# SESAR Engage KTN – PhD final report

| | |
|---|---|
| PhD title: | Machine learning for aircraft trajectory prediction: a solution for pre-tactical air traffic flow and capacity management |
| Candidate's name: | Manuel Mateos |
| Lead supervisor's name: | Xavier Prats |
| Co-supervisor's name (if applicable): | Oliva García-Cantú |
| Proponent institute: | Nommon Solutions and Technologies |
| Consortium institutes (if applicable): | Technical University of Catalonia - UPC |
| Thematic challenge: | TC2 Data-driven trajectory prediction |
| Edition date: | 21 July 2022 |
| Edition: | 1.1 |
| Dissemination level: | Public |



This project has received funding from the SESAR Joint Undertaking under the European Union's Horizon 2020 research and innovation programme under grant agreement No 783287.

# 1. Abstract

The goal of air traffic flow and capacity management (ATFCM) is to ensure that airport and airspace capacity meet traffic demand while optimising traffic flows to avoid exceeding the available capacity when it cannot be further increased. In Europe, ATFCM is handled by EUROCONTROL, in its role of Network Manager (NM), and comprises three phases: strategic, pre-tactical, and tactical. This thesis is focused on the pre-tactical phase, which covers the six days prior to the day of operations.

During the pre-tactical phase, few or no flight plans (FPLs) have been filed by airspace users (AUs) and the only flight information available to the NM are the so-called flight intentions (FIs), consisting mainly of flight schedules. Trajectory information becomes available only when the AUs send their FPLs. This information is required to ensure a correct allocation of resources in coordination with air navigation service providers (ANSPs). To forecast FPLs before they are filed by the AUs, the NM relies on the PREDICT tool, which generates traffic forecasts for the whole European Civil Aviation Conference (ECAC) area according to the trajectories chosen by the same or similar flights in the recent past, without taking advantage of the information on AU choices encoded in historical data.

The goal of the present PhD thesis is to develop a solution for pre-tactical traffic forecast that improves the predictive performance of the PREDICT tool while being able to cope with the entire set of flights in the ECAC network in a computationally efficient manner. To this end, trajectory forecasting approaches based on machine learning models trained on historical data have been explored, evaluating their predictive performance.

In the application of machine learning techniques to demand trajectory prediction, three fundamental methodological choices have to be made: (i) approach to trajectory clustering, which is used to group similar trajectories in order to simplify the trajectory prediction problem; (ii) model formulation; and (iii) model training approach. The contribution of this PhD thesis to the state-of-the-art lies in the first two areas. First, we have developed a novel route clustering technique based on the area comprised between two routes that reduces the required computational time and increases the scalability with respect to other clustering techniques described in the literature. Second, we have developed, tested and evaluated two new modelling approaches for route prediction. The first approach consists in building and training an independent machine learning model for each origin-destination (OD) pair in the network, taking as inputs different variables available from FIs plus other variables related to weather and to the number of regulations. This approach improves the performance of the PREDICT model, but it also has an important limitation: it does not consider changes in the route availability, thus being unable to predict routes not available in the training data and sometimes predicting routes that are not compatible with the airspace structure. The second approach is an airline-based approach, which consists in building and training a model for each airline. The limitations of the first model are overcome by considering as input variables not only the variables available from the FIs and the weather, but also route availability and route characteristics (e.g., route cost, length, etc.).

The airline-based approach yields a significant improvement with respect to PREDICT and to the OD pair-based model, achieving a route prediction accuracy of 0.896 (versus PREDICT's accuracy of 0.828), while being able to deal with the full ECAC network within reasonable computational time. These promising results encourage us to be optimistic about the future implementation of the proposed system.

# 2. Objective of the study

The overall goal of the PhD thesis is to develop and evaluate innovative approaches to air traffic demand forecasting based on artificial intelligence and machine learning techniques, focusing on the pre-tactical phase of the ATFCM process. The specific objectives of the PhD are:

1. to characterise in detail the full range of factors that determine air traffic demand, with particular focus on the variables related to stakeholder behaviour (e.g., airline preferences);
2. to develop demand forecasting models able to incorporate the whole range of identified factors through the use of data-driven techniques;
3. to identify the forecasting approach that provides more accurate and robust results depending on the available information;
4. to perform an initial validation of the proposed concept.

The PhD was initially intended to develop a “coherent and seamless probabilistic prediction throughout pre-tactical and tactical ATFCM”. Nevertheless, following the advice from EUROCONTROL experts gathered during several working sessions at the beginning of the PhD study, the thesis has finally focused exclusively on the pre-tactical phase. The reasons are mainly two:

- According to EUROCONTROL experts, pre-tactical traffic prediction, in particular the prediction of the FPLs, containing the route (horizontal 2D trajectory) and the Requested Flight Level (RFL), has historically received less attention by researchers, so advances in this area can deliver important benefits from the operational point of view. Since EUROCONTROL has been carrying out an internal project aimed at improving its pre-tactical demand forecasting tools, it was agreed to share the outcomes of both projects in order to compare different approaches and exploit possible synergies.
- The correct prediction of the tactical phase would have required information on short time ATFCM measures (STAMs), which are usually managed without a formal log.

The proposed research topic is directly linked with the Engage KTN Thematic Challenge 2 “Data-driven trajectory prediction” and, more tangentially, to Thematic Challenges 3 and 4, “Efficient provision and use of meteorological information in air traffic management (ATM)” and “Novel and more effective allocation markets in ATM”.

# 3. Motivation

## 3.1. General introduction to ATM

ATM is an aeronautical concept that includes all systems, procedures and human resources necessary to ensure the safe and efficient transit of aircraft during all operational phases. ATM consists of three main activities:

- **Air traffic services (ATS),** which encompass alert services, flight information services and air traffic control (ATC).
- **Air traffic flow and capacity management (ATFCM),** for which the term air traffic flow management (ATFM) is also used. This activity consists in balancing capacity and demand by providing the necessary information, modifications and logistics so that the Air Traffic Controllers (ATCOs) can operate nominally, without exceeding their workload limits. Most of the procedures involved take place before flight departure.
- **Airspace management (ASM),** which is in charge of the design of the airspace.

The focus of this PhD is on the ATFCM domain. The PhD study is based on the European ATFCM process, commonly called 'Network Management'. Network Management is currently performed by EUROCONTROL, which plays the role of 'Network Manager', according to the European Commission implementing regulation 2019/123[1].

ATFCM entails a continuous process from strategic planning to the execution of the operations. According to [1], this process is divided in three phases: strategic, pre-tactical and tactical, each one facing a different time horizon. One of the key elements for ATFCM provision are traffic forecasts. The scope of ATFCM activities, and therefore the type of traffic forecasting required to support such activities, are different for each of the three ATFCM phases:

1. **Strategic phase.** This phase takes place from one year and a half to one week before operations. In this phase, aggregated predictions of flows are made to identify major demand-capacity imbalances due to upcoming events. The predictions made are based on historical data, economic trends and seasonal effects, together with the available FIs. The outputs are consolidated in the generation of the network operation plan (NOP).
2. **Pre-tactical phase.** The pre-tactical phase takes place from six days until the day before operations. The objective of this phase is to elaborate the Daily Plan, based on a more refined traffic forecast, which aims to provide an optimal scenario configuration in order to minimise delay and cost for AUs. Traffic forecast is already focused on individual flights. In this phase most FPLs are not available yet, so the main source of information are the FIs. Pre-tactical traffic forecast is performed by the PREDICT software, which generates a prediction of FPLs based on historical data and the available FIs. With this prediction, the ANSPs, the Network Manager Operations Centre (NMOC) and the AUs participate in a collaborative decision making (CDM) process that results in the so-called ATFCM Daily Plan.
3. **Tactical phase.** The tactical phase is carried out during the day of operations and involves using real-time information to adapt the ATFCM Daily Plan. Predictions are based on FPLs.

## 3.2. ATFCM: a prediction problem

The European ATFCM service is provided by the NMOC to all the AUs throughout the ECAC states (currently 44 states), with the purpose of using the available airspace capacity as efficiently as possible. The cornerstone of European ATFCM is the demand and capacity balancing (DCB) process [1]. The main goal of the DCB service is to ensure that airspace capacity and traffic demand match in order to avoid (unsafe) overloaded sectors or airports.

In order to estimate the expected demand, the ATFCM service makes a prediction of the airspace demand by computing the expected trajectories and their evolution over time from the information of each individual FPL. Also, from the pre-declared information of the ANSPs it is possible to make a prediction of the expected capacity that will be available at every airspace sector or airport. These predictions are refined as the day of operations becomes closer, when the quantity and quality of information used for predictions usually increases.

ATFCM consists in balancing capacity and demand to facilitate ATC operations. Theoretical capacity is known (for a given airspace configuration), while the exact demand is only known in real time. Since corrective actions to balance demand and capacity have to be taken in advance, the prediction of the expected demand is a key element of ATFCM.

[1] https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019R0123&from=EN

In the case of the pre-tactical planning, AUs do not usually file their FPLs up to a few hours before the flight takes place, so the only information available to predict the demand is the list of scheduled flights, also called FIs. FIs are not extracted from a standard data source, but they are obtained as a compendium of data from many sources, such as airline schedules and airport slot allocation. Once this information is compiled, the FIs typically contain the following information: flight identifier (FID), origin and destination airports, estimated departure/arrival time, airline and aircraft type.

Ideally, the ANSPs would need to know the demand at each airspace sector to select the most appropriate configuration at each time. Nevertheless, the information contained in the FIs is insufficient to calculate the demand because the flight trajectory is not included. Therefore, one of the most important tasks in pre-tactical ATFCM is the estimation of an FPL given its FI. Currently, the NM uses the PREDICT tool for such purpose.

### 3.2.1. The PREDICT tool

The PREDICT system transforms historical traffic data into predictions for the next 6 days. This process is performed according to the diagram in Figure 1, following the steps below:

1. **Enrichment**: the FIs gathered by EUROCONTROL's Demand Data Repository (DDR) are compared with historical traffic demand. Those flights operated in the past (in principle the week before) with FIs to be flown in the future are confirmed. The off-block time of confirmed flights is also aligned to the FIs off-block times; the FIs that do not correlate with historical data are considered as new flights, and the flights present in the historical data but not appearing on the FIs are considered candidates to be deleted.

2. **Route assignment**: the route of the confirmed flights is assumed to be the same as that of the corresponding historical flights. As for the deleted flights, there is obviously no need to assign an FPL. Route assignment for the new flights is performed as follows:
    a. the system checks the historical FPLs for the same OD pair in the last 28 days (regardless of the airline, if necessary). If more than one FPL is available, it selects the most used. If none is available,
    b. the route of the FPL is searched in the NM catalogue. If it is not available,
    c. the shortest route is generated using a "path finder" engine.

3. **North Atlantic Traffic (NAT) flows substitution**: NAT flows are strongly affected by meteorology; therefore, this is the only aspect in which PREDICT takes into account weather conditions to estimate FPLs. Instead of following the usual approach, NAT flights are assigned a historical FPL from a day with a similar meteorological scenario, based on the weather predictions from UK NATS received 3 days before operations.

4. **Upload to DDR**: the data is made available in the DDR portal for all authorised parties.

The PREDICT procedure is clear, robust, and scalable. Moreover, it has been proved in operations for many years. Nevertheless, it has some limitations that will be addressed in next section.

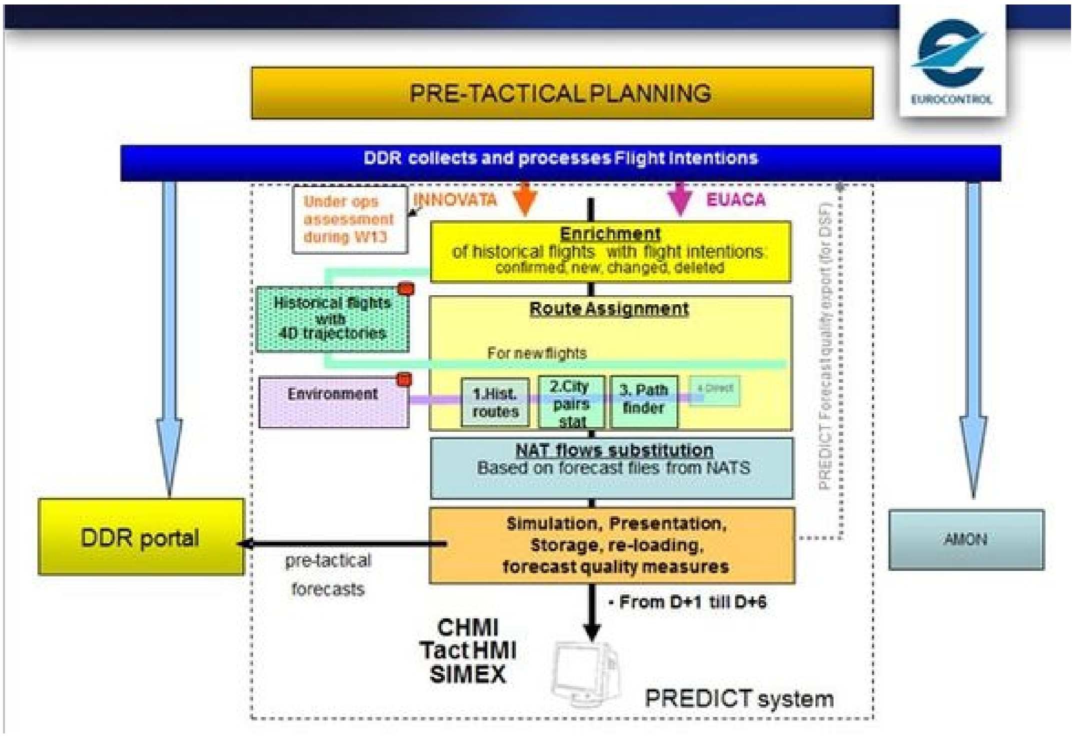


Figure 1: PREDICT tool diagram (source: EUROCONTROL)

### 3.2.2. Pre-tactical traffic forecast: opportunities for improvement

While research on demand prediction in the tactical phase has received much attention, the pre-tactical phase has not received so much interest, and PREDICT has not been significantly improved in the recent years. However, it is widely accepted that the PREDICT software could benefit from some recent advances in trajectory prediction. Potential improvements include the following:

- **Inclusion of new sources of information**: PREDICT only uses FIs plus some environmental information which is basically restricted to procedures and configuration of airspace, overseeing crucial variables such as weather conditions (specially wind), airline preferences, and other factors affecting AU decisions (route charges, fuel price, etc.).
- **New forecasting methods**: the method followed by PREDICT does not follow any cause-effect logic.
- **Invalid predictions**: although PREDICT tool has access to the airspace structure, it does not use it to validate its own predictions. According to the NM experts, the number of non airspace-compatible predictions currently generated by PREDICT is significant, specially the first week of the Aeronautical Information Regulation and Control (AIRAC) cycle.
- **Uncertainty quantification**: another limitation of PREDICT is the lack of statistical information about the uncertainty associated to predictions, which would be very helpful to anticipate the risk of DCB imbalances.

These improvements should increase PREDICT's accuracy but they may also provide more flexibility in unexpected situations (e.g., route closure).

# 4. Advances this work has provided with regard to the state of the art

The main research question in the present PhD thesis is whether the use of machine learning models that rely on historical FPLs are able to identify patterns in AU behaviour regarding the specification of their FPLs and use this information to improve the predictive performance of the PREDICT system.

The use of machine learning for trajectory prediction has been addressed by numerous publications in recent years. Nevertheless, some limitations have been identified. First, while there is abundant research on the prediction of 2D routes, the work on RFL prediction is scarcer [2]. Second, many of the recent studies lack an analysis of the scalability of the proposed approach: a pre-tactical FPLs prediction system is intended to cover an entire network, such as the ECAC area, to facilitate resource allocation and planning. PREDICT manages more than 30,000 flights daily across more than 10,000 OD pairs. Therefore, analysing the applicability at network level is an essential condition for the practical use of new forecasting approaches. For example, the work done in [3] presents results for 5 OD pairs, [4] analyses 3 pairs and [5] uses data for 183 flights.

This PhD has addressed these two limitations:

- We have demonstrated the performance of the proposed methodology to predict the FPLs (both route and RFL), showing that the inclusion of new variables and forecasting methods enable us to outperform EUROCONTROL's PREDICT tool.
- We have developed a tool able to cope with pre-tactical traffic forecast for the entire ECAC network. As part of this effort to deliver a computationally efficient solution, the PhD has proposed a new approach to route clustering based on the area comprised between trajectories, building on the work reported in [6].

# 5. Methodology

The main purpose of the research was the accurate prediction of the FPLs during the pre-tactical phase. To this end, two different types of models have been developed:

- **OD pair-based models:** this was the initial attempt to predict the FPLs. Following this approach, an independent machine learning model is trained for each OD pair to predict the FPL. Independent models were used for the route and the RFL.
- **Airline-based models:** this approach generates an independent machine learning model per airline. The model estimates the probability for the airline to choose a particular route given the characteristics of that particular route. Being based on the route characteristics, this approach enables the prediction of non-observed routes.

The methodology followed to build the models comprises four steps: data acquisition; characterisation of the elements to be predicted: route and RFL; feature construction; and model training and validation.

## 5.1. Data acquisition

This step consists of the extraction of data from DDR and other external data sources.

The FPL data has been obtained from DDR. In particular, data from AIRAC cycles 1801 to 1813 and 1906 to 2002 has been used. Additionally, the DDR provides the required sectors, military zones, route charges, airport locations, and regulations information.

The external data sources used in the study include meteorological data and different socioeconomic data (e.g., fuel prices). The different data sources are described in further detail in Section 6.

## 5.2. Route and RFL characterisation

### 5.2.1. Route characterisation

Routes are complex elements composed by an undetermined number of 2D points. This kind of elements is unlikely to be directly predicted using supervised machine learning algorithms. Hence, state-of-the-art approaches resort to clustering techniques to simplify the route prediction problem. Trajectory clustering algorithms comprise three main elements: the attributes used for the clustering, the distance metric that determines the similarity of trajectories, and the clustering techniques employed.

The main objective of the route clustering proposed in this PhD is to group those routes which are equivalent from the pre-tactical ATFCM point of view. The NM collaboration has been key to derive the requirements for route clustering, which are summarised below:

- The effect of the manoeuvres around the terminal area is not relevant.
- Close enough trajectories that cross the same sectors are expected to have a similar impact.
- Differences in a relatively small part of the FPL are admissible.
- Flight times in the FPL are not relevant for route clustering.

These requirements have helped identify the elements of the clustering which best fit the objectives of the research:

- The attributes used for the clustering are the 2D routes from the FPLs. In order to avoid the effect of the terminal area, the segments of the route closer than 40 nautical miles to the origin and destination airports have been discarded.
- The selected clustering technique aims to include trajectories with small variations as part of the same cluster. Theoretically, density-based spatial clustering of applications with noise (DBSCAN) is the most suitable technique for such purpose. This is consistent with the state-of-the-art review carried out during the PhD thesis, which shows that DBSCAN is the technique most commonly used.
- The distance metric shall provide a clear sense of geometric similarity between routes.

Initially, symmetrised segment path distance (SSPD) was identified as a good candidate for route distance. Nevertheless, the calculation of the SSPD was becoming the main bottleneck from a computational point of view, so the use of SSPD as a distance metric would have limited the scalability of the results. Therefore, another metric was tested, based on the area between routes. Conceptually, the idea of using the area to cluster similar routes makes sense: the area between two routes that are geometrically similar is smaller than the area between two routes that are significantly different. Additionally, the calculation of this area has proved to be around 100 times faster than the SSPD and, unlike the SSPD, it is not affected by the route sampling.

For each calculated cluster, the route with the lowest intra-cluster distance is selected as the representative route for the cluster, also called “central route”.

### 5.2.2. Requested flight level characterisation

Although aircraft can fly at any altitude within their performance range, 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. Consequently, the prediction of the RFL can be seen as a supervised classification problem, where classes are the potential RFLs each aircraft can fly.

The prediction of the RFL has usually been studied by means of physical models that look for an optimal trajectory (e.g., by optimising fuel consumption) (see [7], [8], and [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 (or most up to date) information to compute the optimal trajectory.

## 5.3. Feature construction

Feature construction includes all the methods and algorithms applied to transform the raw data into a dataset that can feed the training of the machine learning model. Feature construction will be independently described for the OD pair-based and the airline-based models.

### 5.3.1. OD pair-based model

Two different models have been developed: a basic model, which considers a reduced number of attributes, and an enhanced model, which considers a larger number of attributes. Both models have been developed for route and RFL prediction.

#### *5.3.1.1. OD pair-based basic model*

The basic model was the first step in the transition from a model based on similarity (PREDICT) to a model that takes into account the correlation between the selected trajectories and different observable features. The basic model takes as inputs the day of the week (i.e., Monday, Tuesday, etc.), the time of flight, the day of the year, the AU, and the aircraft model maximum take-off weight (MTOW); this information is directly obtained from the FPL. Although this information is already available for PREDICT, the historical FPL records are not exploited.

The machine learning algorithms employed in the PhD study require certain transformation of the data before feeding the models. Two main techniques we used for this purpose:

- **One-hot encoding:** it is used to transform a categorical variable with finite categories into a numerical form. To do so, each category becomes a feature with value "1" when the categorical value takes the value of this category and "0" otherwise. One-hot encoding has been used to pre-process the day of week (DoW) and the airlines.
- **Sin-cos transformation:** this technique has been applied to capture the continuity between consecutive days or years, i.e., the fact that a flight departing at 23:55 will behave similar to another departing at 0:10. A sin-cos transformation has been applied to the time of flight and the day of year (DoY). The DoY is the ordinal position of any day of the year starting from the 1st of January (e.g., 1st of May 2018 is DoY 121). The sin-cos transformation consists in the generation of two new features for each variable, so they are always continuous:

$$h_c = \cos\frac{2\pi V}{T} \qquad h_s = \sin\frac{2\pi V}{T} \qquad (1)$$

  where V is the variable to transform (i.e., the time of flight) and T the period. This is T=24 for the hour transformation and T=365 (366 for a leap-year) for the DoY transformation.

#### *5.3.1.2. OD pair-based enhanced model*

The enhanced model has been built on top of the basic model by including additional information that was found relevant for the FPL. According to the findings of previous work ([10], [11], [12], [13]), 4 types of external variables have been considered for this model:

- **Local Wind:** local wind is extracted from the origin and destination airports METAR files for the expected departure and arrival time. The direction and magnitude of the wind for both airports are assigned as features for each flight.
- **Along Track Wind:** the along track wind feature is calculated for the central route of all the clusters. It is computed as the average wind projection along the flight path at specific points of each central route. It may be positive (tailwind) or negative (headwind), with the magnitude indicating the strength of the wind component along the flight path. Although it may be relevant in certain wind scenarios, crosswind has been neglected and left for future research.
- **Convective phenomena:** raw data is extracted from the Climate Data Store (CDS)[2]. Features are calculated again along the central routes. For each meteorological indicator, the average and the maximum value observed along the route are calculated. The meteorological indicators used are:
  - K-index: this index, also known as George's index, is a measure of thunderstorm potential. It is a function of Temperature and Dew Point at several altitudes.
  - CAPE: convective available potential energy; it is a measure of the instability in the atmosphere.
  - Humidity: the presence of a relatively high fraction of water in the atmosphere is a necessary condition for some events such as storms to happen.
- **Past Regulations:** the use of regulations to predict the AU's behaviour has to take into account that regulations are not known during the pre-tactical phase. The hypothesis proposed is that recent past regulations might condition AU's choice. To this end, 3 different scopes have been considered: 1 day before, 7 days before, and during the last 28 days.

### 5.3.2. Airline-based model

The airline-based model has been developed only for route prediction (not RFL).

#### *5.3.2.1. Airline-based model: construction of cluster variables*

Cluster variables are dependent on the route under study. A simple route characteristic, e.g., ground distance, cannot provide information to the model by itself, as it is actually the distance difference with other available routes what is relevant for the route choice problem. In other words, the model needs a reference.

It is also important to highlight that route variability in the FPLs is relatively low. Around 80% of the flights of an airline for a given OD pair follow the same route, i.e., airlines tend to consistently choose the same route and select a different one only under specific conditions. It thus seems logical to take the most flown route as reference. For each AIRAC cycle, we have considered as a reference route the most flown in the previous cycle. For example, if the ground distance of a route is 1,000 km and the ground distance of the most flown route is 1,100 km, the reference value for the first route will be -100 km.

The cluster variables considered in the model are described below:

- **Ground distance:** it is calculated by summing the projected ground length of the different segments composing the route (waypoints closer than 40 NM to the origin and destination airports have been discarded).

[2] https://cds.climate.copernicus.eu

- **Air distance:** it is calculated by adjusting the ground distance with the wind extension. The wind extension is calculated using the average wind projected along each segment of the flight path (weighted by the segment length) for each central route and multiplying this average wind by the central route flight duration.
- **Fuel consumption:** air distance is used as a basis to calculate fuel consumption. Although climbs and descents are typically longer than 40NM, this research assumes that the computed air distance is entirely flown in cruise conditions. Under this assumption, fuel consumption can be approximated by multiplying the air distance by the typical economic cruise fuel consumption. The typical economic cruise consumption for the Boeing 737-800, obtained from Boeing[3], has been taken as a reference value; for other aircraft models, fuel consumption has been assumed to vary linearly with the maximum MTOW.
- **Fuel cost:** kerosene, the standard fuel in commercial aviation, presents a high volatility in its price. This research assumes that the airline is calculating fuel cost according to the actual (spot) price. Therefore, fuel cost is estimated according to daily kerosene price multiplied by the fuel consumption.
- **Route charges:** AUs pay different charges to cover different ATM services. These charges can be airport charges or route charges. As origin and destination airports are already fixed for the prediction, the only possible differences are in the route charges. European route charges are calculated according to the entry and exit points in the different national airspaces that the flight navigates in and they are adjusted monthly.
- **Direct cost:** this variable aggregates the charges and the fuel cost.
- **Convective phenomena:** convective phenomena features are calculated along the central routes in the same way as for the OD pair-based model.
- **Local wind at origin/destination airport:** it is extracted from the origin and destination airports METAR files for the expected departure and arrival time. The effect of this variable was not expected to be seen clearly in all OD pairs, as it appears to be related with those cases in which arrival/departure points are rather separated in the terminal area, the ground distances are almost equally large for both options, and the convenience of using one of them depends on the airport configuration.
- **Military zones:** the European ATM system works under the flexible use of airspace (FUA) concept, which means that airspace is no longer designated as purely “civil” or “military” and any necessary segregation is temporary, based on real-time usage within a specific time period. The airspace information included in the DDR repository contains the geographic description of the different military zones in Europe. Yet, it does not include the schedule of activation/deactivation of these zones, so this activation had to be estimated based on the observed traffic. Once the closure of military zones was estimated, each of the available routes is intersected with the active military zones at each given time and they are discarded as an option if any of the crossed military zones was active.

#### *5.3.2.1. Airline-based model: construction of general variables*

General variables are those which do not depend on the route cluster. Therefore, they do not need a reference.

The variables used are described below:

[3] http://www.boeing.com/-assets/pdf/commercial/startup/pdf/737ng_perf.pdf

- **Time of flight, DoY and MTOW:** the variable construction process is identical to the one implemented for the OD pair-based model.
- **Day of week:** it is broadly accepted that air traffic has a strong weekly component. The DoW has been used in two ways:
  - Model feature: an integer number from 0-Monday to 6-Sunday, similarly to the OD pair-based model.
  - Route filter: routes only flown during weekdays were not considered on weekends and vice versa.
- **Flight direction:** airline behaviour is not expected to be uniform for all geographies. Flight direction is composed by two variables, the geodesic longitude difference between the origin and destination airports and the latitude difference. For example, the flight direction for the OD pair Roma Fiumicino (LIRF) – Amsterdam Schiphol (EHAM) is (-10.51, 7.47).
- **Airport socioeconomic variables:** this variable considers the local population and GDP in the origin and destination airports as proxies for the amount of business trips.
- **OD pair competition:** two proxy variables are considered: the OD pair frequency (computed as the number of flights) and the share of flights for each particular airline. Additionally, two Boolean variables have been created to indicate if the origin airport or the destination airport is a hub for the airline.

## 5.4. Model training and validation

This step is intended to train and evaluate the machine learning models used in each case.

### 5.4.1. OD pair-based model

Models have been independently developed for route and RFL prediction.

Model training and validation are similar for route and RFL prediction. Nevertheless, the approach proposed is slightly different depending on the use of the basic or the enhanced model, so they are discussed separately.

#### *5.4.1.1. Basic model*

The most common machine learning approaches to trajectory prediction found in the scientific literature (linear regression, decision trees, random forests, support vector machines) were tested for a subset of OD pairs. These tests show that the Random Forest algorithm provides better results for the basic model, both for route and RFL.

#### *5.4.1.2. Enhanced model*

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. To avoid these problems, recursive feature elimination (RFE) has been used to automatically reduce the feature set to the most relevant. RFE is a method used for feature selection that fits a model and removes the weakest feature (or features) until the specified number of features is reached. Features are ranked by feature importance.

The four algorithms most commonly encountered in the literature have been explored: Multinomial Logistic Regression, Decision Trees, Random Forest and Support Vector Machines (SVM).

### 5.4.2. Airline-based model

Regarding the training of the airline-based models, there are two relevant aspects to be analysed:

- **Temporal scope**: considering the relatively high observations/features ratio available for the airline-based model (in comparison to the OD pair-based), the hypothesis is that not all the AIRAC cycles will contribute equally to the model performance.
- **Machine learning algorithm selection:** this analysis, as in the OD pair-based model, will select the most appropriate algorithm for the airline-based model.

We have analysed the performance obtained for a single airline, KLM, in order to select the most appropriate temporal scope for model training and the best performing algorithms. The selected airline is KLM. The reason is that KLM has a significant number of flights with heterogeneous characteristics (length, zones, schedules, etc.), enabling the exploration of a wide range of situations.

All short and medium range flights of KLM with origin and destination inside the ECAC area have been considered. Flights larger than 5,000 km have been discarded as they involve information that is not available for the experiments (navigation charges, airspace structure, etc. outside ECAC). To select the best temporal scope for model training, different sub set of the AIRACs 1802-1812 have been used to train a decision tree model and AIRAC 1813 has been chosen as the validation dataset.

Table 1 shows the results obtained with the decision tree model using different sets of AIRAC cycles for model training. These results show that model accuracy does not increase consistently with the number of AIRAC cycles used for training. The explanation to this behaviour seems to be related with the airline's winter/summer seasonal strategies. Our hypothesis is that airline behaviour is different in each season, so the performance is better when training only with AIRAC data from the same season as the testing dataset.

*Table 1:* Airline-based model results for KLM flights: influence of the AIRAC cycles selected for training. The training datasets providing the best performance are highlighted in bold font.

| Model ID | Training AIRACs | Validation AIRACs | Accuracy |
|---|---|---|---|
| 1 | 1812 | 1813 | 0.814 |
| 2 | 1810-1812 | 1813 | 0.831 |
| 3 | 1807-1812 | 1813 | 0.834 |
| **4** | **1802-1812** | **1813** | **0.852** |
| 5 | 1802,1811,1812 | 1813 | 0.849 |
| **6** | **1802,1803, 1811,1812** | **1813** | **0.854** |
| 7 | 1802,1803, 1804,1810, 1811, 1812 | 1813 | 0.844 |
| **8** | **1802,1803, 1804,1811,1812** | **1813** | **0.860** |

The optimal training data set (AIRACs 1802, 1803, 1811 and 1812) has been used to train different machines learning algorithms. The prediction accuracy obtained is shown in Table 2 below.

*Table 2*: Airline-based model results for KLM flights: comparison of different machine learning algorithms. The best performing algorithms are highlighted in bold font.

| **Algorithm** | **Accuracy** |
|---|---|
| Logistic regression | 0.807 |
| **Decision tree** | **0.854** |
| **Random forest** | **0.879** |
| Support vector machine | 0.829 |

### 5.4.3. Benchmark model: PREDICT

In order to evaluate the performance of the proposed models, their accuracy has been compared against that of PREDICT, the tool currently used by the NM. The functioning of PREDICT has been emulated following the information available from the NM documentation and the indications from EUROCONTROL experts. For each flight, the following workflow has been 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.

# 6. Description of the data the study relies on

The data used for both models (OD pair-based and airline-based) has been obtained from EUROCONTROL's DDR and other external data sources.

## 6.2. DDR data

FPL data from AIRAC cycles 1801 to 1813 and 1906 to 2002 has been used. Additionally, the DDR provides the required sectors, military zones, route charges, airport locations, and regulations information.

## 6.2. External data sources

The following external data sources have been considered:

- CDS provides geospatial weather information contained in different products. The ERA5 data product has been used. ERA5 data contains dozens of weather variables, particularly wind and severe weather variables, among others.
- The IOWA MESONET[4] provides access to the airports METAR files. METAR files contain an historic log of the airport's meteorological station.
- Gross Domestic product has been obtained using the gridded dataset provided by [14], which combines national and regional data and is provided with 0.5 geodesic degree resolution.
- Population density has been obtained from the NASA Socioeconomic Data and Applications Center (SEDAC)[5]. The data is based on counts consistent with national censuses and

[4] https://mesonet.agron.iastate.edu/
[5] https://sedac.ciesin.columbia.edu/data/set/gpw-v4-population-density-adjusted-to-2015-unwpp-country-totalsrev11

population registers with respect to relative spatial distribution and it is also provided with 0.5 geodesic degrees resolution.

- Kerosene daily prices are extracted from the Federal Reserve Economic Data (FRED)[6].

# 7. Computational experiments

## 7.1. OD pair-based model experiments

The selected training dataset comprises 16,174 OD pairs. Nevertheless, not all the pairs were suitable to generate a machine learning model:

- 1,744 OD pairs have only one route.
- 1,914 OD pairs do not have observations on the testing dataset, so it is not possible to analyse the model performance.
- 1,709 OD pairs did not provide enough observations (a threshold of 50 observed flights was established).

The OD pair-based models have been developed and tested for the remaining 10,807 OD pairs, which account for around 90% of European flights. Four independent machine learning models for each OD pair have been generated (route-basic, route-enhanced, RFL-basic, and RFL-enhanced).

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 route cluster or RFL matches the one actually observed.
- The global accuracy is defined as the number of correct guesses divided by the number of total flights.
- The combined prediction is considered correct when the predicted route and the predicted RFL are both correct for the same flight.

## 7.2. Airline-based model experiments

The validation of the airline-based model has been performed using two different datasets, covering AIRACs 1801-1813 and 1907-2002.

As the airline model does not impose any requirement regarding data availability, all flights observed in the testing dataset whose flight distance is below 5,000 km have been considered. Following the initial analysis performed for KLM flights, the selected machine learning algorithms are the decision tree and the random forest. Results have been evaluated using as primary metric the accuracy of the system, which is computed according to the same principles as for the OD pair-based model.

The airline-based model generates an independent machine learning model for each airline. In practise, this means that it is necessary to define a list of airlines to be modelled. According to the data analysed (most recent AIRAC cycles before COVID-19 outbreak), the number of unique airlines is over 2,465 nevertheless, the flight sharing is clearly unequal.

The airlines' flight sharing distribution reveals that 10% of the airlines account for more than 95% of the flights. As the number of flights performed by one airline decreases, there is an increasing risk that the number of observations is too low for the proper training of a machine learning model. Therefore, we have developed separate models for the 200 airlines with the highest number of

[6] https://fredhelp.stlouisfed.org/fred/about/about-fred/what-is-fred/

flights, and an additional model for the remaining airlines, which have been grouped into a single "low volume airlines group" identified with the fictitious code "AAA".

# 8. Results

## 8.1. OD pair-based model results

As previously explained, separate models have been developed for the prediction of the route and the RFL. Nevertheless, the ultimate goal is to predict both of them correctly. Therefore, in terms of the added value of the models what is relevant is to evaluate how the two models behave together. Table 1 shows the evaluation of the route prediction models, the RFL prediction models, and the combined models, and the improvement obtained with respect to the PREDICT system.

Table 1: Accuracy of the OD pair-based route prediction model, RFL prediction model and combined models: comparison with PREDICT

| | Accuracy | | | | |
|---|---|---|---|---|---|
| **Model** | **PREDICT** | **Basic** | | **Enhanced** | |
| | | **Value** | **Improvement** | **Value** | **Improvement** |
| 2D route | 79.8% | 80.2% | 0.5% | 81.5% | 2.0% |
| RFL | 58.1% | 59.8% | 2.9% | 61.8% | 5.9% |
| Combined | 49.6% | 50.8% | 2.3% | 52.7% | 6.2% |

The performance comparison against the PREDICT tool is quite satisfactory. The *basic model* provides a 2.3% increment on accuracy, while the enhanced model achieves a 6.2% improvement.

These results are also presented by OD pair in Figure 2. The fraction of OD pairs achieving a higher accuracy than PREDICT is 54.5%, 28.8% of them perform worse than PREDICT, and for the remaining 16.7% the performance is equivalent.

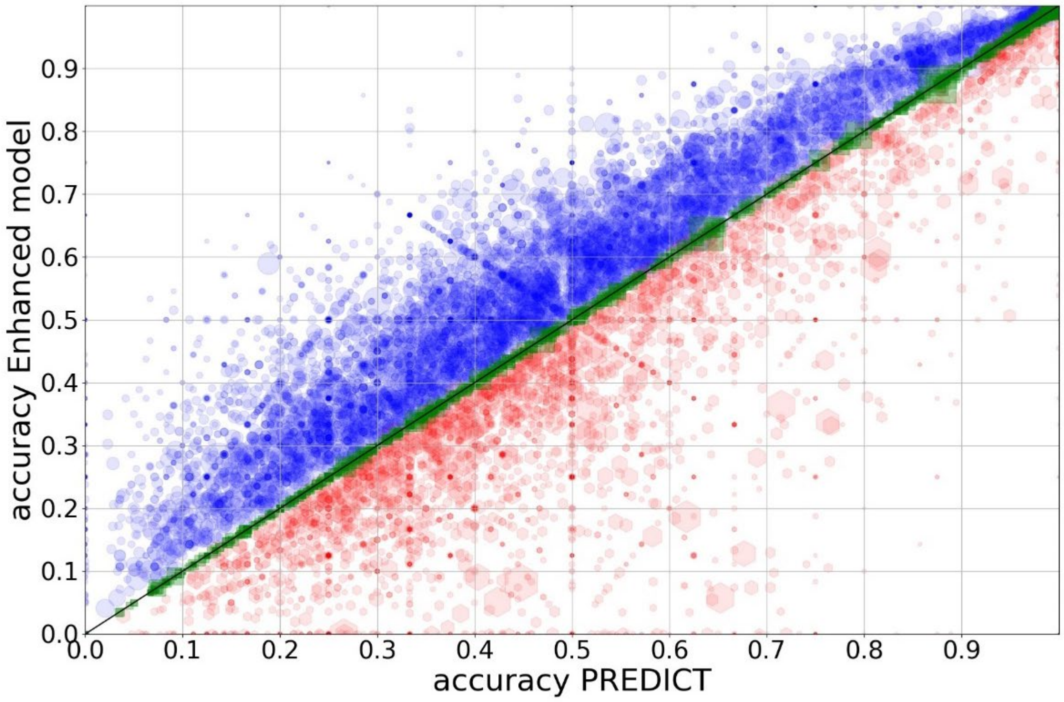


Figure 2: Accuracy of the combined enhanced models by OD pair: comparison with PREDICT. Each point represents an OD pair; the size of the point represents the number of flights

## 8.2. Airline-based model results

Table 2 shows the results provided by the airline-based mode. Accuracy is higher for the random forest algorithm. Both random forest models provide similar results, being the metrics slightly better for model R_2002. The improvement achieved is significantly higher than that provided by the OD pair-based model.

Table 2: Full ECAC airline-based model results: comparison with PREDICT

| Model ID | Training AIRACs | Validation AIRACs | Number of pairs | PREDICT accuracy | Airline-based model accuracy | Improvement |
|---|---|---|---|---|---|---|
| R_1813 (r. forest) | 1802,1803, 1804,1811,1812 | 1813 | 10,369 | 0.825 | 0.892 | 8.1% |
| R_2002 (r. forest) | 1911,1912, 1913,2001 | 2002 | 9,794 | 0.828 | 0.896 | 8.2% |
| T_1813 (tree) | 1802,1803, 1804,1811,1812 | 1813 | 10,369 | 0.825 | 0.883 | 7.0% |
| T_2002 (tree) | 1911,1912, 1913,2001 | 2002 | 9,794 | 0.828 | 0.888 | 7.2% |

Figure 3 compares the results of the AIRAC 1813 random forest model against PREDICT by OD pair. The OD pairs performing worse than PREDICT are a minority (6%) and they also present smaller differences in accuracy (i.e., the red hexagons are closer to the bisection than the blue circles).

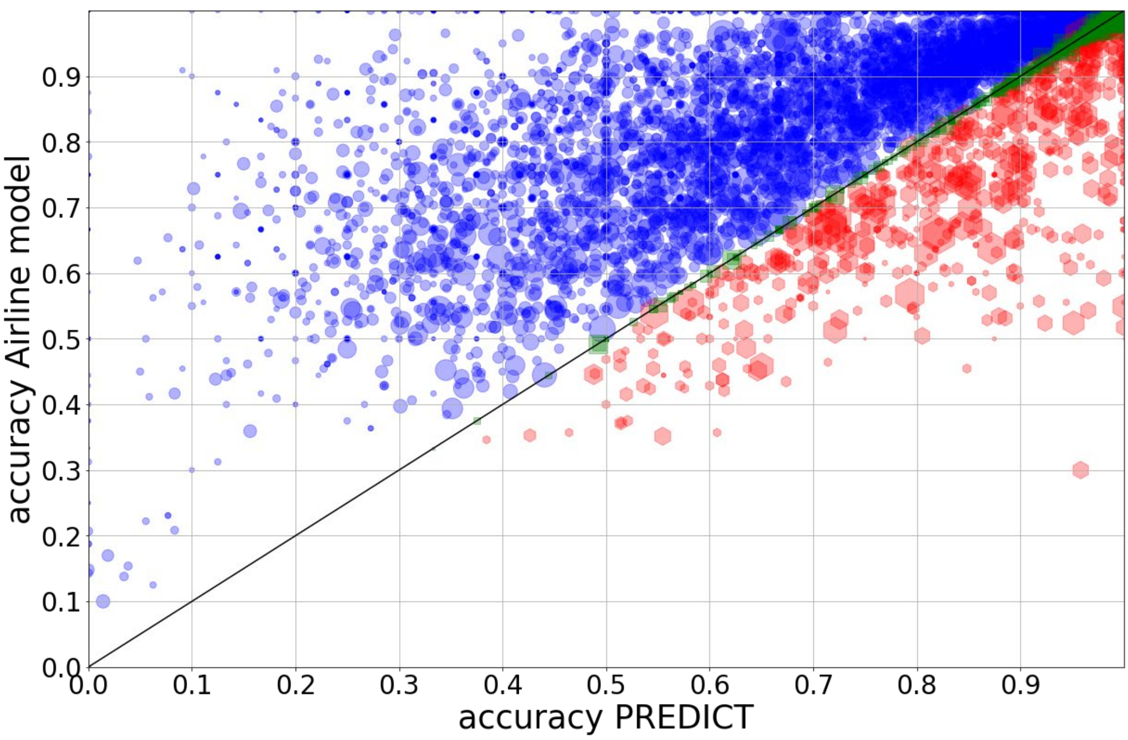


Figure 3: Accuracy of the airline-based model R_1813 by OD pair: comparison with PREDICT. Each point represents an OD pair; the size of the point represents the number of flights

# 9. Analysis of the results

## 9.1. OD pair-based model: feature analysis

The inclusion of new variables in the OD pair-based enhanced model required the application of the RFE technique to keep a reasonable number of variables and avoid overfitting. This approach has achieved a significant improvement in the models.

A detailed analysis has been performed for a subset of OD pairs in order to derive additional insights on the different features considered by the model. To ensure a proper representation of the whole network, a selection of several representative OD pairs has been studied: LTAI-EDDK, EDDT-LEPA, LGAV-LFPG, EHAM-LIRF, LPPT-LFPO and UUEE-EDDF. The main conclusions from the analysis are summarised below:

- RFE leads to picking different variables for each OD pair (as expected).
- Local wind variables seem to be relevant in most cases, in particular for the destination airport.
- Convective event variables are also relevant in all OD pairs. These variables represent more than half of the RFE-selected variables for almost every pair.
- En-route wind seems to be relevant in general terms, although the effect is more relevant for certain pairs.
- Regulation-based variables appear to be less relevant.

## 9.2. OD pair-based model limitations and proposed solution: Bollinger Bands

Although the OD pair-based models provide some improvement with respect to the current PREDICT tool, the proposed approach still presents a major drawback: performance improvement is inconsistent across OD pairs.

A first analysis of the OD pairs showing poor performance revealed that this behaviour seems to be related with sudden changes in the usual selections (route or RFL). Conceptually, a sudden change may justify a drop in the machine learning model performance: for example, if an AU starts using a new route that has not been observed in the training dataset, the machine learning model will not be able to predict such route because it has not been observed, while the PREDICT tool would only fail during the first week of the AIRAC cycle, since PREDICT will just select the route from the previous week. The analysis revealed that PREDICT was performing better than the Enhanced model in those OD pairs showing an anomaly during the last week of the training dataset. Following this analysis, an alarm system was implemented to use PREDICT instead of the Enhanced model when an anomaly was detected.

If there is an observable cause for the machine learning models to underperform PREDICT, corrective measures can be implemented. A possible approach is using the so-called Bollinger Bands. Bollinger Bands or trading bands is a common technique used in stock pricing analysis. This technique is based on the use of a moving average and the standard deviation to establish a moving confident interval for time series. When the price goes beyond the bands, it is considered to have a relevant growing/decreasing momentum and therefore, it theoretically indicates an adequate time to buy/sell. To analyse the changes in the FPL, the concept of "cluster share" has been defined. The cluster share is defined for each OD and for each route cluster or RFL as the number of flights using this route cluster (or RFL) in a week divided by the total number of flights during that week. Treating the cluster share as a time series enables the application of the Bollinger Bands to detect potential anomalies. A very simple example is shown in Figure 4.

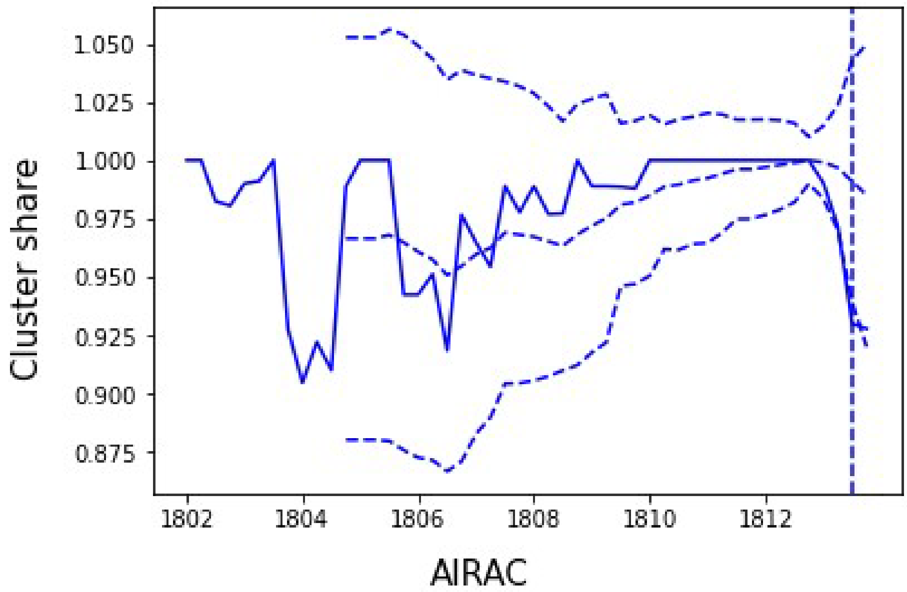


Figure 4: Application of the Bollinger Bands anomaly detection to the route 0 cluster share in the OD pair EDDT-LEPA. Full line represents the cluster share, doted lines represent the bands and the vertical dotted line marks an anomaly detection

Table 3 results demonstrate that the implementation of a Bollinger Bands alarm system can help improve the FPL prediction accuracy. The global increment on accuracy is almost uniform for the route models, the RFL models and the combined models. The combined enhanced model achieves a 7.2% accuracy increment against PREDICT when the alarm system is included. Figure 5 illustrates the changes achieved by the Bollinger Bands alarm system in the combined enhanced model. The effect of the alarm system is clearly visible. Nevertheless, there are still some cases which have not improved with this system.

Table 3: Enhanced models results with and without the Bollinger Bands alarm system

| | Accuracy | | | | |
|---|---|---|---|---|---|
| **Model** | **PREDICT** | **Enhanced model** | | **Enhanced model with Bollinger Bands** | |
| | | **Value** | **Improvement** | **Value** | **Improvement** |
| 2D route | 79.8% | 81.5% | 2.0% | 81.8% | 2.5% |
| RFL | 58.1% | 61.8% | 5.9% | 62.1% | 6.3% |
| Combined | 49.6% | 52.7% | 6.2% | 53.2% | 7.2% |

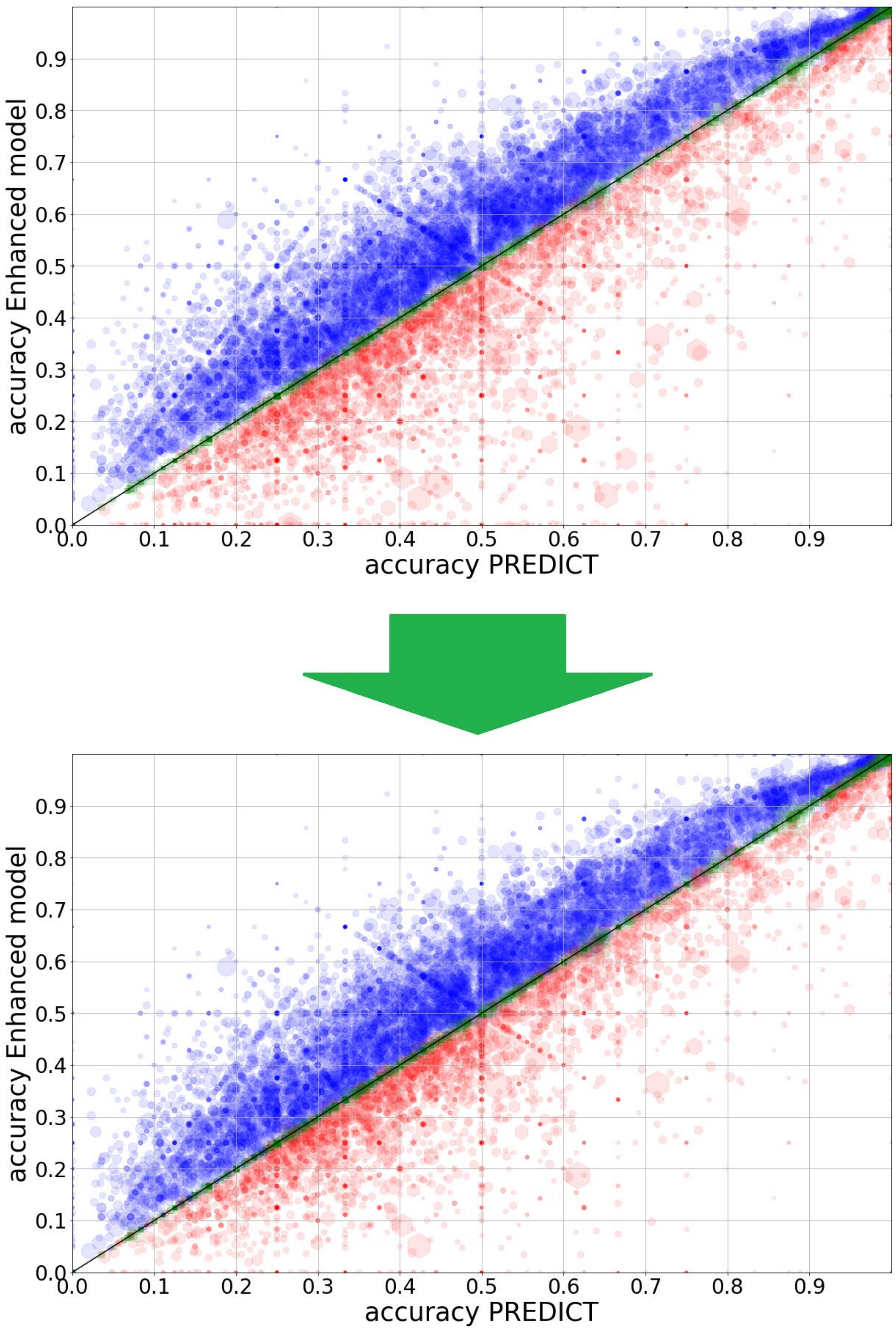


*Figure 5: Evolution of the combined enhanced ML models accuracy plots with the application of the Bollinger Bands system. Each point represents an OD pair, the size of the point represents the number of flights*

## 9.3. Airline-based model: non-observed routes

One of the key improvements brought by the airline-based modelling approach is the capability to predict new routes not previously observed in the training set: since it is not necessary to include such routes in the training, the airline-based model is capable of calculating the probability of flying any new route just by deriving its features. While these cases are quite rare (0.2% of the flights), EUROCONTROL experts have shown a particular interest in them.

To exemplify this feature, we have chosen the OD pair connecting Kristiansand, Norway, and Amsterdam (ENCN-EHAM). This OD pair shows a new route in AIRAC 1813 that has not been flown previously in the training dataset. This new route, shown as Route 3 (in purple) in Figure 6, is used twice during AIRAC 1813.

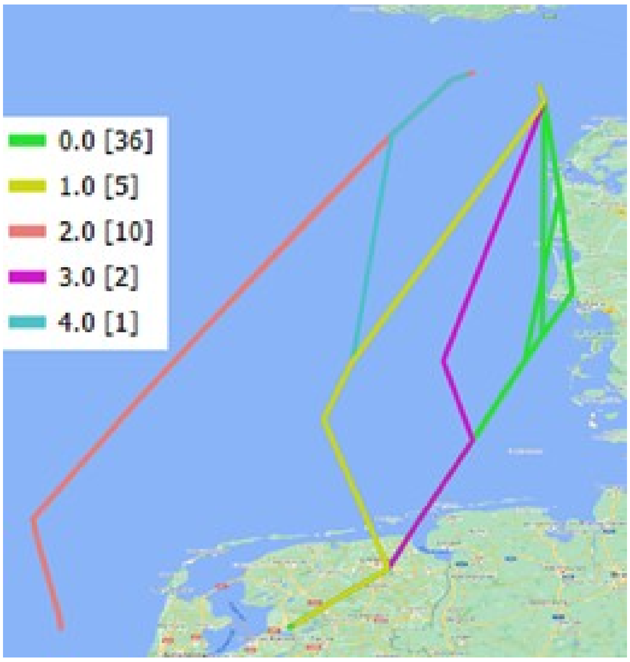


Figure 6: ENCN-EHAM OD pair routes for AIRAC 1813.
The number of times the route has been used is indicated in brackets.

Route 3 was correctly predicted by the model, while PREDICT does not forecast it correctly and the Enhanced model cannot even consider this prediction outcome. The accuracy for all the predicted flights for the ENCN-EHAM OD pair shows an outstanding performance (75.9%) in comparison with the route enhanced model (63.0%) and PREDICT (51.9%).

## 9.4. Airline-based model: feature analysis

The airline-based models have achieved a noticeable improvement in accuracy. As differences in accuracy between the decision tree and the random forest models are not very high and tree models directly provide feature importance, we have used the tree (T_1813) to perform the feature analysis.

The analysis has been performed using the feature importance for each airline model. The feature importance is a normalised value that reflects the weight of each variable in the model decision. The analysis of features importance reveals some interesting insights:

- Most airline models seem to be mainly driven by the direct cost or the ground distance.
- The air distance, the charges, the fuel consumption and the fuel cost are the dominant variable for certain airlines.

- The rest of the features (e.g., MTOW, sin of hour, wind factor, etc.) show lower feature importance values.

### 9.5. OD pair-based and airline-based model comparison

It is important to highlight that the predictions made with the OD pair-based and the airline-based models do not cover the same flights. This is due to the intrinsic limitations of each model (e.g., the airline model does not consider OD pairs over 5,000 km and the OD pair model does not consider pairs with less than 50 flights in the training dataset). Therefore, the comparison of both models needs to ensure that the data used cover exactly the same flights. The intersection covers 9,301 OD pairs, which is a reasonable proportion of the pairs considered in the airline model (10,369).

The comparison is summarised in Table 4. As expected, the improvement of the airline model against PREDICT is much more significant than the improvement of the OD pair-based model against PREDICT.

Table 4: Comparison between the airline-based model, the OD pair-based model, and PREDICT

| Number of pairs | PREDICT accuracy | OD pair enhanced model accuracy | Airline model accuracy (R_1813) |
|---|---|---|---|
| 9,301 | 0.822 | 0.834 | 0.884 |

## 10. Conclusions and look ahead

### 10.1. Summary of contributions of this PhD

The main contributions of this PhD thesis are summarised below:

- We have implemented and validated a new clustering metric based on the area between routes that increases computational efficiency, thus enhancing the scalability of the tool.
- The OD pair-based model has proven that the PREDICT tool can be improved by introducing a machine learning algorithm, without the use of any external variable. Additionally, it has demonstrated that the inclusion of external variables yields higher accuracy, encouraging the introduction of such data sources in operation.
- The effect of ATFCM regulations, which was expected to play a major role in the AU decision making process, has shown a negligible effect on the experiments performed. After a detailed analysis with some experts, regulations have been discarded from the machine learning models. It is important to highlight that the conclusions regarding the irrelevance of the ATFCM regulations are not extensible to other phases of the flight (e.g., tactical ATFCM and operations) where they are expected to play an important role.
- The experiments performed with the airline model show a significant improvement of the prediction accuracy with respect to PREDICT, from the 83% accuracy shown by the PREDICT tool to more than 89%. In practical terms, this improvement means that more than one out of three flights currently erroneously predicted could be correctly predicted using the newly developed models.
- A common shortcoming of recent research work in this field is the lack of a thorough scalability analysis. A pre-tactical FPLs prediction system is intended to predict the flights of an entire network, such as the ECAC area for European ATFCM, to facilitate resource allocation

and planning. To the best of our knowledge, there is no previous work that analyses the applicability of their solutions in this context. This research has proven to develop a solution that improves the current system and it can be applied to the whole network.

## 10.2. Future research

The following elements could potentially help increase the performance of the solution developed in the PhD thesis:

- A higher number of scenarios should be tested in the future to validate the models proposed. It is especially interesting to validate the observed trend regarding the use of training data from the same season. It would be interesting to perform a continuous analysis, validating the model over all the AIRAC cycles throughout a year.
- The proposed airline-based models could be significantly improved if they could be fed with certain airline-related latent variables that are usually not accessible because they are business sensitive (e.g., TOW, thrust settings, cost index, etc.). Machine learning could also be of help for the estimation and prediction of these variables.
- Other machine learning and deep learning algorithms could be explored. The random forest has provided significant accuracy improvement and it is computationally efficient. Nevertheless, other algorithms such as neural networks might improve the prediction performance.
- The experiments performed have been based on the ability of the models to predict individual trajectories. Future research should consider the aggregation of the trajectories in order to compute error compensations and network effects.
- The present thesis focuses on the pre-tactical phase. The proposed solution could be adapted to the tactical phase with some minor adjustments.
- The experiments performed can only provide a glimpse of the improvement reachable by these models. The correct evaluation of the proposed solution should be tested in an operational environment (in shadow mode, for instance) to accurately know their actual impact.

# 11. References

## 11.1 Link to PhD thesis / repository

The dissertation will be available through these public repositories:

https://www.tesisenred.net/

https://upcommons.upc.edu/

## 11.2 Associated outputs and publications


- Mateos, M., Martín, I., García Cantú-Ros, O., & Prats, X. (2022). Full-scale pre-tactical trajectory prediction: Machine Learning to increase pre-tactical demand forecast accuracy. Submitted to IEEE: Transactions on Intelligent Transport Systems.
- Mateos, M., Martín, I., Alcolea, R., Herranz, R., García Cantú-Ros, O., & Prats, X. (2021). Unveiling airline preferences for pre-tactical route forecast through machine learning. An innovative system for ATFCM pre-tactical planning support. In Proceedings of the 11th SESAR Innovation Days.

- Mateos, M., Martín, I., Herranz, R., García Cantú-Ros, O., & Prats, X. (2020). Predicting requested flight levels with machine learning. In Proceedings of the 10th SESAR Innovation Days.
- Mateos, M., Martín, I., García, P., Herranz, R., García Cantú-Ros, O., & Prats, X. (2020). Full-scale pre-tactical route prediction. In: 9th International Conference for Research in Air Transportation (ICRAT).


## 11.3 References cited in this report


[1] EUROCONTROL, “ATFCM operations manual. edition number: 22.1”, EUROCONTROL, Bretigny, France, Tech. Rep., 2018.

[2] Gallego, C. E. V., Comendador, V. F. G., Nieto, F. J. S., Imaz, G. O., and Valdés, R. M. A. (2018). Analysis of air traffic control operational impact on aircraft vertical profiles supported by machine learning. Transportation research part C: emerging technologies, 95, 883-903.

[3] Y. Liu, M. Hansen, D. J. Lovell, and M. O. Ball, “Predicting aircraft tra-jectory choice–a nominal route approach”, inProc. of the InternationalConference for Research in Air Transportation, 2018.

[4] Tastambekov, K., Puechmorel, S., Delahaye, D., & Rabut, C. (2014). Aircraft trajectory forecasting using local functional regression in Sobolev space. Transportation research part C: emerging technologies, 39, 1-22.

[5] Yang, Y. (2017). Practical method for 4-dimentional strategic air traffic management problem with convective weather uncertainty. IEEE transactions on intelligent transportation systems, 19(6), 1697-1708.

[6] H. Naessens, T. Philip, M. Piatek, K. Schippers, and R. Parys, “Predicting flight routes with a deep neural network in the operational air trafficflow and capacity management system”, EUROCONTROL Maastricht Upper Area Control Centre, Maastricht Airport, The Netherlands, Tech.Rep, 2017.

[7] Dalmau Codina, R. (2019). Optimal trajectory management for aircraft descent operations subject to time constraints. Ph.D. thesis. 18

[8] Prats Menéndez, X., Dalmau Codina, R., & Barrado Muxí, C. (2019). Identifying the sources of flight inefficiency from historical aircraft trajectories. A set of distance-and fuel-based performance indicators for post-operational analysis. In Proceedings of the 13th USA/Europe Air Traffic Management Research and Development Seminar.

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

[10] Marcos, R., Ros, O. G. C., & Herranz, R. (2017). Combining Visual Analytics and Machine Learning for Route Choice Prediction. 7th SESAR Innocation Days, Beograd.

[11] Ayhan, S., & Samet, H. (2016, August). Aircraft trajectory prediction made easy with predictive analytics. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 21-30).

[12] Ayhan, S., & Samet, H. (2016, October). Time series clustering of weather observations in predicting climb phase of aircraft trajectories. In Proceedings of the 9th ACM SIGSPATIAL International Workshop on Computational Transportation Science (pp. 25-30).

[13] Wang, Z., Liang, M., & Delahaye, D. (2018). A hybrid machine learning model for short-term estimated time of arrival prediction in terminal manoeuvring area. Transportation Research Part C: Emerging Technologies, 95, 280-294.

[14] Kummu, M., Taka, M., & Guillaume, J. H. (2018). Gridded global datasets for gross domestic product and Human Development Index over 1990–2015. Scientific data, 5(1), 1-15.

# Annex I: Acronyms

| Term | Definition |
|---|---|
| ANSP | Air Navigation Service Provider |
| APM | Aircraft Performance Model |
| ASM | Airspace Management |
| ATC | Air Traffic Control |
| ATCO | Air Traffic Control Officer |
| ATFCM | Air traffic Flow and Capacity Management |
| ATFM | Air Traffic Flow Management |
| ATM | Air Traffic Management |
| ATS | Air Traffic services |
| AU | Airspace User |
| CDM | Collaborative Decision Making |
| CI | Cost Index |
| DBSCAN | Density-Based Spatial Clustering of Applications with Noise |
| DCB | Demand and Capacity Balancing |
| DDR | Demand Data Repository |
| DoW | Day of Week |
| DoY | Day of Year |
| ECAC | European Civil Aviation Conference |
| ETFMS | Enhanced Tactical Flow Management System |
| FI | Flight Intention |
| FID | Flight Identifier |
| FPL | Flight Plan |
| FUA | Flexible Use of Airspace |
| GDP | Gross Domestic Product |
| ML | Machine Learning |
| MTOW | Maximum Take-Off Weight |
| NASA | National Aeronautics and Space Administration |
| NM | Network Manager |
| NMOC | Network Manager Operations Centre |

| Term | Definition |
|---|---|
| NOP | Network Operations Plan |
| OD | Origin-Destination |
| RFE | Recursive Feature Elimination |
| RFL | Requested Flight Level |
| RMSE | Root-Mean-Square Error |
| SSPD | Symmetrised Segment Path Distance |
| SVM | Support Vector Machine |
| TBO | Trajectory Based Operations |
| TOW | Take-off Weight |
| UPC | Technical University of Catalonia |