Predicting Requested Flight Levels with Machine Learning

Manuel Mateos, Ignacio Martín, Pedro García,
Rubén Alcolea, Ricardo Herranz, Oliva G. Cantú-Ros
Nommon Solutions and Technologies
Madrid, Spain
nommon@nommon.es

Xavier Prats
Department of Physics - Aerospace Division
Technical University of Catalonia - BarcelonaTech
Castelldefels, Spain
xavier.prats@upc.edu

Abstract— **The objective of this paper is to present a machine learning approach for the prediction of the Requested Flight Level received during the pre-tactical phase of the Air Traffic Flow and Capacity Management process. A set of machine learning models are proposed in order to determine which Requested Flight Level is the most likely to be filed by an airspace user for a certain origin-destination pair. Results show that the proposed system outperforms the pre-tactical traffic forecasting approach currently used by the European Network Manager in 60% of the 14,465 origin-destination pairs considered in the study, reducing the error of the current solution by 4.8%.**

Keywords-component; pre-tactical ATFCM, RFL prediction, machine learning.

I. INTRODUCTION

The goal of Air Traffic Flow and Capacity Management (ATFCM) is to make airport and airspace capacity meet traffic demand and, when capacity opportunities are exhausted, optimise traffic flows to meet the available capacity, in a seamless process that spans from strategic planning to operations. In Europe, ATFCM is handled by the Network Manager and comprises three phases: strategic planning covers the planning phase between 18 months and 7 days before operations; pre-tactical flow management is applied during the six days prior to the day of operations; finally, tactical flow management takes place in the day of operations.

In order to detect demand and capacity imbalances, the Network Manager forecasts the expected demand at a given timeslot and for all possible airspace sectors (3D airspace volumes) according to the information available at each planning horizon. During the pre-tactical phase, when few or no flight plans (FPLs) have been filed, the only flight information available to the Network Manager are the so-called Flight Intentions (FIs), which include the flight call sign, the airline, the origin and destination airports, the estimated departure time, and aircraft model to be used. The information about the lateral route and the Requested Flight Level (RFL) is not available until Airspace Users (AUs) send their FPLs. To estimate this information, the Network Manager relies on the PREDICT tool, which is used to predict the FPL before it is filed and provide the Network Manager Operations Centre (NMOC) with the information required to ensure a correct allocation of resources in coordination with Air Navigation Service Providers (ANSPs). PREDICT generates traffic forecasts according to the trajectories chosen by the same or similar flight codes in the recent past, without taking advantage of the information potentially encoded in historical FPLs and trajectory data.

Recent work has explored how to use this information to build machine learning models for the prediction of the lateral route during the pre-tactical phase ([1],[2],[3]). Some other work has also applied machine learning methods to predict the complete trajectory (4D). Relevant examples are [4] and [5], which consider wind and temperature as features for the prediction of the 4D trajectory in the tactical phase, and [6], which performs a short-term 4D trajectory prediction based on the initial position and velocity of the aircraft and the local wind. On the other hand, data-based RFL prediction has seldom been specifically addressed.

The prediction of RFL has usually been studied by means of physical models that look for an optimal trajectory (e.g., by optimizing fuel consumption) ([7],[8],[9]). However, AUs do not always request the optimal flight level, either because it is not available (e.g., due to route restrictions, ATC limitations, etc.) or because they do not have all the required information to compute the optimal trajectory. In this paper we will focus on developing a machine learning approach for the prediction of the RFL based on historical FPL data.

The rest of this paper is structured as follows: Section II describes the proposed modelling approach and the machine learning models to be tested; Section III describes the experiments designed to evaluate the performance of the proposed models; Section IV presents and discusses the main results of the experiments; finally, Section V summarizes the main conclusions of the study and discusses future steps.

II. Machine Learning Models for RFL prediction

A. *Problem Statement*

Although aircraft can fly at any altitude within their performance range, Air Traffic Management (ATM) imposes conditions on the allowed flying altitudes. Flight levels are described by a number, which is the nominal altitude, or pressure altitude, in hundreds of feet. ATM establishes rather rigid rules to ensure vertical separation, which in practice means that most intra-European flights require a single cruise flight level at a very specific altitude. Consequently, the prediction of the RFL can be seen as a supervised classification problem, where classes are the potential RFLs each aircraft can fly.

These rules reduce the number of possible flight levels to a few dozens in the majority of the cases, of which only a few are recurrently used.

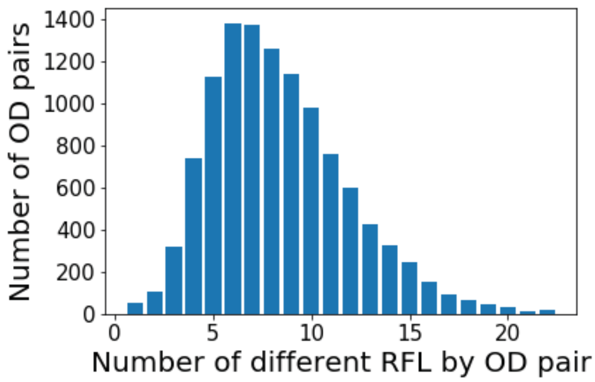

Figure 1. Distribution of the number of different RFLs per OD pair

Figure 1 depicts how many origin-destination (OD) pairs in the 14,465 OD pairs considered in the study account for each value of different RFLs during the year 2018. The number of OD pairs for which more than 15 different RFLs are used is relatively small. For most OD pairs, the number of RFLs is obviously lower. Figure 2 shows the example of the flights between Amsterdam-Schiphol and Rome-Fiumicino.

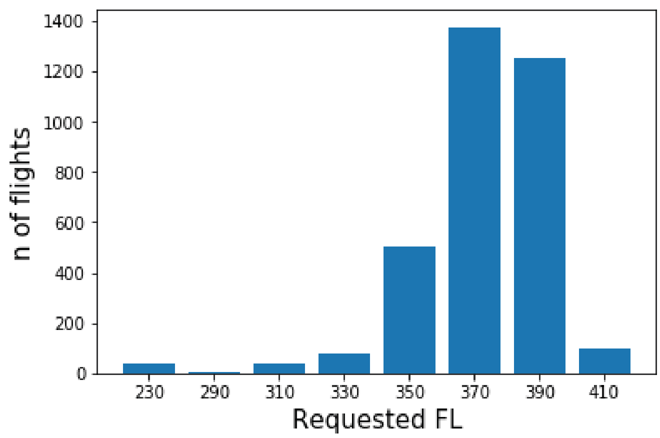

Figure 2- RFL distribution for the OD pair EHAM-LIRF

B. *Machine Learning Models*

According to the features considered, two different models have been developed:

- **Basic Model**, which takes as inputs the time of the flight, the day of the week (i.e., Monday, Tuesday, etc.), the day of the year, the AU (airline), and the aircraft mass.
- **Enhanced Model**, which is built on top of the Basic model by including weather and data on past regulations.

The rationale behind the **Basic Model** is simple: the tool currently used by the Network Manager, PREDICT, is based on the weekly recurrence of FPLs. The idea behind the Basic Model is to leverage such weekly patterns and enrich them with additional information to increase the prediction capabilities of the resulting model. The attributes considered by the model are described below:

- **AU:** one-hot encoding of the airline ICAO code.
- **Day of week:** one-hot encoding of the day of the week when the flight takes place.
- **Hour:** sine and cosine of the expected take-off time (ETOT) hour.
- **Day of the year:** sine and cosine of the day of the year.
- **Aircraft mass:** maximum take-off weight (MTOW) of the aircraft model.

Since there is no previous work that analyzes which machine learning algorithm is the most adequate for the proposed problem, the following classification algorithms were implemented and tested:

- K-Nearest-Neighbours (KNN)
- Multinomial logistic regression
- Decision tree
- Random Forest
- Support Vector Machine (SVM)

The **Enhanced Model** is built by extending the Basic Model with some external variables that might influence AUs behaviour. Following the findings of previous work ([2], [4], [5] and [6]), 4 types of external variables have been considered for this model:

- **Wind:** it is measured as the average wind projection along the flight path, estimated by computing the "along path" wind at specific points of each observed traffic flow. It may be positive (tailwind) or negative (headwind), with the magnitude indicating the strength of the wind component along the flight path. Although it can sometimes have an influence on the RFL, in the model described in this paper crosswind has been neglected and left for future research.

- **Convective phenomena probability:** the average and maximum values of the relative humidity, convective available potential energy (CAPE) and k-index at each traffic flow are used as proxy variables of the probability of occurrence of a storm.
- **Past regulations:** the regulations observed during the expected duration of the flight in the previous day, seven days before, and during the last 28 days are used as an indicator of the expected congestion levels.
- **Local wind at airports:** this is included to account for the fact that the selection of the most convenient trajectory might be affected by the airport runway configuration.

Even though new predictive features can contribute to improving the prediction performance of the model, an excessive number of features could undermine the model training process and lead to overfitting issues. To avoid these problems, Recursive Feature Elimination (RFE) has been used to automatically reduce the feature set to the most relevant features.

C. *Benchmark Model: PREDICT*

In order to evaluate the performance of the proposed models, their accuracy has been measured and compared against that of PREDICT, the tool currently used by the Network Manager. Although the tool is mentioned in numerous EUROCONTROL documents, its implementation details are not publicly available. The functioning of PREDICT has thus been emulated following the information available from the Network Manager documentation [10] and the indications from EUROCONTROL experts. For each flight, the following workflow is applied: (1) look for previous flights with the same call sign on the same day of the week. If this is not possible, the flight operated by the same company at the closest time of the day is selected; (2) if no previous flight for the company is available, the same operation is repeated regardless of the company; (3) if no flight has met the previous requirements yet, the most recent FPL for the same OD pair is selected.

III. Data and Experimental Design

A key condition for the proper training of machine learning models is the use of large datasets, especially when the feature space is large. For the experiments reported in this paper, we have used data from EUROCONTROL's Demand Data Repository (DDR) [11]. In particular, data from AIRAC cycles 1801 to 1813 have been used. These data cover 52 weeks of traffic, from which the first 48 weeks have been used to train the models and the last 4 weeks for testing.

Validation experiments have been conducted for a total of 14,465 OD pairs (practically the whole network). Model evaluation has been undertaken using as primary metric the accuracy of the system, which is computed according to the following principles:

- A flight is considered as correctly predicted when the predicted flight level matches the RFL.
- For each OD pair, accuracy is defined as the number of correct guesses divided by the number of total flights.
- The result of each OD pair is weighted by the number of flights in that OD pair.

IV. Results

All five algorithms mentioned in Section II.B have been tested for both the Basic and the Enhanced models in a reduced

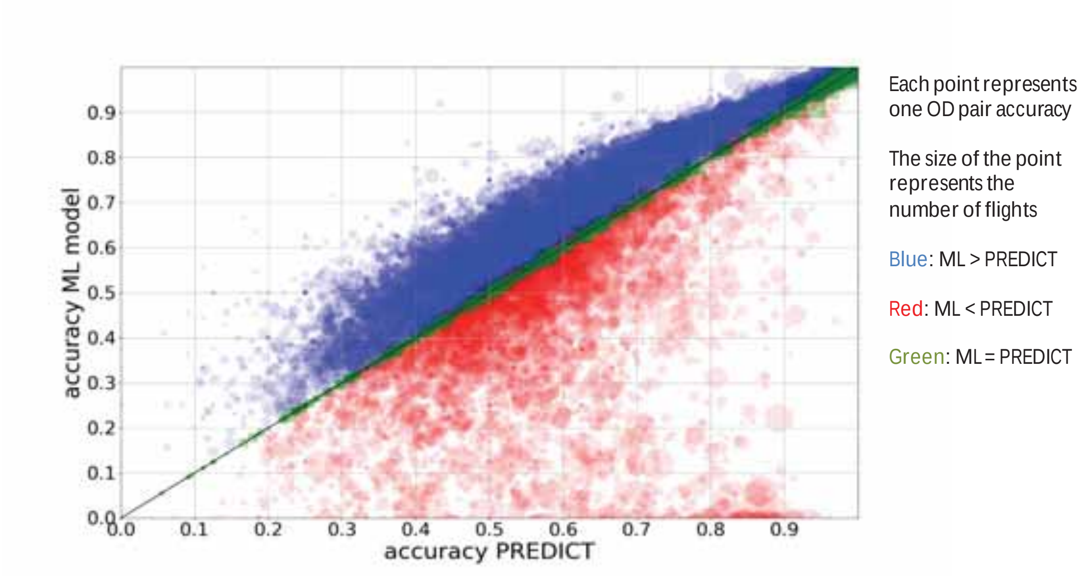

Figure 3 - Accuracy of the Basic Model by OD pair

set of OD pairs. Preliminary results show that the Random Forest algorithm provides slightly better results. Therefore, hereinafter we only report the Random Forest results.

A. Basic Model Results

The Basic Model was tested on a total of 14,465 OD pairs extracted from the available dataset. From these OD pairs, 10,787 OD pairs (75.6%) resulted in a model, while the rest were discarded either because the number of flights was too low to train a machine learning model or because they have only one class (i.e., all flights have the same RFL). Figure 3 summarizes the results of the experiment:

- For 60.33% of the OD pairs, the model shows better performance than PREDICT.
- For 30.11% of the OD pairs, the proposed model provides worse results than PREDICT.
- For 9.56% of the OD pairs, the proposed model shows similar accuracy than PREDICT.

In global terms, this translates into an increase of accuracy of 3.2%. If, for each OD pair, the best approximation was selected, the increase in the accuracy of the resulting hybrid system would be around 8%.

B. Enhanced Model Results

TABLE 1 – ACCURACY OF THE ENHANCED MODEL

OD pair	PREDICT	Basic Model	Enhanced Model
EDDF-UUEE	0.605	0.685	0.741
EDDK-LTAI	0.440	0.360	0.320
EDDT-LEPA	0.519	0.541	0.526
EGLL-OMDB	0.568	0.498	0.540
EHAM-LIRF	0.381	0.422	0.471
LEPA-EDDT	0.598	0.591	0.705
LFPG-LGAV	0.465	0.496	0.574
LFPO-LPPT	0.503	0.523	0.530
LGAV-LFPG	0.504	0.579	0.653
LIRF-EHAM	0.567	0.580	0.482
LPPT-LFPO	0.531	0.622	0.646
LTAI-EDDK	0.422	0.356	0.356
OMDB-EGLL	0.422	0.468	0.489
UUEE-EDDF	0.578	0.566	0.572
Average	**0.507**	**0.521**	**0.543**

At the moment of writing this paper, the Enhanced Model is still under development, so it has only been tested for a small subset of OD pairs, selected according to two main criteria: (1) data availability; and (2) ensuring a variety of OD pairs with different characteristics (length, congestion, etc.). The following pairs have been selected:

- Antalya – Cologne Bonn (LTAI-EDDK)
- Berlin Tegel – Palma de Mallorca (EDDT-LEPA)
- London Heathrow – Dubai (EGLL-OMDB)
- Athens – Paris Charles de Gaulle (LGAV-LFPG)
- Amsterdam Schiphol – Roma Fiumicino (EHAM-LIRF)
- Lisbon Portela – Paris Orly (LPPT-LFPO)
- Moscow Sheremetyevo – Frankfurt (UUEE-EDDF)

The results obtained are presented in TABLE 1. For the OD pairs under study, the Enhanced Model provides a 7% increase of accuracy with respect to PREDICT.

In order to provide a more detailed analysis on the kind of variables used in the model of each OD pair, the number of variables of each type are displayed in TABLE 2.

TABLE 2 - NUMBER OF CONSIDERED VARIABLE BY TYPE

OD pair	Wind	Convective events	Past Regulations	Local wind
EDDF-UUEE	4	6	0	2
EDDK-LTAI	3	7	0	2
EDDT-LEPA	4	5	1	3
EGLL-OMDB	2	7	0	3
EHAM-LIRF	1	7	0	4
LEPA-EDDT	1	8	0	4
LFPG-LGAV	3	7	0	2
LFPO-LPPT	3	6	0	3
LGAV-LFPG	1	7	0	4
LIRF-EHAM	3	5	0	4
LPPT-LFPO	3	5	0	4
LTAI-EDDK	1	9	0	2
OMDB-EGLL	2	7	1	2
UUEE-EDDF	3	7	0	2

The main observations from the table are the following:

- RFE leads to picking different variables for each OD pair.
- Local wind variables seem to be relevant in most cases, in particular for the destination airport.
- Convective event variables are also relevant in every OD pair. It is worth noting that these variables represent more than half of the RFE-selected variables for almost every pair. The most important of these features is the k-index.
- On-route wind seems to be relevant in general terms, although the effect is more relevant in certain pairs.

- Regulation based variables appear to be less relevant as they seldom are selected for the model.

C. *Requested Flight Level versus optimal Flight Level*

Finally, we have studied how close the RFL predicted with the machine learning approach is from the optimal RFL. To this end, the vertical profile of each flight has been simulated using the DYNAMO tool, developed by the Technical University of Catalonia [7], and the accuracy of both approaches has been compared.

While the optimal RFL (calculated using DYNAMO) only corresponds to the actual RFL in 10% of the cases, the Basic Model achieves an accuracy of 60%. Moreover, the average distance to the actual RFL records is higher for the optimization-based approach (3,240 ft) than for the machine learning approach (1,580 ft).

V. CONCLUSIONS AND NEXT STEPS

In this paper we have proposed and compared two machine learning models for RFL prediction: a Basic Model, based on a reduced set of features (time of the flight, day of the week, day of the year, airline, aircraft mass), and an Enhanced Model, which incorporates a number of additional variables aimed at capturing the influence of wind, convective phenomena, congestion, and airport runway configuration.

When tested at the level of the full European network, the Basic Model has proven its ability to outperform the accuracy of the current Network Manager solution by 3.2%. The Enhanced Model has been tested on a reduced number of OD pairs, showing promising results that outperform the Basic Model.

The following research questions have been identified as interesting for future research:

- Explore if the magnitude of the error (distance between failed prediction and actual RFL) has any impact on the demand or, on the contrary, all failed predictions have a similar impact.
- Investigate the applicability of the Enhanced Model to the entire network and the resulting prediction accuracy.
- Explore the inclusion of new features, such as crosswind or the proximity of the route to a severe weather event, in the Enhanced model.
- Develop algorithms for the automatic selection of the most accurate algorithm for each OD pair.
- Explore the feasibility of developing models covering several OD pairs, instead of training a single model for each one of them.
- Integrate these results with previous work on route prediction in order to compute demand indicators and evaluate Demand and Capacity Balancing (DCB) problems.

ACKNOWLEDGMENTS

Manuel Mateos' PhD is funded by SESAR's Engage Knowledge Transfer Network, which has received funding from the SESAR Joint Undertaking under the European Union's Horizon 2020 research and innovation programme under grant agreement No 783287. The opinions expressed herein reflect the authors' view only. Under no circumstances shall the SESAR Joint Undertaking be responsible for any use that may be made of the information contained herein.

The authors would also like to acknowledge the support of the Spanish Centre for Industrial Development (CDTI) through the PRETA project (Grant no. IDI-20190029).

Finally, the authors would like to acknowledge the support provided by EUROCONTROL, by facilitating data access and expert advice. Particularly, the authors would like to thank Ms. Stella Saldana, Mr. Stefan Steurs, Mr. Eric Allard and Mr. Francis Decroly for their advice on the definition of the models and the design of the evaluation experiments.

VI. REFERENCES

[1] M. Mateos, I. Martín, P. García, R. Herranz, O. Cantú-Ros, and X. Prats, "Full-scale pre-tactical route prediction", 9th International Conference for Research in Air Transportation (ICRAT), 2020.

[2] R. Marcos, "Combining Visual Analytics and Machine Learning for Route Choice Prediction: Application to Pre-Tactical Traffic Forecast", SESAR Innovation Days, 2017.

[3] Y. Liu, M. Hansen, D.J. Lovell, and M.O. Ball, "Predicting Aircraft Trajectory Choice-A Nominal Route Approach", 8th International Conference for Research in Air Transportation (ICRAT), 2018.

[4] S. Ayan and H. Samet, "Aircraft trajectory prediction made easy with predictive analytics", 22nd International Conference on Knowledge Discovery and Data Mining, 2016.

[5] S. Ayan and H. Samet, "Time series clustering of weather observations in predicting climb phase of aircraft trajectories", 9th ACM SIGSPATIAL International Workshop on Computational Transportation Science, 2016.

[6] Z. Wang, M. Liang and D. Delahaye, "Short-term 4D trajectory prediction using machine learning methods" SESAR Innovation Days, 2017.

[7] R. Dalmau, M. Melgosa, S. Vilardaga, and X. Prats, "A Fast and Flexible Aircraft Trajectory Predictor and Optimiser for ATM Research Applications", 8th International Conference for Research in Air Transportation (ICRAT), 2018.

[8] APACHE Consortium, Deliverable D4.1, Report on the availability of the APACHE Framework, 2018.

[9] C. Allignol, N. Barnier, and A. Gondran, "Optimized flight level allocation at the continental scale", 5th International Conference on Research in Air Transportation (ICRAT), 2012.

[10] EUROCONTROL, ATFCM Operations Manual, Edition 22.1, 2018.

[11] EUROCONTROL,.Demand Data Repository, retrieved from https://www.eurocontrol.int/ddr.