{\text{total \# of seats}} \quad (2)$$

$$\text{Flight number} = \frac{\text{total \# of departures}}{\text{total \# of days}} \quad (3)$$

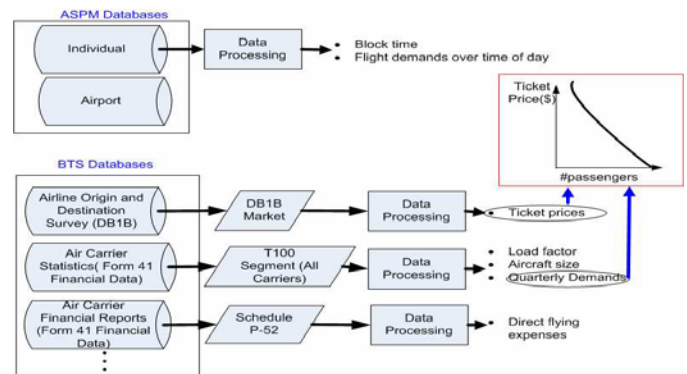


Figure 6. Data processing flow chart

2) Calculations of segment prices using DB1B Market Dataset

DB1B market data table contains directional market characteristics of each domestic itinerary of the Origin and Destination Survey, such as the reporting carrier, origin and destination airport, prorated market fare, number of market coupons, market miles flown, and carrier change indicators. Therefore, DB1B can be used to extract ticket price for market A to B.

The ticket prices in DB1B market table are for itinerary tickets. To get ticket price from airport A to airport B, two situations should be considered. In the first situation, A is origin and B is destination. Therefore, A to B ticket price can be fetched directly from DB1B. In the second situation, A is origin; however B is a transfer airport for flying A to C. Under this situation, the ticket price from A to B cannot be directly got from DB1B. Segment fares are traditionally prorated from itinerary fares. However, there is a fixed cost in any flight leg. This part of fixed cost is large in flight legs of short distance, and decreases in legs of longer distance. We compute segment fares proportionally to the squared root of distances of segments in the itinerary [4]. Therefore, Eq. (4) will be used to get segment fare from A to B.

$$T_{AB} = T_{AC} \times \frac{\sqrt{d_{AB}}}{\sqrt{d_{AB}} + \sqrt{d_{BC}}} \quad (4)$$

Where T_{AB} is the ticket price form A to B, and d_{AB} is the distance form A to B.

Specifically, if a flight has two legs of 100(=10²) miles and 225(=15²) miles, and has the one-way ticket price of \$100, then

leg one is allocated \$40 ($=100 \times \frac{\sqrt{100}}{\sqrt{100} + \sqrt{225}}$) and leg two 60\$ (=100-40).

3) *Extracting airport capacity from ASPM Airport Dataset*

ASPM airport database can provide detail information by quarter hour or hour on the airport, which includes AAR, ADR, wind speed, visibility, runway configuration, scheduled departures, scheduled arrivals, efficiency of departures, and ETMS departures etc.

Therefore, from this database, we can directly fetch airport capacity of AAR and ADR.

4) *Determining demands and delays data from from ASPM Individual Dataset*

ASPM individual database can provide detail schedule information on a flight including carrier, origin, destination, aircraft type, departure date, arrival date, scheduled in time, actual in time, scheduled out time, actual out time, scheduled taxi-out time, actual taxi-out time, scheduled taxi-in time, actual taxi-in time, delays, and block time etc.

Hence, from ASPM individual database, schedule information such as departure and arrival demands, delays in each quarter or hourly can be extracted. Eq. (5) calculates the demands per quarter. ASPM individual database can also provide information scheduled block time. Eq. (6) shows how to calculate delays per quarter.

$$\bar{d}_i = (\sum d_i) / (\sum I_{d_i > 0}) \tag{5}$$

Where $I_{d_i > 0} = \begin{cases} 1 & \text{if } d_i > 0 \\ 0 & \text{otherwise} \end{cases}$

$$\bar{L}_i = \sum_j L_{ij} / d_i \tag{6}$$

Where d_i is the demands in the quarter i , $I_{d_i > 0}$ is to determine whether the there is demand in quarter i or not, and L_{ij} is the delays of flight j at quarter i .

IV. DATA ANALYSIS RESULTS

This part will discuss the data analysis results for 2007 summer at NYC airports. The data from major airlines are only counted in T100. A major airline is defined as the airline which has more than 60 arrival flights to the studied airport in 2007 (Jun 01-Aug 30). Next, we will discuss the detail statistics results such as number of airports NYC served, average flight number per day, number of competing airlines, airfares, the load factor, and the average aircraft seat size for the identified airports etc.

A. *Number of Airports to NYC*

After processing the data, we got that the NYC severs 104 domestic airports in 2007 summer (Figure 7). And EWR serves 81 of them, JFK severs 62, and LGA serves 68 (Table 2).

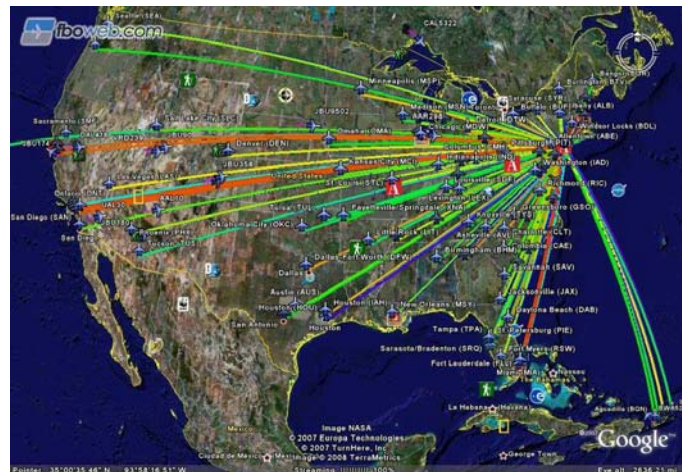


Figure 7. NYC metroplex served airports route map

Table 2. Number of airports served by NYC three airports

Airport Code	# of Airports Served
EWR	81
JFK	62
LGA	68

Table 3 also lists the details about how NYC airports sever the 104 airports. 38 (36.5%) of the identified 104 airports are served by all 3 airports; 37 (35.6%) are served by two of three airports; and only 29 (27.9%) are served by only one of three airports

Table 3. Redundancy of service in NYC airports

# of NYC serving airports	1	2	3
# of airports served	29	37	38
% of airports served	27.9%	35.6%	36.5%

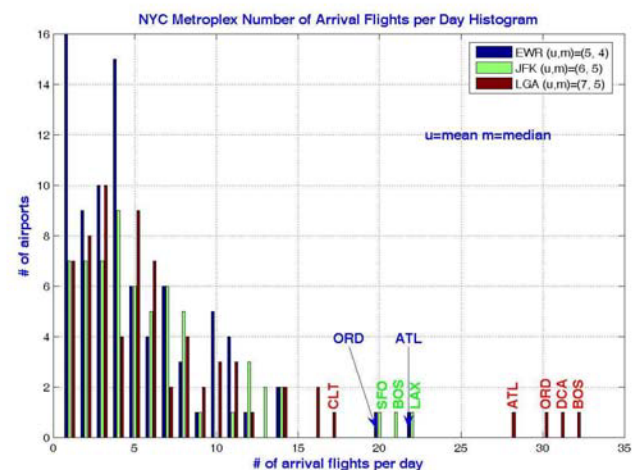


Figure 8. NYC metroplex arrival flight number/day histogram

B. *Flight Number per day for NYC Airports*

Figure 8 shows the average flight number per day histogram for NYC metroplex airports. Figure 8 shows that the average flight number per day per market are 5 to EWR, 6 to JFK, and 7 to LGA. And the medians are 4 to EWR, 5 to JFK and LGA. Therefore, approximate $(5 \times 81 + 6 \times 62 + 7 \times 68) = 1253$

flights each day are flying to/from NYC metroplex airports. In addition, most of airports have less than 15 flights per day. Table 4 also lists the top 20 airports to NYC, and 53% of flights are from these top 20 airports.

Table 4. Top 20 airports to NYC airports

Airport	Code	# of arrival flights/day			
		EWR	JFK	LGA	NYC
Boston Logan MA	BOS	10	21	32	63
Chicago O'Hare IL	ORD	20	11	30	61
Atlanta Hartsfield GA	ATL	22	5	28	55
Ronal Reagan National DC	DCA	7	8	31	46
Raleigh Durham NC	RDU	11	13	16	40
Fort Lauderdale FL	FLL	12	13	14	39
Charlotte NC	CLT	14	7	17	38
Orlando FL	MCO	14	14	9	37
Detroit MI	DTW	10	22		32
Los Angeles CA	LAX	11	5	16	32
Dallas Ft. Worth TX	DFW	10	20		30
San Francisco CA	SFO	11	5	14	30
Dulles VA	IAD	10	12	6	28
Buffalo NY	BUF	5	14	7	26
Pittsburg PA	PIT	8	6	11	25
Miami FL	MIA	11	4	9	24
Houston Bush Int. TX	IAH	8	8	8	24
Port Columbus OH	CMH	5	7	11	23
Cincinnati OH	CVG	8	4	10	22
Cleveland[Hopkins Intl] OH	CLE	6	4	11	21

C. Number of Airlines for NYC Airports

The airline number is got from T100, and only the major airline is counted. Figure 9 shows the number of airlines histogram for the NYC metroplex airports where each route are served by 2 airlines in average. Route CMH-LGA, BOS-JFK, SFO-JFK are most competitive routes, with 5 airlines severing on these three routes.

D. Airfares for NYC Airports

Table 5 also lists the segment prices to top 20 airports. This table illustrates that the price can vary substantially among the three NYC airports. For example, BOS to EWR has an average segment fare of \$200, for BOS to JFK, the average segment fare is \$86, and for LGA is \$148. Therefore, BOS-EWR is most expensive way to NYC. This can be explained by airline competition which drives the price down. From BOS to EWR, only one airline flies on this route, however five airlines fly from BOS to JFK, and 3 airlines are flying from BOS to LGA. In addition, Figure 10 is the airfare histogram for NYC metroplex airports. The average segment airfare to EWR is \$175, to JFK is \$149, and to LGA is \$151. In addition, virtually all airports have segment airfares below \$250.

Table 5. Airfares of top 20 airports to NYC airports

Airport	Code	Airfares(\$)			
		EWR	JFK	LGA	NYC
Boston Logan MA	BOS	200	86	148	140
Chicago O'Hare IL	ORD	164	128	157	156
Atlanta Hartsfield GA	ATL	152	161	147	150
Ronal Reagan National DC	DCA	192	102	142	143
Raleigh Durham NC	RDU	134	112	127	125
Fort Lauderdale FL	FLL	131	132	133	132
Charlotte NC	CLT	127	117	127	125
Orlando FL	MCO	134	135	142	136
Detroit MI	DTW	290	361		335
Los Angeles CA	LAX	193	127	136	146
Dallas Ft. Worth TX	DFW	309	326		320
San Francisco CA	SFO	249	189	239	237
Dulles VA	IAD	81	90	106	93
Buffalo NY	BUF	105	95	97	97
Pittsburg PA	PIT	152	137	146	146
Miami FL	MIA	264	188	212	228
Houston Bush Int. TX	IAH	122	88	125	111
Port Columbus OH	CMH	146	95	115	115
Cincinnati OH	CVG	173	141	181	174
Cleveland[Hopkins Intl] OH	CLE	198	137	150	168

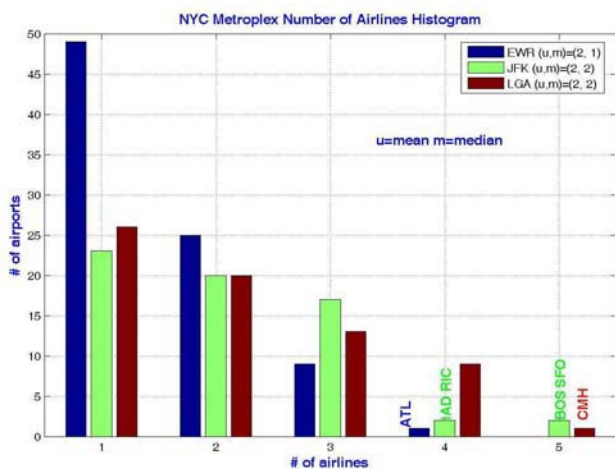


Figure 9. NYC metroplex airports number of airlines histogram

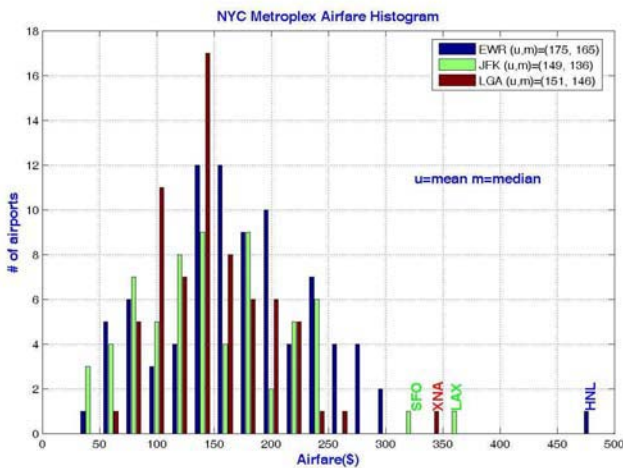


Figure 10. NYC metroplex airports airfares histogram

E. Aircraft Seat Size for NYC Airports

Figure 11 shows the mean aircraft size histogram for the NYC metroplex airports. The average aircraft seat size to LGA is the smallest (75) of the three airports, while JFK is the largest (113). In addition, at EWR and LGA, over one third of the airports are serving with aircrafts having 50 seats or less.

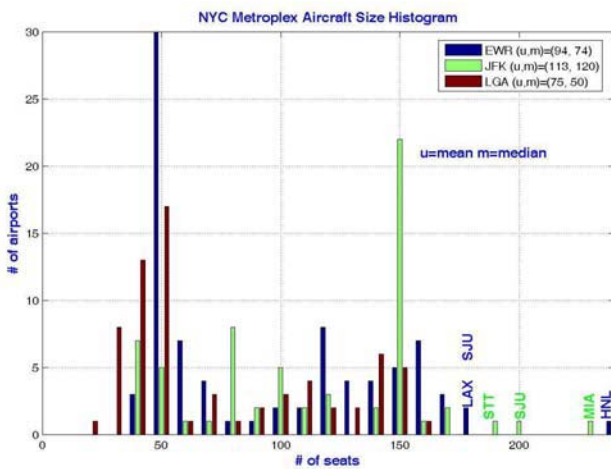


Figure 11. NYC metroplex airports aircraft seat size histogram

F. Load Factor for NYC Airports

Figure 12 presents a load factor histogram for NYC metroplex airports. The graph illustrates that EWR and LGA has better load factors than LGA. LGA-LWB (Lewisburg, West Virginia) has the smallest load factor (0.26). At EWR and JFK, over 90% of airports have load factors greater than 0.6, assuring the profitability of most flights during the summer of 2007.

G. Seat Size vs. Load Factor for NYC Metroplex

Figure 13 shows the relationship between aircraft size and load factor for NYC metroplex. For most airports, aircraft size and load factor are positively correlated. However, for high frequent airports such as DCA, BOS, CLE, the load factors are

not so high (less than or equal to 0.7). Thus, for these highly-competitive airports, if each of the airlines serving these regions wishes to maintain frequency, they must choose smaller aircrafts to maintain profitability. Alternatively, passengers could be equally well-served with larger aircraft but fewer airlines serving these airports. It is on these high-demand routes where upgaging would help rationalize aircraft and runway capacity.

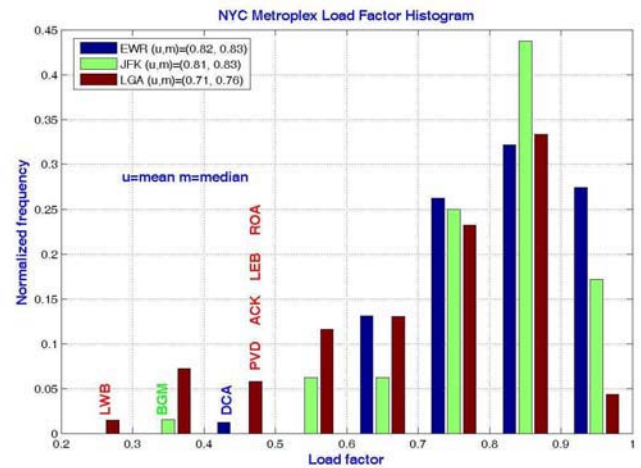


Figure 12. NYC metroplex airports load factor histogram

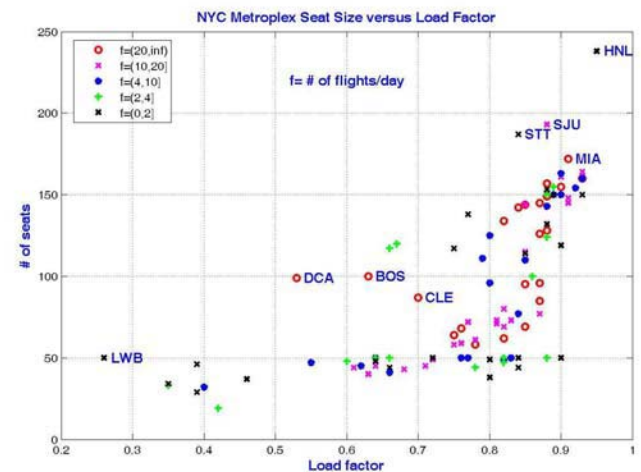


Figure 13. NYC metroplex seat size vs. load factor classified by flight frequency

H. Unit Revenue vs. Load Factor for NYC Metroplex

Figure 14 presents the relationship between unit revenue (=airfare/distance) and load factor. It shows that there are negative correlations between the unite revenue and load factor. MVY (Martha's Vineyard Airport, MA) to NYC has the largest unite revenue (0.83\$/mile), however, TUS (Tucson, AZ) to NYC has the smallest unite revenue (0.08\$/mile).

I. Flight Frequency vs. Seat Size for NYC Metroplex

Figure 15 shows the relationship between frequency (# of arrival flights per day) and seat size, and it is classified by load factor. It shows that most low load-factor aircraft are small and

providing service infrequently. Not surprisingly, flights to Hawaii (HNL) are serviced by large airplanes and have very high load factors.

Moreover, Figure 16 shows the relationship between frequency (# of arrival flights per day) and seat size, and it is classified by unit revenue. It shows that long distance routes usually have lower unit revenue than some short, low load factor routes.

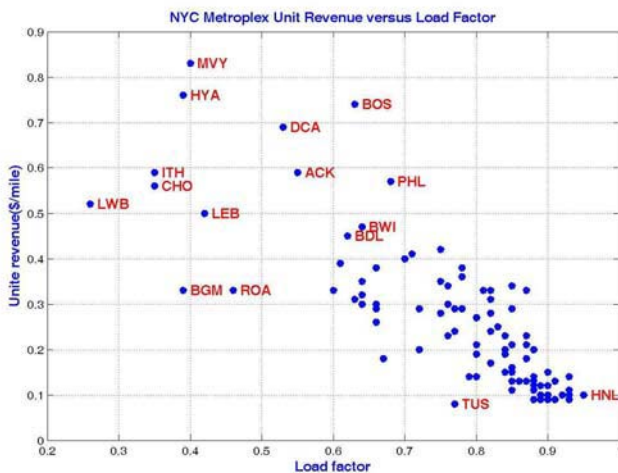


Figure 14. NYC metroplex unit revenue vs. load factor

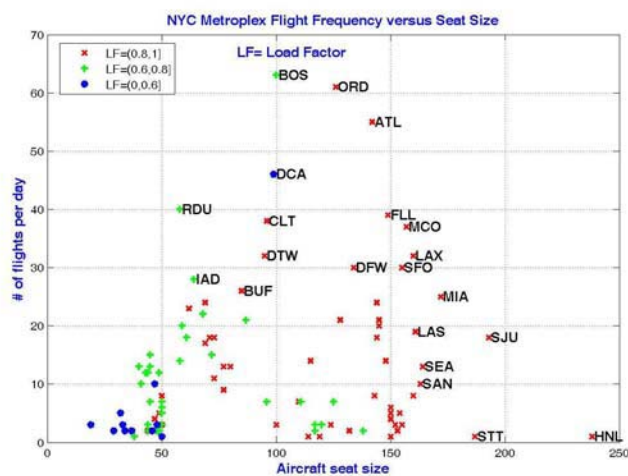


Figure 15. NYC airports frequency vs. seat size and classified by load factor

one. Demands such as the number of arrival flights per day per market (EWR: mean=5 and median=4; JFK: mean=6 and median=5; LGA: mean=7 and median=5) and load factors (EWR: mean=0.82 and median=0.83; JFK: mean=0.81 and median=0.83; LGA: mean=0.71 and median=0.76) imply the heavy passenger demands to NYC airports. The number of airlines serving a market (EWR: mean=2 and median=1; JFK: mean=2 and median=2; LGA: mean=2 and median=2) presents that NYC airports are competitive, which also forces the airfares (EWR: mean=\$175 and median=\$165; JFK: mean=\$149 and median=\$136; LGA: mean=\$151 and median=\$146) down. The average aircraft sizes (EWR: mean=94 and median=74; JFK: mean=113 and median=120; LGA: mean=75 and median=50) presents us the opportunity of upgaging in order to reduce congestion for the crowded NYC airports. The relationship between seat size and load factor discloses a possible way to help reduce congestion by proper upgaging small airplanes to big ones. The data will also be used in our future research that will examine the impacts of both regulation and cost on these airports with the goal of determining how best to allocate the scarce runway capacity that exists within this region.

ACKNOWLEDGMENTS

Work was supported by NASA Grants NRA NNH06ZEA001N-AS. Thank you for technical comments and suggestions to David Smith, Bengi Manley, John Shortle, Rajesh Ganesan, and Vivek Kumar (GMU).

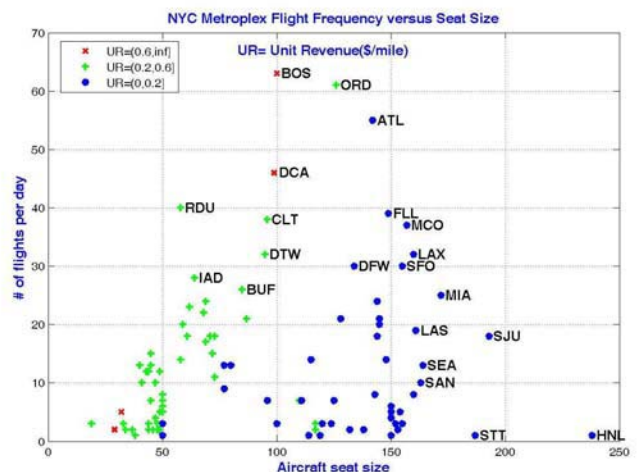


Figure 16. NYC airports frequency versus seat size and classified by unit revenue

V. CONCLUSION

The paper is the first in a series of papers that studies the NYC Metroplex of airports. This paper took a first look at the status of NYC airports in terms of the airports served, seat capacity, airfares, load factors etc. The results identify the NYC metroplex serving 104 domestic airports (EWR:81, JFK:62, LGA:68). The results also indicate that NYC metroplex exhibits redundant service. 38 (36.5%) of the identified 104 airports are served by all 3 NYC airports; 37 (35.6%) are served by two; and only 29 (27.9%) are served by

REFERENCES

- [1] Summary of airline on-time performance year-to-date through december 2007. http://www.bts.gov/programs/airline_information/airline_ontime_tables/2007_12/html/table_01.html. Accessed December 2007.
- [2] Airport snapshot. <http://www.transtats.bts.gov/airports.asp?pn=1/>. Accessed December 2007.
- [3] X. Ning, "Method for deriving multi-factor models for predicting airport delays," Ph.D. dissertation, George Mason University, December 2007.

- [4] L. Le, "Demand management at congested airports: how far are way from utopia?" Ph.D. dissertation, George Mason University, August 2006.
- [5] Bureau of transportation statistics (BTS) databases and statistics. <http://www.transtats.bts.gov/>. Accessed December 2007.
- [6] Aviation system performance metrics (ASPM)–complete. FAA. <http://www.apo.data.faa.gov/aspm/entryASPM.asp>. Accessed December 2007.

Doctoral Consortium

Demand for Low-Cost Airlines in Australia

Panarat Srisaeng, Cees Bil, Margaret Tein

School of Aerospace Mechanical and Manufacturing Engineering
Royal Melbourne Institute of Technology University
Melbourne, Australia

Abstract — the purpose of this study was analysis of low-cost airline demand in Australia. As part of this project, an econometric method was applied to develop a regression model for forecasting demand. The research hypothesis being that low-cost airline demand in Australia is based on the following variables: domestic airfares, price of other transport modes, population, disposable income and tourist numbers. It was found that demand for low-cost airlines is primarily a function of domestic airfare and population while tourist numbers and price of other transport modes did not have a significant influence.

Keywords-low cost airlines; demand modelling

I. INTRODUCTION

The low-cost airline concept has been very successful in North America and Europe. Emergence of low-cost airlines has significantly stimulated demand for air travel. In Australia, after deregulation in 1990, several low-cost airlines were established but most were unsustainable. In 2000 Virgin Blue entered the market and is the first low-cost airline in Australia to be both successful and profitable (Forsyth, 2003). Virgin Blue's success has attracted other low-cost airlines such as Jetstar to enter the market. The study aims to develop a demand model for low cost airlines in Australia and identify variables affecting the demand model significantly. As demand for air travel is related to and affected by one or more economic, social or supply factors, this study used an econometric method to develop a demand model. Econometric models attempt to measure causal relationships allowing forecasting of the impact of change implementation on any variable and consequent prediction of demand level impact (Doganis, 2002). Knowing demand for low-cost airlines helps government and business sectors arrange adequate air service infrastructure to meet future demand and provides more accurate information on which to base strategic plans and decisions. The hypothesis of this research is that the total number of low-cost airline passengers depends on the following

independent variables: domestic airfares, price of other transport modes, population, disposable income and tourist numbers. A multiple linear regression analysis was used to test this hypothesis. This paper starts with a general background of the Australian airline industry, followed by a review of previous research studies. The methodology is then introduced and finally results are presented followed by conclusions.

II. LITERATURE REVIEW

Studies of air travel demand have used a variety of methodologies and variables. Battersby and Oczkowski (2001) analyzed a demand model for domestic air travel in Australia using the regression method. Four independent variables: airfares, income, substitute prices and seasonality were considered. Both price and income elasticity were found to be lower than in previous studies. Savage and Dykstra (1995) studied demand elasticity for air travel to and from Australia. The model was separated into two parts: leisure travel and business travel. The study found variables determining leisure travel were airfares, income and relative prices while income and relative prices were found to be the most important determinants for business air travel. Ghobrial and Kanafani (1995) estimated air passenger demand between various city pairs in the United States by using regression analysis. Population and per capita income were selected to represent the socioeconomic variables while the supply variables included airfare, travel time, city specific variables and level of service parameters including aircraft size and number of flights. Although the model explained only 50% of the variation in data, results suggested that air passenger demand was highly dependent on the frequency of flights, travel time and airfare. Tretheway and Oum (1992) identified variables which possible affect

demand for air travel, price income price and other modes of transport, frequency of service, timing of service, day of the week, season of the year, safety and company goodwill, demographics, distance, in-flight amenities, customer loyalty and travel time. The results showed that the most significant of these determinants of demand for air travel are price and income.

III. METHODOLOGY

Previous studies focused on demand elasticity for full-service air travel in Australia. However demand for low cost airlines has not yet been studied in detail. This study therefore will focus on developing a demand model for low cost airlines in Australia by using secondary data sources from the Australian Bureau of Statistics (ABS) Bureau of Transport and Regional Economics (BTRE) and low cost airlines' annual reports. In air travel demand, analysis many explanatory variables influence passenger numbers. The procedure for developing regression models and forecasting air passenger traffic can be illustrated by the flowchart shown in figure 1. The major objective of regression analysis is to study the relationship between selected variables by measuring the response of one variable by a set of variables then use the regression model to estimate the dependent variable by given independent variables. The first step in demand model development is reviewing past travel trends. To identify patterns in the relationship between these variables a scatter diagram is plotted. The next step is to identify factors influencing travel in the past and those which may affect it in future. This step mainly relies on previous research studies. A literature review can assist variable selection by highlighting general characteristics influencing demand for travel. Variable selection also depends on availability of empirical data and operational costs. Next the function form of the demand model is established from the list of variables selected. Plotting a scatter diagram can assist to identify variable relationships whether linear or non-linear forms. The model is then used to fit historical data using the ordinary least squares (OLS) method to estimate coefficients. The next step in model development is to evaluate model accuracy. Statistics such as the coefficient of determination (R^2) and analysis of variance (ANOVA) are used to evaluate the model. The coefficient of determination (R^2) is a measure of the

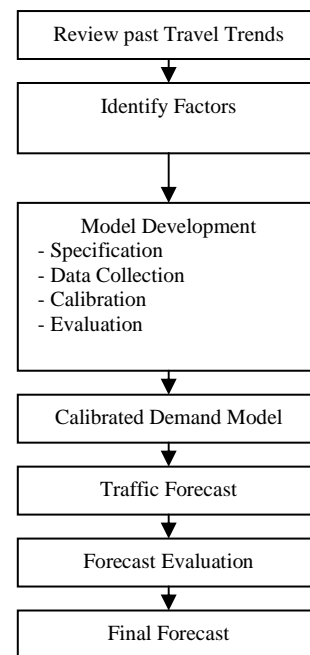


Figure1: Regression Model Development (Taneja,1978)

adequacy of the regression equation. In other words, this statistic shows how well the model fits the data. If the R^2 value is near 0 it implies a relationship does not exist between dependent and independent variables. Conversely, if the R^2 value is near 1 it implies there is a strong relationship between dependent and independent variables. Analysis of variance (ANOVA) is often presented in connection with the regression model. This is the breakdown of the total sum of squares into the explained sum of squares and the residual sum of squares. The purpose of presenting the table is to test the coefficient significance (Muddala, 2001). The next step is to estimate air travel demand. Demand forecast is produced by using this verified demand model and the trend projection. With this method the independent variables will be projected and the dependent variable can be estimated.

IV. VARIABLE SELECTION

The dependent variable used is the total number of passengers who traveled on Virgin Blue Airways and Jetstar, the two low cost carriers in Australia, in a particular year. Since Virgin Blue has operated since 2002 and Jetstar has flown since 2004 historical data on low cost airlines passengers was

only available from 2002. In this study the series of total number of passengers are represented per quarter. Based on previous research, most air passenger demand studies selected socioeconomic factors and transport supply factors to analyse the relationship between variables. In this study, the following variables are selected as socioeconomic variables: population, disposable income and tourist numbers while airfares and price of other transport modes are chosen as transport supply factors. Tourist numbers were not selected by previous demand model development, however this variable might impact a demand model for low cost airlines as low fares may attract tourists who would otherwise travel by other transport modes. The total population of Australia is collected from the Australian Bureau of Statistics covering the second quarter of 2002 and extending to the final quarter of 2004. Disposable income is the net income from which taxes have been deducted to represent purchasing power. This is an important airline industry variable because it give a sense of the amount of money people can spend on air tickets (Ellsworth, 2000). To adjust for inflation, disposable income is divided by the consumer price index (CPI) taken from the Australian Bureau of Statistics with 1990 as the baseline. Tourists are people whose main purpose for the trip is holiday include those who not only travel by air but also by other transport modes. Domestic full-service airline fare is an important factor in passenger decisions which influence low cost airline demand. This study uses the domestic airfares price index for economy class for full-service airlines taken from the Bureau of Transport and Regional Economics (BTRE) to calibrate the demand model. The airfares price index for economy class uses July 2003 as the baseline and is adjusted for the CPI using the Australian Bureau of Statistics Consumer Price Index. The price index of other transport modes is selected to analyse the relationship between numbers of low cost airlines passengers and price of alternative travel options. Like previous variables, the price index of other modes of transport is adjusted for inflation and has 1990 as the baseline.

V. DEMAND MODEL

The demand model is developed to estimate the total number of low cost airline passengers. The statistical software, SPSS, is used to determine the

final forecast model. From the scatter diagram of these variables, the relationship between dependent and independent variables in this model is determined to be a linear relationship.

The general form of the regression model is:

$$Y_t = b_0 + b_1P_t + b_2I_t + b_3T_t + b_4A_t + b_5O_t$$

Where Y is the dependent variable, total number of low cost airlines passengers, P is the population of Australia, I is the disposable income, T is the total number of tourists, A is the domestic airfare index and O is the fare of other modes of transport index. The SPSS is used to perform the regression analysis. The ordinary least-square method (OLS) is used to determine the parameters b_i in the regression model and the STEPWISE method is used to select the variables. Stepwise regression begins by entering variables into the model one at a time. The first variable to be selected is the parameter which shows the strongest correlation with the dependent variable. Each time a new parameter is considered for entry into the model, the program simultaneously tests the variables in the model for removal. The t statistic is used to test the hypothesis that the coefficient of the variable is 0. If the significance of adding a variable to the equation is less than or equal to 0.05, in other words the confident interval is less or equal to 95%, then the variable is included in the model. However, a parameter will be removed if its significance level exceeds 0.10. The stepping procedure ends when the significance of the dependent variable does not improve (Ellsworth, 2000).

VI. RESULTS

A stepwise multiple regression analysis was executed with five exogenous variables. The adjusted coefficient of determination, Adjusted R^2 , is a measure of model adequacy or how well the model fits data corrected for bias. Statistical results indicated the model explained 98% of variation in the endogenous variable "total number of low cost airlines passengers". The model is tested for overall significant by using F-test. The F-test is a formal hypotheses test that is designed to deal with a null hypothesis that contains multiple hypotheses or a single hypothesis about a group of coefficient. Alternatively, there is a measure called the p-value or marginal significant value which is used to test the hypothesis of regression. If p-value is less than

critical value the null hypothesis will be rejected. T-test is the test usually used to test a hypothesis about the individual regression slope coefficient. If p-value is less than critical value the null hypothesis will be rejected. Overall the model is statistically significant (p-value < 0.000). It can be implied that there is at least one independent variable which relates to the total number of low-cost airlines passengers. To determine which independent variables have a relationship to the dependent variable, T-statistics were used as a criterion. If p-value of each variable is less than or equal to 0.01, the variable will be included in the model. We find two variables, population (P) and domestic airfares index (A) are statistically significant (p-value < 0.01). The signs of the regression coefficients of these variables are positive which is in line with expectations. If fares of full-service airlines increase, passengers are more likely to fly with low cost carriers. However, three other variables, disposable income (I), total number of tourists (T) and price index of the other transport modes (O) did not prove to be significant at a 0.1 probability level and were removed from the model. The resulting simple demand model for low cost airlines is:

$$Y_t = -80,923,650.611 + 3.125P_t + 11,146.106A_t,$$

The coefficient of the population variable is positive (b = 3.125) and statistically significant (p-value < 0.01). This indicates the higher the greater the Australian population, the more likely demand will be for low cost airlines. According to the stepwise regression, it is determined that disposable income, the total tourist numbers and price index of other transport modes are not statistically significant for estimating the number of low cost airlines passengers.

VII. CONCLUSION

It can be concluded that demand for low cost airlines in Australia is a function of population and full-service domestic airfares price index. This study shows exogenous demand variables for low cost airlines are different from demand variables for full-service air travel in Australia. Disposable income, tourist numbers and the price index of other transport modes are not significant to the model. When domestic airfare increases the number of low cost airline passenger will increase and also when number of population increases the number of low

cost airline passenger will increase as well. This model is based on data for the past three years, and can be verified in future as more data becomes available.

REFERENCES

- [1] Australian Bureau of Statistics 2003-2006, 2003 year book Australia, cat. no. 1301.0, ABS, Canberra.
- [2] Australian Government Department of Transport and Regional Services 2006, Domestic Airfares indexes
- [3] Battersby, B. & Oczkowski, E. 2001, An Econometric analysis of the demand model for domestic air travel in Australia, *International Journal of Transport Economics*, Vol. 28, No. 2, pp.193-204.
- [4] Doganis, R. 2002, *Flying Off Course: The Economics of International Airlines*, Routledge, London.
- [5] Ellsworth, S. 2000, Estimating Air Passenger Travel in the Atlantic Region, master of science in engineering Thesis, The University of New Brunswick.
- [6] Forsyth, P. 2003, Low-cost carriers in Australia: Experiences and impacts, *Journal of Air transport Management*, Vol. 9, pp. 277-284.
- [7] Maddala, K. 2001, *Introduction to econometrics*, 3rd edition, Wiley, England
- [8] Qantas Airways Limited, 2005, Preliminary monthly traffic and capacity statistics, (<http://www.qantas.com.au/infodetail/about/investors/trafficStats/July2005.pdf>)
- [9] Qu, H. & Lam, S. 1997, A travel demand model for mainland Chinese tourists to Hong Kong", *Tourism Management*, Vol. 18, No. 8, pp. 593-597.
- [10] Savage, S. & Dykstra, C. 1995, Demand Elasticities for Air Travel To and From Australia, working paper 20, Bureau of Transport and Communications Economics, Canberra.
- [11] Taneja, N. 1978, *Airline Traffic Forecasting*, Heath, Toronto.
- [12] Virgin Blue Holdings Limited, 2005, Summary of operating statistics, (<http://www.virginblue.com.au/pdfs/investors/December2004.pdf>)

Information Design for Collaboration in Distributed Team Work

Simone Rozzi
Innovative Research
EUROCONTROL
Bretigny sur Orge
France

Simone.Rozzi.Ext@eurocontrol.int

Paola Amaldi Bob Fields
Interaction Design Center
School of Computing Science
Middlesex University
Hendon Campus
London, UK
p.amaldi-trillo; b.fields@mdx.ac.uk

Abstract—This paper describes a doctoral research plan on collaborative practices in ATC. The research is in its early phase and intends to investigate ATC collaborative practices under the Target Time of Arrival Project currently under development at EUROCONTROL. Expected outcome will fall in the area of display design and/or validation. Current efforts are allocated to a review of theories and models that characterize human activities in relation to the context. Such review will inform later data collection and design phases.

Keywords-component: human factors; collaborative work; distributed team work; target time of arrival (TTA) concept.

I. INTRODUCTION

Information technology is known as having great potential to improve performance and safety, where that is appropriate, in virtually any domain, including complex and safety critical ones. I focus on a particular area for potential improvement: how can information technology better support the coordination of co-operative work? This seems to be a particularly delicate issue in highly distributed work settings such as air traffic control, where the safe management of operations depends heavily on the ability to share in a timely fashion the relevant information.

In particular problems might arise when people use common information in ways different from the intended ones. The Überlingen accident (July 2002) could be regarded as an example of a failure to co-ordinate safety critical information, in the presence of an advanced information technology application (TCAS).

A set of studies have been developed and proposed in the CSCW literature that looked at these issues [1, 2]. According to these works designing new technology to support collaborative environments, cannot be limited to considering the information flow and formal procedures but it has to take into account how people construct shared interpretation of information [3]. In other words an approach requiring an analysis of how Common Information Spaces (CIS) are constructed and maintained seems more appropriate to avoid the risk of “disrupting cooperative work by computerizing formal procedures” [2].

This aspect is particularly true in ATC, where controllers have to coordinate their actions, take real time decisions extract information effectively under often high time pressure.

II. OBJECTIVE

This research proposes to characterize collaboration and coordination in ATC with the particular focus on how distributed operators construct a common knowledge representation that supports mutual understanding of goals and intentions.

The candidate application is the Target Time of Arrival (TTA) operational concept. TTA consists in associating to each flight a time windows in order to meet a target time of arrival, thus ensuring improved predictability and reduced traffic buncing. This concept will shift the attention from a sector based perspective to a process perspective, where all of the distributed actors must work together in function of common goal (TTA), instead of sector goal. The level of maturity of the TTA concept is currently between level V1(idea) and V2 (Prototype) of the ATM concept of maturity scale. Initial validation exercises [4] carried out at ECC have indicated that many Human Factors issues are still open, for instance it is not clear how the responsibility between controllers and pilots is going to be shared, how controllers can work with the TTA, what are the HMI information requirements for pilots and controllers. For example controllers felt TTA information was insufficient and need to be further specified. In general the challenge presented to operators appear to be how to achieve effective synchronization between pilots and controllers and controllers of diverse en-route centers to respect the time of arrival.

Research outcomes related to the TTA are expected in the area of display design, i.e. how to portray TTA information in the light of constraints as emerged from a system level analysis, and/or validation, i.e., how to validate, and/or analyze data validation of the distributed display concept.

III. PROPOSED APPROACH

I intend to carry out my work by following a system-centred, rather than “user-centered” approach [5]. Two main assertions are that design has to be grounded on an understanding not

only of the specific tasks being in the focus of investigation, but also on an understanding of the context where the action takes place. This view suggests going beyond traditional views of human activities as sequential actions, with an understanding of the relations existing between the overall system/organizational goals and the “purposeful” actions carried out in everyday practice.

The second assertion is that such complex inter-relations can only be apprehended through iterative learning cycles. As shown in the operational validation literature [e.g. 6, 7] evaluating an artifact often implies obtaining feedback on its operational impact, thus looking at (i) how the artifact will affect current methods of working; (ii) whether it will introduce new tasks; (iii) how it will relate and co-ordinate with other tasks that although not in the focus of the evaluation, show to be connected in the current working practice.

While some understanding of such functional relationships will come from qualitative and quantitative research methods, it is postulated that evaluation is a *learning* cycle where initially the most valuable and usable feedback is not so much or exclusively on the features of the prototypes but on the structure of the work practice, how tasks are related and organized to achieve the goals that are partly defined by the organization and partly are worked out by the operators (see the notion of “finishing the design” [8]. In this respect, it is our initial model of the operational environment and not only the prototype, which is “tested” during the cyclical evaluation and iteratively revised. The better we understand the activity the more our evaluation can be focused on the right “unit of analysis” that is likely to go beyond what we have originally focused on.

IV. PROGRESS TO DATE

The research is now in its early stages. An on going literature review is currently looking at theories and models that study air traffic controller activities as a system coupled with their context, rather than as a linear set of operations. This latter approach has known spread diffusion in Human factors and HCI communities despite studying tasks in isolation can lead to oversimplification of the real life operating conditions, thus introducing potential for erroneous actions.

The objective of the review is to provide an inventory list of principles to study collaborative human activities in relation to the context. Such approaches go beyond the traditional focus on individual action and present one or more of the following characteristics:

- (a). Go beyond a component de-composition of human activities, typical for instance of Task Analysis;
- (b). Characterize the contextual factors where activity takes place;
- (c). Might include System Theory concepts such as functional relationship and self regulation.

So far the review has covered works on literature on work group design, accident models and Human Reliability Assessment. The review is covering, but is not limited to, the following theories/models:

- Structural Systemic Theory of Activity (STST) [9]
- Soviet Cultural Psychology [10];
- Systemic Accident Model (STAMP)[11];
- Contextual Control Model (COCOM)[12, 13];
- Cognitive reliability and Error Analysis Method (CREAM)[14];
- Socio Technical System Theory (STST)[15].

V. ACKNOWLEDGEMENT

This study is part of an ongoing doctoral research programme sponsored by EUROCONTROL Experimental Centre, Brétigny sur Orge, France. Mentoring and progress tracking are provided by Mr Alan Drew and Mr Marc Bourgois.

REFERENCES

- [1] D. Randall, "What is Common Information?," in *Workshop on Cooperative Organization of Common Information Spaces*, Denmark, 2000.
- [2] K. Schmidt and L. Bannon, "Taking CSCW seriously: supporting articulation work," *Journal of Computer Supported Cooperative Work. J Collaborative Computing*, vol. 1, pp. 7-40, 1992.
- [3] B. R. Winthereik and S. Vikkello, "ICT and integrated care: Some dilemmas of standardising inter-organizational communication," *Computer Supported Cooperative Work*, vol. 14, pp. 43-67, 2005.
- [4] C. Chalou, K. Zeghal, P. Martin, and F. Dowling, "NOP/TTA Validation Strategy," EUROCONTROL, Brétigny sur Orge 2007.
- [5] P. Checkland, *Soft systems methodologies: A 30-year retrospective*, 1999.
- [6] P. Amaldi, R. Lane, S. Pilverdier, and M. Stoner, "RVSM4. Final Report," EUROCONTROL, Brétigny, France 1999.
- [7] S. Rozzi, P. Woodward, P. Amaldi, B. Fields, and W. Wong, "Evaluating Combined 2D/3D Displays for ATC," in *5th INO EUROCONTROL Innovative Research Workshop*, Brétigny sur Orge - France, 2006, pp. 167-174.
- [8] K. Vicente, J., *Cognitive Work Analysis: Toward safe, productive, and healthy computer-based work*. NJ: Erlbaum, 1999.
- [9] G. Bedny and W. Karwowsky, *A Systemic-Structural Theory of Activity*. Boca Raton, FL: CRC Press Taylor & Francis Group, 2007.
- [10] L. S. Vygotsky, *Mind in Society. The Development of higher Psychological Processes*. Cambridge, MA: Harvard University Press, 1978.
- [11] N. Leveson, "A New Accident Model for Engineering Safer System," in *MIT Engineering System Division Internal Symposium*, Boston, US, 2002.
- [12] E. Hollnagel, "Contextual Control Model (COCOM)," 2005.
- [13] E. Hollnagel and D. Woods, D., *Joint Cognitive System. Foundations of Cognitive Systems Engineering*. Boca Raton, FL: CRC Press Taylor & Francis Group, 2005.

- [14] E. Hollnagel, *Cognitive reliability and error analysis method (CREAM)*. London, UK: Elsevier Science Ltd, 1998.
- [15] F. E. Emery, *Characteristics of Socio-Technical Systems*, 1959.

Optimising the Predictability and Flexibility of Dynamic System: Case of 4D Aircraft Trajectory of Air Traffic Management

Trung-Tuyen Hoang
 EUROCONTROL Experimental Centre
 Bretingy sur Orge cedex, France
 Email: trung-tuyen.hoang.ext@eurocontrol.int

Abstract—The 4D trajectory is envisioned as the kernel of the future Air Traffic Flow Management system. In this research, we propose an approach dealing with the predictability and flexibility of the system using 4D trajectory. A mix integer programming problem is proposed to minimize the deviation from actual flown 4D trajectory in relationship to the reference trajectory.

Key words: Air Traffic Flow Management, 4D Trajectory, mix integer programming.

I. INTRODUCTION

In Air Traffic Management (ATM), punctuality is essential to the smooth operations for the safety of flights knowing that most flights are subjected to operational uncertainties due to quality of weather forecast and/or technical and logistics issues. Traditional ATFM systems are flight-based, i.e. the schedule are established with discrete events and determinist approach. Because of operational uncertainties, there exist gaps between scheduled and executed traffic [3]. Removal of these gaps can lead to a better use of airport resources and improve the punctuality of the system.

The SESAR (Single European Sky ATM Research) documents identified the sources of uncertainties and defined the future system based on the notion of 4D trajectory in order to reduce the uncertainties, increase the flight punctuality and safety of flight,... A 4D contract is a set of couple space-time $(O_1, t_1), \dots, (O_n, t_n)$ where O_1, \dots, O_n is the space coordinate of waypoints and t_i is the estimated time for arrival (ETA) at which aircraft must reach the waypoint O_i . The time constraint increases the predictability of the system and but degrades the flexibility of the system.

In SESAR, the Key Performance Areas called Predictability and Flexibility are defined as follows:

- Predictability is the ability of the ATM system to ensure a reliable and consistent of 4D Trajectory performance. i.e the ability to control the variability of the deviation between the actually flown 4D trajectory of aircraft in relationship to the Reference Business Trajectory
- Flexibility is the ability of the ATM system and airports to respond to "sudden" changes in demand and capacity: rapid change in traffic patterns, last minute notifications or cancellations of flights, change to the Reference Business Trajectory, late aircraft substitutions,

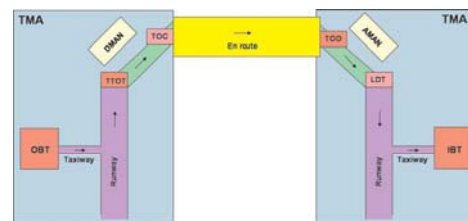


Fig. 1. Flight scheme

sudden airport capacity changes, late airspace segregation request, weather, crisis situation, etc.

Typically a flight time is decomposed of three stages:

- Taxi-out time ($TOT - OBT$), the time elapsed between departure from the origin airport gate (Pushback time: OBT) and wheels off (Takoff Time: TOT). This stage is assisted by a Departure Manager which aimed at optimizing the sequence of traffic in the terminal area including the minimization of flight times by increasing the accuracy of estimated time of departure.
- Taxi-in time ($IBT - LDT$), time time elapsed between wheels down (Landing Time: LDT) and arrival at destination airport gate (In Block Time: IBT). This stage is assisted by an Arrival Manager which aimed at optimizing the sequence of arrival traffic based on available runway and preferred aircraft arrival time data.
- In-flight time ($LDT - TOT$), the total time an aircraft is in the air between an origin-destination airport pair, i.e from the wheels-off at the origin airport to wheels-down airport at the destination.

Amongst the three stages of flight, the predictability at the first and last stages are not good enough: Taxi out time and taxi-in time. Therefore improving the predictability at the first and the last stage of the flight are important. For the better use of airport resources (gates and runways) and aircraft, it is very important in ATFM to be able to predict the takeoff time from pushback time and top of descent from takeoff time for each flight. In other words, it is very important to be able to control the uncertainties at takeoff time and top of descent to optimise the available of runways and to schedule the next flight. Currently, the time window at pushback time is between

fifteen minutes ahead of planned time and fifteen minutes after the planned pushback time. Similarly the time window at takeoff time is between five minutes prior to and ten minutes after the planned takeoff time. One may suggest that the large margin at pushback time (and takeoff time respectively) are the sources of the large uncertainties at the takeoff time (and Top of Descent time respectively). Therefore a reduction of time window at pushback time and departure time may have an positive impact on the uncertainties of the system.

The aim of this research is twofold:

- Study the influence of the time window at pushback time on the takeoff time and the impact of time window at takeoff time on top of descent time. Then we can reasonably modify the time window to reduce the uncertainties of the takeoff time and at the top of descent time.
- Build the predictable model for the top of descent. The predictable model has two important properties: (1) Predictable property, i.e. It will provides the top of descent time and (2) Flexible property, i.e. it must take into account the operational uncertainties

A. Related work

Previous work related to this research topic includes result obtained by [3]. The authors proved that systematically there are gaps between planned and executed traffic. This result means that it is impossible to eliminate the uncertainties in scheduling of flight. Concerning to the prediction takeoff time, the most common approach is use the statistical model to establish the probability distribution of departure delay then to deduce the taxi-out time. This approach used by [5] in their work as they used the queuing theory to estimate the taxi-out time. The authors identified the takeoff queue size as an important factor affecting the takeoff time and validated the model by calculating the taxi-out time for each runway configuration while [10] focused on the distribution of departure time to deduce the model for the estimation takeoff time.

The research carried out by [8] used the Markov decision process to determine the optimal trajectory of multi-aircraft under uncertainty.

None of these studies had investigated the impact of the margin of pushback time on takeoff time and margin of takeoff time on the top of descent.

Our research is not to provide only the takeoff time predictive model but also aims at incorporating the "sudden change" i.e. we investigate a trade-off between the predictability and flexibility.

II. MODELING

A. Assumptions and Notations

To model this problem, our assumptions are as follows:

- Flight Plan: 4D Reference Trajectory including:
 - Origin airport, destination airport.
 - Scheduled departure time (TakeOff Time: TOT), Top of Descent Time (TOD).

- 4D trajectory defined as a set of consecutives space-time, denote $(O_1, t_1), \dots, (O_n, t_n)$ where O is the space coordinates, and t is estimated time that the flight will reach O at t .

- The aircraft flies along the line between two waypoints with constant speed.

Notations:

- F is set of flights.
- For each flight f :
 - Q_f is the set of all possible routes.
 - For $r_f \in Q_f$, $(O_1^{r_f}, t_1^{r_f}), \dots, (O_{n_{r_f}}^{r_f}, t_{n_{r_f}}^{r_f})$ is the set of consecutive waypoints. In our problem, t_1^f is the Takeoff time and $t_{n_f}^f$ is top of descent time, we can extend our model for further flight path but for the moment we will focus on this segment of flight. So n_f is the number of waypoints on the f 's contract Trajectory.
 - $v_i^{r_f}$ is the aircraft speed between two waypoints $O_i^{r_f}$ and $O_{i+1}^{r_f}$. $v_{max}^{r_f}, v_{min}^{r_f}$ are the boundaries of v_f . These values can be obtained from database of Aircraft performance(BADA) of EUROCONTROL Experimental Centre.
 - $m_f(x_t, y_t, z_t, t)$ is the position of aircraft f at t .

From these notations, we have:

- 1) The flight f takes off at $t_{r_f}^1$ will reach the waypoints $O_2^{r_f}, \dots, O_{n_{r_f}}^{r_f}$ at $\frac{d(O_1^{r_f}, O_2^{r_f})}{v_1^{r_f}}, \frac{d(O_{n_{r_f}-1}^{r_f}, O_{n_{r_f}}^{r_f})}{v_{n_{r_f}-1}^{r_f}}$.
- 2) The time elapsed from take off time to top of descent is $\sum_{i=1, \dots, n_{r_f}, r_f \in Q_f} \frac{d(O_i^{r_f}, O_{i+1}^{r_f})}{v_i^{r_f}}$. Then the delay time is $d_{r_f} = \sum_{i=1, \dots, n_{r_f}, r_f \in Q_f} \frac{d(O_i^{r_f}, O_{i+1}^{r_f})}{v_i^{r_f}} - (t_{n_f}^f - t_1^f)$.
- 3) The total delay for all flight is:

$$TD = \sum_{r_f \in Q_f, f \in F} d_{r_f} \quad (1)$$

$$= \sum_{r_f \in Q_f, f \in F} \left(\sum_{i=1, \dots, n_{r_f}, r_f \in Q_f} \frac{d(O_i^{r_f}, O_{i+1}^{r_f})}{v_i^{r_f}} - (t_{n_f}^f - t_1^f) \right). \quad (2)$$

B. Constraints

- Separation minimum: Two any aircrafts must be separated by a minimum distance. i.e $d_{f,g}(t) = \|m_f(x_t, y_t, z_t, t) - m_g(x_t, y_t, z_t, t)\|_2 \geq r$, with r is minimum separation and $f, g \in F$. The real number r depends on the aircraft types: heavy, medium or light.
- Sector capacity: The number of aircrafts in the sector at any moment should not exceed the sector capacity. We define the indicator function $I_{fj}(t)$, whose value is 1 if the aircraft is in the sector j at t and 0 otherwise. This constraint can be represented by: $\sum_{f \in F} I_{fj}(t) \leq C_j(t)$, $\forall j$ and t

The goal is to minimize the total delay (TD).

But we can not solve this problem of optimisation with real decision variables v_t .

We denote v_{ref}^f the reference speed of flight f . This speed depends on the aircraft characteristics like the engine types, the flight level,... This speed is used to establish the Reference Trajectory of flight f . We can rewrite the total delay under the form:

$$TD = \sum_{f \in F} \left(\sum_{i=1, \dots, n_{r_f}, r_f \in Q_f} \frac{d(O_i^f, O_{i+1}^{r_f})}{v_{ref}^f} \frac{v_{ref}^f}{v_i^{r_f}} - (t_{n_f}^f - t_1^f) \right). \quad (3)$$

In order to keep the aircraft along the reference trajectory, we use the speed ajustement and alternative route as tools. Theoretically, the ratio of real speed to reference speed can vary from 0.001 to infinity but in this framework this ratio is limited from 0.8 to 1.2. Note x_i^f is the integer number, $x_i^f \in 0, \dots, 40$ and we can approximate the ratio of speed variation as $\frac{80+x_i^f}{100}$. The problem to solve is:

$$Min \sum_{f \in F} \left(\sum_{i=1, \dots, n_{r_f}, r_f \in Q_f} \frac{d(O_i^f, O_{i+1}^{r_f})}{v_{ref}^f} \frac{80+x_i^f}{100} - (t_{n_f}^f - t_1^f) \right). \quad (4)$$

under the constraints sector capacity and minimum separation.

To simplify the problem, we can divide the segment into two phases: climbing phase and cruising phase. During the cruising phase, we can suppose the ratio of speed variation within 0.95 to 1.05 and during the climbing phase, we limit this ratio of speed variation within 0.9 to 1.1.

- According to the simplification above, the speed variation during the cruising phase is limited between 0.95 and 1.05. The ratio can be represented by $\frac{v_i^f}{v_{ref}^f} = \frac{95 + \sum_{j=1, \dots, 10} x_{ij}^f}{100}$ for $i = 2, \dots, n_{r_f}$.
- This ratio during the climbing phase can be represented by $\frac{v_i^f}{v_{ref}^f} = \frac{90 + \sum_{j=1, \dots, 20} x_{ij}^f}{100}$, with $i = 1$.
- x_{ij}^f is the decision variables 0 – 1. And $x_i^{r_f} = \sum_{j=1, \dots, 10} x_{ij}^f$, with $i = 2, \dots, n_{r_f}$ is the level of speed modification during cruising phase of flight f along the route r_f
- $x_i^{r_f} = \sum_{j=1, \dots, 20} x_{ij}^f$, with $i = 1$ is the level of speed modification to minimize the deviation from the reference path

$$MinTD = Min[x + y] \quad (5)$$

where

$$x = \sum_{f \in F} \left(\sum_{i \geq 2, r_f \in Q_f} \frac{d(O_i^f, O_{i+1}^{r_f})}{v_{ref}^f} \frac{90 + \sum_{j=1, \dots, 10} x_{ij}^f}{100} - (t_{n_f}^f - t_1^f) \right) \quad (6)$$

and

$$y = \sum_{f \in F} \left(\sum_{i=1, r_f \in Q_f} \frac{d(O_i^f, O_{i+1}^{r_f})}{v_{ref}^f} \frac{90 + \sum_{j=1, \dots, 20} x_{ij}^f}{100} - (t_{n_f}^f - t_1^f) \right) \quad (7)$$

We have to solve the problem (5) under the constraints of sector capacity and minimum separation.

C. Problem complexity

This is a Integer programming problem, the constraints are sector capacity and minimum separation.

- For each flight we have: $\sum_{r_f \in Q_f} n_{r_f}$ decision variables
- Total decision variables: $\sum_{f \in F} (\sum_{r_f \in Q_f} n_{r_f})$
- This sum is majored by $R||F|||N||$ with $N = \max_{r_f \in Q_f, f \in F} n_{r_f}$ and R is the number of possible routes that an aircraft can fly along of.

III. FUTURE WORK

The research has been initiated and real data is currently under investigating for initial testing of the model. In the Doctoral symposium the author wishes to be advised on the topics such as:

- Investigation of the current taxi-out time and taxi-in time.
- Simulation with different airports.
- Incorporation of the "random walk" into model.
- The model will be validated by comparing the results obtained from simulation and real results recorded by radars.

REFERENCES

- [1] SESAR Consortium. The atm target concept, d3, September 2007.
- [2] F. Ferchaud. *Etude des Zones d'Absorption pour la Gestion de Flux du Trafic Arien*. PhD thesis, Universit de Bordeaux 1, 2006.
- [3] C. Gwiggner and V. Duong. Average, uncertainties and interpretation in flow planning. *Proceedings of the 2th International Conference Research in Air Transportation*, 2006.
- [4] A. Haraldsdottir, J. Scharl, M. Berge, E. Schoemig, and M. Coats. Arrival Management with Required Navigation Performance and 3D Paths. 2007.
- [5] H. Idris, J.P. Clarke, R. Bhuvu, and L. Kang. Queuing model for taxi-out time estimation. *Air Traffic Control Quarterly*, 10(1):1–22, 2002.
- [6] E. Mueller and B. Chatterji. Analysis of aircraft arrival and departure delay characteristics. Technical report, AIAA's Aircrafts Technology, Integration, and Operations, Los Angeles, California, 1-3, October 2002.
- [7] A. Mukherjee. *Dynamic Stochastic Optimization Models For Air Traffic Flow Management*. PhD thesis, University of California, Berkeley, 2004.
- [8] A. Nilim, L. El Ghaoui, and V. Duong. Multi-aircraft routing and traffic flow management under uncertainty. In *Proceeding ATM Seminar 2003*, 2003.
- [9] R.A. Shumsky. *Dynamic statistical models for the prediction of aircraft take-off times*. PhD thesis, Massachusetts Institute of Technology, 1995.
- [10] Y. Tu, M. Ball, and W. Jank. Estimating Flight Departure Delay Distributions A Statistical Approach With Long-term Trend and Short-term Pattern. *Journal of the American Statistical Association*, 2006.

The Integrator Market

Actors and Their Strategies

Evy Onghena

Department of Transport and Regional Economics
University of Antwerp
B-2000 Antwerp, Belgium
evy.onghena@ua.ac.be

Abstract— Integrators – companies that offer vertically integrated, time-definite, door-to-door transport – have a significant market power. However, insight into the market structure and the cost structure of these companies, as well as into the consequences of their expansion and cooperation strategies, is lacking. The purpose of this paper is to analyse the integrator market from an organizational perspective, describing the strategic behaviour of the market participants. This paper provides a clear insight into the major actors of the integrator market and their expansion and cooperation strategies.

Integrators; strategic behaviour; expansion; cooperation; industrial economics

I. RATIONALE AND SETTING

The globalization of the world economy involves that companies are reorganized, resulting in a worldwide spread of their production systems. In the ‘global village’, where production and consumption are increasingly internationalized, transport and logistics services are crucial to reduce cycle times and increase products’ speed to market. Logistics service providers with an extensive network allow to keep the transport costs of international trade down, while realizing a more efficient use of production factors.

However, during the past decade, the requirements for transport and logistics have become stricter and more numerous. The share of high-value and/or time-sensitive goods with a short economic life cycle (e.g. high-tech products, textiles, pharmaceuticals, etc.) has increased. Therefore, there is a growing need for fast and reliable transport that allows companies to gain access to global markets and supply chains.

Integrators are crucial for the delivery of those transport services. Since they are able to control the total supply chain (‘one-stop shopping’), the strategic and operational importance of integrators for the commercial and production processes of shippers is extremely high. Inventory costs and total distribution costs can be minimized by factors such as speed, reliability, guaranteed delivery within a certain time, visibility and flexibility [1]. In addition, integrators enlarge

companies’ catchment area and their options for the location of production and assembly facilities [2].

Besides the importance of integrators for shippers, the competitiveness of a region is also partly determined by the presence of one or more integrators. This can be illustrated very well by the relocation of DHL’s hub activities from Brussels National Airport to Leipzig in 2008, which will have far-reaching microeconomic and macroeconomic consequences for the airport and for the Belgian economy in general.

Over the past decade, international express has grown at more than twice the rate of total worldwide air cargo¹ traffic, averaging 12.9%² annually [3]. The integrators account for almost 85% of the world’s express shipments [4]. Currently, only four players are fully integrated across all transport modes, including air transport: UPS, FedEx, DHL and TNT.

A market structure in which only a few sellers account for a substantial proportion of total sales, is an oligopoly. The main challenge for the analysis of an oligopoly is strategic behaviour or strategic interdependence between competitors, which means that each firm’s optimal behaviour depends on its assumptions about its rivals’ likely reactions [5].

Despite the importance of integrators for shippers and economic regions (embedment), the strategies of integrators are hardly examined in literature. Insight into the market structure and the cost structure of these companies, as well as into the consequences of their expansion and cooperation strategies, is lacking.

This paper aims at an analysis of the integrator market from an organizational perspective, focusing on the growth and relational strategies of the market participants. This analysis fits in a broader industrial-economic analysis of the

¹ According to Boeing, air cargo consists of freight, express and air mail. In some publications, air cargo is used as a more general term than air freight. In this thesis however, both terms will be used without any distinction.

² Measured in revenue tonne-kilometers (RTK).

integrator market. Section 2 gives an overview of the research questions and the methodology. The results of this paper correspond with the first step of a four step methodology. Section 3 provides a definition of integrators. Section 4 outlines the business positioning of integrators within the air cargo industry. In section 5, the Big Four integrators and their main expansion and cooperation strategies are considered. Section 6 deals with new competitors in the market. Section 7 finally summarizes the main conclusions of this research.

II. RESEARCH QUESTIONS AND METHODOLOGY

This paper gives the framework in which four main hypotheses will be tested:

1. Each integrator will finance its growth independently.

2. New players will enter the integrator market.
3. Total costs for vertically integrated companies (integrators) offering an integrated, door-to-door supply chain are lower than the sum of costs for companies offering parts of the supply chain.
4. Integrators will get involved into different types of horizontal and vertical cooperation outside the integrator market.

For testing these hypotheses, a four step methodology will be used. The research methodology, including inputs and outputs, is depicted in figure 1.

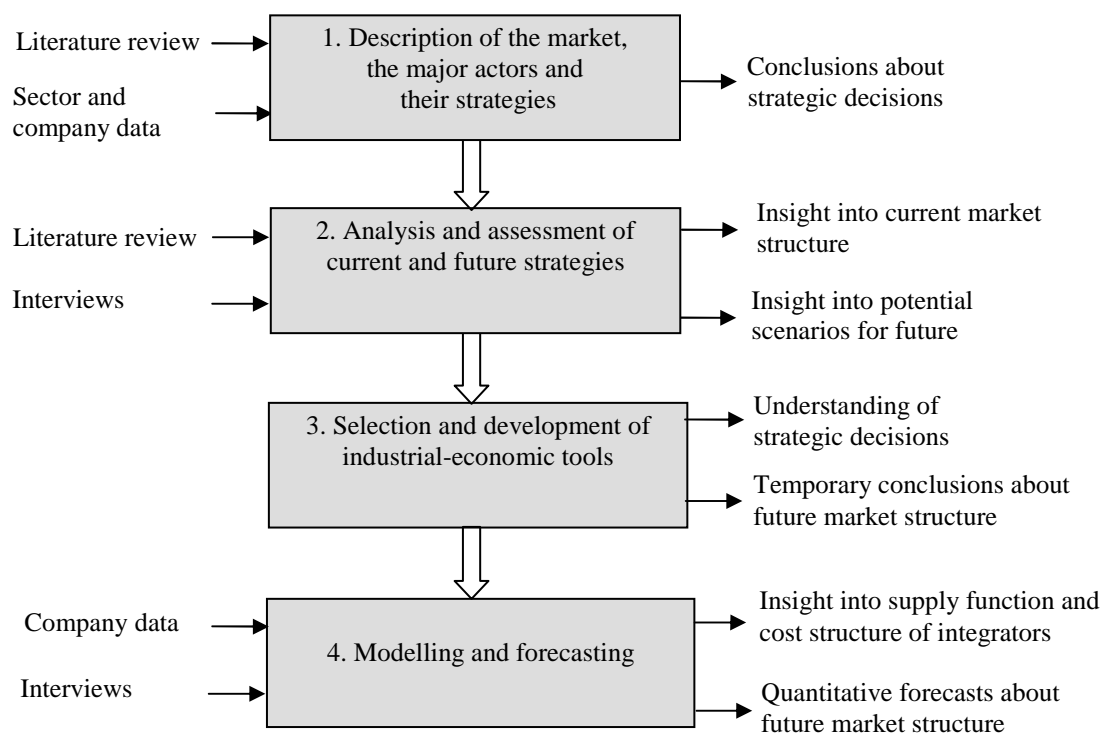


Figure 1. Research Methodology

First, the integrated express market will be analysed from an organizational perspective, describing the behaviour of the market agents. The first step will result in some important conclusions about the expansion and cooperation strategies adopted by the integrators in the past. In the second step, the integrators' expansion and cooperation strategies will be analysed from an industrial-economic perspective. Based on this analysis, an insight will be gained into the current market structure and the future development of the market. In a next step, industrial-economic tools will be selected and developed in order to understand the expansion and cooperation strategies adopted by the integrators in the past. Subsequently, an industrial-economic simulation model applicable to the integrator market will be developed. This model will allow understanding, explaining and forecasting strategic behaviour of the current market players. Finally, the model will be used

to make forecasts about the future market structure and the economic consequences of strategic decisions.

III. INTEGRATED EXPRESS CARRIERS OR INTEGRATORS: A DEFINITION

Many definitions of integrators can be found in the literature. In this paper, the following definition, based on [6], will be used: "integrators are vertically integrated express companies that provide time-definite, door-to-door services and, for that purpose, perform their own pick-up and delivery services, operate their own fleet of aircraft and trucks and tie it all together with advanced information and communication technologies."

This definition contains the most important characteristics of integrators, namely:

- integrated door-to-door service;
- own transport assets;
- strongly developed ICT-skills (e.g. tracking and tracing)

IV. BUSINESS POSITIONING OF INTEGRATORS WITHIN THE AIR CARGO INDUSTRY

Within the air cargo industry, two major cargo types can be distinguished, namely general cargo (heavy lift) cargo and express cargo [7].

The air cargo industry consists of different actors, which can be categorized into three main sub-industries:

1. general or traditional air cargo industry;
2. air express industry;
3. postal services industry.

This categorization is presented in figure 2.

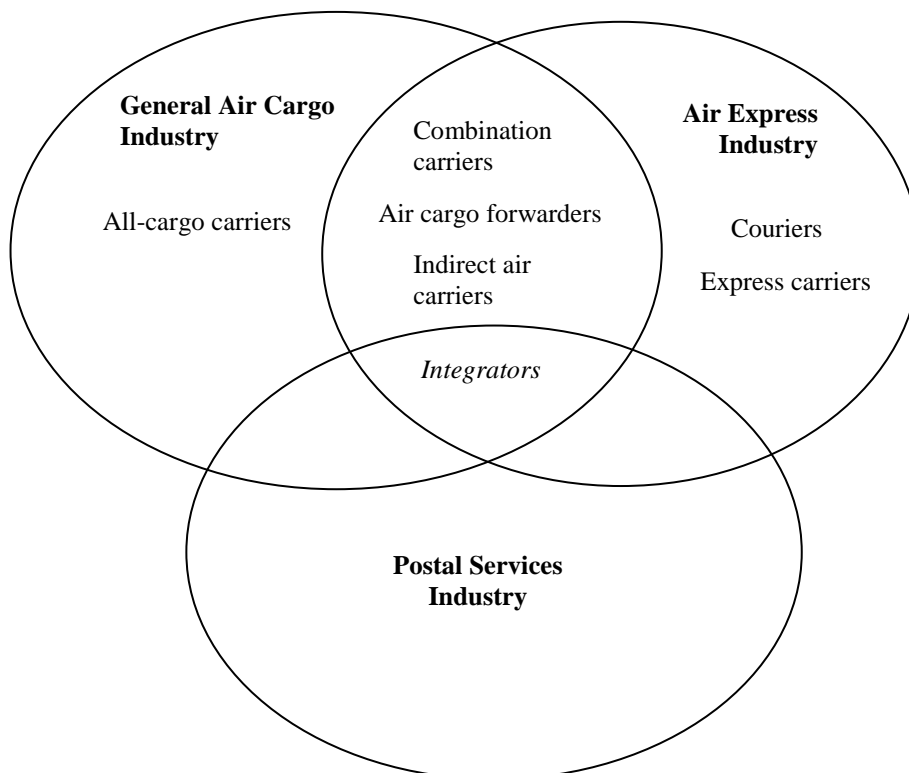


Figure 2. Categorization of the air cargo industry

The general or traditional air cargo industry consists of actors providing airport-to-airport services. General air freight is the most common type of air cargo movement. General air freight companies focus on the transportation of specialized and/or consolidated freight, consisting of individual shipments from many different customers grouped together and transported as one large shipment in an air container [8]. All-cargo carriers, such as Cargolux, Polar Air Cargo, Nippon Cargo Airlines and MK Airlines, belong to the general air cargo industry.

The air express industry is described as “an industry of which the core business is the provision of value-added, door-to-door transport and deliveries of next-day or time-definite shipments, including documents, parcels and merchandise goods” [9]. Typical characteristics of the air express industry are time-definite delivery of goods, door-to-door delivery and

full tracking control of shipped goods. Two main types of players are part of the air express industry, namely couriers and express carriers. Examples are GeoPost, General Logistics Systems and Ziegler Express.

Combination carriers, air cargo forwarders and indirect air carriers are positioned at the intersection of the general air cargo industry and the air express industry. Due to the strong competition of integrators, some combination carriers try to stay competitive by offering time-definite, door-to-door services themselves and, thus, are entering the air express industry. This strategy can be illustrated by the case of Cathay Pacific, which expanded its services to the express industry by offering its Wholesale Courier and Cargo Express services. Despite the fierce competition between integrators and combination carriers, both types of market players can also be complementary to one another. This is illustrated very well by

the case of Northwest Airlines, which will lose DHL Express, its most important cargo customer, as from the end of 2008. Northwest Airlines is obliged to reconsider its cargo activities due to this loss.

Air cargo forwarders and indirect air carriers are also positioned at the intersection of the general air cargo and the air express industry since they are increasingly offering value-added, logistics services and door-to-door delivery. Whereas in the past they were merely considered as intermediaries between shipper and airline, they are increasingly acting nowadays as integrated logistics service providers. Examples of these market actors are Kuehne + Nagel, Schenker and Panalpina.

Public postal operators, which originally belonged only to the postal services industry, are currently positioned at the intersection of the postal services industry and the air express industry. The reason for this is the entering of postal service providers (e.g. Deutsche Post World Net, TPG, La Poste, etc.) in the air express market.

Integrators are part of all three sub-industries since they are offering a broad service portfolio including general air freight services, express services and mail services.

The most important conclusion of this section is that the boundaries between the different segments of the air cargo

industry are blurring since the services offered by the different players start to overlap. This intensifies competition and leads to various forms of cooperation. An example of the overlap between different segments of the air cargo industry is the case of Austrian Post, which is increasingly developing its B2B express parcels business by the acquisition of parcel carriers such as trans-o-flex (Germany, 2006), Scherübl (Austria, 2007), Merland Express (Hungary, 2007) and Road Parcel (Hungary, 2007). Austrian Post aims at becoming a niche player in the European B2B express market. Another example is Lufthansa Cargo, which created time:matters in 1992, a subsidiary specialized in same-day, door-to-door express transport.

V. EXPANSION AND COOPERATION STRATEGIES OF THE BIG FOUR

The Big Four integrators have very different origins and were founded in different eras. This influences their corporate culture and, hence, results in distinctive ways of strategic thinking and decision-making. Therefore, to understand strategic decisions made by the key players in the industry, the historical development of these companies was investigated, as well as their recent acquisitions and cooperative agreements. The impact of strategic decisions on the integrators' service portfolio and geographical coverage is shown in tables 1 and 2.

TABLE I. ORIGINAL AND CURRENT SERVICE PORTFOLIO OF INTEGRATORS

	Original focus	Main reasons for shift in focus	Current focus
UPS	Ground	Creation of UPS Airlines (1985) Acquisition of Fritz Companies (2001) & Menlo (2004)	Ground Air Ocean
FedEx	Air	Acquisition of Caliber (1997)	Air Ground
DHL	Air	Acquisition by DPWN (2003) Acquisition of Danzas (1999) & Exel (2005)	Air Ground Ocean
TNT	Ground	Creation of European air network (1987)	Ground Air

Source: Company Websites

TABLE II. ORIGINAL AND CURRENT GEOGRAPHICAL FOCUS OF INTEGRATORS

	Original focus	Main reasons for shift in focus	Current focus
UPS	US	Cooperation with Sinotrans (China, 1988) Acquisition of LYNX Express (UK, 2005) Acquisition of Challenge Air (US, 1999)	US, Europe, Asia-Pacific, Latin America
FedEx	US	Acquisition of Flying Tigers (China, 1989) and route authority from Evergreen International Airlines (US, 1987) Acquisition of Flying-Cargo (Hungary, 2007) Acquisition of Prakash Air Freight (India, 2007)	US, Asia-Pacific, Europe
DHL	US	Acquisition by DPWN (Germany, 2003)	Europe, US, Asia-Pacific
TNT	Europe	Acquisition of Speedage (India, 2006) Acquisition of Hoau (China, 2007) Acquisition of Mercúrio (Brazil, 2007)	Europe, Emerging Markets (China, India, Brazil, etc.)

Source: Company Websites

As illustrated in table 1, UPS was originally founded as a ground-based delivery company. The company entered the overnight air delivery business in the 1980s and created its own airline, UPS Airlines, in 1985. Since its primary focus was on ground delivery, its current ground network is still larger than that of its rivals. On the contrary, FedEx originally focused on overnight air express parcel services, a niche that did not exist at the time. The acquisition of Caliber in 1997 diversified FedEx's product portfolio and brought it in direct competition with UPS's ground-based services. DHL was founded as a company shipping documents by airplane. Through the cooperation with Deutsche Post, it gained access to the most extensive network of ground-based operations in Europe. Until now, the main strength of DHL is its air network, rather than its ground-based operations. TNT was created as a company focusing on road and rail transport services. In 1987, TNT started to operate its European air network.

UPS, as well as DHL, offer ocean freight services. UPS and DHL both entered the ocean freight segment by means of strategic acquisitions. UPS started to offer ocean freight services in 2002 after its acquisition of Fritz Companies in 2001. The purchase of Menlo in 2004 further strengthened UPS's position in the ocean freight market. The most important motive to add ocean freight to its product portfolio was not growth but customer satisfaction. DHL entered the ocean freight business in 1999 after the acquisition of Danzas Holding. The purchase of Exel in 2005 reinforced its logistics division.

UPS and DHL have been able to position themselves in logistics and supply chain management, in contrast with FedEx and TNT. UPS decided to become active in supply chain management on the request of its customers. Once

decided to enter that business, UPS started to take over other companies with a suitable product portfolio in the first place and the right geographical coverage in the second place [10].

FedEx and TNT have a similar business model, as they mainly focus on transport and transport-related value added services. FedEx is less interested in logistics and 3PL services since the margins on these services are small. TNT even decided to exit the logistics and freight forwarding businesses [11].

While UPS and DHL want to offer all types of products (express and logistics) on a worldwide scale, FedEx and TNT are focusing more on specific products (particularly express transport) and specific markets. One of TNT's strategic choices is the development of special services, namely the transport of shipments with special requirements concerning volume, weight or other specifications which fall outside the normal standards for express transport [12].

Table 2 indicates that both UPS and FedEx have their origins within the US. UPS expanded to Europe, Asia-Pacific and Latin America through acquisitions and cooperative agreements. FedEx entered the Asia-Pacific market through the purchase of Flying Tigers in 1989 and the acquisition of Chinese route authority from Evergreen International Airlines in 1995. FedEx became active on the European market in the 1990s but its European operations were loss-making. The company decided to withdraw from Europe in 1993 and to concentrate on its core business, namely transatlantic services. The acquisition of Flying-Cargo Hungary Kft. in 2007 demonstrates FedEx's strategy to re-enter the European market and fits in the company's Eastern European expansion strategy. Flying-Cargo Hungary is FedEx's Hungarian global service participant since 2003. Table 3 provides an overview

of the key facts of the Flying-Cargo Group, a logistics service provider offering solutions along the entire supply chain. FedEx will continue with acquisitions in Europe the coming years. It is expected that its acquisitions in Europe mainly are in the air freight business, since FedEx is not interested in offering ground-based services in Europe itself. Currently, FedEx offers ground-based services in Europe through alliances with e.g. La Poste. This is in contrast with DHL's strategy to establish its own ground network in the US.

TABLE III. FLYING-CARGO GROUP: KEY FACTS

Flying-Cargo Group	
Founded	1982
Headquarters	Lod, Israel
Employees	1200
Network	operations in 13 locations

Source: own composition based on http://www.flying-cargo.com/pages_eng/about.asp

DHL was founded in the US but rapidly extended its geographical network. The cooperation between Deutsche Post and DHL entailed a convergence of services, comparable to the service convergence resulting from the acquisition of Caliber by FedEx, which shifted FedEx into ground-based operations. DHL, originally focusing on air express services, gained access to the most extensive network of ground-based operations in Europe through the cooperation with Deutsche Post. Deutsche Post, on its turn, obtained access to an international air express network in Europe, Asia and North America [13].

TNT originally focused its business on Europe. Nowadays, TNT is the market leader in the intra-European express market. The company seeks to reinforce its intra-European market position, as well as its position in the European domestic markets, by means of organic growth. TNT aspires to a leading position in the emerging markets, of which two are located in the Asia-Pacific region, namely India and China. TNT is expanding its position in India and China mainly through acquisitions. Besides the connection of these markets with Europe, TNT plans to develop a domestic express network in these countries, both on the ground and in the air. This strategy differs from that of UPS and FedEx, which focus on international express and air freight in India and not on domestic traffic. DHL also has a different strategy in these markets as this integrator concentrates on the development of a domestic air network. Concerning TNT's strategy on the world's largest express market, the US, the company does not strive for a large market share on that market. TNT reconciles itself to the dominance of UPS and FedEx in the US and considers the US express market as a mature market with limited growth opportunities. In the US express market, TNT focuses on intercontinental express traffic, in cooperation with local partners, as well as on the development of services for niche markets such as the fashion industry or the financial industry [12].

VI. ENTRANCE OF NEW COMPETITORS IN THE MARKET

The Big Four integrators are facing increased competition from new market entrants. These are mainly traditional freight forwarders which are evolving towards integrated logistics service providers. ABX Logistics Worldwide, Kuehne + Nagel and Panalpina are considered as the main new competitors.

ABX Logistics Worldwide, acquired by 3i in 2006, ranks in the top 12 air and sea freight forwarders worldwide. The company divested its domestic road distribution activities in several European countries to concentrate on its core business: international air, sea and road freight forwarding. Since ABX Logistics is in the first place a freight forwarder and no pure logistics player, it forms alliances with large logistics service providers, e.g. alliance with Penske Logistics in 2006. This strategy allows the company to become one of the world's largest freight forwarders and logistics service providers and to enter into competition with the Big Four integrators. Two years after its acquisition, the private equity group 3i will sell ABX Logistics again. The company's total value in 2008 is estimated at € 600 million, compared with € 80 million in 2006. Table 4 provides some key figures of ABX Logistics Worldwide.

TABLE IV. ABX LOGISTICS WORLDWIDE: KEY FACTS

ABX Logistics Worldwide	
Founded	1993
Headquarters	Brussels (Belgium)
Revenues 2006	€ 2.3 bn
Employees	8000
Network	250 sites under its own name in more than 35 countries. Representation in a further 60 countries by its own exclusive agents

Source: own composition based on <http://www.abxlogistics.com>

Kuehne + Nagel (K+N) has evolved from a traditional international freight forwarder to a leading global provider of integrated supply chain solutions. In 2005, the company acquired ACR Logistics, one of Europe's leading contract logistics providers. The deal fits into K+N's growth strategy in contract logistics. Table 5 provides an overview of some of the company's key facts.

TABLE V. KUEHNE + NAGEL: KEY FACTS

Kuehne + Nagel	
Founded	1890
Headquarters	Schindellegi, Switzerland
Revenues 2007	CHF 20975 million
Employees	51075
Network	830 offices in more than 100 countries

Source: own composition based on <http://www.kn-portal.com/>

Panalpina is described as a leading integrated freight forwarding and logistics company. The company has the intention to expand its global network further, through organic growth as well as by targeted acquisitions and alliances. More details about Panalpina are provided in table 6.

TABLE VI. PANALPINA: KEY FACTS

Panalpina	
Founded	1954
Headquarters	Basel, Switzerland
Revenues 2007	CHF 10592 million
Employees	15301
Network	500 branches in 90 countries. Cooperation with selected partners in other 60 countries

Source: own composition based on
<http://www.panalpina.com/www/global/en/about.html>

VII. CONCLUSIONS

The analysis in this paper shows that the services offered by integrated and non-integrated service providers start to overlap, which leads to a blurring of the boundaries between sub-industries of the air cargo industry. This intensifies competition and leads to various forms of cooperation among service providers.

Concerning the strategies of the Big Four, it can be concluded that DHL and UPS are growing closer to each other as they want to offer all types of products (express and logistics) on a worldwide scale. On the other hand, FedEx and TNT have a similar business model, as they are focusing more on specific products (particularly express transport) and specific markets.

As far as the second research hypothesis is concerned, the results of this paper show that new players are entering the integrator market.

Regarding the fourth hypothesis, it turns out from this paper that it is crucial for integrators to have a global network. They aim to reach a global presence via acquisitions and cooperation agreements outside the integrator market.

According to the next step in the methodology, an industrial-economic analysis of the integrators' current and future strategies will be made. This involves a determination

of the market structure. Although the integrator market is an oligopoly at first sight, the networks of the agents may overlap so that the aggregate market structure may be different. In order to complete the second step, industrial-economic literature will be reviewed, both theoretical and applied to comparable business sectors. In addition, interviews will be done with industry stakeholders in order to better understand the strategic decisions of the industry participants made in the past, as well as to assess the planned strategies and developments for the future.

REFERENCES

- [1] L. Sowinski, "Air Cargo Integrators Do It All", 2006, http://www.worldtrademag.com/CDA/Articles/Feature_Article/4f07bdd039aae010VgnVCM100000f932a8c0.
- [2] H. Baum and A. Henn, "Produktivitäts- und Wachstumseffekte der Kurier-, Express- und Paketdienste für die arbeitsteilige Wirtschaft", Institut für Verkehrswissenschaft, Universität zu Köln, 2004, http://biek.de/download/studie/BIEK_Studie_2004.pdf.
- [3] Boeing, "World Air Cargo Forecast 2006/2007", 2006, <http://www.boeing.com/commercial/cargo/wacf.pdf>.
- [4] US-ASEAN Business Council, "Express Delivery Services: Integrating ASEAN to Global Markets", 2005, http://www.global-express.org/doc/EDS_Report.pdf.
- [5] J. Lipczynski, J. Wilson, and J. Goddard, Industrial Organization, Competition, Strategy, Policy, Harlow, Pearson Education, 2005.
- [6] W.J. Zondag, "Competing for Air Cargo, A Qualitative Analysis of Competitive Rivalry in the Air Cargo Industry", 2006, http://www.fapaa.org/pdf/News/Jun06_CompetingforAirCargoThesis.pdf.
- [7] A. Zhang, C. Lang, and L. Leung, "Intermodal alliance and the rivalry of transport chains: The air cargo market", Transportation Research Part E, vol. 43, pp. 234-246, 2007.
- [8] M. Pirenne, "Worldwide Express Transport", guest lecture for the course Van Breedam and Van de Voorde, academic year 2006-2007, Logistiek en expeditie, Universiteit ANtwerpen, Antwerpen, 2007.
- [9] Oxford Economic Forecasting, "The Economic Impact of Express Carriers in Europe", 2004, <http://www.oef.com/Free/pdfs/euroexpressfinal091104.pdf>.
- [10] De Lloyd, 2007, "UPS laat de boel ook varen", edition 6 July 2007, p. 8
- [11] P.A. Evans, 2007, "Business model of FedEx," interview with Vice President Sales & Marketing Europe, Middle East & Africa FedEx Express, 2 October 2007, Courier and Parcel Logistics Summit 2007 Barcelona.
- [12] G. Meeussen, 2007, "TNT mikt op Europa en groeimarkten", De Tijd, edition 14 September 2007.
- [13] M. Taylor and A. Hallsworth, "Power relations and market transformation in the transport sector: the example of the courier services industry", Journal of Transport Geography, vol. 8, n° 4, pp. 237-247, 2000.

A Construction Rationale to Tailor Crew Resource Management Training to Target Audiences

Vera Hagemann, Sandrina Ritzmann and Annette Kluge

Research Institute for Organizational Psychology
University of St. Gallen
9000 St. Gallen, Switzerland
vera.hagemann@unisg.ch
sandrina.ritzmann@unisg.ch
annette.kluge@unisg.ch

Abstract— This paper gives an overview on the first steps of a 3-year Crew Resource Management and Human Factors training project. A construction rationale consisting of a training needs assessment phase and of theory driven reflections on training design is presented. For the needs assessment, a careful choice and application of methods to gather information is vital, because this information will form the base of training design. Furthermore, a learning theory (instance-based learning theory), training methods, legal requirements and training strategies (cross training, guided team self-correction and team coordination and adaptation training) as well as their contributions to training design are described. The intention to generate a training theory and the development of a classification of training methods along the criteria knowledge, skills and attitudes and theory- or experience-based learning are presented.

Keywords: *Crew Resource Management; Teamwork; Teamtraining; Human Factor; Needs Assessment; Training Theory*

I. INTRODUCTION

Crew Resource Management (CRM) trainings have been utilized in civil and military aviation for more than 20 years now [1]. But although CRM training and the like, as Human Factors (HF) training in aircraft maintenance or Team Resource Management (TRM) training in air traffic control, are well established and, depending on the industry, mandatory, the topic continues to be of great relevance. Recent examples like the lucky ending crash-landing of a British Airways Boeing 777 in January 2008 at Heathrow Airport demonstrate this. CRM trainings have been defined as programs and instructional strategies to train crews to effectively use all their available resources - information, equipment and people - in order to improve safety and performance [2,1].

Although there exists a substantial amount of research in the field, [1] emphasize that "the full impact of CRM training on safety cannot yet be ascertained (p. 393)" and is still not understood. They report findings on the impact of CRM trainings on trainees' *reactions, learning and attitudes, behaviors and/or its impact on the organization*. Participants generally showed positive reactions towards trainings, but the results concerning learning, attitude change, transfer of behavior and organizational effects were mixed. What are the

reasons for this mixed picture? Reference [1] state amongst other reasons that firstly, trainings are *often not tailored to the target audience* and secondly, *programs are often designed by subject matter experts, who know what to teach, but not necessarily how*. Here, the knowledge of training experts is needed.

II. AIM AND RESEARCH QUESTIONS

The *aim of our work* is to tackle these deficiencies in a 3-year CRM and HF training project started in January 2008. The project is a cooperation with a training providing company in the aviation sector which is owned by an airline. It provides pilot and cabin crew training as well as training for maintenance personnel.

In this paper, we present a construction rationale for training and follow *two goals*:

1) We describe the research questions, method and preliminary results of the training needs assessment phase of our project. A sound needs assessment is the first step to well-tailored and effective training programs, as people, tasks, behaviors and the organization have to be taken into account. [3]. Our first four *questions* are:

(a) What is being done by researchers and by practitioners in the field of civil aviation and other areas where team training is conducted?

(b) What is the current practice of CRM training in our partner company?

(c) What contents, i.e. what knowledge, skills and attitudes (KSA) have so far been trained in the area of CRM? How could they be ameliorated by including behavioral descriptions that specify theoretical concepts?

(d) How tight are the legal constraints when it comes to the design of training?

2) We outline results and insights from training research literature that will guide training development on the basis of the results of the needs assessment phase. In this second phase of our project, the focus clearly lies on *how* teaching of CRM could be done. Our last three *questions* are:

(e) How can we translate a learning theory that focuses on experience-based learning into a training theory?

(f) And within this training theory, which methods/tools (lecture, case-study, role-play, exercise, simulation, video,

LOFT or behavioral role modeling) are adequate for gaining the specific required KSA and how big should be the portion of each of these methods in a CRM training?

(g) How can we combine the most successful elements of several training strategies (e.g. cross training, guided team self-correction and team coordination and adaptation training) in order to create the most efficient training intervention to train CRM?

III. PROCEDURE

At the moment, we are doing an exploratory interview study on the actual practice of CRM and similar trainings in different industrial sectors, e.g. air traffic control, military aviation (jets and helicopters), swiss army armoured corps and mechanised units, flight schools and airlines. The underlying reason is that, although numerous theoretical articles have been written about the evolution, evaluation and effectiveness of CRM trainings and empirical studies have been published, *there is very little information on what is actually done in CRM trainings by practitioners, and how trainings are planned and composed.*

In parallel, we started the needs assessment process in our partner company and already evaluate strategies for training design (see figure 1). The next section describes the methods we employ in our needs assessment stage.

IV. METHOD

The goal is to gain a clear picture on accurate, multiple photographs of the current situation [3]. The use of several different techniques helps to avoid methodological biases. In the following, we will shortly describe what techniques we mainly plan to apply to answer our research questions.

A. Best industry practice

Interviews. Our exploratory study is done in an interview study to shed some light on what is done by CRM and HF training practitioners in different companies and industry sectors. We question key informants, i.e. heads of training or responsible persons in charge of CRM in semi-structured interviews. We explicitly ask our interview partners to express their concerns of and personal attitude toward CRM. The goal of these interviews is to get an overview on models and methods and tools used in CRM trainings and on the assumptions and concerns of trainers and training developers.

This shall help us to incorporate best practice approaches, but also to avoid pitfalls that were experienced by our interview partners. First results show that practitioners are generally convinced of the importance of CRM and similar training strategies, but are confronted with several problems as well. Especially in military aviation, it is difficult to motivate trainees for CRM topics, which are often judged as "psycho-babble" or as "charm schools" [4]. This might be a result from the early days of human aspects training, where the focus was much more psychological and less applied.

B. Current practice in our partner company

Interviews. We will conduct interviews with trainers and trainees to gather information on the design and conduct of and the participation in existing CRM and HF trainings. These interviews will shed light on problems and well-working aspects of actual trainings.

Observation. To get insights into the current practice of training, we will host sessions of all trainings with CRM or HF content. The focus will be on *how* training is conducted. Ideally, those trainers will be interviewed whose sessions we host, and those trainees who participate in these sessions. This approach allows us to compare our observations with the statements from trainers and trainees.

C. Legal requirements

Document analysis. The important legal documents for training of human factors aspects are JAR-OPS and JAR-FCL (Joint Aviation Requirements, Operations/Flight Crew Licensing) [5,6] for pilots and cabin crew and Annex 2 of EASA-Part 145 [7] for maintenance staff. The different trainings that have to be conducted (e.g. introductory CRM course, annual recurrent training), their content and prescribed training elements (also called "core elements"), as well as their repetition cycle are specified. Examples of training elements are "assertiveness" for cabin crew members or "stress and stress management", "communication" and "decision-making" for both cabin crew members and pilots. *Although it seems clear what assertiveness means, it is vital for training design to define specific behaviors that are judged as "assertive".* This is part of the specifications of KSA and training requirements described below.

D. Specifications of KSA and training requirements

Literature analysis. Efforts to specify behavioral patterns within training elements have been made before by researchers as well as by airlines themselves and can be found in literature, but mainly to construct rating systems used in performance appraisal of CRM skills. The NOTECHS rating system, for example, which is the European taxonomy of pilot's non-technical skills was composed from three sources: existing rating systems, research findings reported in literature and discussion with subject matter experts [8]. However, behavioral markers used to judge performance should also be applied to construct training, because before assessing performance, the desired competencies have to be trained [9]. But information on behavioral specifications used to build training programs is less readily available.

Training needs assessment <ul style="list-style-type: none"> • Best industry practice • Current practice in our partner company • Legal requirements • Specification of KSA and training requirements
Development of a training concept <ul style="list-style-type: none"> • Special emphasis on experience-based learning • Fit of training methods and training goals • Combination of successful team training strategies

Figure 1. Intended course of action.

This project is sponsored by the Swiss Confederation's innovation promotion agency (CTI)

Document analysis. We will analyze existing training manuals and other training material like movies, cases etc. to get an overview on how training requirements have been translated into training measures.

Interviews. We will conduct Critical Incident interviews with experienced job incumbents to adapt or derive behavioral descriptions and training objectives of training elements. Critical Incident interviews aim at gathering information on situations where CRM behavior played a crucial role. Critical incidents shall contain descriptions of the situation, the task at hand, the actions and their results.

Questionnaires. We will employ questionnaires to gather further information on the behavioral descriptions of training elements we derive from the literature analysis and the Critical Incident interviews. The questionnaire will contain behavioral descriptions like "encourages inputs and feedback from others" (example from NOTECHS; [8]). Pilots will rate these statements on three dimensions: a) importance, b) learnability of this behavior, and c) frequency with which this behavior is demanded in daily work. Our sample will consist of first officers and captains from the airline owning our partner company.

Group discussion. To synthesize the information from Critical Incident interviews and questionnaires, a choice of interview partners and questionnaire respondents will be invited for a group discussion. Group discussions allow it to share the problem and data analysis with participants. The goal will be to challenge our preliminary conclusions. We plan two sessions with ten participants each.

Preliminary results of the activities described will be presented at the doctoral symposium at ICRAT'08. But as the design of training programs is not done with the needs assessment, we also want to provide an outlook on the second phase of our project.

V. TRAINING DESIGN

The focus lies on *tailoring the contents, strategies and tools of a CRM training to the audience*. This procedure is based on the results from training research literature and our needs assessment in a previous phase of this project.

A. Legal requirements

First of all, one characteristic that has to be kept in mind when developing training is the high degree of *regulation* in the aviation industry. Demands on training prescribed by law have to be met, as trainings have to be approved by the regulatory body. As mentioned above, "Core elements" of CRM training, for example, are defined in the JAR-OPS and JAR-FCL [7,6]. These elements have to be taken into account when developing the contents. Moreover, training design must be tailored to the specifications and behavioral markers (that we will get from our needs assessment) of these elements in order to train the desired competencies. Therefore the methods and strategies described below are very essential.

B. Training theory

A superior goal will be the *transfer of a learning theory into a training theory*. This novel and innovative course of action within our study takes *experience-based learning* into account, because we have a close look at the *Instance-based*

learning theory (IBLT) that assumes five sequenced learning mechanisms within the context of dynamic decision making situations [10]. One important learning mechanism and the first one is the formation of "instances", which contain triplets of (1) the situations, (2) the decisions that have been made in these situations, and (3) the utility of these decisions. These instances are often retrieved and re-used in moments of decision-making instead of learned rules or heuristics. Therefore, it is important to give trainees the chance to gain experiences, i.e. instances, during training. Within this setting they can explore their decisions and learn from doing mistakes. Learning is terminated by a feedback update, the last learning mechanism, that helps to understand what had happened and what kind of new strategies are necessary for a successful outcome. A training theory should describe, explain and predict how learning can be enhanced with regard to defined learning objectives by using well considered methods. Our training theory for social decision making situations should devise how relevant aspects of such decision making situations could be trained most successfully by creating "instances". Simulations and role-plays for example, are of importance.

C. Training methods

Furthermore, it is also indispensable to establish a perfect fit between *training methods* and *learning outcomes* [9], in order to achieve the intended goal of a CRM training (e.g. error prevention, decision making, coordination, leadership). The methods and tools used in trainings must be tailored to the tasks and competencies of the trained team to enhance teamwork [11,12]. Competencies combine different KSA necessary to succeed in an organization [13]. A lot of methods like lecture, video based demonstration and practice are declared to be effective in enhancing teamwork [11]. Reference [9] developed a classification if a method supports knowledge (e.g. how to communicate), skill (e.g. giving feedback or being assertive) and/or attitude (e.g. valuing my crews comments). According to these distinctions and the fact that learning can be theory- and experience-based, we *developed a classification*: Methods are differently categorized whether they enhance knowledge, skill and/or attitude and whether they are theory- or experience-based (see table 1). Lecture, lesson, case study, exercise and Line Oriented Flight Training (LOFT) are supposed to enhance knowledge. Skills should be developed through role-play, exercise and LOFT and these three methods plus case study and videos are supposed to alter attitude. Moreover, simulators are often used to facilitate technical/task related and team related competencies in order to reduce human failure and accidents [14].

TABLE I: CLASSIFICATION OF TRAINING METHODS

	<i>Knowledge</i>	<i>Skill</i>	<i>Attitude</i>
<i>Experience-based learning</i>	exercise, LOFT, simulation	role-play, exercise, LOFT, simulation	role-play, exercise, LOFT, simulation
<i>Theoretical-based learning</i>	lecture, lesson, case-study	behavioral role modeling	case-study, video

Using simulators supports error learning and developing shared mental models among team members with different tasks and duties. We suppose that simulation enhances knowledge, skill and attitude during experience-based learning. Behavioral role-modeling is also a method that leads to significant performance and behavior improvement in trained teams [15], but a match between the behavior model, the role play and the real work situation must exist. This method is supposed to alter skills during experience-based learning. If an indented goal of the CRM training would be just to know, for example, how to lead, the selection of the method would be a different one as compared to the goal that the trainee should be able to show a trained behavior (e.g. being assertive at work).

D. Training interventions

Within a training intervention, methods will be applied and, if suitable, combined. Two team training strategies that showed promising results in enhancing team performance are, for example, *Guided Team Self-Correction* or *Cross-Training*. The first strategy focuses on the leader, who helps the team in diagnosing and solving problems whereas the second one enables team members to use more efficient communication and coordination strategies and to built up shared mental models [16]. A third successful team training strategy is called *Team Coordination and Adaptation Training* [17]. Here, team members learn to improve team work during periods of high stress by anticipating and discussing potential challenges during low-workload periods.

VI. CONCLUDING REMARKS

A lot of research on CRM trainings has been done before [11,18,1], but results show several limitations: Training studies don't report evaluations on all necessary levels (reaction, learning/attitude, behavior, organization) and don't specify what was done in training interventions. Furthermore, a significant amount of research was conducted within the military aviation context. These factors hinder generalization and application of results for teams in other contexts outside (military) aviation. The goal of our project is to detect mechanisms responsible for the success of CRM trainings within civil aviation. By developing training based on a sound needs assessment and on insights from theory and research and by evaluating implemented training measures, we want to contribute to the research field and enable a transfer of the findings to other fields of application outside aviation.

REFERENCES

- [1] E. Salas, K. A. Wilson, C. S. Burke and D. C. Wightman, "Does crew resource management work? An update, an extension, and some critical needs," *Hum Factors*, vol. 48, pp. 392-412, 2006.
- [2] R. L. Helmreich and H. C. Foushee, "Why crew resource management? Empirical and theoretical bases of human factors training in aviation," in *Cockpit Resource Management*, E. L. Wiener, B. G. Kanki and R. L. Helmreich, Eds. San Diego: Academic Press, 1993, pp. 3-45.
- [3] I. L. Goldstein, *Training in Organizations*, 3rd ed. Pacific Grove: Brooks/Cole, 1993.
- [4] R. L. Helmreich, A. C. Merritt and J. A. Wilhelm, "The evolution of crew resource management training in commercial aviation," *Int J Aviat Psychol*, vol. 9, pp. 19-32, 1999.
- [5] Joint Aviation Authorities (JAA), "JAR-FCL (Aeroplane)," Colorado: Global Engineering Documents, 2006.
- [6] Joint Aviation Authorities (JAA), "JAR-OPS 1," Colorado: Global Engineering Documents, 2007.
- [7] European Aviation Safety Agency (EASA), "Regulation (EC) 2042/2003 Annex II-Part 145," http://www.easa.eu.int/doc/Regulation/reg_2042_2003_Part145.pdf, (2003)
- [8] R. Flin, L. Martin, K. M. Goeters, H. J. Hörman, R. Amalberti, C. Valot and H. Nijhuis, "Development of the NOTECHS (non-technical skills) system for assessing pilots' CRM skills," in *Contemporary issues in human factors and aviation safety*, D. Harris and H. C. Muir, Eds. Aldershot: Ashgate, 2005, pp. 133-154.
- [9] N. MacLeod, *Building safe systems in aviation: A CRM developer's handbook*. Aldershot: Ashgate, 2005.
- [10] C. Gonzales, J. F. Lerch and Ch. Lebiere, "Instance-based learning in dynamic decision making," *Cognitive Sci*, vol. 27, pp. 591-635, 2003.
- [11] R. J. Stout, E. Salas and J. E. Fowlkes, "Enhancing teamwork in complex environments through team training," *Group Dynam*, vol. 1, pp. 169-182, 1997.
- [12] C. R. Paris, E. Salas and J. A. Cannon-Bowers, "Teamwork in multi-person systems: a review and analysis," *Ergonomics*, vol. 43, pp. 1052-1075, 2000.
- [13] I. Goldstein and J.K. Ford, *Training in organizations*, 4th ed. Belmont: Wadsworth, 2002
- [14] K. Sonntag and R. Stegmaier, *Arbeitsorientiertes Lernen*. Stuttgart: Kohlhammer GmbH, 2007.
- [15] K. Sonntag and N. Schaper, "Förderung beruflicher Handlungskompetenz," in *Personalentwicklung in Organisationen*, K. Sonntag, Eds. Göttingen: Hogrefe Verlag, 2006, pp. 270-297
- [16] J. A. Cannon-Bowers and E. Salas, "Team performance and training in complex environments: recent findings from applied research," *Curr Dir Psychol Sci*, vol. 7, pp. 83-87, June 1998.
- [17] E. Salas, D. R. Nichols and J. E. Driskell, "Testing three team training strategies in intact teams: a meta analysis," *Small Gr Res*, vol. 38, pp. 471-488, Aug 2007.
- [18] B. B. Morgan, E. Salas and A. S. Glickman, "An analysis of team evolution and maturation," *J Gen Psychol*, vol. 120, pp. 277-291, July 1993.

Technology Assessment and Prioritization for Small and Medium Airports: A Methodological Approach

Olivia J. Pinon

Graduate Research Assistant

Aerospace Systems Design Laboratory

School of Aerospace Engineering

Georgia Institute of Technology

Atlanta, Georgia 30332-0150

Email: olivia.pinon@asdl.gatech.edu

Prof. Dimitri Mavris

Boeing Professor, Director

Aerospace Systems Design Laboratory

School of Aerospace Engineering

Georgia Institute of Technology

Atlanta, Georgia 30332-0150

Email: dimitri.mavris@ae.gatech.edu

Dr. Elena Garcia

Research Engineer II

Aerospace Systems Design Laboratory

School of Aerospace Engineering

Georgia Institute of Technology

Atlanta, Georgia 30332-0150

Email: egarcia@asdl.gatech.edu

Abstract—The air transportation industry is a significant source of employment for millions of people around the world. It is also an indispensable part of the economic infrastructure and as such, the gridlock experienced and forecast at large airports may have major negative impacts on the economy. This research aims to address the increase in demand and resulting capacity issues by considering the implementation of operational concepts and technologies at underutilized airports. The objectives of this work are primarily to off-load the busiest airports by increasing operations at smaller airports, reduce door-step to destination travel time, and provide transportation alternatives. More particularly, this work proposes a methodology to help in the assessment and prioritization of equipment packages and technologies necessary to enable that increase in operations. By associating multi-criteria technology selection techniques to ongoing small airport simulation effort, this work aims at helping airport managers make more informed decisions with regards to equipment offers in order to meet their future technological needs.

I. INTRODUCTION

Though very sensitive to rising fuel prices, and political and economical crises [1], the air transportation industry has not stopped growing over the last decades, both in terms of passengers and aircraft movements [2]. However the passenger traffic is far from being uniform and is mainly concentrated over a few airports, generally in metropolitan areas, meaning that most of the airport infrastructure is currently underutilized. In addition, since 1990 and more significantly after 2000, the major legacy carriers in the U.S. underwent major restructuring and gradual downsizing of their fleet, replacing large aircraft with smaller regional jets. The emergence of regional jets, along with the significant growth in low-cost carriers experienced during these years [3], resulted in the number of operations growing faster than the passenger traffic [4]. This increase in the number of operations is also expected to be reinforced within the next 10 to 15 years with the entry into the market of Very Light Jets (VLJs).

A. The Problem

The trends in passenger and aircraft movements is likely to continue within the next decades [5]. As a matter of fact,

the Federal Aviation Administration (FAA) is forecasting a 40 percent increase over today's passenger demand by 2010 [6], with a "45 percent increase in passengers between 2005 and 2017 being accomplished by a 33 percent increase in air carrier operations" [3]. This statistic would imply that the number of operations would grow slower than the number of passengers. This would then worsen the already existing disparity between demand and capacity and reinforce congestion levels at some major airports [4].

Tomorrow's air transportation system will be characterized by an increase in the types of airspace users (regional jets, very light jets, unmanned aerial vehicles, etc.) as well as very few new airports development projects. This leads to the realization that the forecast demand and resulting capacity needs will have to be addressed with innovative uses of the existing airport infrastructure ([4], [7]).

Airports have been identified as the major constraint to growth [8] and different strategies have been proposed to address the capacity issue. These strategies can mainly be divided into capacity increase strategies (addition of new runways, use of new/additional equipment, implementation of new operational concepts, etc.) and demand management strategies (peak period pricing, shifting flights from congested airports to less-busy secondary and regional airports, etc.) [9]. However, no strategy alone can solve the problem. Eurocontrol, for example, in "*The Challenges to Growth*" study published in December 2004 [10] found that even if they use every runway to its maximum capacity, "airports will still be unable to cope with the demand if traffic continues to increase in line with the higher estimates of future growth" [11]. Similarly, there exist divergent opinions with regards to the capability of new technologies and operational concepts to resolve the congestion issue.

II. THE NEED FOR A DIFFERENT APPROACH

Most of the research conducted in the past has only considered either demand management or capacity increase strategies. However, it has recently been acknowledged that the improvement of the air transportation system should come from the implementation of a combination of

solutions and strategies. Furthermore, each combination of solutions should be evaluated with economic and policy factors/impacts/analysis in mind [12], and not only from a technical perspective.

Secondary and underutilized airports have also been the focus of recent studies ([13], [4], [12], [14], [3], [2]). The development and increase of operations at smaller, underutilized airports now appear as a viable and key means to meet travel demand in congested metropolitan areas. The NGATS Report for example states that “it is essential to enable increased operations at smaller airports in the same region to offload some of the demand on the busiest airport(s) where practical (e.g. air taxi operations)” [14]. In the same report, it is acknowledged that “significant growth at the busiest airports as well as regional and smaller airports is needed to achieve the capacity goal of the NGATS” [14]. In its Report to Congress, the FAA also mentions that “redistribution of traffic among airports to make more efficient use of facilities is another measure that can be used to reduce delays” [3]. In that same report, the FAA stresses that “another factor that helps to limit delay is the ability of carriers to introduce service to outlying, suburban airports, using them to relieve congestion at the principal airport” [3]. The THENA Consortium also recognized that “new secondary airports that are adjacent to main population areas might constitute an additional air traffic channel (with even more rapid growth rates than the hub), especially for short haul, point-to-point routes” [2]. Further, some governments are also more interested in developing secondary airports as illustrated by the British government who refused to expand London Heathrow but gave the priority to the expansion of Stansted airport, the London metropolitan region secondary airport [15].

Finally, the growing interest for secondary airports also comes from the travelers themselves. More and more travelers are flying from alternate or secondary airports and are motivating their choice by citing reasonable driving time, competitive air fares and time savings [16]. As mentioned in a recent newspaper article, “from 1996 to 2002, the number of passengers departing from Manchester Airport almost quadrupled, to 1.85 million from 500,332. During the same period, passengers leaving from Logan declined by about 10 percent, to 11 million” [16]. This trend has also been observed at other airports such as Fort Lauderdale or Midway [16], confirming that this type of airports offers a viable option to air travel. The growing interest for secondary airports, particularly due to the presence of low cost carriers, also exists in Europe. Brussels South Charleroi airport, for example, saw its passenger traffic increasing from 200,000 travelers to more than 2 million annually in only four years, primarily due to the presence of two of the busiest low-cost airlines [17]. However, while secondary and regional airports may be part of the solution, they often lack the appropriate equipment and technology that would allow for an increase in their number of operations.

Some studies have been focussing on the impact of new technologies and operational concepts on both the National Airspace System (NAS) and airports. However, little work exists that considers both operational concepts implementation and spatial shifting of flights from busy and congested airports to close-by less used secondary and regional airports. Hence, very little work has been carried out that focuses on the technological and operational impacts of technologies on small and medium airports.

Furthermore, small or regional airports differ greatly from large airports, as their budgets, needs and constraints are different. As such, the benefits identified in previous studies to large airports may not be applicable to small and medium airports. Definition and selection of technology or equipment portfolios must be based on thorough benefits assessment as these decisions will require suitable investment strategies. However, the benefits considered shouldn't be limited to performance only, but should include cost and monetary benefits as well. This is particularly important when considering small and medium airports, as their budget and ability to finance equipment investments are more limited than for large airports. Finally, current work on the topic lacks a methodical approach for airport technology evaluation and selection. Smaller airports have different needs and constraints, and all existing or future technologies may not be suitable. Technology and equipment selection should thus be made based on airports' needs, constraints and requirements with regards to their current equipage and operations, but also with regards to the future type of aircraft mix that will be operating at these airports.

A. Approach

This research proposes to address the increase in demand and resulting capacity issues by considering the implementation of operational concepts and technologies at underutilized airports. The methodology developed offers a simulation and multi-criteria decision-making framework to assess and prioritize equipment packages and technologies, based on both performance and economic metrics. The operational concepts and related technologies/equipment considered in the scope of this work are mainly related to Communication, Navigation and Surveillance, both on board and on the ground.

This methodology is divided into three main steps:

- Identifying the gap: this step consists in identifying airport needs, constraints and requirements to reduce the gap between the forecasted demand and the airport capabilities
- Addressing the gap: this step consists in assessing the benefits and degradations of candidate technologies and equipment against both performance and financial factors. A candidate portfolio of technologies is then obtained through a Multi Criteria Decision Making framework.
- Closing the gap: this step consists in ensuring that the technology portfolio defined in the previous step provides the airport with the desired capabilities while remaining within the budget considered.

B. Research Goals and Objectives

The goal of this research is to provide a parametric, robust, and multi-criteria environment to help in the evaluation and prioritization of technologies and equipment packages for small and medium airports. Such an environment will enable "what-if" games and trade-off analyses to be conducted and will provide scenario-based solutions to the airport managers, hence allowing them to make more informed decisions with regards to equipment offers.

The objectives of this work are four-fold:

- Unburden the demand at the busiest airports by increasing operations at smaller airports
- Improve mobility by reducing door-step to destination travel time
- Provide transportation alternatives
- Gain a better understanding of the functional and emergent relationships between the different technologies and operational concepts at the airside level

C. Challenges

Airports are complicated and complex systems exhibiting many interacting, interrelated and interdependent components. As such, the challenges of this work are directly related to the characteristics of such systems. Challenges can be divided into research and technical challenges. Research challenges include:

- Identifying the appropriate airport measures of performance
- Identifying the appropriate technical and financial factors that the equipment packages/technology portfolios have to be evaluated against
- Identifying the different sources of uncertainties
- Obtaining a proper understanding of technology and equipment interdependencies and interactions

Technical challenges mainly concern the development of the Multi Criteria Decision Making environment and the lack of information available with respect to the different technologies and operational concepts. As a matter of fact, limited work has been conducted on the impact of existing equipment or technology on airport performance. Hence, this lack of data makes the equipment/technology evaluation and impact assessment difficult. Finally, the Multi Criteria Decision Making tool developed in the framework of this research should provide appropriate fidelity and robustness without requiring excessive computer time and resources.

D. Benefits

This research proposes to address some of the recommendations and research gaps mentioned in previous studies. Particularly, this work considers the need expressed by many to:

- Develop "models/tools, operational concepts and methodologies to assist in assessing airport operations efficiency, (...), exploring trade-offs, implications, and interdependencies between several airport performance metrics" [2]

- To account for emerging technologies [18]
- To incorporate both ground-based and airborne systems capabilities ([5], [18])
- And to develop "tools to monitor and quantify implications, measures and effectiveness of the new strategies and solutions proposed" [19]

This research is also relevant to the following NextGen goals and strategies [20]:

- Satisfy future growth in demand (3X current levels) and operational diversity
- Develop airport infrastructure to meet future demand: integrate airport, airspace and air traffic management design, development and deployment
- Develop cost-effective concepts, technologies, and procedures for providing comprehensive air traffic services at small airports

Finally, this work will provide scenario-based robust solutions to the issue of capacity and delay and deepen the understanding of the functional and emergent relationships between the different technologies and operational concepts at the airside level.

III. CONCLUSION

This research proposes to address the increase in demand and resulting capacity issues by focussing on the implementation of operational concepts and technologies at underutilized airports. By doing so, this work considers both demand management and capacity increase strategies. This work also addresses the lack of a structured approach for airport technology evaluation and selection by proposing a methodology for the assessment and prioritization of technology for small and medium airports, with technologies and operational concepts being evaluated with respect to both airports performance, and economic needs and constraints. The multi-criteria decision making environment proposed will provide airport managers with the ability to conduct tradeoff analyses, and make more informed decisions with regards to technology offers. Finally, this work is relevant to NextGen goals and strategies and proposes to address some of the recommendations and research gaps mentioned in previous studies.

ACKNOWLEDGMENT

This research is supported by THALES Air Systems.

REFERENCES

- [1] C. Gresnigt, "Future needs and trends from an airline point of view," in *Presentation of the 1st THENA Workshop: European Airports: Current Problems and Challenges*. Brussels, Belgium: IATA Europe, February 2002.
- [2] T. N. on Airport Activities (THENA), "Airport capacity and efficiency - final synthesis," Fifth Framework - Promoting Competitive and Sustainable Growth, Tech. Rep., 2003.
- [3] FAA, "Report to congress - national plan of integrated airport systems (npias) (2007 - 2011)," U.S. Department of Transportation - Federal Aviation Administration, Tech. Rep., 2006.
- [4] P. A. Bonnefoy and R. J. Hansman, "Emergence of secondary airports and dynamics of regional airport systems in the united states," MIT International Center for Air Transportation., Department of Aeronautics and Astronautics - Massachusetts Institute of Technology - Cambridge, MA 02139 USA, Tech. Rep. ICAT-2005-02, May 2005.

- [5] S. Consortium, "Air transport framework - the current situation," SESAR Consortium, Tech. Rep. DLM-0602-001, July 2006.
- [6] G. L. Donohue, "U.s. air transportation: An approaching perfect storm," *Aerospace America*, pp. 26-32, August 2006.
- [7] T. M. C. US Department of Transportation, Federal Aviation Administration and C. for Advanced Aviation System Development, "Capacity needs in the national airspace system - an analysis of airport and metropolitan area demand and operational capacity in the future," U.S Department of Transportation, Federal Aviation Administration, The MITRE Corporation and Center for Advanced Aviation System Development, Tech. Rep., June 2004.
- [8] S. Hasan, "System wide modeling for the jpdo," November 2006.
- [9] K. G. Zografos and M. A. Madas, "A critical assessment of airport demand management strategies in europe and u.s.: A comparative perspective," in *Transportation Research Board Annual Sunday Simulation Workshop*, Washington, D.C, January 2003.
- [10] Eurocontrol, "Piata - performance indicators analysis tool for airports," 2004.
- [11] Eurocontrol and A. Europe, "A vision for europe aviation," EURON-TROL and ACI EUROPE, Tech. Rep., 2005.
- [12] G. Hunter, K. Ramamoorthy, P. Cobb, A. Huang, M. Blake, and A. Klein, "Evaluation of future national airspace system architectures," in *AIAA Modeling and Simulation Technologies Conference and Exhibit.*, no. AIAA Paper 2005-6492, San Francisco, CA, August 2005.
- [13] P. A. Bonnevoy and R. J. Hansman, "Investigation of the potential impacts of the entry of very light jets in the national airspace system," MIT International Center for Air Transportation (ICAT), Massachusetts Institute of Technology, Cambridge, Massachusetts 02139 USA, Tech. Rep. Report 2006-02, 2006.
- [14] H. Swenson, R. Barhydt, and M. Landis, "Next generation air transportation system (ngats) air traffic management (atm) - airspace project," National Aeronautics and Space Administration, Tech. Rep., 2006.
- [15] P. Forsyth, "The impacts of emerging aviation trends on airport infrastructure," *Journal of Air Transport Management*, vol. 13, pp. 45-52, 2007.
- [16] B. Estabrook, "Smaller airports are growing in stature," *The New York Times*, December 2003.
- [17] W. Underhill, "Low-cost airlines redrawing europe's map," *Newsweek International*, March 2006.
- [18] Eurocontrol, "Air traffic management strategy for the years 2000+: Executive summary," Eurocontrol, Tech. Rep., 2003.
- [19] T. N. on Airport Activites (THENA), "Synthesis report on airport simulation modeling," Tech. Rep., September 2003.
- [20] JPDO, "Next generation air transportation system integrated plan," Joint Planning and Development Office, Tech. Rep., 2004.

Evaluating aeronautical regulations using rigorous specifications

Safeguarding against unintended consequences

Eduardo Rafael López Ruiz

PhD Student

Long-term Design and Systems Integration Department

ONERA

Toulouse, France

eduardo.lopez-ruiz@onera.fr

Abstract—The purpose of this thesis project is to present an innovative methodology (consisting of methods, tools and procedures) which seeks to improve the rulemaking processes currently used to develop aeronautical safety and security regulations. The two main contributions of this methodology are: its use of rigorous methods and tools to help improve the regulation's validation process and its capacity to help identify the impact of proposed amendments on enacting regulation (while helping mitigate regressions).

Keywords; *Rigorous modeling, Very Light Jet, aeronautical regulations, safety, security.*

I. INTRODUCTION

The chief objective of Civil Aviation Authorities (CAA) worldwide is to continuously guarantee the safety and security¹ of civil aviation. To ascertain this, they have implemented a set of complementing and hierarchical regulations at the international, national and local level.

These regulations impose standards and recommended practices specifically targeting the prevention² of either **accidental events** or **unlawful acts of interference** within a given domain. This "regulation enforcement" approach to safety and security imposes that the regulation's *innate quality* and its *homogenized* and *ubiquitous implementation* become effectual factors to the achievement of their objective.

In what concerns a regulation's innate quality, [1] identified that aviation **security regulations** have three esteemed traits steering their effectiveness, which are: *consistency*, *robustness* and *unambiguousness*. Presently, rulemaking procedures include a consultation and validation phase. In it, proposed drafts are analyzed and discussed until they are considered mature for adoption and publication. For this, special attention is placed in: verifying their compatibility with existing rules, attesting the exhaustiveness of their scope and limiting their

¹ **Safety** relates to the prevention and mitigation of accidental events, which can affect material or people while **security** is the prevention and mitigation of intentional acts, which aim to affect planes or people.

² Regulations seek to prevent accidental events and unlawful acts, whereas reactive/emergency procedures dictate actions that help mitigate their consequences.

equivocalness (given the inherent ambiguity of natural languages). Nonetheless, operational feedback has proven that the current process can benefit from improvements. This is of great significance since -analogously with safety-critical software- these properties ensure that the benchmark regulation being enforced by the CAAs is intrinsically effective.

Moreover, over the past few decades, technological and ideological changes have prompted amendments within the industry's established *Regulatory Framework* (see Section 4). The purpose of such amendments has been the continual assurance of safe, secure and efficient³ commercial operations under an enhanced state-of-affairs (through the exclusion, inclusion and/or evolution of affected regulations and/or procedures). The predicament is that, independently of their origin and dimension, regulatory and procedural amendments inevitably lead to the (unwilling) introduction of new errors [2] and the obsolescing of sanctioned workarounds. In other words, amendments lead the framework from an "error-cognizant" state, where (an indicative part of) its inherent errors have been identified (and possibly solved or circumvented), to an "error-incognizant" state.

Therefore, the aim of this thesis project is to present an innovative methodology (consisting of methods, tools and procedures) that will help improve the rulemaking processes currently used to develop aeronautical safety and security regulations.

To better recognize this situation, this thesis also seeks to appreciate the *Regulatory Framework* being implemented in aeronautics, by identifying how the prevailing ideas (principles, objectives and policies) within civil aviation are linked to the regulatory infrastructure (regulations and procedures) being implemented. This will help identify and bound the "impingement zone" of the regulatory amendments, opening the path for the subsequent development of the methodology that will help identify possible regressions arising from such type of amendments. Finally, the methodology that is proposed shall be appraised by implementing it in the study of the

³ A global efficiency attained through the prioritized optimization of various factor such as: environmental and economic costs, performance ...

discerning operational certification requirements for Very Light Jets (VLJs) in Europe and the United States of America.

II. THE CHOSEN APPROACH

In 2003, a group of French universities and research laboratories proposed the implementation of rigorous methods to assist in the specification, design and validation of regulation documents [1]. They named their project EDEMOI.

Rigorous methods had already been used within other domains of aeronautics, "to enhance the current practice of procedures development" [3] and for the analysis and verification of aeronautical safety critical systems. However, the formal methodology developed by EDEMOI (see Figure 1) sought to enhance the rulemaking process by incorporating simulation and counterexample checking tools into the validation phase, to better ensure the regulation's innate quality. This methodology is centered on a two-step approach involving two stakeholders: the Certification Authorities, which establish International Standards concerning Civil Aviation Security, and the Model Engineers, who translate these natural language documents into formal models that can be tested.

In the first step of this approach, a *model engineer* extracts the security goals imposed in the *International Standard* and translates them into a semiformal model that faithfully represents their structure and relations (while reducing the use of inherently ambiguous terms). This *graphical model*, comprehensible to both stakeholders, is later revised and validated by the *certification authority*, giving way to the second step of the 'EDEMOI approach' in which the model engineer performs a systematic translation of the semiformal model to produce a *formal model* that can be analyzed through *test scenarios*.

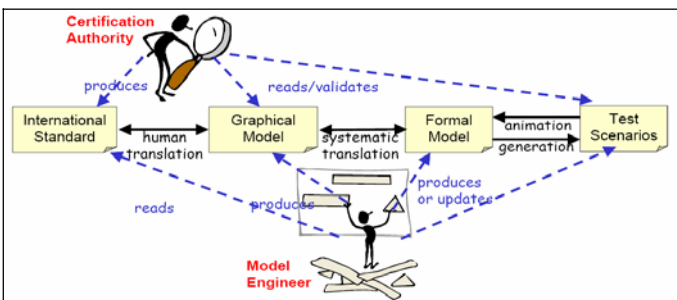


Figure 1. The EDEMOI methodology.

Having obtained positive results from these experiences, this thesis proposes a twofold expansion of the EDEMOI methodology by (1) broadening its scope to include aviation safety regulations and (2) by extending its usability throughout the regulation's "lifecycle".

More concisely, the extended methodology will devise methods and provide tools to help validate and enhance aviation regulations throughout their development, verification, validation and amending phases. This presents a large challenge as there are fundamental and operational differences between safety and security regulations. At the outset, safety regulations need to be more adaptive to the industry's constantly evolving state-of-affairs, helping steer developments

instead of contriving their progress. Therefore, will their consistency and robustness still be driving traits? And can their adaptability be inferred through the models?

For example, given the current technological and economical trends in the aeronautical sector, the industry is preparing itself to the challenge of successfully extending single crew operations to jet aircraft. More concisely, the Very Light Jets (VLJs). The stakeholders to this undertaking -such as the aircraft manufacturers, service providers and safety regulators (which in the case of Europe would encompass EUROCONTROL and EASA)- are concerned with determining the regulatory enhancements (exclusions, inclusions and/or evolutions) that will be required to ensure safe operations under this new state-of-affairs.

The extended use of this methodology, in this case, would be focused in helping identify the impact of regulatory enhancements by modeling and comparing the regulations and procedures, before and after an amendment is enacted (see Figure 2).

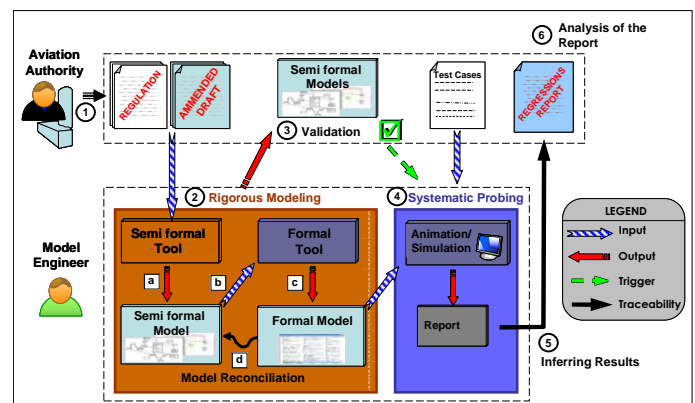


Figure 2. The extended methodology.

Therefore, analogously with the work done in [4], rigorous formal models of the affected aeronautical regulations will be studied and animated to: (1) help identify the impacts of proposed technologies on the regulation (influencing the aircraft's operation and flight-cabin design), and (2) to infer possible solutions for such incompatibilities.

It is important to say that this methodology does not aim to provide the consequences of the "inaction" with respect to the *adjusting factors*⁴, but rather to prevent the regression of the regulatory framework by assessing the impact that the amendments themselves have on the system (e.g. loss of consistency and/or robustness, introduction of ambiguous statements or ideas...).

III. RIGOROUS SPECIFICATION

As was done in our appraisal of Regulation 2320/2002 [4], an interpretation of the regulation is captured using (part of) the UML language (*Figure 2, Step 2.a*). The use of this graphical

⁴ An adjusting factor is any operational, ideological and/or technological change whose introduction, into the civil aviation system, obliges a change in the contemporary regulations to preserve the appropriate overall functioning of the system.

notation helps tackle the text's innate ambiguity, while proposing a conceptual layout that can be validated/invalidated by officials from the certification authority.

The validation of this conceptual layout (*Figure 2, Step 3*) helps establish the adhesion of the semiformal model to the convened international standard. For this, the appraisal and feedbacks provided by aviation authority officials are integrated into the model by way of amendments. Once validated, the semiformal model is translated (*Figure 2, Step 2.b*) to a rigorous formal model using translation rules between the semiformal and formal notations (in this case, UML → Z).

This ensures that our methodology benefits fully from the integration of both approaches: the intuitive structured notation of the semiformal approach and the precise semantics of the formal approach.

Finally, when the models have been deemed mature enough (both in their notation and their faithfulness to the regulation) an animation or verification tool (*Figure 2, Step 4*) is used to test the formal model's consistency (through simulation) and robustness (through counterexample checking). The results of the tests and simulation are stored to enable regression analysis after further evolutions of the regulation and the models (*Figure 2, Step 5*). Currently, two formal method targets are being considered: RoZ + Jaza Animator [5] and Alloy Analyzer [6]. However, independently of the software option that will be ultimately chosen, it is important to fully apprehend the context of these regulations and of the approach privileged by the CAAs.

IV. THE REGULATORY FRAMEWORK

An important aspect of this thesis work is to correctly understand the regulatory framework being implemented today in the civil aviation domain. Therefore, the first part of this thesis sought to propose a model of the idealistic regulatory framework which is in place today.

As stated before, CAAs worldwide strive to continuously guarantee the safety, security and efficiency of the civil aviation system. For this purpose they define *policies* and administer *regulations* which are in line with globally embraced *objectives* and *principles*. In addition to this, concerned aviation stakeholders develop *procedures* that dictate safe and practical methods for the successful performance of aeronautical-related tasks⁵.

Combined, these five elements make up what we will refer to as the *Regulatory Framework*. So, in broad terms, this framework is a mix of: (1) the underlying ideas/concepts (i.e. *principles*, *objectives* and *policies*) deemed essential for the development of civil aviation and (2) the regulatory infrastructure required for their implementation (i.e. *regulations* and *procedures*).

As shown in Figure 3, the Regulatory Framework is rooted from the *Principles* defined by the various concerned States. However, it is the task of the ICAO to translate these political goals into attainable *Objectives* and to define its *Policies*. The

different CAAs then align their policies with those of the ICAO and define new policies for the domains outside of ICAO's competence.

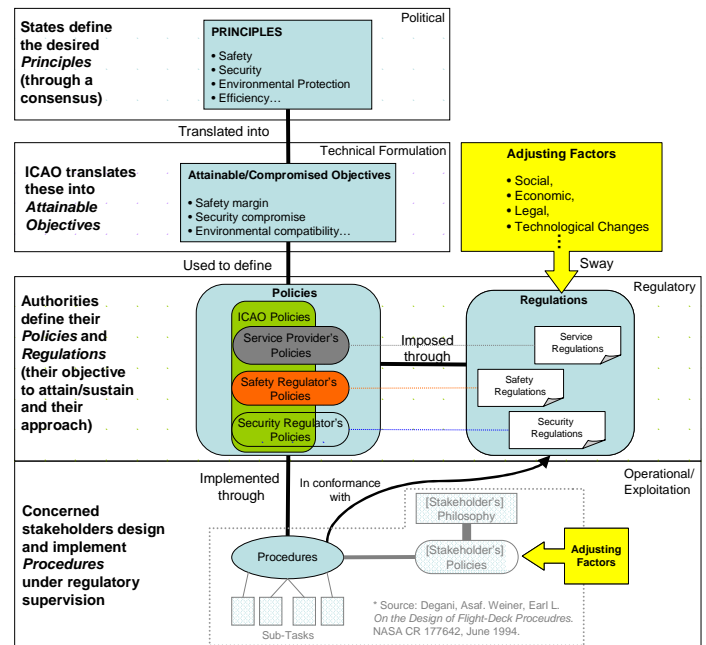


Figure 3. The Regulatory Framework.

The Policies defined by the CAAs are then imposed through local *Regulations*. These regulations must be consistent with the standards stated in the ICAO Annexes. This has led to a comprehensive base of compatible regulations. Concerned stakeholders then design and implement *Procedures* in conformance with these regulations.

However, these five elements cannot be considered as simple aggregates of the Regulatory Framework. As discussed in [7], the link between these elements is important and can be exploited to facilitate the detection of possible discrepancies and conflicts. Furthermore, a clear and structured representation of these links will provide: (1) an understanding of the Regulatory Framework and (2) insight on its interactions with the different adjusting factors (which impose a change in its state-of-affairs).

It should be noted that Figure 3, reflects the hierarchy of the different civil aviation organizations/agencies and their respective documents/rules by highlighting that CAA's derive their regulations from ICAO policies and adopt its standards. Furthermore, it identifies the regulations and the procedures, as the "impingement zone" for the *adjusting factors*. Because, given the industry's need for a flexible Regulatory Framework, its underlying principles, objectives and policies use abstract wording to convey what is deemed essential for the development of civil aviation. This makes them less susceptible to sways (owing to adjusting factors) than the regulations and procedures, which require detailed wording given the special need for unambiguity at their level.

⁵ Under the guidance and supervision of the Civil Aviation Authorities.

V. APPRAISING THE METHODOLOGY

The methodology proposed will be appraised by implementing it on the study of the discerning operational certification requirements for Very Light Jets (VLJs) in Europe and the United States of America. These small but relatively high-performance airplanes are a potential adjusting factor to a large part of the regulatory infrastructure presently established within civil aviation. This is due, in part, to the considerable contrast between their small size/weight (seating between 5-8 passengers with an MTOW under 4,536 kg) and their relative high performance (Cruise speed: ~ 0.62 M). But more particularly, by the fact that they were designed to fly within the same flight band (and terminal airspace) as that of commercial-aviation airplanes (FL 330-350) with an "in-design" compatibility for both single and double flight-crew operations.

Concerned with this situation, both EUROCONTROL and the European Aviation Safety Agency (EASA) have taken steps to ensure the smooth entry of this new "technology" while seeking to alleviate the ripple effects that it will have on the civil aviation system. For example:

Under current mandates, VLJs are not required to be equipped with an Airborne Collision Avoidance System (ACASII) to operate within the EUR region.

But, given their forecasted growth and their incompatible speed (with respect to large commercial airplanes), EUROCONTROL may seek to impose the mandatory equipping of VLJs with an ACASII system; to continue ensuring a high level of safety and efficiency in the pan-European Air Traffic Management (ATM) system.

EASA, on its part, has opted to limit the airplanes' operational envelope by restricting it to double-crew operations. This decision was based, in part, on the increased likelihood of: level busts, airspace incursions, runway incursions and fatigue in single pilot operations [8]. This is in clear contrast with the Federal Aviation Administration's (FAA) decision of certifying single flight-crew operations under a special scheme⁶.

These concerns not only demonstrate some of the regulatory enhancements that will ensue the VLJ concept, they also hint the (possible) need for a larger and more comprehensive regulatory enhancement; namely the evolution of the regulations' applicability criteria. A shift from the current criteria is required; the aircraft's weight and passenger seating capacity can no longer be regarded as the main parameters for determining its regulatory requirements. New criteria must be adopted, to effectively highlight that it is the aircraft's operating environment and its performance which are determinant.

VI. CONCLUSIONS AND FUTURE WORK

Having already applied the methodology to the modeling of Regulation 2320/2002, we have tested the compatibility between the security guidelines imposed and the different notations used to represent them. Consequently we shall

implement the extended methodology to the case of VLJ aircraft, in order to illustrate the main ideas and contributions of both the model and the methodology.

By our part we believe that the introduction of VLJs into the civil aviation system represents an excellent opportunity to appraise this methodology. For this, we shall benefit from the discerning operational certification requirements for VLJs in Europe and the United States of America. Since VLJs in Europe are not going to be certifiable for single crew operations (in contrast with the FAA's decision for Part 135 operations), this gives us a " Δ " (delta) between two state-of-affairs which can be used as a "before" and "after" state to compare (and tweak) the performance of our proposed tools. The systematic probing would focus on the regressions introduced by the amendments implemented in the USA, with regards to its VLJ stance in a bid to facilitate the detection of unintended consequences.

However, the work will not be a clear-cut translation of the requirements (into a graphical and formal model). There is a complexity in specifying all of their aspects; with a potential loss of connotation during the conversion. This problematic is inherent to the passage from a natural language to the semi-formal and formal notation. Nevertheless, the converse is also true; the translation to formal notation helps enrich the requirements by imposing precision in terms and relations.

Additionally, work is being pursued to determine the possibility (and the interest) of extending this same methodology onto other aspects of civil aviation, such as flight procedures and manuals.

REFERENCES

- [1] R. Laleau, et al. "Adopting a situational requirements engineering approach for the analysis of civil aviation security standards," *The Journal of Software Process: Improvement and Practice (SPIP)*, Vol. 11, Issue 5, Pages 487-503. July 2006.
- [2] S. Deutsch and R.W. Pew, "Single pilot commercial aircraft operation," BBN Report No. 8436, Cambridge, USA, November 2005.
- [3] A. Degani, M. Heymann, and I. Barshi, "A formal methodology, tools and algorithm for the analysis, verification and design of emergency procedures and recovery sequences," NASA Internal white paper, 2005.
- [4] E.R. López Ruiz, "Formal specification of security regulations: The modeling of European civil aviation security," Master Thesis, SUPAERO, Toulouse, France, December, 2006.
- [5] Y. Ledru, "Using Jaza to animate RoZ specifications of UML class diagrams," IEEE Columbia, April, 2006, [The 16th International Z User Meeting, ZUM 2006].
- [6] D. Jackson, "Dependable software by design," *In Scientific American*, June, 2006. Volume 294 Number 6. Page 68.
- [7] A. Degani, and E.L. Weiner, "On the design of flight-deck procedures," NASA CR 177642. USA, June 1994.
- [8] F. Woods, "Very Light Jets. The qualification challenge," EASA, Brussels, Belgium, May 2007, [Very Light Jet (VLJ) Workshop].

⁶ Limited to Part 135 operations. Requires an experienced professional-pilot licence holder that has undergone special training.

Author Index

Abernethy, J.....	429	Hackney, B.....	39
Aloysius, S. S.....	413	Hagemann, V.....	497
Amaldi, P.....	281, 295, 481	Halbherr, T.....	317
Angulo, G. C.....	159	Hansen, M.....	51, 59, 127, 403
Atkins, D.....	437	Hardmeier, D.....	311
Badanik, B.....	389	Hay, L.....	317
Balakrishna, P.....	355	Helman, S.....	317
Ball, M.....	13, 95	Henderson, J.....	117
Belyavin, A.....	317	Hoang, T. T.....	485
Benitez, M.....	159	Hurter, C.....	263
Bil, C.....	477	in't Veld, A.....	151, 235
Blom, H. A. P.....	177	Janic, M.....	169
Boccalatte, A.....	295	Joyekurun, R. T.....	281
Bolfing, A.....	317	Kageyama, K.....	69
Bolic, T.....	215	Kapp, V.....	263
Boussouf, L.....	103	Kim, A.....	59
Bradley, E.....	429	Klopfenstein, O.....	111
Bui, M.....	271	Klophaus, R.....	459
Calderón-Meza, G.....	3	Kluge, A.....	497
Castelli, L.....	397	Kotegawa, T.....	87
Causse, M.....	303	Kumar, V.....	445
Chaloulos, G.....	137	Loomes, M.....	295
Chen, X.....	127	Lorenz, B.....	191
Churchill, A. M.....	21	Lovell, D.....	21, 95
Colageo, M.....	207	Lygeros, J.....	137
Conversy, S.....	263	Manley, B.....	29
de Jong, H. H.....	177	Martin, P.....	295
De Laurentis, D. A.....	87	Mavris, D.....	501
de Leege, A.....	151	Medina, M.....	287
Deau, R.....	345	Meert, M.....	363
Dehais, F.....	303	Mijatovic, D.....	363
Delahaye, D.....	103	Mulder, M.....	151, 235
Di Francesco, A.....	207	Nace, D.....	111
Djokic, J.....	191	Nagaoka, S.....	69
Donohue, G.....	3, 39, 185, 467	Nazeri, Z.....	185
Durand, N.....	345	Nejjari, F.....	421
Elferink, N. H.....	253	Netjasov, F.....	169
Fakih, H.....	389	Nguyen, A.....	95
Fields, B.....	295, 481	Onghena, E.....	489
Fricke, H.....	191, 199, 335, 371	Oseguera-Lohr, R.....	467
Fujita, M.....	143	Pagliari, R.....	271
Fuller, I.....	451	Pastor, J.....	303
Ganesan, R.....	355	Pearce, D.....	451
Ganji, M.....	95	Pinon, O.....	501
Garcia, E.....	501	Prats, X.....	421
Gaukrodger, S.....	295	Puechmorel, S.....	103
Ghijs, S. S.....	245, 253	Puig, V.....	421
Giannazza, D.....	77	Quevedo, J.....	421
Goman, A.....	437	Rakas, J.....	215, 403
Gotteland, J-B.....	345	Ranieri, A.....	397
Groenouwe, R.....	235	Ritzmann, S.....	497
Groppe, M.....	271	Rozzi, S.....	295, 481
Gupta, G.....	403	Ruiz, L.....	505
Gwiggner, C.....	69	Schaar, D.....	327

Scholte, J. J.	177
Schultz, M.	335
Schwaninger, A.	311, 317
Sengers, P. A.	245
Sengstacken, A.	87
Sharman, R.	429
Sherry, L.	3, 29, 39, 185, 227, 287, 327, 355, 445
Shultz, M.	371
Shulz, C.	371
Smirti, M.	127
Smith, D. A.	227
Srisaeng, P.	477
Stefanik, M.	389
Stroeve, S. H.	177
Subramanian, B.	13
Tein, M.	477
Thiel, C.	199
Thompson, T.	445
Tien, S-L.	13
Trani, A.	117, 379
Uang, G.	215
van Paassen, R.	151, 235
Vlachou, K.	21
Wang, D.	39
Wang, L.	467
Wong, W.	281, 295
Wrobel, L.	413, 451
Xu, Y.	379
Yoon, Y.	51