Resilient Arrival Runway Occupancy Time prediction for decision-making tool in Barcelona (LEBL) airport.

Miguel García Martínez, Javier García Moreno & Rubén González Sendino
CRIDA
Madrid, Spain
mgmartinez@e-crida.enaire.es
jgamoreno@e-crida.enaire.es
rgsendino@e-crida.enaire.es

Álvaro Rodríguez Sanz
Technical University of Madrid (UPM)
Madrid, Spain
alvaro.rodriguez.sanz@upm.es

Abstract— When trying to maximise the use of the airport airside, it becomes a key factor for tower Air Traffic Controllers (ATCos) to optimise the runway occupancy time for landing aircraft (AROT) while maintaining the required safety levels and/or reducing unsafe events such as missed approaches and runway incursions. Accurate tools to detect and predict unsafe events are becoming necessary in order to assist ATCos in their tasks. In aviation, there are many different data sources and the paradigm of data sharing between different stakeholders is not fully implemented. This paper proposes the use of novel machine learning techniques to model AROT prediction in Barcelona airport (LEBL), using ten models trained with more than 270,000 flights, exploiting different airport and Air Navigation Service Provider (ANSP) data sources to design a resilient decision-making tool that assists tower ATCos in real operation environment to optimize runway occupancy while increasing the safety level by decreasing human error.

Keywords-component; machine learning; AROT; neural network; prediction; airport; capacity; safety.

I. INTRODUCTION

Airport capacity is considered as one of the main bottlenecks in the air traffic network [1]. Air traffic demand is growing rapidly, and at the same time, airport throughput is limited by the existing infrastructure and operational capabilities [2]. Using the available airport capacity effectively will help to solve imbalances and to achieve a sustained increase in throughput performance, therefore airports need to be recognized as being part of the whole Air Traffic Management (ATM) system in a “gate-to-gate” environment [3].

EUROCONTROL defines demand as the airlines’ schedule of flight plan submitted on the day of operation, the throughput is the airport actual air traffic movement on the day of operation (arrivals and landings) and the capacity is defined as the theoretical air traffic movement capability of the airport [3]. ICAO defines capacity as the number of movements per unit of time that can be accepted during different meteorological conditions [4].

The AECFA (the Spanish agency for airport slot allocation and monitoring) [5], which is in charge of the coordination of the time slots for all flights in the Spanish airports, specifies the airport runway capacity indicators. For LEBL airport, the runway capacity is defined for different periods of time of the day, e.g. 38 arrivals and 40 departures during the busiest period (05:00 to 20:59 Local time) for summer season 2019. To release latent capacity of an airport there is a range of options depending on the nature of the inefficiency and in the budget constraints.

The minimum separation prescribed by ICAO Doc 4444 is 5 NM (unless otherwise stated by the appropriate ATS authority) [6]. In LEBL, separation between aircraft in the landing sequence is based on a minimum separation distance of 3.0 NM, as the surveillance systems’ capabilities permit this. Many studies show that reducing the separation minima between aircraft in the landing sequence could lead to a potential increase on the runway capacity performance [7]. For instance, Time Base Separation (TBS) approaches, commonly based in RECAT-EU sequencing (that takes into account the aircraft performances and their wake vortex category) seek to enhance throughput [8]. These methods rely on better Runway Occupancy Time (ROT) predictions to maintain or increase the safety levels (e.g. avoid missed approach procedures or runway incursion) [9]. ROT is a key driver of airport runway throughput, especially when low airborne separation minima are applied [10]. Therefore, it is important to provide ATCos and operators with the appropriate tools for ROT estimation.

Furthermore, the air transport industry is evolving from concentrating the effort on reactive safety initiatives to a

A. State of the Art

The evolution of technology as well as the improvement in the field of Machine Learning (ML) and Artificial Intelligence (AI) such as Neural Networks (NN), are currently allowing the easy generation of intelligent models that are able to create predictions with past data in the field of ATM [12][13][14]. These models can be used to predict forthcoming operational risks during real time operation environment, as for example a probabilistic forecast of the Runway Occupancy Time in real time [15].

However, to be used in a real operational environment the intended system must be resilient. The findings of past studies can be complemented with models that allow real-time operational tools to optimise runway capacity. Why resilient? Due to the importance on critical operations, the system needs to consider different threaten situations to remain active, the solution provided needs to be accurate enough to be useful and be available when requested by the user [16].

In ATM, there are many data source generators, which are not perfect (e.g. contain incomplete data or not accurate enough) to be used in real operation environments. Therefore, there is a need of high accuracy to be operationally exploitable as well as a high availability of the information (have the information always and as soon as possible). Many times, this lack of data can be attributed to the complex nature of the ATM system as a whole, where hundreds of stakeholders are involved and where data-sharing paradigms are considered rare [17]. There are thousands of data sources but there is a simpler hierarchical categorization with a top-level approach, composed by Flight plan data, Surveillance data and Meteorological that gathers most of the available information sources which are critical for ATM operational problems resolution and prediction.

B. Aim

The main goal is to generate a resilient system able to predict ROT for Arrival aircraft (AROT), capable to work in real operation environment in real time as well as to validate the models using as case study Barcelona Airport (LEBL) with historical data, being this the first step to draft a decision-making system for tower ATCos. Based on past analyses [10][15], the AROT prediction methodology should move from a statistical approach to ML techniques for coping with the variability of AROT behaviours.

First, for an ideal scenario, an iterative and progressive model, where the last prediction (i.e. at 4NM from threshold) will include all the data sources, is proposed. The system will be able to work with different data sources and to adapt itself to work only with the available resources.

Finally, a flexible system with different temporal horizons as prediction goal will be developed. This system will update its predictions as soon as a new source is available or just after an update on the data sources (i.e.: the flight plan, the airport information and the weather reports). This system is summarized in Fig. 1. Some of these temporal horizons could change dynamically depending on the source and different flights could have different number of predictions, for example if a new weather report appears, the result of the prediction must be calculated again. Nevertheless, at least every flight processed through the system will have 4 predictions, usually with this chronology:

- When the flight plan is created.
- When the stand information is created.
- When the last weather report is created (a report that has around 30 minutes of validity).
- When the flight reaches the speed limit approach point (for LEBL is located at 4NM to the runway threshold (THL), i.e. around 2 minutes for landing).

Additionally, with the use of ML techniques all predictions will be presented with a confidence value, taking into account the final user of a real scenario (the ATCo).

When each source is available, the data necessary for the different prediction horizons will be accessible in a timely manner, so the inputs for a real-time support tool can be provided.

C. Structure

The paper is structured as follows. In Section II we describe the data sources as well as the use case where the prototype is intended to be implemented, including a descriptive analysis. In Section III the methodology is widely explained, describing the data preparation activities, the AROT feature selection stage and models construction. In section IV all the models are
outlined, showing their main results. Finally, conclusions and future work are described in section V.

II. DATA SOURCES AND USE CASE

First, a top-level data requirement list is proposed by considering the different parts of the ATM system. Therefore, to feed the system, we have appraised the most common data sources used in airport and air navigation environment. i.e.: Surveillance data, Flight Plan data, Meteorological data and Airport data. These generic items will be detailed on section III with concrete data sources for each one of the top-level categories.

When developing a predictive model, it is important to tailor the algorithms to the selected use case, in this case LEBL airport. The airport has three runways: i.e. six operational directions, five of which are used for landing. In terms of air traffic operations, LEBL is the second busiest airport in Spain [18] and the 27th in the World, according to ACI Annual Traffic Report for 2018 [19].

From January 2017 to August 2018 (included), 289,631 arrivals were registered, the arrival distribution per runway was: 25R (66.36%); 25L (4.02%); 07R (0.14%); 07L (14.77%); 02 (14.71%).

A. AROT (Arrival Runway Occupancy Time) and RE (Rapid Exit taxiway) understanding

The Arrival Runway Occupancy Time, (AROT) is defined as the elapsed time in seconds that an aircraft is occupying the runway during the landing phase, starting when the aircraft is crossing the runway threshold (THL) until its tail vacates the runway [20].

To compute the amount of time that certain aircraft is occupying the runway, it is needed to define those temporal milestones, first when the aircraft crosses the runway THL and second when the tail vacates the runway. For this purpose, we have used the aircraft position in the airport environment, obtained from airport’s Multilateration (MLAT) system, which accuracy is of ± 5 meters and its precision is of 1 second. With this accuracy it is not possible to detect where the tail of the aircraft is, therefore, in this paper we have assumed that a single position point of the MLAT system is considered as the whole aircraft.

To select the time instant when the aircraft crosses the runway THL, similar to what was done in [10], we have developed an algorithm capable of detecting the first point from the surface radar of an aircraft that crosses the runway THL. This is done by defining the runway as a polygon and the THL as a line intersecting the polygon, this intersection defines de start time to measure the AROT as shown in Fig. 2.

In the same way, to compute when the aircraft vacates the runway, we select the first point of the trajectory of MLAT that is outside the polygon of the runway. In addition, we have extracted from the Spanish AIP (Aeronautical Information Publication) for LEBL [21], the polygons for the possible Rapid and non-rapid Exit (RE) taxiways for every landing runway.

After computing the AROTs and REs for every landing aircraft, we can plot their historic distribution. Fig. 3, shows, with a heat map, the average AROT distribution in seconds for all aircrafts landing and their final airport stand position, it also shows that in some areas of the airport, Terminal 2, the average AROT tends to be higher, more than 70 seconds, when compared with the Terminal 1, around 45 seconds.

III. METHODOLOGY

A. Data Sources and Acquisition

During the data preparation phase, the first task is to look for data fulfilling the top-level data hierarchy, the 6 selected data sources are listed here from (1) to (6):

- **Surveillance data**: this mainly includes the aircraft 4D position (latitude, longitude, altitude/flight level and time) from different radar systems, two different systems are used, radar tracks from Secondary Surveillance Radar (SSR) (1) and MLAT systems (2) all provided by ENAIRE, the Spanish ANSP.
• **Flight Plan data:** Aircraft description and planned route detail, obtained from the Spanish Integrated Flight Plan System (IPV) (3) provided by ENAIRE.

• **Meteorological data:** Airport weather based on public METAR (MÉTéorologique Aviation Régulière) reports (4) provided by the Iowa State University repository [22].

• **Airport data:** Airport information (e.g. stand and terminal assigned for each aircraft) obtained from LEBL airport resources allocation and monitoring system (SCENA) (5) and the Spanish AIP (6), all provided by ENAIRE.

Under the CRIDA-ENAIRE collaboration umbrella, CRIDA has created one of the biggest data sources integration in the ATM world. All input data sources used in this paper come from a data warehouse (DWH) as a queryable structured database with an underlying dimensional data model formed by dimensions (entities) and facts (measures), hosted on a Microsoft SQL Server. The output data source used to generate the models (information containing the ROT by flight metric) has been computed from MLAT radar tracks with frequency of 1 second (raw JSON files in ASTERIX category 19 format [23]). Then, for the purpose of this paper, a mixed-approach data solution is achieved with a dimensional data model (hosted on a Microsoft SQL Server) and a document data model (hosted on a MongoDB NoSQL Server). Both models are merged on demand by flight information (using callsign and landing time).

As a result, we obtain firstly, SSR radar information stored in the data warehouse covering the position of all aircrafts flying within the Spanish airspace with a frequency of 5 seconds (218,491,746 points), and second, MLAT radar information covering only the position during the approach and landing phases for each flight with a frequency of 1 second (534,221,737 points).

**B. Data Engineering**

Since ML algorithms only work with numerical variables, one of the biggest challenges is to transform categorical features to numerical ones in an effective way, especially on the Flight Plan case where 7 of the 8 features have literals values. Some ML libraries manage automatically these transformations, but usually they produce inefficient solutions as these require expert knowledge about the current data you are translating to and that you cannot infer from the data itself. One Hot Encoding and Ordinal Coding are the techniques combined by expert criteria for this paper [24]. These techniques transform variables from a range of literal values to vectors of binary values. Explicit features transformations are described in Table 1.

Another challenge is to transform radar information describing a time series into a non-time series prediction. To characterize the aircraft approach track as a controlled number of features to maintain the complexity of ML model at a desirable level (i.e. less than 50 dimensions for radar, less than 100 for the full model), a polynomic function of maximum degree of 10 is the chosen candidate for each of the representative dimensions; i.e. Vertical evolution (speed z) and Horizontal evolution (speed x, speed y). The methodology to choose the final function consist on computing the different functions from 3 to 10 degrees and then test them with R-squared metric [25].

Finally, a 4th degree polynomic function was chosen after checking R^2 values with a threshold of 0.98, for instance, Fig. 4 shows an example of a flight with R^2 = 0.988 for the magnitude of the velocity function (vel_mod) and R^2 = 0.994 for the flight level (FL). Additionally, to enrich the characterization of the descend phase, summary statistics such as minima, average and maxima values for selected dimensions are also included into the model to ease the anomaly detection and deviations caused by human factors as the pilot skills.

Date and time features are included to give the chance to the ML model to better understand the patterns of the airport operation affecting the AROT. On the SCENA source, the stand is translated to the related ramp. The ramp is a physical aggregation of a group of stands more representative for the operation.

**C. Cleansing process**

To obtain the required quality and accuracy in the results, it is important to detect abnormal values and filter them from training and modelling algorithms. After analysing the available datasets, minor errors at data source level were detected and filtered. An example of abnormal values are aircraft with zero engines equipped or flights with a mismatching on the landing runway.

![Figure 4 polynomial function to describe aircraft landing trajectory.](image-url)
**D. Feature Selection**

Different results are obtained, Table 2 shows top-15 features ordered after Pearson and Mutual Information algorithms by coefficients (importance). Being Mutual Information more accurate, as explain in [26], for this paper’s purpose good results are obtained with both algorithms. At this stage, the goal is to clearly identify useless features, not establishing a perfect classification from the more to less important.

All available features from each data source are included, except for those that do not provide useful information, as lack of data (e.g. source error or unavailability for LEBL), redundant data (e.g. dates and datetimes) or data is out of the study’s scope (e.g. the actual in block time). After the feature selection by data source, some of their dimensions are reduced or maintained. For example, in the case of SCENA source, only one feature is selected, i.e. the ramp. METAR has promising features regarding AROT, e.g. runway state, unfortunately for the dataset used in this paper, their features suffer from lack of data available, and only the cloud state information related to the approach phase is included, filtered by altitude.

**Table 2 Feature Selection, Pearson and Mutual Information Methods**

<table>
<thead>
<tr>
<th>Feature Selection</th>
<th>Pearson</th>
<th>Top Features</th>
<th>Mutual Info.</th>
</tr>
</thead>
<tbody>
<tr>
<td>RWY02 (FP)</td>
<td>0.50666</td>
<td>R [15]</td>
<td>0.13114</td>
</tr>
<tr>
<td>vel_y (R)</td>
<td>0.39192</td>
<td>R [10]</td>
<td>0.10817</td>
</tr>
<tr>
<td>avg_vel_y (R)</td>
<td>0.37452</td>
<td>min_vel_y (R)</td>
<td>0.10281</td>
</tr>
<tr>
<td>max_vel_y (R)</td>
<td>0.34868</td>
<td>min_vel_y (R)</td>
<td>0.09956</td>
</tr>
<tr>
<td>Heading (R)</td>
<td>0.34038</td>
<td>min_vel (R)</td>
<td>0.09835</td>
</tr>
<tr>
<td>min_vel (R)</td>
<td>0.33993</td>
<td>vel_x (R)</td>
<td>0.08955</td>
</tr>
<tr>
<td>Temperature (M)</td>
<td>0.21827</td>
<td>max_vel (R)</td>
<td>0.08358</td>
</tr>
<tr>
<td>RWY25R (FP)</td>
<td>0.21812</td>
<td>avg_vel (R)</td>
<td>0.08169</td>
</tr>
<tr>
<td>max_vel (R)</td>
<td>0.21433</td>
<td>max_vel (R)</td>
<td>0.07413</td>
</tr>
<tr>
<td>vel_x (R)</td>
<td>0.19340</td>
<td>min_vel (R)</td>
<td>0.06757</td>
</tr>
<tr>
<td>max_vel (R)</td>
<td>0.19318</td>
<td>hour_of_day (FP)</td>
<td>0.03809</td>
</tr>
<tr>
<td>avg_vel (R)</td>
<td>0.17495</td>
<td>RWY25R (FP)</td>
<td>0.03147</td>
</tr>
<tr>
<td>kickoff (M)</td>
<td>0.15051</td>
<td>Feel (M)</td>
<td>0.02956</td>
</tr>
<tr>
<td>min_vel (R)</td>
<td>0.13824</td>
<td>Temperature (M)</td>
<td>0.02879</td>
</tr>
<tr>
<td>wake_cat (FP)</td>
<td>0.13626</td>
<td>company_cat (FP)</td>
<td>0.02817</td>
</tr>
</tbody>
</table>

**E. Modelling and Machine Learning (ML) Techniques**

Neural network (NN) is the type of ML algorithm selected for this paper. NN have been widely proven effective on other disciplines, especially on medicine, since decades ago [27]. ATM science is starting to discover their benefits, Neural Networks provide advantages when implicitly detecting complex nonlinear relationships between dependent and independent variables. This paper tries to take advantage and quantify these benefits compared to other techniques aimed to ATM problems as done in [28][29].
Best results on ML problems came from the merge of human hints based on expertise and on what data suggests. None of the two sides should be ignored. For this paper, a team of ATM experts established a starting point in feature selection with filters over data and dimension reduction or expansion to better characterise the context regarding the approach and landing phases. Later, all human hints are tested and validated with deep data analysis using correlation functions and other machine learning algorithms. Additionally, to contextualize the ML models results, a simple mean model will be proposed as reference to assess the ML models.

The goal is then 10 different models listed on Table 3, 9 formerly named by the data sources they include and lastly a simple mean model. For the scenario where all sources are available, there are 2 different models proposed. The difference between these models is the RADAR part, where 2 approaches will be proposed: the simplest one (information only about the status at the 4NM point) and the most complex one, with the suffix “FULL” (information about the evolution of all the approach phase). The mean model is created for 5 dimensions i.e.: wake turbulence, aircraft type, runway, company and the hour of the estimated landing; and calculates the mean AROT for each dimensional combination.

For this paper, Java based library, Neuroph, and the framework Neuroph Studio, were used to create the Neural Networks [30]. The training process has been achieved by batches of 30,000 and 80,000 iterations depending on the time of converging to a solution, with different values for learning rate (0.2, 0.1, and 0.05).

The dataset included is divided in 3 parts: training, testing and validation, i.e. first the dataset is split into 2 pieces, from January 1st, 2017 to March 31st 2018 (subsequently separated in train and test dataset) and the second dataset from April 1st 2018 to August 31st 2018 (as a validation dataset).

<table>
<thead>
<tr>
<th>Number of sources included</th>
<th>Model name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MEAN; FP</td>
</tr>
<tr>
<td>2</td>
<td>FP_METER; FP_RADAR; FP_SCENA</td>
</tr>
<tr>
<td>3</td>
<td>FP_METER_SCENA; FP_METER_RADAR; FP_SCENA_RADAR</td>
</tr>
<tr>
<td>4</td>
<td>FP_METER_SCENA_RADA;</td>
</tr>
<tr>
<td></td>
<td>FP_METER_SCENA_RADAR_FULL</td>
</tr>
</tbody>
</table>

IV. RESULTS
To evaluate each model, we have measured the error, defined as the difference between the predicted AROT and the real AOTP. Then, trying to define the tolerance of the system based on the error, we have established the absolute accuracy of the models for different thresholds; i.e.: the accuracy for 7 seconds is the percentage of flights having a difference between the predicted and the real AROT ≤ 7 seconds. In addition, we have established the percentage accuracy of the model, i.e.: the accuracy for 20% is the percentage of flights having a difference between the predicted and the real AROT ≤ 20% of the real AROT value. The way of calculating accuracy is important for this problem, as an error of 4 seconds in a 40 seconds AROT sample is different than the same error (4 seconds) in a 60 seconds AROT sample.

The absolute accuracy is shown in Fig. 5 and Fig. 6 for all models, where the y-axis is the accuracy of the model (absolute and percentage accuracy respectively) and the x-axis is the tolerance. Both results indicate that the best model is the one that predicts the AROT at the closest moment to landing and uses all data sources, FP_METER_SCENA_RADAR_FULL.

The results obtained for this model were:
- Absolute accuracy: 76.31% with ±7 seconds
- Percentage accuracy: 87% with ±20% over real AROT

The MEAN model (basic statistical approach) presents the worst result, this proves that AI could help in ATM tasks especially when the problem complexity goes farther than a linear regression, as in the ROT problem. Approximately 80% of the time resources when solving these kind of problems (statistical based) is spent preparing the data [31] and it is necessary both for the ML approach and for the mean model, so ML techniques increase their applicability (better results with comparable resources).

Analysing the distribution of each prediction in comparison with the real AROT, we can see in Fig. 7 that, the more relevant data sources are included in the model, the better its results fit with the real distribution. As we were expecting from the data exploratory analysis, the model tends to predict the average AROT, which is 46 seconds in LEBL.

In Fig. 5 and Fig. 6 we can see that all models are very similar regarding to the precision obtained. But when we analyse the distribution of the results in Fig. 7, the model that include most features, fits better to the real distribution. This is very important, as only with the flight plan you can have a useful prediction in a very early time of operation. Therefore, if any important connection like radars is lost, this tool can work automatically with less information to help during airport operation.

In addition, for the best model, we had divided the results per runway as shown in Fig. 8, to establish a confidence value per runway. Same procedure is done to establish confidence intervals around wake turbulence, aircraft type, company and the hour of the estimated landing. For confidentiality reasons, only runway confidence can be published. Analysing the
validation results in Fig. 9 (April 1st, 2018 to August 31st 2018), we discovered that at periods with higher number of flights (busy timeframes), the models produce better results.

That means the model’s absolute accuracy increases with demand at LEBL, and both trend lines converge. This is interesting because the AROT predictions of the system are more valuable when the airport is operating close to the capacity limit. To conclude, the results analysed per day are between 70% and 80% of accuracy, most of the cases are around 76%, meaning this model is very robust and consistent along time.

![Figure 5 Models evaluation using absolute accuracy](image)

![Figure 6 - Models evaluation using percentage accuracy](image)

![Figure 7 Distribution of the predicted AROT against the real AROT](image)

**V. FINDINGS AND COVERED GAPS**

**A. Conclusions**

The main contribution of the study is the provision of a robust methodology to generate resilient AROT predictions using Machine Learning techniques that can be used during real time operations. It is an iterative and dynamic approach. The outcome system is able to adapt itself to the available data sources, in an incremental way, (i.e. with more data sources, the better the results obtained are), and is scalable to other airports. We highlight that the Flight Plan Information is clearly decisive, especially on regular conditions of airport load and weather for LEBL, only with this information, the FP model can predict more than 70% of the cases with a tolerance of ±7 seconds. In addition, these models will help decision-making processes for tower ATCos during real operations, allowing a potential increase of runway capacity utilisation, while maintaining or enhancing the safety levels by anticipating controllers missed approach situations. For capacity calculations, in strategic planning phases, our analysis provides better predictions of ROT values than the basic mean estimation. Additionally, we have challenged our algorithm to be validated against the period of the year with the highest number of operations for LEBL, finding that these models work better during traffic movement peaks.
B. Limitations & Future Work

Even if the amount of data is considerably high, it would have been desirable to have at least 2 years for training and a whole year for validation. For this paper, April 2018 to August 2018 period has been selected as validation period, because even if weather is worst during winter rather than summer, Barcelona has a soft weather along the year. In further steps, is expected to integrate these prediction into a decision-making system to aid ATCos during real operations and to evaluate which separation measurement (time or space) can be applied, as well as to evaluate the use of AROT predictions as a complement for future ATM architecture programmes (e.g. TBS and RECAT-EU). Future work will also address the possible impact of departure aircraft and the interactions between arrivals and departures. This methodology will be tested in other airports in Spain, being LEMD (Madrid) the next one, in addition we will incorporate new features, i.e. a better airport load metric and a Rapid Exit prediction to feed the AROT models, since the analysis performed indicates that there is a high correlation between AROT and RE.

REFERENCES


[22] Iowa State University Public Repository (2020). Available at: https://mesonet.agron.iastate.edu/ [Last accessed February 8, 2020].


