

SESAR Engage KTN – catalyst fund project final technical report

Project title:	Probabilistic information Integration in Uncertain data processing for Trajectory Prediction ('PIU4TP')
Coordinator:	Centro Italiano Ricerche Aerospaziali (CIRA)
Consortium partners:	-
Thematic challenge:	TC2 Data-driven trajectory prediction
Edition date:	22 July 2021
Edition:	1.0
Dissemination level:	Public
Authors:	Gianpaolo Romano / CIRA
	Antonio Vitale / CIRA
	Edoardo Filippone / CIRA
	Gaetano Zazzaro / CIRA
	Francesco Martone / CIRA

The opinions expressed herein reflect the authors' views only. Under no circumstances shall the SESAR Joint Undertaking be responsible for any use that may be made of the information contained herein.



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 and executive summary

1.1 Abstract

The objective of the PIU4TP project is the development of a data-driven methodology for the trajectory prediction from long to short term before scheduled time of flight. Specifically, the methodology uses machine learning and data mining techniques to perform data analysis and to learn from past experience the aircraft future behaviour in terms of flight path selection. Therefore, it exploits historical data and uncertainties of current forecasts of some relevant mission and aircraft parameters to compute trajectory prediction outcomes enriched with associated probabilistic information. The project's final aim is to build a methodology that can support the Network Manager with air traffic flow and capacity management, allowing the optimization of flight distribution among sectors and flight routes, the anticipation of air traffic flow requests and the identification in advance of potential conflicts.

1.2 Executive summary

The PIU4TP project aims to develop a data - driven methodology named P4T (Prediction for Trajectory) for the prediction of the flight trajectory in terms of selection of the most likely sequence of waypoints in the strategic and pre-tactical phases, starting weeks before the flight execution with the declared flight intention of the airspace users and ending few hours before the estimated off blocks time.

The tactical management of an ordered, efficient and safe air traffic is currently highly affected by a number of uncertainties, which will finally require many modifications to flight plans and can produce relevant delays on the schedule of flights.

The P4T methodology aims to investigate how flight plans can be better predicted, from long to short term before scheduled time of flight, by considering historical data and uncertainties on current forecasts of some relevant parameters. This methodology, providing reliable predictions of flight plans, is expected to allow ATFCM centres to perform a sound management of the uncertainties affecting the air traffic and to limit changes to the plans in the tactical phase, so improving ATM efficiency, punctuality and reducing environmental impact. Safety will be also affected by limiting ATCO workload and reducing the risk of hotspots occurrence.

Actually, there is a large number of parameters that can affect the optimal flight plan selection. A few of these, among the most relevant ones, have been considered in the project, specifically the weather forecast and the estimated take-off weight. Indeed, the project is a proof of concept. Its objective is to consider some parameters that mainly affect the selection of the optimal flight plan and to investigate how the information about these parameters and related uncertainties, which characterize the parameters forecast before the flight, can be exploited in an integrated approach to perform in advance a reliable prediction of the flown trajectory. Although the developed methodology is demonstrated considering few uncertain inputs, it is generic and applicable also to a wider set of uncertain inputs. Indeed, all the steps that define the methodology developed in this project, described in section 2.3.2, can be applied to different use cases, which consider different sets of uncertain inputs (if the historical data and current prediction of these uncertain inputs are available). Obviously, the models obtained with the applied techniques need a new training, considering the new input variables, and this need for re-training is a primary requirement of Machine Learning, due to its data-driven character. This re-training step allows us to always maintain a high prediction accuracy.

The idea of P4T is to build the predictive model of flight trajectories by applying Data Mining and Machine Learning (ML) techniques. Instead of programming explicitly a computer to solve

a difficult problem, ML uses algorithms to learn from past experience (historical data) how to obtain behavioural models based on complex but statistically reliable rules. This model, once implemented, will use as input the weather forecast and take-off aircraft mass estimation, with related uncertainties. Indeed, the exploitation of the uncertainties on the inputs allows associating probabilistic information to the predicted trajectories and this is the main innovative feature of the proposed methodology. Therefore, PIU4TP represents a change of perspective. The project aimed to demonstrate that the **uncertainty** inherently present in a weather forecast and that also normally affects also the take-off weight data before the flight, could augment the knowledge base available to the Network Manager (NM) transforming the uncertainties in a valuable information for a more efficient flight trajectory planning and allocation.

The trajectory prediction capability can allow the network manager anticipating air traffic flow requests, supporting the decision-making process of flight distribution among sectors and flight routes, and thus avoiding an excessive tactical management of the flights. On this path, the project starts from very low TRL and moves from the concept to TRL 2, first designing a methodology for flight plan prediction and then performing a preliminary demonstration based on simulated data and a simplified use case. This allows evidence to be provided of the proposed methodology applicability, and potential benefits arising from its use.

2. Overview of catalyst project

2.1 Operational/technical context

Trajectory Prediction (TP) is one of the most relevant capability and need of the current and, above all, the future management of air traffic, in its expected implementation of the Trajectory Based Operations (TBO) paradigm. The TP process supports the activity of several ATM actors and Airspace Users (AU), which apply different tools and methodologies. The TP is performed iteratively from the initial planning till to the completion of the flight, as sketched in the Figure 1, to support operations in strategic, pre-tactical and tactical phases.

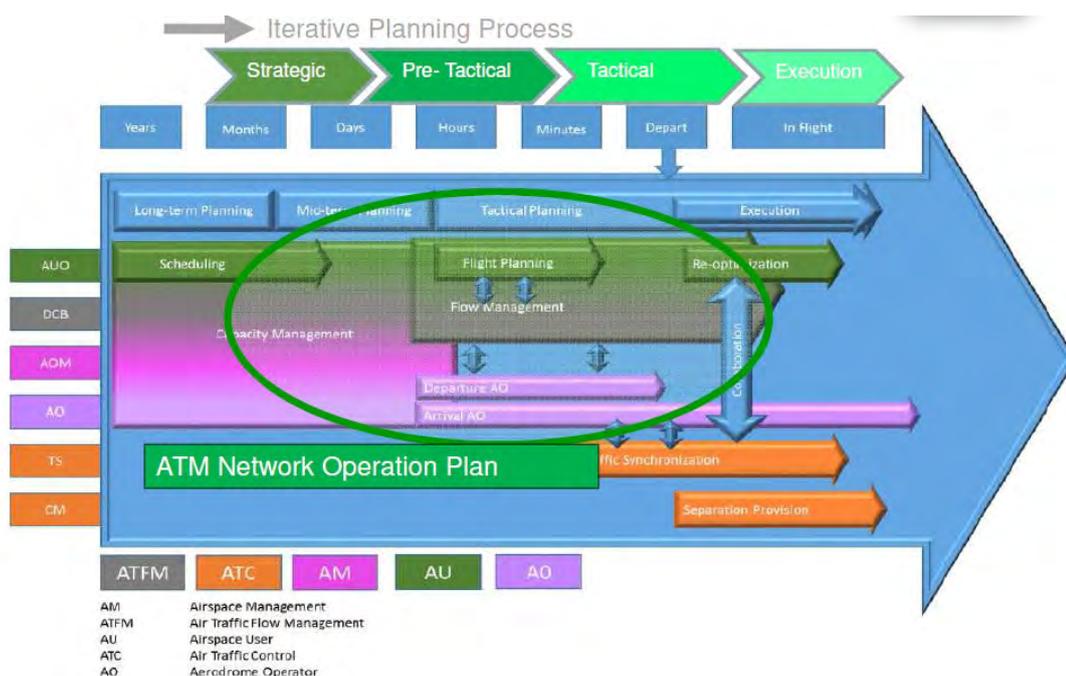


Figure 1 – Collaborative Trajectory Planning (Source: *Ballerini et al. “Trajectory Prediction Network Manager Demand Forecast Key Enablers to TBO”, Engage Workshop Nov. 2018*)

Often TP methodologies and tools provide deterministic forecast of the trajectories without any quantification of the uncertainty affecting the prediction, as highlighted in the Thematic Programme description. However, the TP process is uncertain by its nature, indeed it predicts actual trajectories by using models, which are approximation of the reality affected by a given accuracy, and uncertain input data, such as weather forecast and aircraft actual performance. The relevance of quantification and management of uncertainties on trajectory prediction has emerged in the recent years, and research activities are on-going on the topic [RD1], [RD2], [RD3], [RD4], [RD5], [RD6], [RD7].

The expected outcomes of the PIU4TP project is a methodology which allows using data mining and machine learning techniques for 4D Trajectory Prediction, dealing with the relevant stochastic information inherent to the input data, and providing the stochastic characterization related to the predicted trajectories. The methodology developed in the study is applicable in strategic and pre-tactical phases. It benefits from the progressively reduced level of uncertainties associated with the forecast to improve the trajectory prediction as the scheduled time of flight approaches. Indeed, in strategic forecast, starting from the filing of the flight plan, large uncertainties affect the input data for the forecast and thus the predicted trajectories have low level of confidence; progressing in time, while approaching the flight execution, the uncertainties on the required input data reduce, and improved trajectory predictions could be gotten as well as higher level of confidence associated to the predictions. Actually, the proposed methodology could be further developed and suitably customized as a Decision Support Tool in tactical management of the airspace. In fact, it could support ground decisions (on tactical clearances and trajectory distribution) and airborne decision (such as airborne delegated medium-term separation operations and manoeuvres). However, the development of the methodology for the application in the tactical phase is out of the scope of the present project.

In conclusion, the availability of reliable TP approach could support the improvement of the ATM system performance and the PIU4TP project aims contributing to fit this need. In fact, the integration of TP tool into the Network Manager's, ANSPs' and flight operations centres for 4D trajectory planning tools, leads to enhanced collaboration in trajectory management, such that capacity can be matched to demand by a better anticipation of AU behaviour. Moreover, improved predictability allows avoiding in advance potential conflicts and then enhancing air traffic safety.

2.2 Project scope and objectives

A lot of efforts have been done to develop TP algorithms that can meet the stringent safety requirements typical of the aviation sector. The traditional approach uses a more or less simplified dynamical model of the aircraft based on a number of parameters and then solve a set of differential equations to recover the flown trajectory, taking in consideration the influence of the surrounding weather conditions. This line of research has some drawbacks: the dynamical model can only reproduce a simplified version of the actual aircraft behaviour, the solution of the differential equations may be inaccurate or even unstable since many input parameters are difficult to be measured with sufficient precision, and many of the possible causes affecting the flight of the airplane cannot be adequately modelled, such as the intents of the pilot or airline operator and the directives of ATC. Moreover, the uncertainties on the parameters that influence the trajectory often are not considered at all and the value of these parameters are used simply to perform deterministic on-off decision. Finally, the use of model-based techniques to predict the actual flown trajectories requires high computational burden.

An increasingly interesting alternative to model-based solutions is offered by a data-driven approach: it uses a collection of past flown trajectories to statistically predict the behaviour of future flights by exploiting all the information implicitly included in the historical data. With the improving quality and growing volume of the data collected in ATC systems, data-driven methods have become mainstream in current aircraft trajectory prediction research and may allow overcoming the limitations of model-based approach.

The PIU4TP project aims to contribute to the research activities in the framework of the data-driven approach. Its objective is to define and validate a methodology that provides trajectory prediction enriched with its relevant probabilistic information, by exploiting the uncertainty inherently connected to the data used as inputs by the TP process. The proposed approach is consequently intended to support the planning activities in terms of demand-capacity balance, pre-tactical identification of hotspots and potential conflicts. However, it is out of the scope of the proposed methodology the provision of the best trajectories' allocation. Furthermore, the 4D-trajectories are considered, that is, the time is part of the information included in the output provided by the methodology.

The project responds to some of the main issues highlighted in the Trajectory Prediction Thematic Network description, that are:

- *use of machine-learning techniques to infer airspace users' behaviour, intentions and preferences from historical data and enhance tactical and pre-tactical trajectory prediction;*
- *aggregation of probabilistic predictions into probabilistic traffic counts at a strategic and pre-tactical level;*
- *integrating predictions about factors affecting flight planning and execution, including weather conditions.*

From a data-driven perspective, the objective of this project is to build a predictive model able to make short, medium and long-term predictions of trajectories given a set of uncertain inputs. Machine learning develops algorithms that learn from past experience how to obtain statistically optimal solutions. Much of the effort has been dedicated to the research of the best way to deal with the uncertainties in weather forecasts and aircraft take-off configuration.

The defined methodology has been designed and validated using simulated data. The choice to use simulated data is due to the lack of open access datasets which provide a huge (thousands of flights) and coherent set of data including real aircraft trajectory and related flight plan, actual aircraft take-off weight and its forecast at different time in advance with related forecast uncertainties, weather data experienced during flight and their forecast at different time in advance with related forecast uncertainties. Moreover, the simulated data allow testing the methodology in a fully controlled environment, that is, the value of the parameters of interest and the rules and the assumptions that lead to perform the flight along a specific flight plan are perfectly known. The project's purpose is to demonstrate that the methodology is able to catch all the useful information that are available in the data, including the forecast uncertainties, and to exploit them to perform in advance a reliable prediction of the flown trajectory. Once this result has been achieved on a simplified use case, the P4T methodology could be tested on more complex use cases (however it is not the objective of the present project), also considering actual data, if available. Indeed, P4T takes the form of a Lifecycle Model for the analysis and modelling of flight paths in the context of trajectory prediction. It is iterative and incremental since it allows to add new input variables (such as aircraft type, airline, variables related to passenger connections, restricted areas due to military or national security activity) and external parameters (such as new flight route, new time-frames, etc.) by iterating through the phases of the lifecycle.

2.3 Research carried out

The research activity has been structured into three phases, which performed in sequence allowed achieving the project objective:

Phase 1: Operational Scenario Definition

The first phase concerned the definition of relevant reference scenario, including the investigation about available data, the construction and tuning of models to generate the simulated data, and the realization of needed databases.

Phase 2: P4T Methodology Development

The second project phase dealt with the P4T methodology development. An overview of applied and applicable Machine Learning and Data Mining techniques to the problem of trajectory prediction have been analyzed and the methodology implemented.

Phase 3: P4T Methodology Evaluation

In the third phase, the proposed methodology has been evaluated and the benefits deriving from the methodology application analyzed and discussed.

2.3.1 Phase 1.: Operational Scenario Definition

The definition of the operational scenario is performed through the following steps:

- Route Selection: definition of the airspace and the routes considered for the design and validation of the methodology.
- Relevant Parameters Identification: selection of the parameters considered in the scenario that affect the optimal trajectory selection.
- Forecast Time Window Definition: definition of the time frame in which the methodology is applied.
- Data Generation: generation of the data needed as input to design and validate the trajectory prediction methodology.
- Simulated Dataset Verification: verification of the quality of the generated dataset.
- Assessment Metrics Definition: selection of the metrics for the assessment of the performance of the trajectory prediction methodology.

Since the PIU4TP project is a proof of concept, in order to assess the performance of the proposed methodology and without affecting the validity of the results, the scenario definition is based on simulated data and on some simplifying assumptions. As said above, the use of simulated data is justified because it allows testing the methodology in a fully controlled environment and provide a complete dataset which is not available in the open access datasets of real aircraft trajectories. Concerning the simplified assumptions, the project analysed only two flight routes within the European airspace, as test case for trajectory prediction, and considers few parameters affecting the selection of the optimal flight plan. Moreover, the input data were collected in a predefined time window and in a limited number of dates in advance with respect to the scheduled date of flight. However, it is worthy to highlight that the proposed methodology, once validated, could be applied to any route, at any time, and including any factor affecting the flight plan choice, if the required input data are available.

Details about the operational scenario are provided in the next sub-sections.

Route Selection

To support the methodology development and validation, the European airspace has been considered, and two routes have been selected:

- London Heathrow Airport (ICAO code: EGLL) - Athens Eleftherios Venizelos Airport (ICAO code: LGAV)
- London Gatwick (ICAO code: EGKK) - Malta International Airport (ICAO code: LMML)

Both routes are normally executed by several airlines; they fly through different national airspaces and go across different airspace sectors. Some possible flight plans have been associated to each flight route.

Relevant Parameters Identification

Actually, there is a large number and types of parameters that can affect flight plan selection and requests for a flight plan change both during pre-flight planning and flight execution. As anticipated, the PIU4TP project only considers two of these parameters, that are relevant in the strategic and pre-tactical phases, namely:

- actual aircraft take-off weight
- weather conditions

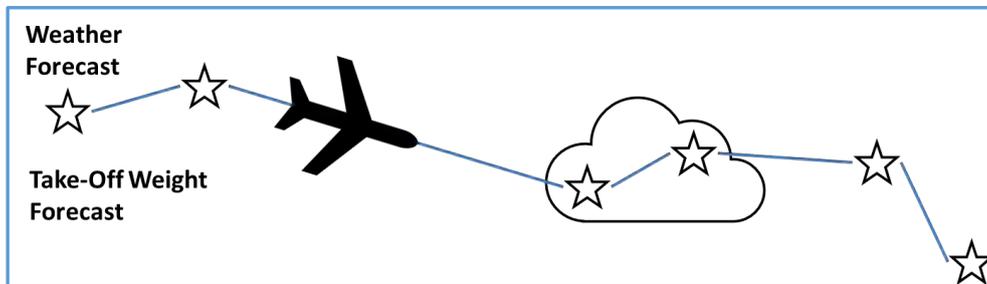


Figure 2 – Input variables into PIU4TP project for trajectory prediction

In fact, the actual take-off weight affects the climbing performance of the aircraft [RD3], [RD8], [RD9], and the selection of the optimal flight level, as described in [RD10]. The effects of weather conditions on the performed flight plan are widely known and reported in several works in the literature [RD6], [RD7], [RD11]; for example, pilots can decide to follow a route because it allows to take advantage of the tail winds making the flight faster, while reducing fuel consumption. Those above mentioned are just two of possible inputs affecting the TP. Other causes of uncertainties such as the pilot intent, FMS performance, ATC tactical intervention, **are excluded** from this project.

Forecast Time Window Definition

Since the project focuses on the strategic and pre-tactical phases, we consider a time window of 15 days before the scheduled date of flight as encompassing the overall study analyses. Consequently, fixed a T_f as the day/time of actual operations, it is expected that the NM starts to manage the flight routes allocation to requiring aircraft, 15 days before T_f , denoted as time T_f-15 . Each aircraft that intends to fly across the European controlled airspace needs to file a flight plan to the NM, starting 15 days and till few hours before the flight. In the planning phase, recurrent flights are normally considered, too, and their flight plans are well known also before 15 days in advance. Anyway, we consider in our study to start the process at T_f-15 days; it is worthy to remark that this assumption has no impact on the validity and the generality of the project results.

Because the methodology intends to demonstrate how the flow capacity management process could benefit from improvements in forecast of the meteorological conditions, two other dates are considered in the pre-tactical phase: 5 days before T_f (denoted as T_f-5) and 1 day before T_f (denoted as T_f-1). In strategic forecast, large uncertainties affect the input data and thus

the predicted trajectories have low level of confidence; progressing in time, while the uncertainties on the required input data reduce, better trajectory forecasts can be obtained as well as a higher level of confidence associated with the flight plans prediction.

Eventually, using the information available on actual meteorological conditions (we can identify this as Tf+), the flown flight trajectory is identified, and this can be compared with recursive forecasts at Tf-15, Tf-5 and Tf-1 in order to assess the methodology performance. The process timeline is sketched in Figure 3 .

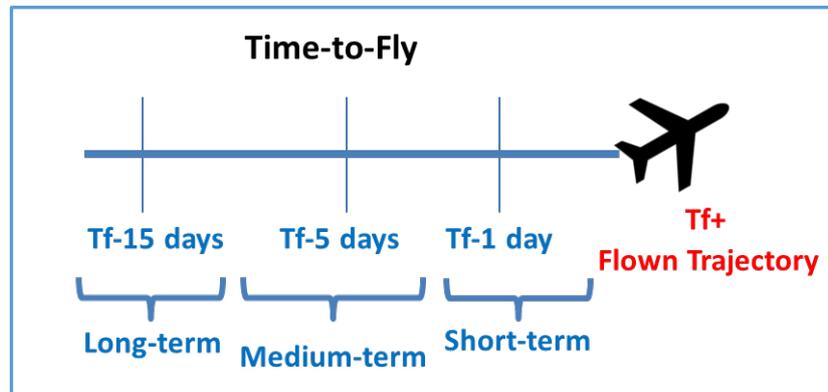


Figure 3 – Flight plan timeline for simulated data generation

Data Generation

The information about a huge number of flights shall be available to design and validate the methodology. For each flight the following data are required:

- the set of possible flight plans that can be flown along the selected route (EGLL-LGAV or EGKK-LMML);
- the weather forecasts (and their probabilistic characterization) along the flight route as above identified, at each date of the trajectory prediction computation;
- the take-off weight estimations (and their probabilistic characterization) at each date of the trajectory prediction computation;
- the actual flown trajectory the day of flight;
- the actual weather conditions experienced in the date of the flight;
- the actual take-off weight during the flight.

All these data have been generated through simulation, bearing in mind the need to be as realistic as possible. To this end, the following open access sources of information have been exploited to generate the needed data:

- flight plans for each route have been downloaded from the website www.flightplandatabase.com. The flight plans are defined in terms of departure and destination airports and the list of waypoints that defines the corresponding route.
- ERA5 database of the European Centre of Medium-range Weather Forecast-ECMWF [RD12] has been used to get 3D (longitude, latitude, air pressure) weather data for a wide range of selectable dates. For each selected day/hour, the database provides the re-elaborated weather information, that is the actual weather information as derived from a complex assessment process of the weather info from several sources. The ERA5 also provides the uncertainties characterization for the weather forecasts [RD13]; many of these characterizations apply back till to 15 days before the date, and this is a sound reason for selecting in 15 days the time range of our application. The data used within the PIU4TP project are wind intensity and direction and atmospheric

temperature. Exploiting the ER5 database, for each examined flight, the forecast of these parameters (including the uncertainty on the forecast) at Tf-15, Tf-5 and Tf-1, and the actual value in the day of the flight are computed.

- The take-off weight forecast and actual values vary among two precise limits, the Operating Empty Weight (OEW) and the Maximum Take-Off Weight (MTOW). These values for most of the aircraft are available in the literature [RD14].

The database OpenSky network has been also investigated to assess its applicability to the project. This database provides real trajectory data but does not provide all the other data needed as input for the PIU4TP methodology design and validation, such as the take-off weight of the aircraft performing a recorded trajectory, neither other source of experimental information can be used to complement the data available in OpenSky. For that reason, the OpenSky data have been not used and the actual flown trajectories were computed through simulation.

Twelve possible flight plans (three different lateral flight plans which can be performed at four different cruise flight levels) for each route have been selected. A generic short/medium range aircraft has been chosen to perform the flights, with take-off weight varying in the range 50-80 tons.

The data generation process comprises the following steps, applicable to each of the flight considered in the project:

- For each date, in which trajectory prediction shall be performed (Tf-15, Tf-5 and Tf-1), compute:
 - the estimated take-off weight and related uncertainties;
 - the weather forecast (atmospheric temperature and wind velocity) and related uncertainties along each flight plan;
 - the potential No-Fly Zones due to weather conditions and of the probability to cross one of them;
 - the estimated time of arrival in each waypoint and related uncertainties.
- For the date of flight (Tf) compute
 - the actual take-off weight
 - the actual weather conditions
 - the flown flight plan and the related 4D trajectory, in terms of time of arrival in each waypoint of the plan.

The take-off weight (TOW) estimation is randomly draw, assuming a Uniform stochastic distribution within the allowable range (from OEW to MTOW). Indeed, data on estimated TOW are not freely available in the literature, as well as information about the stochastic distribution applicable to the TOW estimation uncertainty. When little or no a-priori statistical information about the uncertain parameters is available, the use of the uniform distribution represents a conservative choice, because it guarantees that the probability of performance satisfaction under this uncertainty distribution is smaller than the probability under any other distribution [RD15]. The uncertainty on the estimated value depends on how much in advance with respect to the scheduled flight date the estimation is computed (it decreases while approaching the flight date) and it is defined as a percentage of the whole range of variation (that is, the difference between MTOW and OEW). The process for the take-off weight data generation starts with the first random draw performed on the whole range of variation. It defines the first weight forecast value at Tf-15 days (W_{15}); an uncertainty of 35% of the TOW range (that is, $TOW\ range = MTOW - OEW$) is associated to this value (W_{unc15}). Next, the forecast at Tf-5 (W_5) is computed through a new draw in the range $[W_{15} - W_{unc15}, W_{15} + W_{unc15}]$ and a new uncertainty (W_{unc5}), equal to 15% of the TOW range, is associated to it. The procedure is repeated for the forecast at Tf-1 (associated uncertainty is 5% of the TOW range) and for Tf+. In this last case, only the value of the weight is computed, without associating an

uncertainty to it; indeed, this value represents the actual take-off weight experienced during the flight and not a forecast of it. Of course, a saturation of the TOW value to the range [OEW, MTOW] is applied after each random draw, in order to guarantee that the generated value is realistic. Finally, it is worthy to remark that the numerical values used to characterize the uncertainty at different forecast times are just an attempt used because actual data are unavailable. However, these values do not affect the applicability of the trajectory prediction methodology neither the assessment of its performance because the same rules and parameters are applied to generate both the data used for the design and the validation of the methodology.

As said, the computation of weather data is based on the ERA5 database [RD12], specifically the reanalysis dataset is exploited. The ERA5 database provides for each date and at different hours (with one-hour resolution) the weather parameters defined on a three-dimensional spatial grid. Within the PIU4TP project, we downloaded more than 2000 different datasets, referring to days of October and November from 1979 to 2013 at 2pm hours. The considered months (October and November) and hour (2pm) were selected as a first test case and chosen through a random draw among the possible options. These data are used to design and preliminarily validate the methodology according to the objective of the PIU4TP project, that is, to demonstrate a proof of concept at low TRL. Next step, subject for future research projects, will be the training and assessment of the methodology on a wider set of input data (possibly actual experimental data) with uniform distribution along time. Indeed, weather at each time of the day could present different features that shall be properly modelled by the ML algorithm, in order to get a methodology applicable to predict the flight trajectory, whichever is the aircraft departure time, and to accurately assess its performance.

For each simulated flight, a file is extracted from the weather database (each dataset can be associated to just one flight) and the values of the relevant atmospheric parameters (atmospheric temperature and wind velocity vector) are considered. Weather forecast is evaluated in each waypoint of all the possible flight plans for the selected route, through an interpolation of the grid provided by ERA5. As far as the PIU4TP team is aware, ERA5 does not provide for a given date a complete dataset including the forecast for the selected date at Tf-15 days, Tf-5 days and Tf-1 day. To overcome this limitation, the following process has been implemented. It allows getting consistent data, which are realistic (although not actual), because computed starting from actual weather data, and suitable for a proof of concept demonstration. The reanalysis file for a given date is used as forecast at Tf-15 days. The forecast is characterized with an uncertainty for each variable that depends on the time in advance with respect to the scheduled date of the flight at which the forecast is performed; this uncertainty is provided in [RD13]. An example of such uncertainty on the temperature is presented in the following figure, which is an excerpt of [RD13].

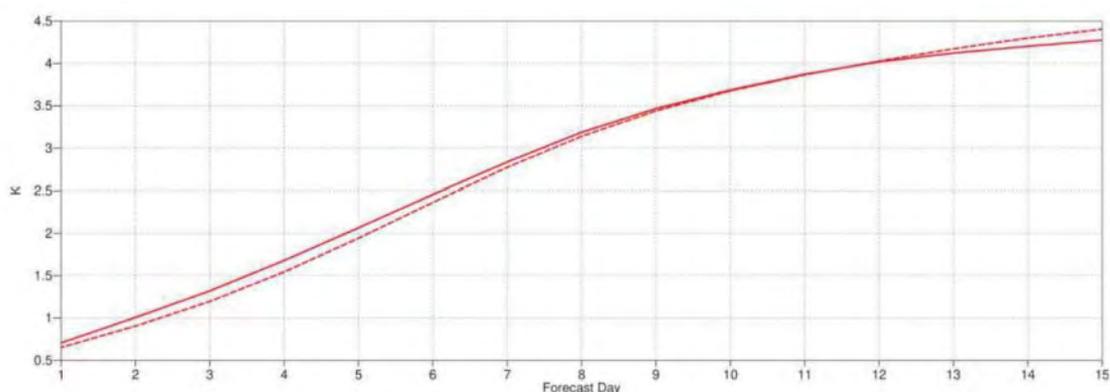


Figure 4 – Uncertainty on atmospheric temperature forecast depending on the forecast day (standard deviation, dashed lines; RMS error solid lines)

We assume that the forecast of the atmospheric parameters are stochastic variables with a Uniform distribution (it is an assumption, but any other distribution can be used), which is defined by the mean value (forecast) and the standard deviation (from figure above). Then we can compute the forecast at Tf-5 days by performing in each waypoint a random draw from the stochastic distribution defined at Tf-15 days. A new standard deviation can be computed at Tf-5 by using again Figure 7 for the temperature and similar plots for the other atmospheric parameters. The procedure is applied also to compute the forecast at Tf-1, starting from the stochastic distribution defined at Tf-5. Finally, a new draw is performed from the distribution at Tf-1 to obtain the atmospheric parameters to use for computing the actual flown trajectory (Tf+).

Once the atmospheric parameters are available, the No-Fly Zones can be computed. Two types of NFZ are considered: hazard NFZ and discomfort NFZ. The former shall be avoided during the flight, because weather conditions do not allow a safe flight. The latter could be avoided, depending on the operator's choice; indeed, crossing this NFZ could be not comfortable for the passengers but without affecting the safety of the flight. Both NFZs are defined with respect to the wind velocity. As far as the occurrence of a hazard NFZ is concerned, we define a threshold $Th1$ for the horizontal wind speed and a second threshold $Tv1$ for the vertical wind speed. Waypoints in which wind is expected to exceed one of these thresholds are included in the NFZ. At Tf-15, Tf-5 and Tf-1 days, the wind speed forecast is defined by a mean value and an uncertainty, which characterize a stochastic distribution. Consequently, defined the stochastic distribution, it is possible to compute the probability to exceed the NFZ thresholds and then the probability the waypoint has to be included in the NFZ. Performing this computation in each waypoint of a flight plan allows computing the probability that the flight plan has to cross the NFZ. The same approach is applied to define the discomfort NFZs; we define another couple of thresholds $Th2$ and $Tv2$, with $Th2 < Th1$ and $Tv2 < Tv1$, and the discomfort NFZ includes the waypoints in which the horizontal wind speed belongs to the range $[Th2, Th1]$ or the vertical wind speed belongs to the range $[Tv2, Tv1]$. Next, the probability to cross the NFZ is computed for each flight plan as above; the obtained probability for discomfort NFZ is halved, because we assume that just 50% of the operators take care to avoid this kind of NFZ. At Tf+, the wind speed experienced during the flight is known without uncertainty, therefore the probability of a flight plan to cross NFZs could only be 0 or 1.

The Estimated Time of Arrival to Each Waypoint is computed through kinematic equations. It is assumed that the flight is performed at the Optimal Mach number, denoted as ECON Mach, that depends on the TOW and the corresponding optimal flight level. The following figure, excerpt of [RD10], shows an example of this relation for the Airbus A340. Once the take-off weight is known, and assuming a cost index for the flight, the ECON Mach can be computed. Then the speed of sound is derived from the atmospheric temperature, the airspeed is calculated from Mach number and speed of sound, and the ground speed is evaluated composing the air speed and the wind speed. Finally, the time of arrival in each waypoint is obtained by dividing the length of the leg preceding the waypoint for the ground speed. Since the weather data are uncertain at Tf-15, Tf-5 and Tf-1, also the estimated time of arrival to each waypoint will be uncertain. In computing the time to reach the waypoints in the first legs of a flight plan, the climb performance of the aircraft is also considered by adding to the estimated time an additional delay. This climb performance depends on the TOW and are available in the open literature for some aircraft models [RD14].

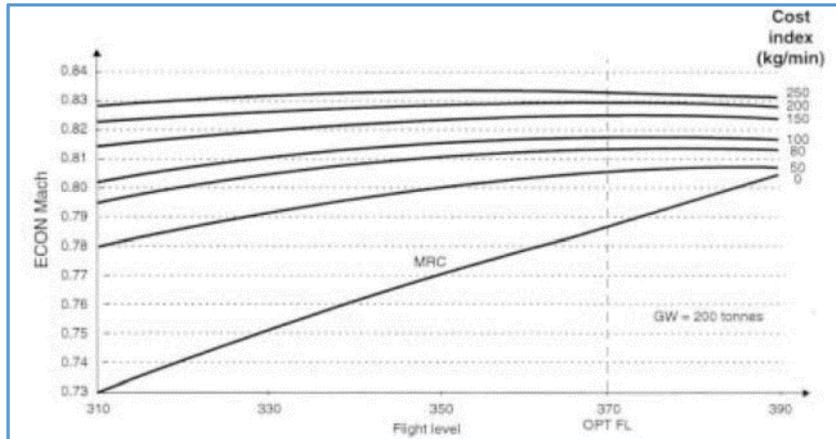


Figure 5 – Optimal Mach number depending on flight level, TOW and cost index

Computed weather conditions and take-off weight are considered as inputs for the selection of the flight plan among the possible options. Specifically, as detailed before, the presence and localization of the No-Fly Zones are determined by the weather, whereas the TOW affects the climbing performance of the aircraft and the optimal cruise altitude. An example of the relation between TOW and optimal flight level is shown in the following figure, from which it is possible to evaluate the flight level, given the TOW and the cost index selected by the operator. The figure is an excerpt of [RD14] and refers to Airbus A340. It is worthy to remark that the weather conditions could also contribute to determine the optimal flight level, because the relation between TOW and FL, for a fixed CI, varies with atmospheric temperature (different relations are defined for different temperature).

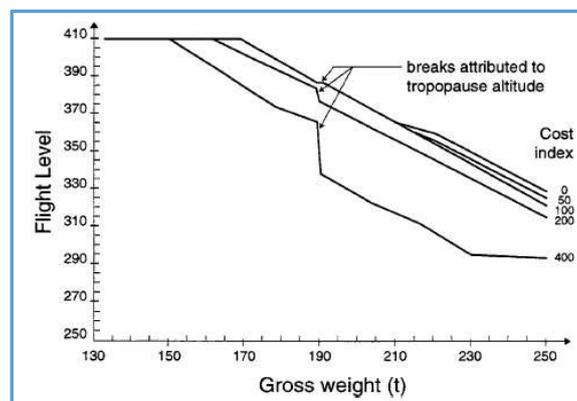


Figure 6 – Optimal flight level depending on TOW and cost index

Based on these considerations, the following rules apply to select the most suitable flight plan (among the available ones for the considered route) when the take-off weight and the weather conditions are known:

- The selection of the lateral flight plan is based on the avoidance of the NFZs.
- The selected flight level (vertical flight plan) is the optimal one with respect to the take-off weight for a given cost index.

At Tf+, the weather conditions and TOW are measured without uncertainties. TOW allows selecting the best vertical flight plan. Concerning the lateral flight plan, if only one of the possible options (for the considered route) avoids the NFZs, then the optimal flight plan is completely defined and added to the generated dataset. Otherwise the obtained data are

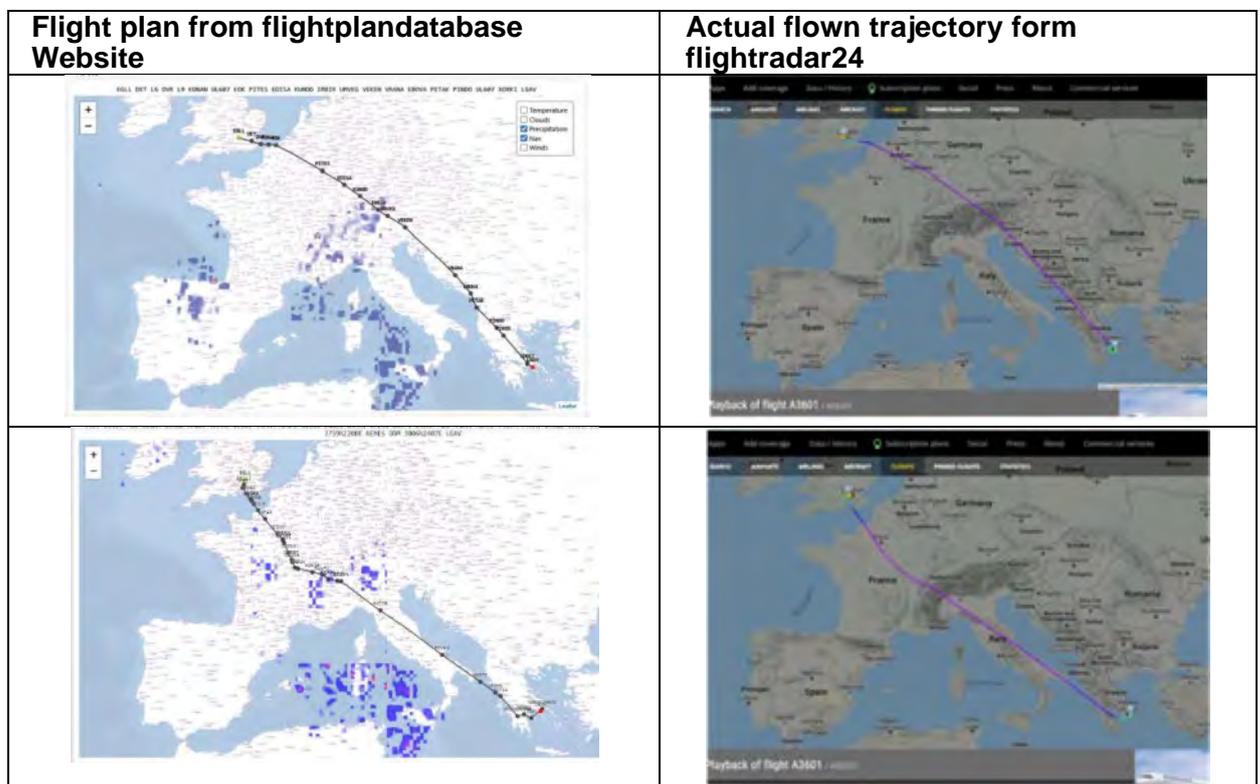
discarded (because they do not provide a unique solution). When forecasts are computed at Tf-15, Tf-5 and Tf-1, the wind, atmospheric temperature and take-off weight values are uncertain and described by a stochastic distribution, as explained in the previous sections. Consequently, also the selection of the flight plan (lateral and vertical) is characterized through a stochastic distribution; specifically, more than one selection is possible and a probability is associated to each of the selected options. The probabilities associated to each flight plan in the forecast time window (from Tf-15 to Tf-1), based on forecasted inputs, are additional information provided by the data generation process, which will not necessary be exploited in the training and validation of the trajectory prediction methodology.

Simulated Data Verification

Two types of verification are performed on the generated data: numerical consistency check and qualitative comparison with real data.

The numerical consistency check is performed on the TOW and weather parameters by verifying that these parameters always belong to a predefined range. The allowable TOW range is bounded by operating empty weight and maximum take-off weight of the considered aircraft. The bounds for each weather parameter are computed as minimum and maximum of the values that the parameter assumes in all the downloaded ERA5 datasets and enlarged by considering the maximum allowable uncertainty for the parameter, also provided by ERA5 documentation. The result of this check was always positive for all the generated data.

The qualitative comparison with real data is performed to check the realism of the selected flight plans. To this end, each possible flight plan is qualitatively compared with actual flown trajectories that are observed on the <https://www.flightradar24.com/data/flights> website. An example of this check is shown in the following figure for the route EGLL to LGAV. Also this check, although qualitative, provided positive results for all the considered flight plans.



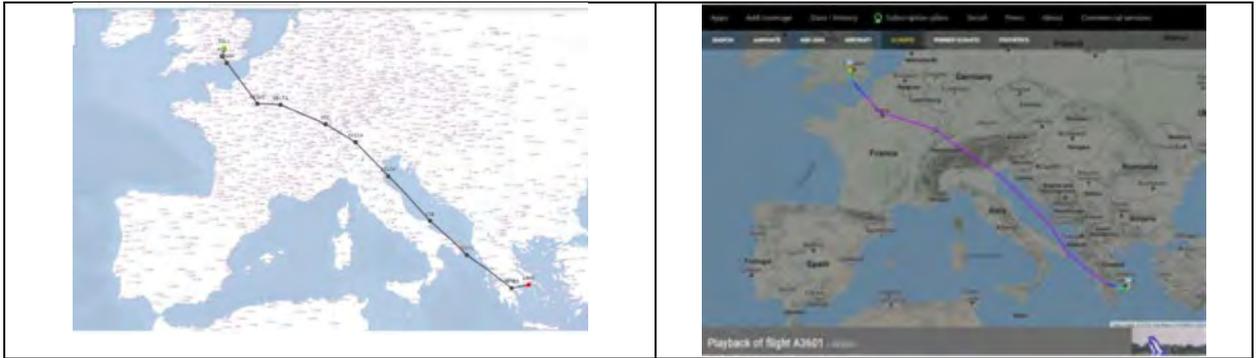


Figure 7 – Flight plans selection procedure

Assessment Metrics Definition

Generally, in machine learning, a predictive model is defined in terms of a number of parameters. In supervised learning, some of these parameters are derived via a learning process (*training*), i.e. using a dataset of labelled samples (*training dataset*) trying to make the model predict the correct label for each of the samples. The training itself is controlled by other parameters called *hyperparameters*, which also must be tuned to build good predictive models.

Optimizing the overall predictive power of a model both respect its parameters and hyperparameters requires the definition of some sort of measure to quantify its performance. A performance metric of a model F (for example, a classifier trained by applying an Artificial Neural Network) is a measuring function that assigns to F a real number m . Mainly, machine learning is used to solve problems that fall within two different categories, i.e. classification and regression. In classification the model has to predict a discrete variable, i.e. the class, among a finite number of classes, the sample belongs to. In regression the variable is continuous. It's possible to define different metrics for the different types of problems in machine learning.

In **classification**, all the considered performance metrics are based on the *confusion matrix*, which is the main and most common method used to show the results obtained by a classifier. The confusion matrix is a table with entries that represent the number of samples classified in a certain class. The rows of the table are indexed by the actual classes and the columns by the predicted classes. So, in a binary classification problem with a positive class and a negative class, the confusion matrix is defined as:

		Predicted Class	
		P	N
Actual Class	P	TP	FN
	N	FP	TN
Estimations			

Table 1 Binary Confusion Matrix.

On the main diagonal there are the numbers of correct classifications, TP is the number of True Positives and TN is the number of True Negatives. The off diagonal elements report the number of the misclassifications, FN is the number of False Negatives, i.e. positive samples incorrectly classified as negative, and FP is the number of False Positives, i.e. negative samples incorrectly classified as positive.

The first performance metric is the accuracy which measures how good the model is in correctly predicting both positive and negative cases:

$$Accuracy = \frac{\text{Number of correct predictions}}{\text{Total number of samples}} = \frac{TP + TN}{TP + TN + FP + FN}$$

Accuracy however is a tricky metric because it can give misleading information about the performance of a model. This is especially the case in the situations where the dataset is imbalanced, i.e. there are many samples of one class and not much of the other. Meaning, if your model is performing well on the class that is dominant in the dataset, accuracy may be high, even though the model might not perform well in other cases.

There are other metrics that can be calculated from the confusion matrix very useful for evaluating the classifier performances, even when the dataset is imbalanced (Table 2).

#	Symbol	Performance Metric	Definition as	What does it measure?
1	<i>TPR</i> or <i>REC</i>	True Positive Rate – Sensitivity or Recall	$\frac{TP}{TP + FN}$	How good model is in correctly predicting positive cases
2	<i>TNR</i>	True Negative Rate – Specificity	$\frac{TN}{TN + FP}$	How good model is in correctly predicting negative cases
3	<i>FPR</i>	False Positive Rate – Fall-out	$\frac{FP}{FP + TN}$	Proportion of incorrectly classified negative cases
4	<i>FNR</i>	False Negative Rate – Miss Rate	$\frac{FN}{FN + TP}$	Proportion of incorrectly classified positive cases
5	<i>PPV</i> or <i>PRE</i>	Positive Predictive Value – Precision	$\frac{TP}{TP + FP}$	Proportion of correctly classified positive cases out of total positive predictions

Table 2 Metrics defined from the confusion matrix.

Precision and recall are two very useful metrics, they answer to two different questions about the performance of the model: the former gives the proportion of positive identifications that are correct, the latter gives the proportion of actual positives that have been identified correctly. Recall is relevant in those contexts where it is important to have a low number of false negatives, whereas precision when it is important to maintain low the number of false positives.

These two metrics are different but related. In fact, increasing precision in general leads to a lowering of the recall and vice versa. For this reason, often, their harmonic mean is used, the so called F1-score:

$$F1 = 2 \frac{PRE \times REC}{PRE + REC}$$

Since the objective of **regression** is to predict a continuous variable, the metrics used to measure the performance of machine learning models are different than those defined for the classification case.

The most popular metric is the mean squared error (MSE) due to its simplicity. Given, for each of the N samples in the dataset, the values of the target variable $\{y_i\}_{i=1,\dots,N}$ and the predicted values $\{\tilde{y}_i\}_{i=1,\dots,N}$, the MSE is the average squared distance between the predicted and actual values:

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \tilde{y}_i)^2$$

The result is a non-negative value and the goal is to get this value as close as possible to zero.

Another very popular metric is the root mean square error (RMSE), a direct variation of the MSE metric since it is simply:

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \tilde{y}_i)^2}$$

The advantage of RMSE is that it is in the same unit as the value to be predicted. Since both MSE and RMSE are very sensible to the presence of outliers in the dataset, their use should be more useful when large errors are particularly undesirable. Moreover, they are both differentiable.

A metric that doesn't require the calculation of squares or square roots, useful when outliers are not a particular issue, is the mean absolute error (MAE):

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \tilde{y}_i|$$

2.3.2 Phase 2: P4T Methodology Development

The development of the P4T methodology was carried out in three phases detailed in the following Figure 8.

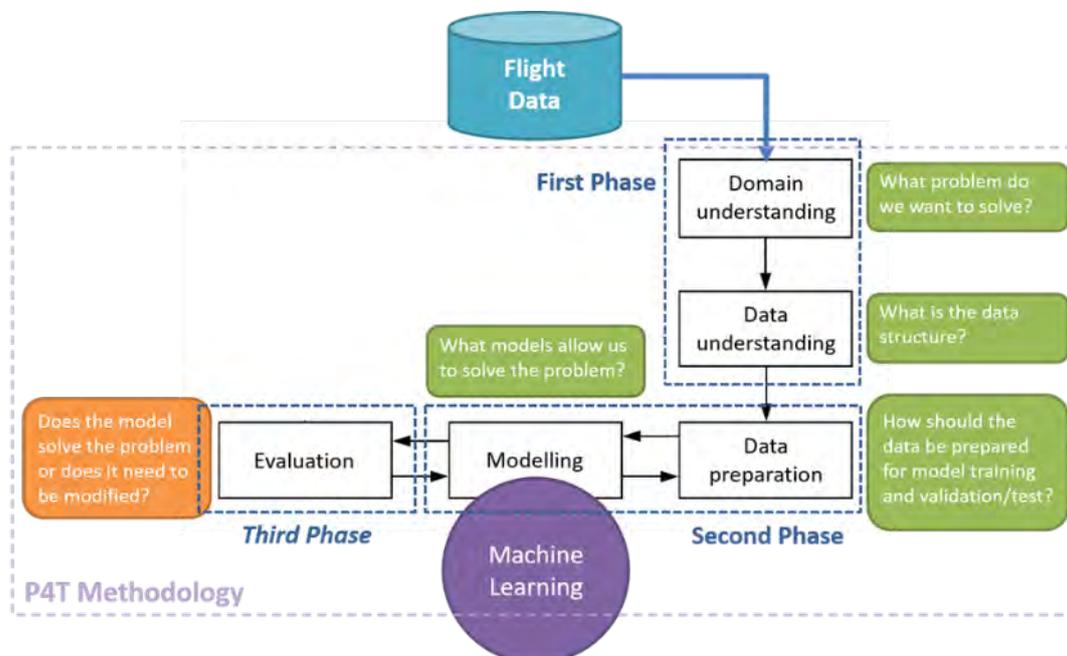


Figure 8 – Data Driven methodology in the PIU4TP project.

In the following paragraphs we are going to describe the activities carried out in each of these different phases.

Overview of machine learning and data mining techniques applied to the aircraft trajectory prediction problem.

As a preliminary step, a research has been conducted in the scientific peer-reviewed literature to gather the most up-to-date information on the application of machine learning and data mining to the problem of aircraft trajectory prediction. Mainly, data mining and machine learning methods have been applied in order to cluster, classify and model large amount of flight trajectories data. Clustering is often used to find similarities in flight paths, for example, clustering has been used to define typical flight paths in groups of trajectories, to characterise the deviations from the nominal flight plans and use this information to represent the flight intent into a trajectory prediction algorithm. In recent times, classical fully-connected and deep neural networks are the most used algorithms to model trajectories and to build machine learning models to predict the flown trajectory from the departure to the arrival airport or during the climbing or descent phase, or to estimate the time of arrival (ETA) of a flight at the airport or in the terminal manoeuvring area (TMA). Many of the examined articles used as input variables the meteorological condition during the flight, but none of them considered the take-off weights and the uncertainties on the data. This overview has been very useful in clarifying which are the most promising algorithms to investigate as prediction models and to implement in the methodology, keeping in mind that the model should be able to manage the uncertainties in the input data. The results of this overview have proven useful also in the following activities in domain and data understanding as well as in modelling.

Domain and Data Understanding

These first steps of the methodology were performed to clarify the properties of the data generated from the WP2 activities before any attempt of modelling.

The **Domain Understanding** phase included the fixing of the objectives of the Data Analysis goals and the assessment of the situation. In particular, the mapping from domain issues to data analysis problems. The analysis also considered the results from the Consultation Exercise Meetings, in fact the methodology has to be developed bearing in mind that it could be used as a tool to support the operations of air traffic management.

As a result of this phase in the P4T methodology the domain objective, which consists in the prediction of the flight path, has been translated into a data analysis objective, which consists of a multiclass classification, regarding the prediction of the flight plan, both horizontal and vertical, and of a regression, regarding the estimation of the time of arrival on the waypoints of the horizontal flight plan. Then, for the flight plan, the problem to address can be stated as:

- Predict which, among N possible flight plans, will be selected for the flight execution.
- Input variables for the predictive model are:
 - Forecast temperature and wind speed and direction (east, north and down components) for each waypoint of the flight plan with relative uncertainties.
 - Forecast take-off weight with uncertainty

The first group of variables will be used for the prediction of the lateral flight plan, i.e. the sequence of the waypoints, the mean temperature in the zone of flight and the estimated take-off weight will be used for the prediction of the vertical part of the flight plan, i.e. the cruise flight level.

The prediction of the time of arrival on the waypoints (ETA) has been stated as a classical regression problem, with input variables the sequence of the waypoints, the forecast temperature and horizontal wind direction (east and north components) for each waypoint and the forecast take-off weight.

To support the methodology development and validation, the data of both the routes identified in the operational scenario definition have been considered: London Heathrow (EGLL) –

Athens (LGAV) (in the following we will refer to this route simply as London – Athens) and London Gatwick (EGKK) – Malta (LMML) (in the following simply London – Malta).

The **Data Understanding** phase includes the initial data collection description, data exploration and the verification of data quality. In particular, the activities in this phase allow to capture and understand the meanings and statistics of the relevant variables for the analysis (features). In particular, there are 20 variables comprised in the simulated data related to weather condition, take-off weight and the relative uncertainties, aircraft position and speed and flight level. The following table shows the list of the variables simulated for the data analysis process.

Variable	Description	Unit
WP_ID	Way Point Identifier	–
Lon	WP Longitude	degrees
Lat	WP Latitude	degrees
Temp	Temperature	K
TempUnc	Uncertainty on temperature	K
VnWind	North component of wind speed in WP	m/s
VnWindUnc	Uncertainty on North component of wind speed	m/s
VeWind	East component of wind speed in WP	m/s
VeWindUnc	Uncertainty on East component of wind speed	m/s
VdWind	Down component of wind speed in WP	m/s
VdWindUnc	Uncertainty on Down component of wind speed	m/s
Weight	Weight at take off	kg
WeightUnc	Uncertainty of weight	kg
FlightLevel	Flight level	ft
PrFlightLevel	Uncertainty of flight level	ft
Mach	Mach number	–
PrMach	Uncertainty of Mach number	–
Vground	Speed with respect to the ground	m/s
VgroundUnc	Uncertainty of speed with respect to the ground	m/s
EstTime	Time needed to cover the distance between 2 consecutive waypoints	s
EstTimeUnc	Uncertainty on time arrival	s

Table 3 – List of the simulated variables.

The simulated data are grouped into two folders of files in Microsoft Excel format, one for each of the two selected routes (London – Athens and London – Malta). For the route from London to Athens there are 2052 simulated flights, while for the other route from London to Malta the simulated flights are 2023.

For each flight 4 different Excel files are produced:

- 3 files contain the forecasted weather conditions along the route and the estimated take-off weight with the respective uncertainties in distinct time frames before estimated off-block time (EOBT):
 - 15 days (Tf-15) before EOBT, for the strategic/long-term scenario;
 - 5 days (Tf-5) before EOBT, for the medium-term scenario;
 - 1 day (Tf-1) before EOBT, for the pre-tactical/short-term scenario.Each of these files contain 3 different sheets reporting the simulated data for each of the possible predefined lateral flight plans. Each sheet may contain up to three replicas of the lateral flight plan referring to different possible flight level, the number of replicas depends on the forecasted take-off weight and the amplitude of its uncertainty, in fact, an ampler interval of uncertainty may encompass more than one valid flight level.
- The remaining file refers to the day of flight and contains the selected flight plan, the weather data along the route and final take-off weight.

Note that the lateral flight plans may have different lengths, i.e. each plan may be defined by a different number of waypoints.

The examination of the simulated data has showed that there are only a limited and discrete number of different possible values for the flight level. This is in full agreement with the normal practice in aviation, *in fact, air space is divided into tracks, with planes flying in a specific altitude range depending on the direction they are going in and the routes they are taking. This standardizes routine air traffic, making it safer to fly.* For this reason, also the prediction of the vertical part of the flight plan has been casted as a classification problem using the mean temperature in the zone of flight and the take-off weight as input variables.

There is also a one-to-one correspondence between the flight level and the optimal cruising Mach number of the aircraft, so that once established the value of the flight level the Mach is uniquely defined. This is perfectly reasonable in a first approximation, taking apart the possible variation due to the necessity to compensate for the effects of the wind speed along the route. So, temperature and take-off weight contain all the information to predict the cruising airspeed, for this reason, in the regression model for the estimation of ETA we consider these variables as input to the model and not the estimated cruising speed or flight level, avoiding also to introduce into the model the dependency on other estimated parameters.

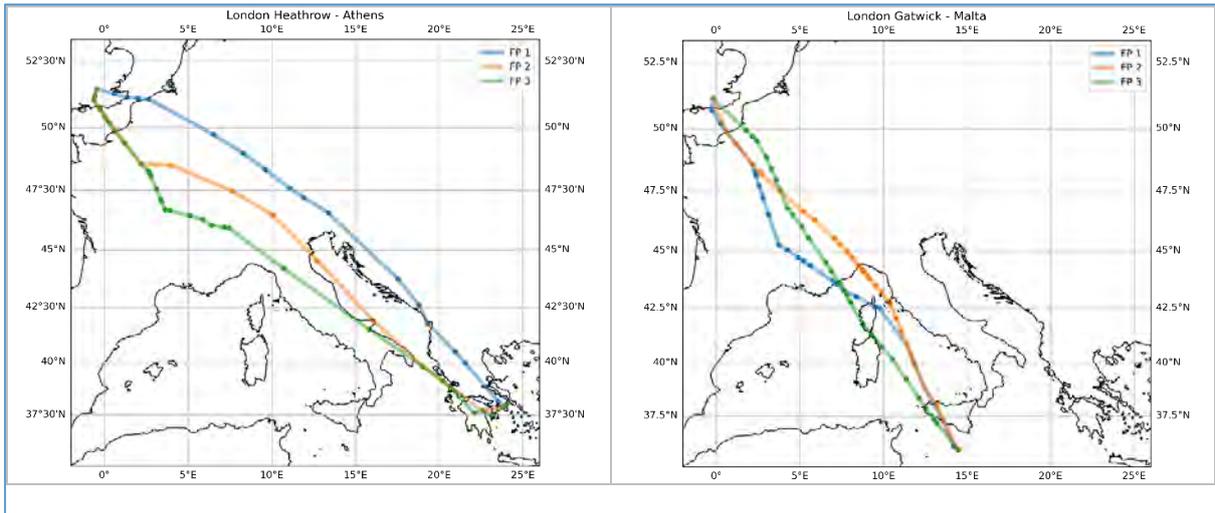


Figure 9 – The two selected routes with the different predefined lateral flight plans

A descriptive analysis has been conducted on the input variables, characterizing their statistical properties to describe and summarize the raw data and gain first insights useful for the subsequent predictive modelling steps.

Data Preparation

With the availability of the simulated data, an important step in the analysis has been to identify and solve eventual problems within the data. In order to obtain the final dataset to be used in the modelling phase, data has been pre-processed to report them in a format usable by modelling algorithms. For example, particular attention has been given to the normalization of the input variables, especially useful for the application of algorithms based on neural networks. In general, data collection and data preparation are the most time-consuming activities in a machine learning project.

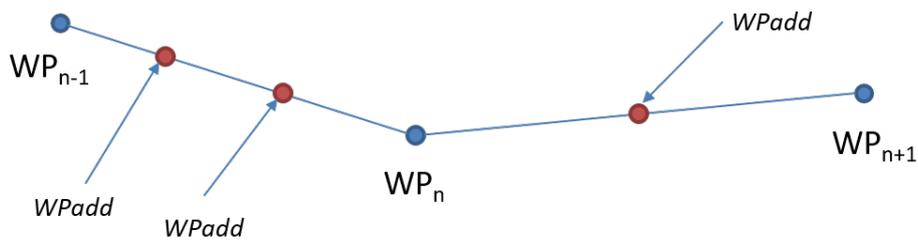


Figure 10 – Waypoints added along the legs of a flight plan.

Since, mainly, machine learning algorithms for classification and regression accept as input vectors of fixed length and in order to avoid possible biases due to the different lengths of the lateral flight plans, as a first step in data preparation, a certain number of dummy waypoints (names *WPadd*) were added to the flight plans to make them of the same length [Figure 10]. The waypoints were added where needed along the legs connecting two successive waypoints, in such a way to not change the direction of flight of the aircraft. The procedure adopted is straightforward:

1. The sequence of waypoints in the flight plan is a list: $[wp_1, wp_2, \dots, wp_N]$.
2. Set a threshold T on the maximum length of a leg between two consecutive waypoints.
3. Start with the first waypoint and let $i = 1$.
4. Compute the length d_i of the leg between wp_i and wp_{i+1} .

5. If $d_i > T$ then add a dummy waypoint in the middle of the leg between wp_i and wp_{i+1} .
6. Increment i by one.
7. If $i < N$, then go to step 4.
8. If $i = N$ and the desired length of the flight plan has been reached, then stop.
9. Repeat the procedure from step 2 using also the added waypoints and a different threshold T .

The weather condition on the added waypoints were also simulated with the same procedure used for the flight plan.

For the modelling step we used different strategies for the construction of the datasets for flight plan and flight level classifications and for the estimation of the time of arrival.

Regarding the prediction of the flight plan, we build separate datasets for each of the selected route and for each time frame before EOBT. Since we are assuming that the definition of the lateral flight plan and the choice of the cruise flight level may be taken as independent, we build different datasets for the prediction of these two target variables [Figure 11].



Figure 11 – Structure of the datasets.

Fixed the route and the time frame, the simulated dataset provides for each flight forecasted weather conditions along the alternative flight plans, as well as the estimated take-off weight.

The datasets for the prediction of the lateral flight plan contain vectors with the following structure:

$$\left(T^{(1)}, V_N^{(1)}, V_E^{(1)}, V_D^{(1)}, T^{(2)}, V_N^{(2)}, V_E^{(2)}, V_D^{(2)}, \dots, T^{(L)}, V_N^{(L)}, V_E^{(L)}, V_D^{(L)} \right) \quad (1)$$

where L is the number of waypoints in the flight plan, including the added dummy waypoints. The components of these vectors are only the weather variables, i.e. the temperature T and the three components of the wind speed along the three directions north-south V_N , east-west V_E and down-up V_D .

To consider the uncertainties, the value of the weather variables used to construct the input vectors for modelling is drawn from a gaussian distribution centred on the simulated value and with standard deviation $\sigma = \Delta/3$, where Δ is the associated uncertainty. The choice of σ is made to have a gaussian ample enough to take all the interval of uncertainty of the weather variable, i.e. $6\sigma = 2\Delta$. This sampling is repeated for a fixed number of times.

Then, for the components of the vector (1), we have:

$$\begin{aligned} T^{(i)} &\sim \mathcal{N}(T_0^{(i)}, \Delta T^{(i)}/3) \\ V_j^{(i)} &\sim \mathcal{N}(V_{0j}^{(i)}, \Delta V_j^{(i)}/3) \end{aligned}$$

where $i = 1, \dots, L, j = N, E, D$, and $T_0^{(i)}$ and $V_{0j}^{(i)}$ are the simulated values of the variables.

The vectors constructed for each flight plan are, then, concatenated to form a unique vector of length given by the product of the number of flight plans, the number of waypoints and the number of weather variables. This procedure is repeated for every simulated flight and all the vectors are collected into the dataset for modelling, whose size is the product of the number of flights by the number of the samples drawn from the gaussian distributions. To define the target variable, an integer number is given as a label to each possible flight plans. The target variable for the training of the models is the label corresponding to the lateral flight plan used for the execution of the flight.

The procedure used to construct the dataset for the prediction of the flight level is similar. The flight level is a characteristic of the flight, not of the single flight plan, and we are assuming that the choice of the flight level depends mainly on the take-off weight and on the mean temperature in the zone of flight.

The dataset for the training of the models for the prediction of the flight level is made up of vectors with the following simple structure:

$$(T_m, W)$$

where T_m is calculated by taking all the waypoints of all the possible lateral flight plans, eliminating all the repeated waypoints and averaging the temperatures on all the remaining waypoints, W is a value repeatedly drawn from a gaussian distribution centred on the simulated value W_0 and $1/3$ of the uncertainty ΔW as standard deviation:

$$W \sim \mathcal{N}(W_0, \Delta W/3)$$

Then, for the target variable, to each possible flight level is given as label an integer from 1 to the number of possible flight levels. The target variable for the training of the model is the label corresponding to the flight level used for the execution of the flight.

The dataset for the regression problem of estimating the time of arrival on the waypoints of the lateral flight plan was built starting from the data of the simulated flights, i.e. those referring to the day of flight. The variables included in this data are a subset of those listed in **Table 3**:

Variable	Description	Unit
WP_ID	Way Point Identifier	–
Lon	WP Longitude	degrees
Lat	WP Latitude	degrees
Temp	Temperature	K
VnWind	North component of wind speed in WP	m/s
VeWind	East component of wind speed in WP	m/s
VdWind	Down component of wind speed in WP	m/s
Weight	Weight at take off	kg
FlightLevel	Flight level	ft

Variable	Description	Unit
Mach	Mach number	–
Vground	Speed with respect to the ground	m/s
EstTime	Time needed to cover the distance between 2 consecutive waypoints	s

These variables refer to the flight plan used during the execution of the flight and carry no uncertainties. This choice of data was dictated by the consideration that, unlike the prediction of the flight plan, where the choice of the target plan does not change by letting the input variables change value in the range defined by the respective uncertainties, we don't have suitable target variables for the regression for all the possible values of the input variables if uncertainties were also considered.

So, we decided to build a data-driven model of the aircraft dynamics building one dataset for each of the two routes considered. Each dataset contains rows with the following structure:

$$(d, b, T, V_N, V_E, W)$$

where d is the distance between two consecutive waypoints of the same flight calculated along a loxodrome, b is the track angle between the two waypoints, T , V_N and V_E are the temperature and the two components of the wind speed along the north-south and east-west directions on the starting waypoint and, finally, W is the take-off weight. We decided to not introduce into the regression model a dependency on other estimated parameters, such as the flight level or the mach number, since the temperature and the take-off weight should contain enough information to let the model gain a knowledge about the cruise speed of the aircraft. We concentrate our attention on the cruising phase of the flight, leaving out from the analysis the climbing from the departure airport to the cruising flight level and the descending phase to the arrival airport.

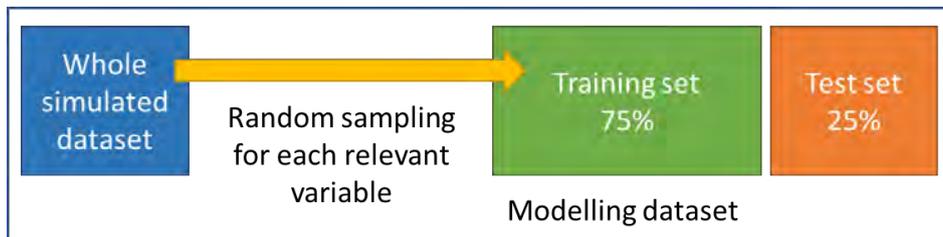


Figure 12 – The process to build the modelling datasets.

The modelling datasets obtained with these procedures are split into training and test sets [Figure 12]. The training sets are used for the construction and optimization of the predictive models, while the test sets are kept apart for the final validation of the performance of the models.

Data Modeling

This kernel phase represents the application of one or more Machine Learning algorithms able to obtain a classification, a regression, or a clustering model, trained on the prepared datasets. This step includes the testing of the obtained models in order to select the best one from a statistical point of view.

In the PIU4TP project, the aircraft trajectory prediction problem has been stated as two classification problems, one for the prediction of the lateral flight plan and one for the prediction

of the flight level, and a regression problem for the estimation of the time of arrivals on the waypoints of the flight plan.

In general, a machine learning model depends on a number of parameters, some of these parameters, for example the weights in a logistic regression model, can be optimized using a training dataset, i.e. a set of examples of pairs of an input vector and the corresponding desired output. But there are some parameters, called hyperparameters, which define the overall architecture of the model and which can only be tuned by repeating the training with different model architectures.

In order to select the best model for the problem at hand, part of the available data is used as a validation dataset. This dataset can be used to obtain an unbiased evaluation of a trained model for tuning the model's hyperparameters.



Figure 13 – Hold-out validation.

Two possible strategies can be used to split the data into training and validation dataset: holdout and K-fold cross-validation. With holdout [Figure 13], the data is simply split into two sets, one is used for training and the other for validation. With K-fold cross-validation [Figure 14] the available data is split into K subsets, one of these subsets is used as a validation set while the remaining are used for training. The process of training and validation is repeated K times, each time using a different subset for validation. The overall performance of the model is evaluated as an average of the performance obtained in each run.

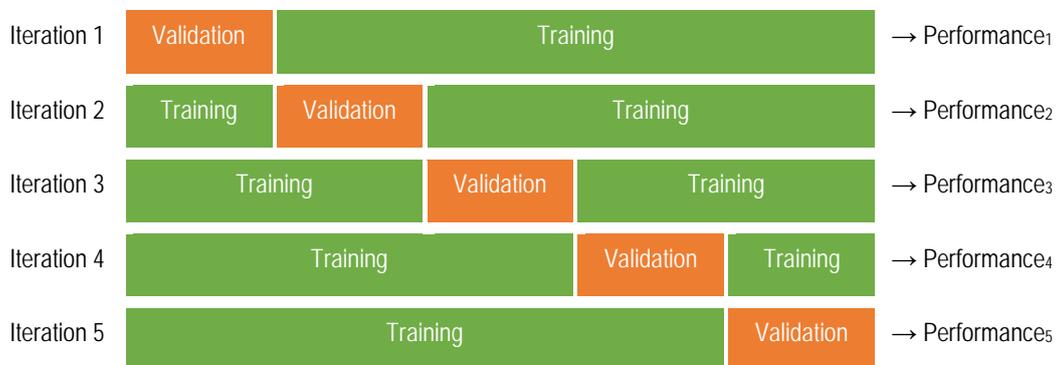


Figure 14 – K-fold cross validation.

The K-fold cross-validation is particularly useful when there is a small amount of data for training or to obtain a more accurate estimate of model prediction performance.

In the development of the methodology we have used both holdout and a k-fold cross-validation (with $k = 10$), obtaining very similar results, so in the following only the results for the cross validation are reported.

In the following paragraph we are going to summarize the results of the modelling separately for the prediction of the flight plan and prediction of the ETAs. The three timeframes have been dealt with much the same procedure and there were not special difficulties and limitations encountered during the training of the models, apart for the low, but increasing as the day of the flight closes in, accuracy in the long-term timeframe.

Prediction of the flight plan

The performance metrics used during training have been described in a previous paragraph about the operational scenario, the starting point is the confusion matrix [Figure 15], that is a table where the diagonal entries represent the number of correct classifications, whereas the off-diagonal elements report the number of misclassified input. Figure 15 is a generalization of Table 1 to a multiclass classification problem:

		Predicted			
		Class ₁	Class ₂	...	Class _N
True/Actual	Class ₁	m_{11}	m_{12}	...	m_{1N}
	Class ₂	m_{21}	m_{22}	...	m_{2N}

	Class _N	m_{N1}	m_{N2}	...	m_{NN}

Figure 15 – Example of a confusion matrix for a multilabel classification problem.

the matrix element m_{ij} is the number of input samples belonging to the i -th class classified by the model as belonging to the j -th class. From the confusion matrix can be defined a number of possible performance metrics, the more used are:

$$\text{overall accuracy} = \frac{\text{number of correct predictions}}{\text{total number of samples}} = \frac{\sum_i m_{ii}}{\sum_{i,j} m_{ij}}$$

$$\text{recall for class } i = \frac{\text{number of correct prediction for class } i}{\text{number of inputs belonging to class } i} = \frac{m_{ii}}{\sum_j m_{ij}}$$

$$\text{precision for class } i = \frac{\text{number of correct prediction for class } i}{\text{number of inputs classified as belonging to class } i} = \frac{m_{ii}}{\sum_j m_{ji}}$$

Accuracy is the most used performance metrics, but in some situation may give misleading results, for example when there is a strong imbalance in the training set and the examples of one class outnumber the other classes. Furthermore, since, in general, precision and recall are somewhat inversely related, increasing one of the two tends to reduce the other, a useful metrics combining the two is the f1-score, given by the harmonic mean of precision and recall. In Figure 16 is depicted the overall process used for the training of the classification models. An information-gain based filter has been used to reduce the number of input variables to the most significant ones. Among those tested in the development of the methodology, the models showing the best performance in classification are the decision trees, as can be deduced from the tables in Annex II. In particular Annex II reports the fine-tuning step and the selection of the hyperparameters of the algorithms used during the modelling phase.

In Table 4 and Table 5 are reported the results obtained for the accuracy in the prediction of the lateral flight plan.

The complete results of modelling for the two routes and for all the timeframes before EOBT are reported in Annex III.

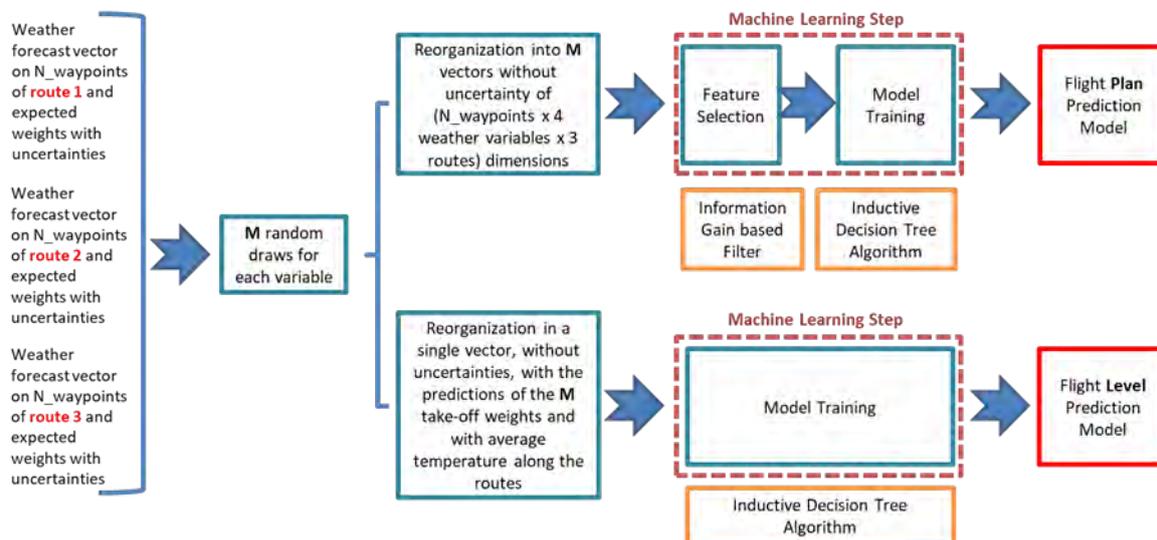


Figure 16 – The modelling process used in the training of the models.

	Accuracy		
	TF-15	TF-5	TF-1
London - Athens	38.5%	53.4%	99.9%
London - Malta	35.8%	81.2%	99.9%

Table 4 - 10-fold cross-validation results for the lateral flight plan prediction.

	Accuracy		
	TF-15	TF-5	TF-1
London - Athens	46.7%	69.9%	88.7%
London - Malta	44.2%	72.3%	90.3%

Table 5 - 10-fold cross-validation results for the lateral flight level prediction.

For the lateral flight plan, we note that the training datasets are substantially balanced, especially the one for the London-Malta route, while the one for the London-Athens route presents a slight imbalance in favor of the first lateral flight plan, as highlighted in Table 6. The ability of the models to make correct predictions is very low for the long term, 15 days before take-off the performance is only slightly better than that of a classifier that assigns labels randomly. In this time frame, for the London-Athens route, the model tends to prefer the first plane of lateral flight, this could be a further sign of imbalance in the dataset. The results improve, however, rapidly as the temporal distance from EOBT decreases, a sign that the models have been able to effectively learn the information useful for the classification.

	London-Athens	London-Malta
Flight plan 1	36.6 %	32.2 %
Flight plan 2	30.1 %	33.0 %
Flight plan 3	33.3 %	34.8 %

Table 6 - Composition of the training dataset for lateral flight plan prediction.

For the flight level, the training datasets show a more marked imbalance, once again lower in the case of the London-Malta route. The models perform better than the random classifier starting from 15 days before EOBT and the rate of correct classification increase steadily approaching the day of the flight.

	London-Athens	London-Malta
Flight level 330	23.5 %	25.0 %
Flight level 350	31.7 %	29.6 %
Flight level 370	27.1 %	27.7 %
Flight level 390	17.7 %	17.7%

Table 7 - Composition of the training dataset for the prediction of flight level.

In modelling we have used one-hot encoding, i.e. we have encoded the numbers i labelling the different lateral flight plans (flight levels) with a vector of dimension equal to the number of alternative lateral flight plans (flight levels), whose components are all zero except the i -th component which is set to 1. Using this encoding for the target variables, the output of the models is a vector of real numbers whose size is, again, equal to the number of alternative lateral flight plans (flight levels). The components of the output vectors are all positive numbers that sum to 1 and thus may be interpreted as a probability distribution over the possible lateral flight plans (flight levels) given the vector of inputs. The output of the model is the lateral flight plan or flight level to which corresponds the highest probability.

Since we may consider the choice of the lateral flight plans and of the flight level as independent, the product of these probabilities gives the overall probability for the selection of a flight plan (lateral + vertical). These overall probabilities can be represented as a heat-map or a bar plot graph, for example:

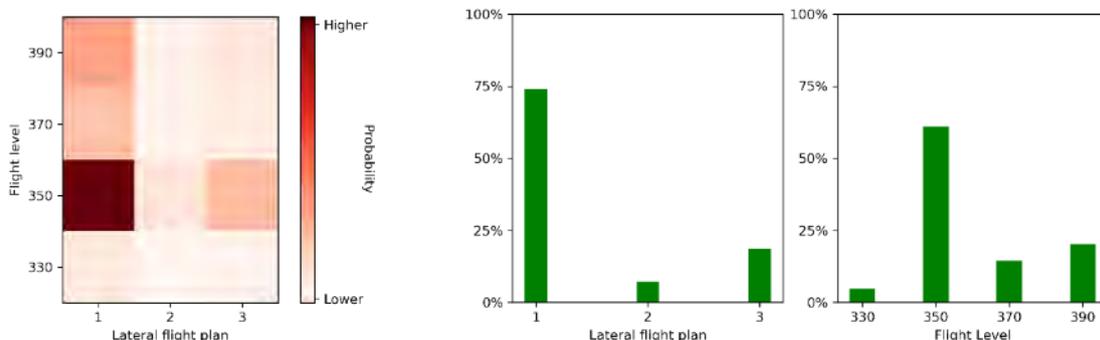


Figure 17 On the left, heat-map of the joint probability for the prediction of the flight plan. On the right, bar plot of the marginal probabilities for the prediction of the lateral flight plan and flight level.

Prediction of the time of arrival

We trained many different models for the estimation of the time of arrivals on the waypoints of the flight plan during the cruising phase of the flight. During training, to optimize model performance, we used the MSE measure:

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \tilde{y}_i)^2$$

where $\{y_i\}_{i=1,\dots,N}$ are the values of the target variable and $\{\tilde{y}_i\}_{i=1,\dots,N}$ are the predicted values, N is the number of samples in the dataset. The model performance is reported using the MAE:

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \tilde{y}_i|$$

which give a measure of performance in the same units of the target variable and is more easily interpreted.

The dataset used were built starting from the data relative to the executed flight and carry no uncertainties. The datasets were almost balanced in terms of the flight plan used for the flights, for both the two selected routes:

	London-Athens	London-Malta
Flight plan 1	36.2 %	32.0 %
Flight plan 2	30.1 %	33.0 %
Flight plan 3	33.7 %	35.0 %

Among those used during modelling, the models giving the best performance for both the London-Athens route and the London-Malta route were two random forest regressor models with 150 estimators (see Annex II for details), with the following performance on the training set:

	MSE (s ²)	MAE (s)
London-Athens	10.5	1.8
London-Malta	4.8	1.4

2.3.3 Phase 3: P4T Methodology Evaluation

The evaluation of the performance of a model is of paramount importance to assess the real capability of the model to be used in a production environment. To this end part of the available data is to be kept apart in a test dataset never used in any step of the training/validation process.

The test dataset for the evaluation of the models trained in the P4T methodology comprise 100 randomly chosen flights for each of the routes selected, with data referring to 15 days, 5 days and 1 day before the EOBT and to the day of execution of the flight. It is worth pointing out a major difference between the training/validation dataset and the test dataset. As

described in the previous paragraph, the training dataset is built by sampling the input variables from certain distributions defined by their respective uncertainty, so from each simulated flight in the training set we get M different records corresponding to the same target flight plan. The dataset obtained by this procedure is then split randomly into a training and a validation dataset, these two sets are disjoint but it may be possible that records referring to the same flight may be present in both sets. The test dataset, instead, is made up by all the records of all the flights taken apart for the evaluation of the models.

We chose two different strategies for the evaluation of the model for the prediction of the flight plan and for the estimation of the time of arrivals, both strategies are detailed in the following paragraphs.

Flight plan prediction

Since we are considering the choice of the lateral flight plan and the choice of the flight level as independent, the prediction of the flight plan is a two steps process that can be performed in parallel [Figure 18].

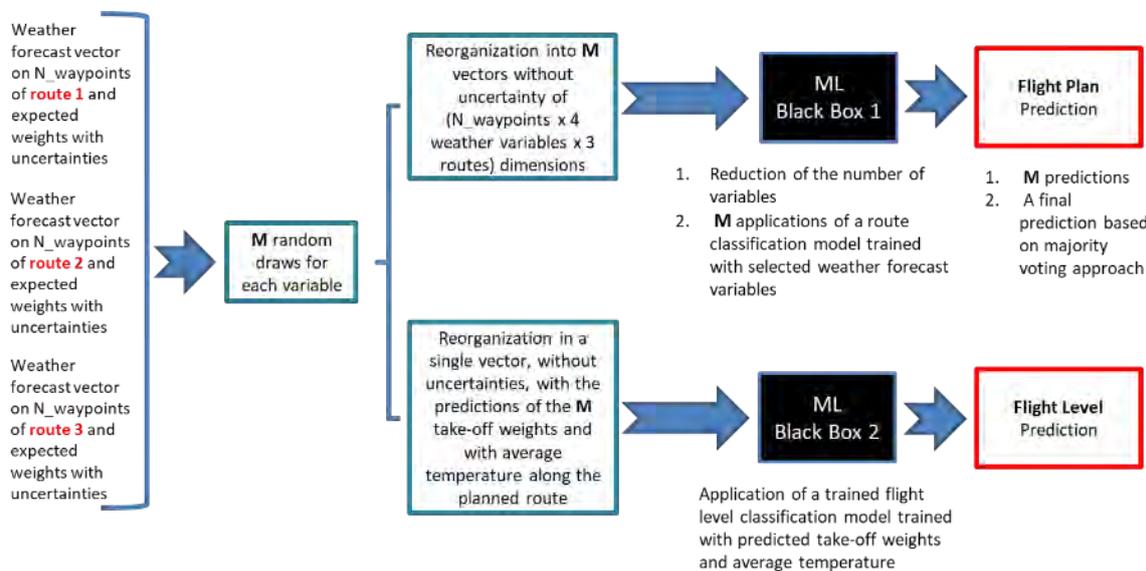


Figure 18 Online phase of the P4T methodology for flight plan prediction.

The M vectors obtained through the sampling procedure are used as input to the model for the prediction of the lateral flight plan to obtain M different predictions, the final output of the model is the one recurring most often (majority voting). For the prediction of the flight level we have a unique vector as input to the model and the output is the most probable flight level (one-hot encoding).

The performance of the predictive models on the test dataset are presented in Table 8 for the route from London to Athens and in Table 9 for the route from London to Malta.

	Lateral flight plan	Flight level	Flight plan (horizontal + vertical)
Tf-15	31 %	48 %	12 %
Tf-5	63 %	67 %	42 %
Tf-1	78 %	88 %	68 %

Table 8 Test results for the route London-Athens.

We explicitly note that a classifier that chooses the lateral flight plan and the flight level completely at random should have an accuracy of about 8.33 %, so even in the long-term case

Tf-15 the accuracy in predicting the flight level (horizontal and vertical) is still significantly better than a random classifier.

	Lateral flight plan	Flight level	Flight plan (horizontal + vertical)
Tf-15	34 %	48 %	13 %
Tf-5	76 %	66 %	50 %
Tf-1	83 %	89 %	74 %

Table 9 Test results for the route from London to Malta.

These results confirm the overall good performance of the classifiers, in particular the accuracy of the prediction increases remarkably as the time of the departure closes in and the forecast values of the input variables get closer to the values experienced during the execution of the flight and the corresponding uncertainties get smaller. The prediction of the flight level shows quite good performance, also at 15 days before the day of the flight. It is mainly due to the information provided by the forecast of the take-off weight, that is uncertain but within a bounded range; if the forecast of take-off weight was perfectly known without uncertainties, then the prediction of the flight level could significantly improve being the TOW one of the main parameters affecting the optimal flight level selection. On the other hand, this prediction degrades if the take-off weight forecast is not available. The performance of the lateral flight plan prediction is determined by the uncertainties on the weather forecast, that have a less significant impact also on the flight level prediction.

In the following figures are reported some of the results obtained by applying the models to the test set for the route from London to Malta, but there are very similar results also for the other route from London to Athens. In the heat-maps, a little circle indicates the correct values for the lateral flight plan and the flight level, the colour is red for wrong prediction of either the two output variables and is green for a correct prediction of both, the values in the graph are the predicted probabilities. Same for the bar plots, the colour red indicates a wrong prediction, green a correct prediction, on the x axis the value corresponding to the actual value is coloured green.

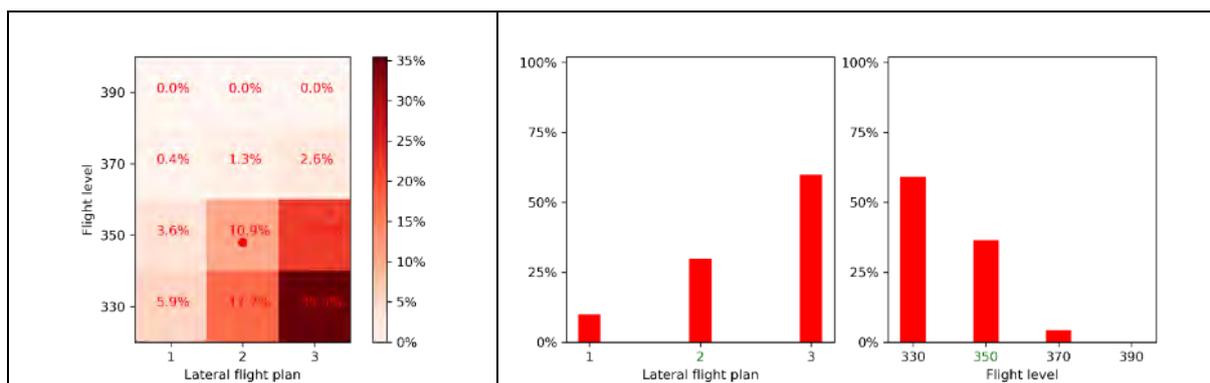


Figure 19 - Both predictions are wrong, but the correct values have the second highest probability.

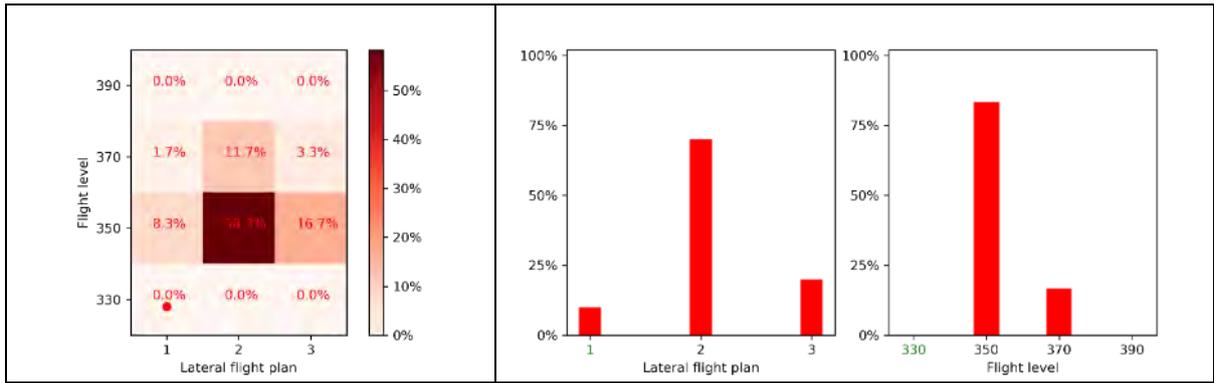


Figure 20 - Both predictions are wrong and the correct values have low or zero prediction probabilities.

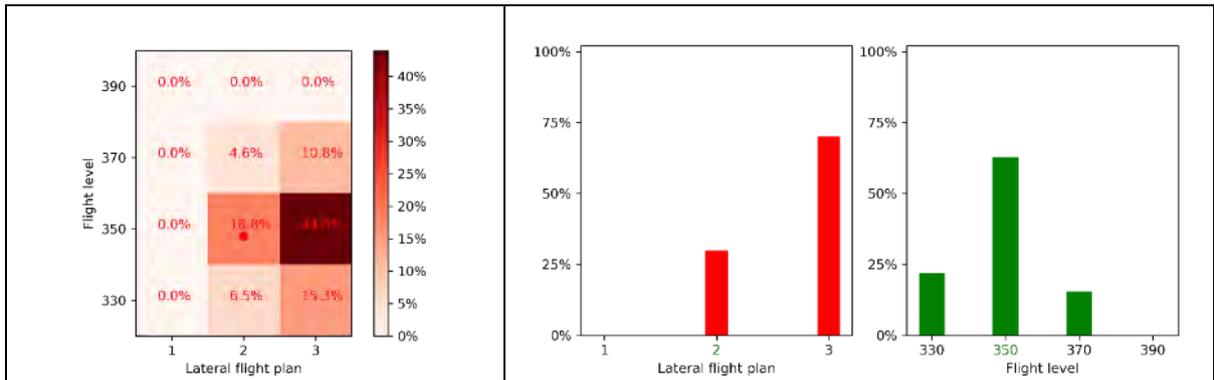


Figure 21 - Only the prediction of the lateral flight plan is wrong with second highest prediction the correct value.

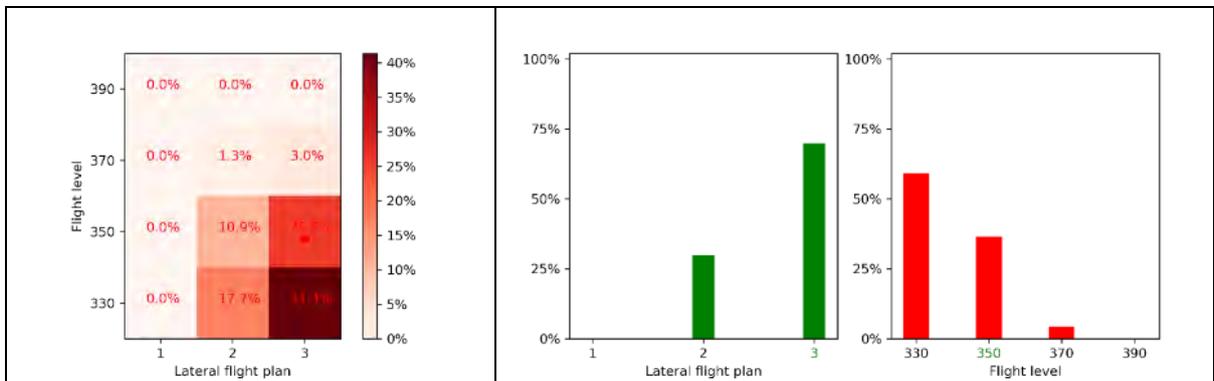


Figure 22 - The prediction of the lateral flight plan is correct, but the one for the flight level is wrong.

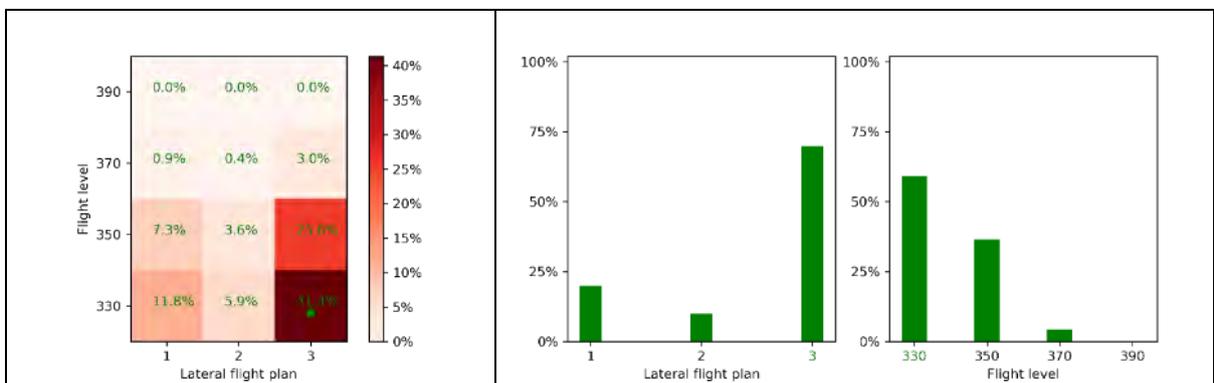


Figure 23 - Both the predictions are correct, but the overall joint probability is rather low.

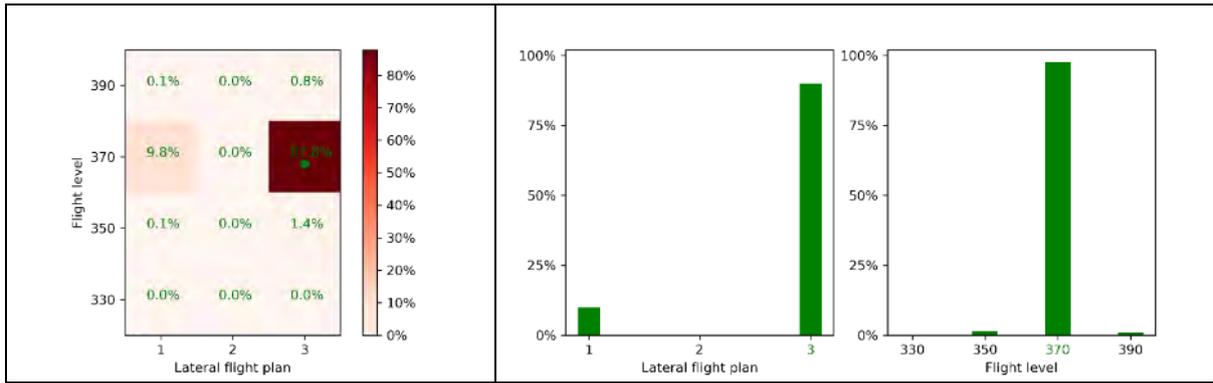


Figure 24 - The flight plan has been correctly predicted and the joint probability is high.

Prediction of the time of arrival

To test the performance of the regression model, we must first use the lateral flight plan prediction model to obtain a prediction of the sequence of waypoints to be used as input to the model for estimating the arrival times. We used as test the same flights used to test the models for the prediction of the flight plans.

To assess the performance of the model we have to compare its predictions with the arrival times reported during the actual flight. But, since the predicted flight plan may differ from the actual flight plan, instead of comparing the arrival times on the individual waypoints, we compared the overall duration of the flight during the cruise phase [Figure 25].

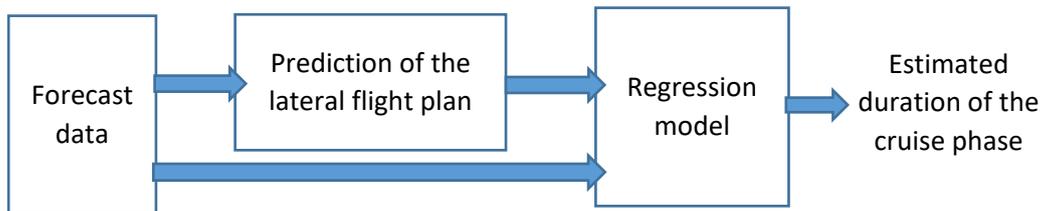


Figure 25 Workflow for the prediction of times.

The results of the tests for the two selected routes and for the different considered time frames are reported in the following figures.

In Figure 26 there are the histograms of the absolute values of the difference between the actual and the predicted duration of the cruise flight for the route London-Athens for all the considered time frames before the EOBT. For this route, the cruise flight extends for 22 waypoints. It is evident that the performance of the model gets better approaching the day of the flight.

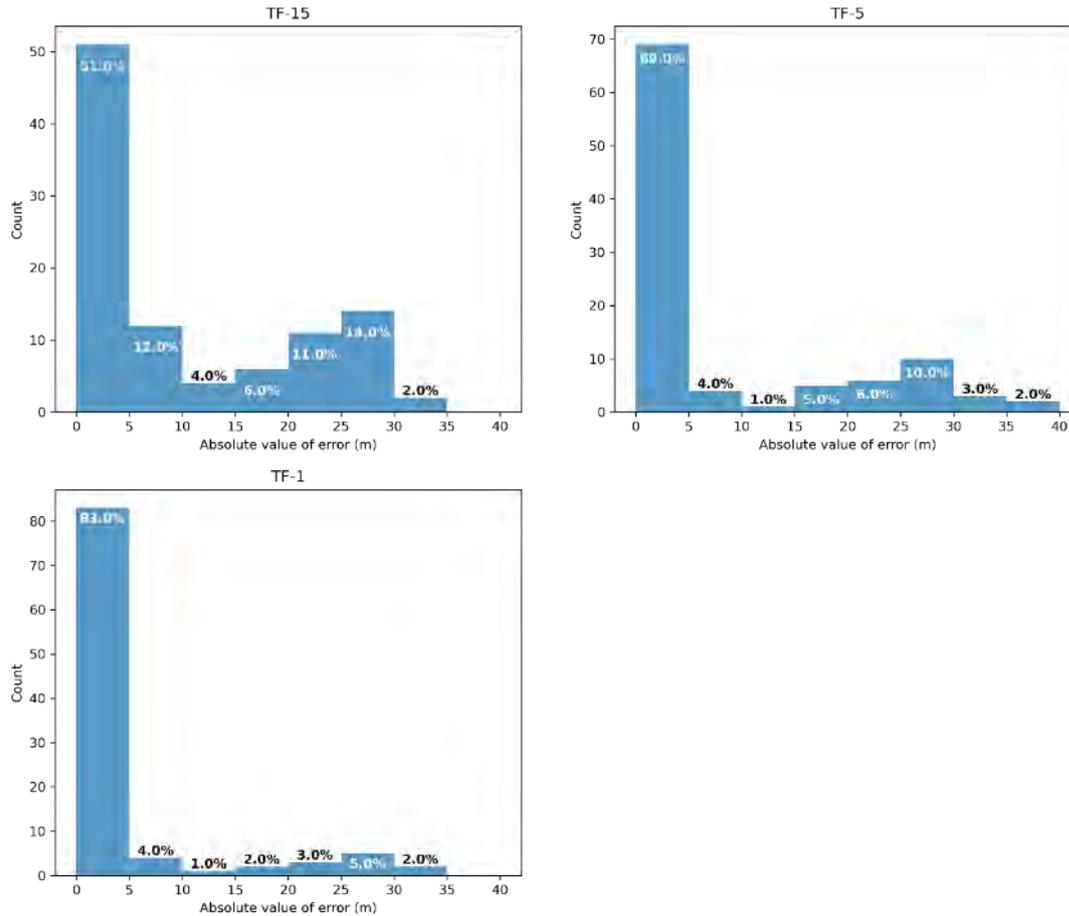


Figure 26 Histograms of the absolute differences between actual and predicted cruise flight duration for the London-Athens route, for different time frames before EOBT.

In Figure 27, Figure 28 and Figure 29 there are three scatter plots, one for each of the timeframe before EOBT for the route London-Athens, on the x axis there is the actual cruise flight duration and on the y axis there is the predicted duration. For a perfect regressor all the points should lie on the bisector (red line), in the figures we see a distribution of points, those further from the bisector are the ones with higher prediction error. The points are coloured according to the correctness of the flight plan prediction. We see from the figure that when the lateral flight plan is correctly predicted, the error in the prediction of the duration of the cruise flight is lower, on the order of few minutes in absolute value over a flight of duration greater than two hours.

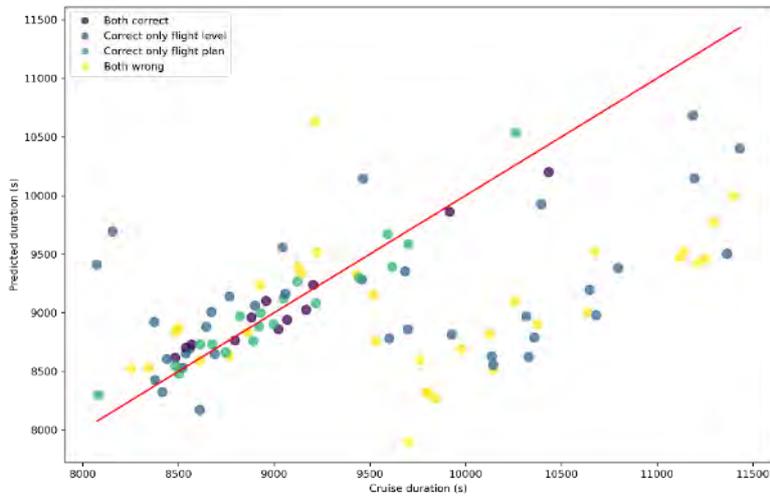


Figure 27 London-Athens, actual cruise flight duration vs. predicted duration for Tf-15.

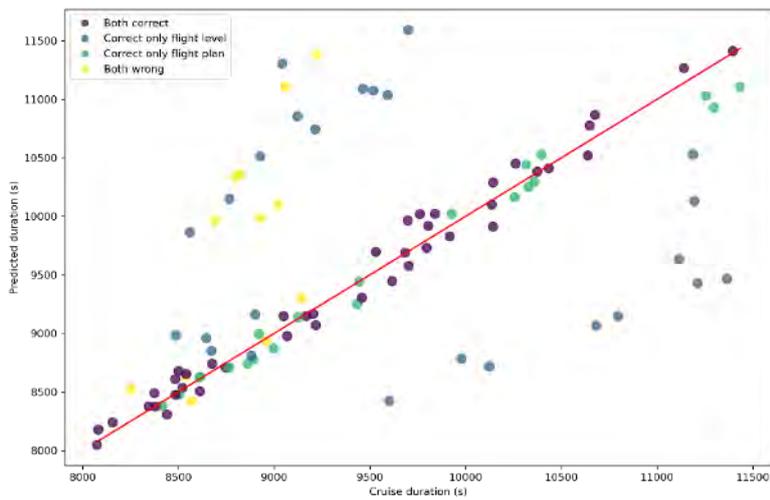


Figure 28 London-Athens, actual cruise flight duration vs. predicted duration for Tf-5.

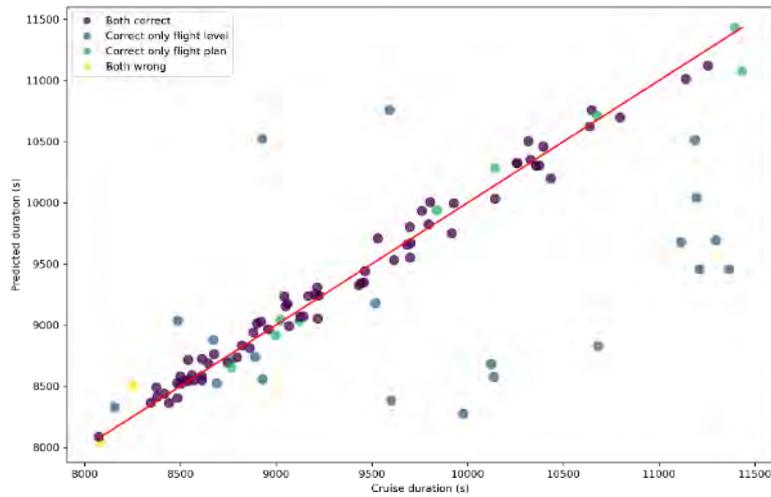


Figure 29 London-Athens, actual cruise flight duration vs. predicted duration for Tf-1.

In the following figures we present the same graphs for the route from London to Malta. For this route, the cruise flight extends for 21 waypoints, with a duration that ranges from about 1.5 to about 2.5 hours.

The histograms in Figure 30 show a similar pattern of that for the route London-Athens: approaching the EOBT the number of flights with a prediction error in the range 0-5 minutes increases steadily, with a corresponding reduction in the number of flights with high prediction errors.

The scatter plots in Figure 31, Figure 32 and Figure 33 are very similar to those reported for the London-Athens route: lower prediction errors are associated specially to those flights with at least a correct prediction of the lateral flight plan.

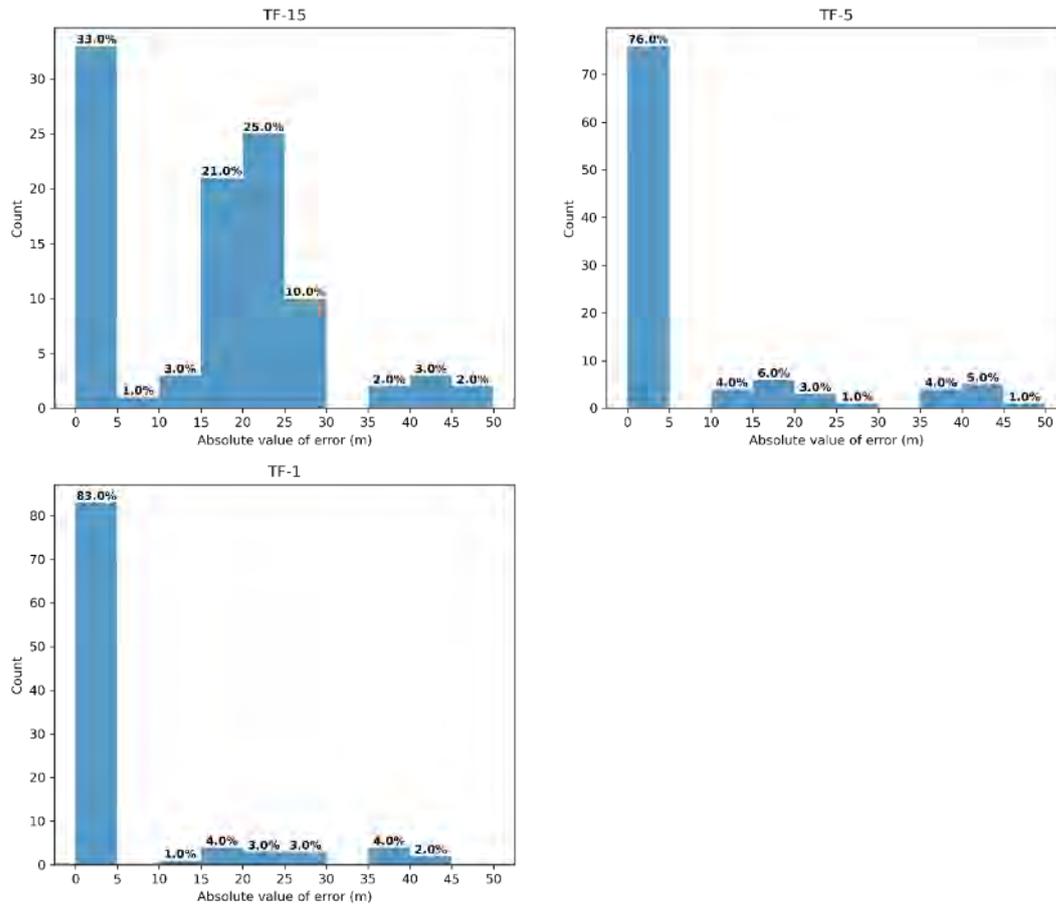


Figure 30 Histograms of the absolute differences between actual and predicted cruise flight duration for the London-Malta route, for different time frames before EOBT.

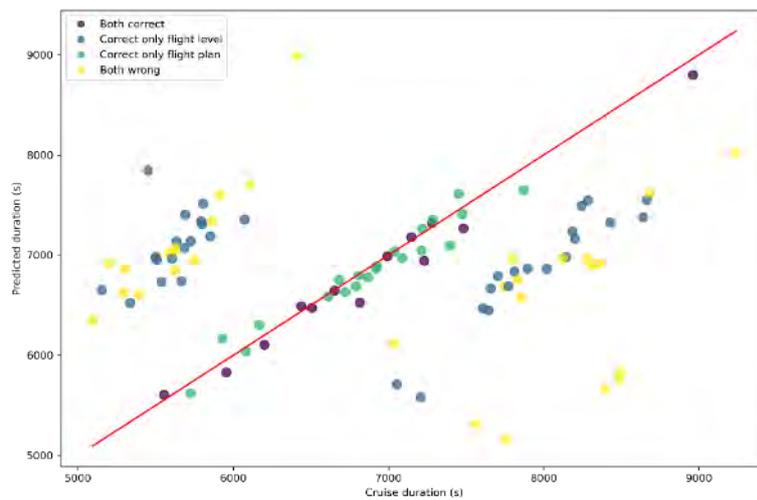


Figure 31 London-Malta, actual cruise flight duration vs. predicted duration for Tf-15.

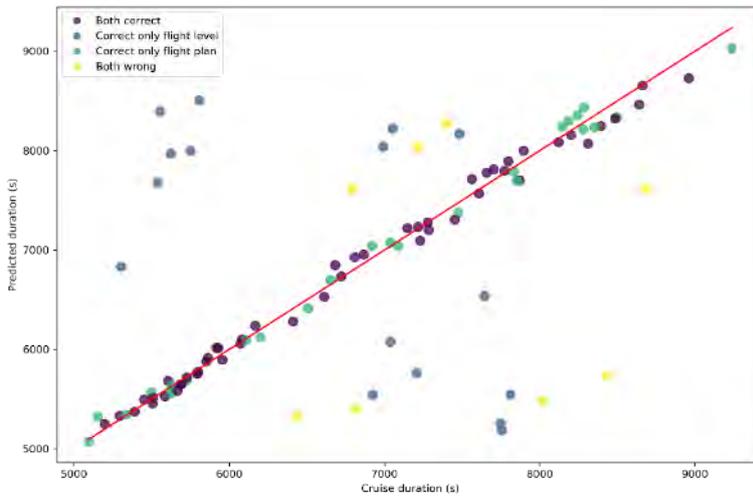


Figure 32 London-Malta, actual cruise flight duration vs. predicted duration for Tf-5.

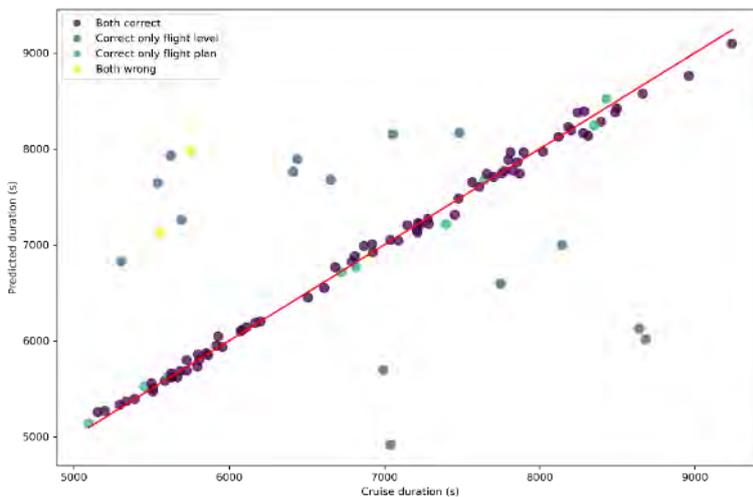


Figure 33 London-Malta, actual cruise flight duration vs. predicted duration for Tf-1.

To give an idea of the improving performance of the regression model when the lateral flight plan is correctly predicted, in the following histograms we show the absolute value of the error limited only to the flights with a correct prediction of the lateral flight plan. For both the routes, the error doesn't exceed 360 seconds (6 minutes), and the number of flights with a value of the error below 120 s increases remarkably approaching the day of the flight.

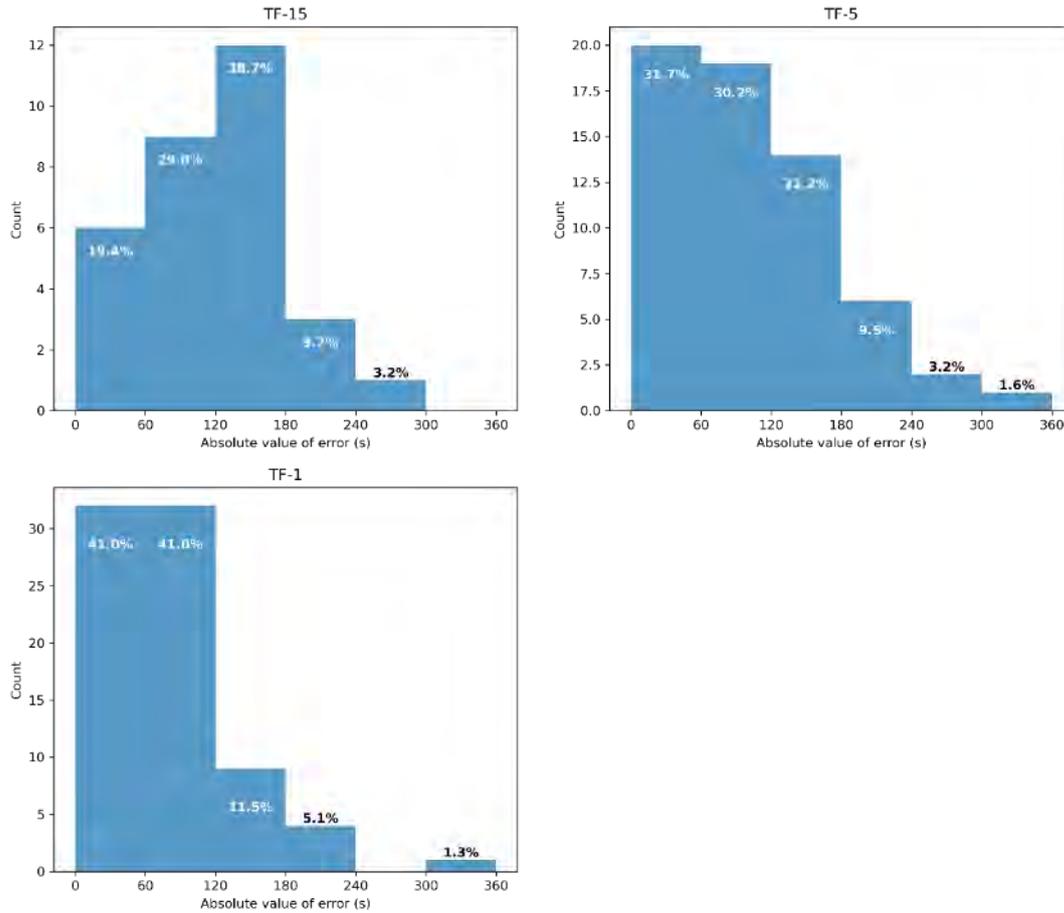


Figure 34 Histograms of the absolute differences between actual and predicted cruise flight duration for the London-Athens route, for the flight with a correct prediction of the lateral flight plan and for different time frames before EOBT.

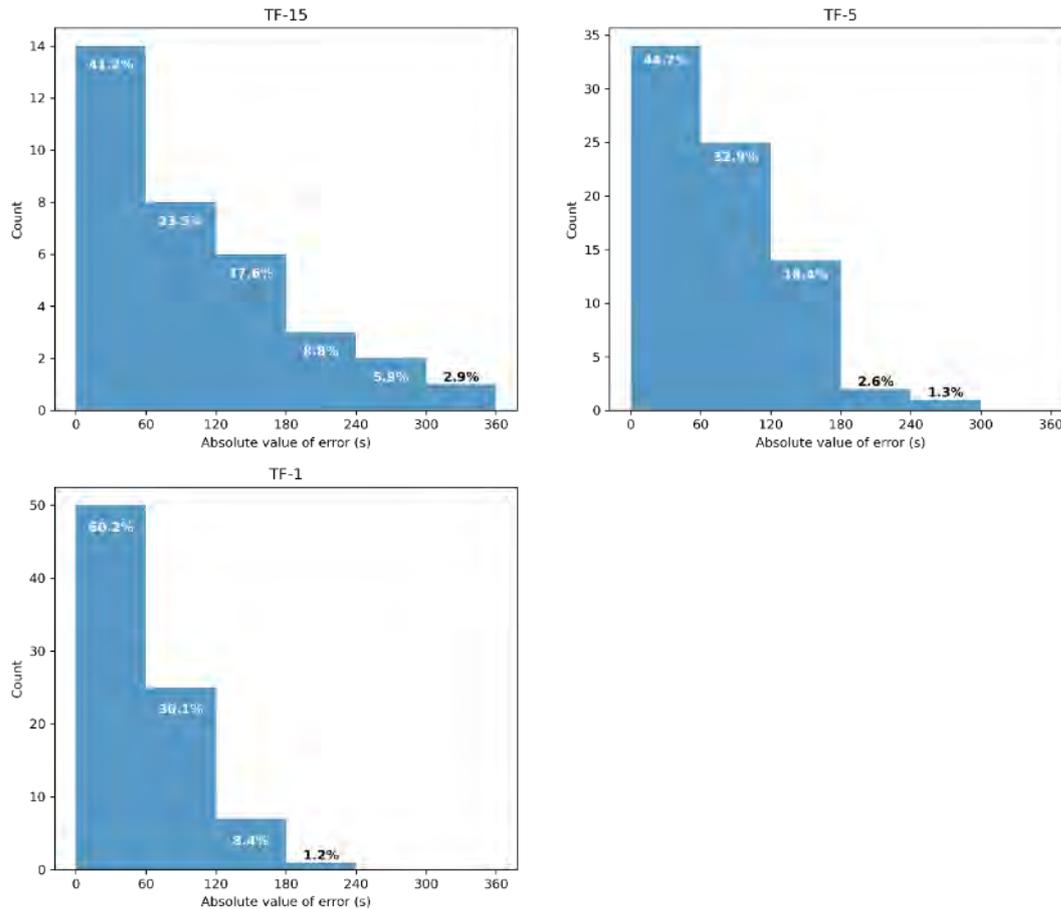


Figure 35 Histograms of the absolute differences between actual and predicted cruise flight duration for the London-Malta route, for the flights with a correct prediction of the lateral flight plan and for different time frames before EOBT.

2.4 Results

The PIU4TP project has defined a methodology for trajectory prediction on long, medium and short term, which is able to manage the uncertainties that by nature affect the input data to the trajectory prediction process.

The methodology has been assessed considering a simplified use case and simulated data. Specifically, only two factors that influence the selection of the optimal flight plan has been considered in the defined scenario, that is, weather conditions and take-off weight. Indeed, the project is a proof of concept and aims at providing evidences of the proposed methodology applicability and potential benefits arising from its use. Concerning the use of simulated data, this choice is due to the lack of open access datasets which provide a huge (thousands of flights) and coherent set of data including real aircraft trajectory and related flight plan, actual aircraft take-off weight and its forecast at different time in advance with related forecast uncertainties, weather data experienced during flight and their forecast at different time in advance with related forecast uncertainties. Moreover, the simulated data allow testing the methodology in a fully controlled environment, that is, the value of the parameters of interest and the rules and the assumptions that lead to perform the flight along a specific flight plan are perfectly known. The simulated data were produced within the framework of the PIU4TP project by defining suitable simulation models and exploiting the data found in the open access databases. The evaluation of the methodology performance has been carried out using some metrics suitably selected in the literature.

The tests performed to evaluate the methodology performance have proven the effectiveness in the prediction of the flight plan of an aircraft in the long, medium and short term before the estimated off-block time. The output includes a complete spatial prediction of the flight plan (horizontal and vertical) enriched with an estimation of the time of arrival on the waypoints of the flight plan (limited to the cruise phase of flight). The probability of the prediction is provided, too. The accuracy of the prediction depends on the time in advance with which it is computed and increases sharply as the time approaches the day of the flight, reaching values around 70% one day before the EOBT. This behaviour is expected because the weather forecasts improve and the uncertainties on the input data reduce as the EOBT approaches.

The obtained results highlighted that, in the considered simplified use case, the methodology is able to catch the information that are available in the input data, including the related uncertainties, and to exploit them to reliably predict in advance the flown trajectory. Next step will be the application of the methodology to more complex use cases, possibly using actual data, to fully assessing its performance.

3. Conclusions, next steps and lessons learned

3.1 Conclusions

The PTU4TP project investigated several data driven approaches to carry out the trajectory prediction task. It selected the most promising one, by using suitable metrics, and implemented it in a methodology applicable to long, medium- and short-term predictions.

Preliminary assessment of the methodology has been achieved considering a simplified use case, with a limited set of parameters affecting the flight plan selection, and analysing simulated data. The obtained results highlighted that the methodology is able to catch the information that are available in the input data, including the related uncertainties, and to exploit them to reliably predict in advance the flown trajectory.

In order to further mature the concept, future research shall focus on more complex use cases, which consider a wider set of input parameters, and analyse actual flight data. The application of the proposed methodology to these new use cases is straightforward, because the methodology takes the form of a Lifecycle Model for the analysis and modelling of flight paths in the context of trajectory prediction. It is iterative and incremental since it allows adding input parameters (such as aircraft type, airline, variables related to passenger connections, restricted areas due to military or national security activity) and external information (such as new flight route, new time-frames, etc.) by iterating through the phases of the lifecycle. This future research will finally assess the relevance of historical data to support the flight plan prediction task and permit to reach a good confidence level about practical usability of such data and of the proposed methodology to the real ATM world.

3.2 Next steps

Next steps listed in this section include the planned outputs related to the PIU4TP project and the potential further development of the research activities, which could be carried out through the participation to future SESAR Calls.

A planned project output concerns the dissemination activities and consists in the contribution to the Engage catalyst project showcases at the SESAR Innovation Days 2021 that will take place virtually on 7-9 December 2021 due to the COVID-19 pandemic crisis. The contribution will summarise the key project results and outcomes.

With reference to future activities to be performed in new projects, very useful recommendations were collected during the two Consultation Exercise meetings, thanks to the participation of several stakeholders. Main suggestions for future development activities of the project results to favour potential large-scale applications are:

- increase the number of input variables to the P4T methodology to consider other sources of uncertainty in the prediction of trajectories;
- perform sensitivity analyses to quantify the effect of the uncertainty in the input data on the uncertainty on the predictions.

Both these recommendations will be taken into account if opportunities for future project arise. It is important to highlight that a data-driven approach is fundamentally based on the use of data characteristic of the study domain. It would be desirable to have some flight data providers among the stakeholders of future projects, in order to have real data on which to apply algorithms and data analytics techniques to build more realistic use cases.

3.3 Lessons learned

The Engage KTN catalyst initiative is very interesting especially for young researchers because the planned duration of one year allows them to carry out study activities on specific thematic objectives. Furthermore, the initiative requires a low managerial load which favours of the project team to keep focus almost exclusively on study and research activities with obvious benefits on the project. Engage's managerial approach based on the assessment of technical documents and with the support of mentors certainly represents an added value of the initiative.

For the future it is hoped that this initiative will be continued.

4. Dissemination

ID	Title	Description
1	P4T: A Methodology to Support the Flight Trajectory Prediction. An introduction to the PIU4TP project. Short paper and presentation	This short paper and the associated poster presented the PIU4TP project at SESAR Innovation Days 2020 held on 7 – 10 December 2020
2	PIU4TP - Probabilistic information Integration in Uncertain data processing for Trajectory Prediction	Project presentation held at Engage workshop – Data-driven trajectory prediction, 25 January 2021, virtual event
3	Operational Scenario Definition	This document is the output of the WP2 Scenario Definition and the deliverable D1. It describes the approach adopted to generate the simulated dataset, which shall be sufficiently realistic in order to pave the way to any further methodology development. The document also defines some metrics applicable to assess the performance of the methodology
4	Methodology Description	This document is the output of the WP3 Methodology and the deliverable D3 of the project. It aims at sharing the approach used in the project for the development of a data-driven methodology applicable to the context of the trajectory prediction.
5	PIU4TP Web page	PIU4TP web page on the CIRA web site
6	Participation to OPTICS2 4 th year assessment	PIU4TP provided contribution to OPTICS2 4 th year assessment devoted to evaluate the extent to which the European Research Community is contributing to the achievement of Flightpath 2050 safety and security goals.

5. References

5.1 Project outputs

1. P4T: A Methodology to Support the Flight Trajectory Prediction. An introduction to the PIU4TP project. Short paper and presentation
<https://www.sesariju.eu/node/3438>
2. PIU4TP - Probabilistic information Integration in Uncertain data processing for Trajectory Prediction
Public presentation
3. D1 - Operational Scenario Definition
4. D3 - Methodology Description
5. PIU4TP Web page at <https://www.cira.it/en/aeronautics/sistemi-di-bordo-e-atm/piu4tp>
6. PIU4TP contribution in the OPTICS2 framework

5.2 Other

- [RD1] A. Rodríguez-Sanz, F. Gómez Comendador, R. M. Arnaldo Valdés, J. A. Pérez-Castán, P. González García, M. N. G. Najar Godoy, “4D-trajectory time windows: definition and uncertainty management”, Aircraft Engineering and Aerospace Technology, Vol. 91 No. 5, pp. 761-782, 2019, [doi: 10.1108/AEAT-01-2018-0031](https://doi.org/10.1108/AEAT-01-2018-0031)
- [RD2] L. Ma and S. Tian, “A Hybrid CNN-LSTM Model for Aircraft 4D Trajectory Prediction,” in IEEE Access, vol. 8, pp. 134668-134680, 2020, doi: 10.1109/ACCESS.2020.3010963
- [RD3] T. Zeh, J. Rosenow, R. Alligier and H. Fricke, “Prediction of the Propagation of Trajectory Uncertainty for Climbing Aircraft,” 2020 AIAA/IEEE 39th Digital Avionics Systems Conference (DASC), 2020, pp. 1-9, doi: 10.1109/DASC50938.2020.9256711
- [RD4] X. Guan, R. Lv, L. Sun, Y. Liu, “A Study of 4D Trajectory Prediction Based on Machine Deep Learning”, 2016 World Congress on Intelligent Control and Automation
- [RD5] X. Zhang, S. Mahadevan, “Bayesian neural networks for flight trajectory prediction and safety assessment”, Decision Support Systems, Vol. 131, 2020, doi: 10.1016/j.dss.2020.113246
- [RD6] D. Rivas, A. Franco, A. Valenzuela, “Analysis of aircraft trajectory uncertainty using Ensemble Weather Forecasts”, 7Th European Conference for Aeronautics and Space Sciences (EUCASS), 2017, doi: 10.13009/EUCASS2017-259
- [RD7] S. Sankararaman, M. Daigle, “Uncertainty Quantification in Trajectory Prediction for Aircraft Operations”, AIAA Guidance, Navigation, and Control Conference, 2017, doi: 10.2514/6.2017-1724
- [RD8] M. Uzun and E. Koyuncu, “Data-Driven Trajectory Uncertainty Quantification For Climbing Aircraft To Improve Ground-Based Trajectory Prediction,” ANADOLU Univ. J. Sci. Technol. - Appl. Sci. Eng., pp. 323–345, Jun. 2017.
- [RD9] R. Vazquez and D. Rivas, “Propagation of Initial Mass Uncertainty in Aircraft Cruise Flight,” J. Guid. Control Dyn., vol. 36, no. 2, pp. 415–429, Mar. 2013.
- [RD10] AIRBUS, “Getting to grips with the cost index”, Flight Operations Support & Line Assistance, Issue II - May 1998
- [RD11] A. de Leege, M. van Paassen, M. Mulder, A machine learning approach to trajectory prediction, AIAA Guidance, Navigation, and Control (GNC) Conference, 2013, p. 4782
- [RD12] ERA5, <https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5>
- [RD13] T. Haiden, M. Janousek, F. Vitart, L. Ferranti, F. Prates, “Evaluation of ECMWF forecasts, including the 2019 upgrade”, 2019
- [RD14] AIRBUS, “Getting to grip with Cost Index”, Issue II, May 1998
- [RD15] B.R. Barmish, C.M. Lagoa, “The Uniform Distribution: A Rigorous Justification for its use in Robustness Analysis,” Mathematics of Control, Signals, and Systems, 10, 203-222, 1997, doi: 10.1007/BF01211503.
- [RD16] Copernicus, <https://climate.copernicus.eu/climate-reanalysis>

- [RD17] ECMWF, <https://www.ecmwf.int/>
- [RD18] Fabec - Eurocontrol, "Maastricht Upper Area Control Centre. Stakeholder Management", 2018
- [RD19] A. Lemetti, T. Polishchuk, R. Sáez, and X. Prats, "[EN-A-31] Analysis of Weather Impact on Flight Efficiency for Stockholm Arlanda Airport Arrivals," ENRI Int. Workshop on ATM/CNS, Tokyo, Japan, EIWAC 2019
- [RD20] GRIB, <https://en.wikipedia.org/wiki/GRIB>

Annex I: Acronyms

Term	Definition
ATC	Air Traffic Control
ATFM	Air Traffic Flow Management
ATFCM	Air Traffic Flow and Capacity Management
ATM	Air Traffic Management
AUC	Area Under the Curve
CI	Cost Index
ECMWF	European Centre for Medium-Range Weather Forecasts
EOBF	Estimated Off-Blocks Time
FL	Flight Level
FP	Flight Plan
GS	Ground Speed
OEW	Operating Empty Weight
MAE	Mean Absolute Error
ML	Machine Learning
MSE	Mean Squared Error
MTOW	Maximum Take-Off Weight
NM	Network Manager
NOC	Network Operation Centre
P4T	Prediction for Trajectory
PIU4TP	Probabilistic information Integration in Uncertain data processing for Trajectory Prediction
RMSE	Root Mean Squared Error
ROC	Receiver Operating Characteristic
SSE	Sum of Squared Errors
SST	Total Sum of Squares
TAS	True Air Speed
TBO	Trajectory Based Operations
Tf	Time of flight
TOW	Take-Off Weight
TP	Trajectory Prediction
WP	WayPoint

Annex II: Summary of the performance of the models tested during the development of the methodology

Flight plan (horizontal and vertical) prediction

Approach	Algorithm name	Algorithm Description	Hyperparameters and Ranges
Neural Network	Multilayer Perceptron (MLP)	A classifier that uses backpropagation to learn a multi-layer perceptron to classify instances	<ul style="list-style-type: none"> HiddenLayers (H) $\in [(\#features+\#classes)/2, (\#features+\#classes)]$ TrainingTime (T) = # epochs to train through. T $\in [100, 1000]$ LearningRate (R) = the learning rate for weight updates. R = 0.3 Momentum (M) = Momentum applied to the weight updates. M = 0.2
Bayesian	Bayes Net (BayesNet)	Bayes Network learning using various search algorithms and quality measures	<ul style="list-style-type: none"> Estimator = Algorithm for finding the conditional probability tables of the Bayes Network = "Simple Estimator" that estimates probabilities directly from data with alpha (A) hyperparameter that is the initial count on each value. A $\in [0.01, 1]$ searchAlgorithm = the method used for searching network structures = K2 that is a Bayes Network learning algorithm using a hill climbing algorithm restricted by an order on the variables with P hyperparameter that is the maximum number of parents a node in the Bayes net can have. P $\in [1, \#features]$
Decision Tree	Inductive Decision Tree (J48)	Class for generating a pruned or unpruned C4	<ul style="list-style-type: none"> reducedErrorPruning = Whether reduced-error pruning is used instead of C.4.5 pruning = YES numFolds (F) = determines the amount of data used for reduced-error pruning. One fold is used for pruning, the rest for growing the tree. F $\in [2,15]$ confidenceFactor = the confidence factor used for pruning (smaller values incur more pruning) = 0.25 minNumObj = the minimum number of instances per leaf = 2
Decision Tree	Random Forest (RandomForest)	Class for constructing a forest of random trees	<ul style="list-style-type: none"> numIterations (I) = the number of trees in the random forest. I $\in [10,300]$ maxDepth = the maximum depth of the tree, 0 for unlimited = 0 numFeatures = sets the number of randomly chosen attributes = $\text{int}(\log_2(\#predictors) + 1)$

Table 10 - Hyperparameters Descriptions and Ranges of Variability

In Table 10, the hyperparameters descriptions of the applied Machine Learning algorithms, with their ranges of variability, are reported. For the sake of clarity, some of them are fixed, in accordance with the default suggestions of the SW tools and libraries used.

Table 11 and Table 12 show the performances of the trained models for the lateral flight plan and for the flight level, with the chosen hyperparameters of the selected models that had the max accuracy, after a fine-tuning step. Going deeper into the results obtained, decision trees (J48 and Random Forest) exhibited the highest maximum accuracies, while Bayesian Networks exhibited on average the lowest performance and often unsatisfactory.

Furthermore, for each time frame and for each route, the algorithmic approaches mostly showed the following and descending ordering of average and max accuracies (with very few exceptions):

1. Decision Tree (J48, Random Forest)
2. Artificial Neural Network (MLP)
3. Bayesian (BayesNet)

As already mentioned, the Bayesian approach with networks has often shown unsatisfactory performances, although their accuracies have shown a lower variability than that of the other algorithmic approaches.

In conclusion, the average performances of the decision trees are very high both considering the 10-fold cross-validation (see Table 11 and Table 12), the hold-out validation method, and the set of the 100 test flights.

10-fold cross-validation results for the lateral flight plan prediction							
Route	Time Frame	Algorithm name	Min Accuracy	Average Accuracy	Max Accuracy	Selected	Hyperparameters of model with Max Accuracy
London – Athens	TF-15	MLP	0.285	0.291	0.294	NO	H=16 ; T=500
London – Athens	TF-15	BayesNet	0.198	0.207	0.210	NO	A=0.5; P=2
London – Athens	TF-15	J48	0.290	0.378	0.385	YES	F=2
London – Athens	TF-15	RandomForest	0.275	0.369	0.382	NO	I = 130
London – Athens	TF-5	MLP	0.430	0.441	0.447	NO	H=19 ; T=750
London – Athens	TF-5	BayesNet	0.441	0.452	0.457	NO	A=0.01; P=2
London – Athens	TF-5	J48	0.481	0.525	0.534	YES	F=2
London – Athens	TF-5	RandomForest	0.452	0.498	0.520	NO	I = 207
London – Athens	TF-1	MLP	0.773	0.810	0.835	NO	H=27 ; T=500
London – Athens	TF-1	BayesNet	0.752	0.768	0.776	NO	A=0.32; P=1
London – Athens	TF-1	J48	0.861	0.874	0.891	NO	F=6
London – Athens	TF-1	RandomForest	0.987	0.998	0.999	YES	I = 250
London – Malta	TF-15	MLP	0.271	0.285	0.303	NO	H=24 ; T=342
London – Malta	TF-15	BayesNet	0.269	0.270	0.272	NO	A=0.41; P=3
London – Malta	TF-15	J48	0.311	0.315	0.358	YES	F=6
London – Malta	TF-15	RandomForest	0.284	0.304	0.320	NO	I = 245
London – Malta	TF-5	MLP	0.604	0.683	0.728	NO	H=18 ; T=682
London – Malta	TF-5	BayesNet	0.518	0.610	0.672	NO	A=0.3; P=1
London – Malta	TF-5	J48	0.592	0.692	0.795	NO	F=4
London – Malta	TF-5	RandomForest	0.712	0.790	0.812	YES	I = 20
London – Malta	TF-1	MLP	0.825	0.883	0.910	NO	H=40 ; T=558
London – Malta	TF-1	BayesNet	0.792	0.827	0.845	NO	A=0.73; P=1
London – Malta	TF-1	J48	0.966	0.970	0.992	NO	F=6
London – Malta	TF-1	RandomForest	0.972	0.991	0.999	YES	I = 250

Table 11 - Results for the lateral flight plan prediction

10-fold cross-validation results for the flight level prediction							
Route	Time Frame	Algorithm name	Min Accuracy	Average Accuracy	Max Accuracy	Selected	Hyperparameters of model with Max Accuracy
London – Athens	TF-15	MLP	0.221	0.310	0.399	NO	H=12; T=625
London – Athens	TF-15	BayesNet	0.311	0.364	0.378	NO	A=0.75; P=1
London – Athens	TF-15	J48	0.415	0.432	0.467	YES	F=2
London – Athens	TF-15	RandomForest	0.402	0.425	0.430	NO	I = 290
London – Athens	TF-5	MLP	0.401	0.494	0.599	NO	H=21; T=500
London – Athens	TF-5	BayesNet	0.338	0.388	0.401	NO	A=0.01; P=3
London – Athens	TF-5	J48	0.551	0.603	0.699	YES	F=3
London – Athens	TF-5	RandomForest	0.552	0.599	0.652	NO	I = 225
London – Athens	TF-1	MLP	0.661	0.725	0.790	NO	H=16; T=350
London – Athens	TF-1	BayesNet	0.620	0.715	0.782	NO	A=0.55; P=3
London – Athens	TF-1	J48	0.775	0.840	0.887	YES	F=2
London – Athens	TF-1	RandomForest	0.777	0.807	0.872	NO	I = 250
London – Malta	TF-15	MLP	0.228	0.301	0.383	NO	H=15; T=550
London – Malta	TF-15	BayesNet	0.350	0.352	0.359	NO	A=0.01; P=2
London – Malta	TF-15	J48	0.415	0.432	0.442	YES	F=6
London – Malta	TF-15	RandomForest	0.416	0.425	0.435	NO	I = 75
London – Malta	TF-5	MLP	0.499	0.591	0.628	NO	H=16; T=355
London – Malta	TF-5	BayesNet	0.511	0.531	0.589	NO	A=0.77; P=2
London – Malta	TF-5	J48	0.624	0.691	0.723	YES	F=8
London – Malta	TF-5	RandomForest	0.630	0.675	0.719	NO	I = 150
London – Malta	TF-1	MLP	0.669	0.729	0.822	NO	H=18; T=850
London – Malta	TF-1	BayesNet	0.617	0.728	0.751	NO	A=0.65; P=2
London – Malta	TF-1	J48	0.792	0.840	0.903	YES	F=4
London – Malta	TF-1	RandomForest	0.798	0.817	0.884	NO	I = 120

Table 12 - Results for the flight level prediction

Flight duration estimation

Approach	Algorithm name	Algorithm Description	Hyperparameters and Ranges
Neural Network	Multilayer Perceptron (MLP)	A regressor that uses backpropagation to learn a multi-layer perceptron to predict time of flight	<ul style="list-style-type: none"> • Number of hidden layers (N) ∈ [1, 2] • Number of units per layer (H) ∈ [8, 16, 32, 64, 128] • TrainingTime (T) = # epochs to train through. T ∈ [50, 500] • LearningRate (R) = the learning rate for weight updates. R = 0.001 • Momentum (M) = Momentum applied to the weight updates. M = 0.0
Decision Tree	Inductive Decision Tree	Class for generating a random tree regressor	<ul style="list-style-type: none"> • numFolds (F) = determines the amount of data used for reduced-error pruning. One fold is used for pruning, the rest for growing the tree. F ∈ [2, 50] • minNumObj = the minimum number of instances per leaf = 1
Decision Tree	Random Forest	Class for constructing a forest of random trees regressors	<ul style="list-style-type: none"> • numIterations (I) = the number of trees in the random forest. I ∈ [25, 400] • maxDepth = the maximum depth of the tree, 0 for unlimited = 0 • numFeatures = sets the number of randomly chosen attributes = auto, i.e. equals the number of input features

Table 13 - Hyperparameters Descriptions and Ranges of Variability

In Table 13, the Machine Learning algorithms used for the estimation of the duration of the cruise flight are reported with a description of the hyperparameters used for optimization and their ranges of variability, some of them are fixed to the default values suggested by the library used.

Table 14 shows the performances of the trained models for the prediction of the duration of the cruise flight, with the chosen hyperparameters of the selected models that had the minimum mean absolute error (MAE).

Note that the variables used for these models refer to the flight plan used during the execution of the flight, so the only time frame in Table 2 is TF-0. The reasons for this choice are described in the final technical report.

Hold-out validation results for the cruise flight duration prediction							
Route	Time Frame	Algorithm name	Min MAE	Average MAE	Max MAE	Selected	Hyperparameters of model with min MAE
London – Athens	TF-0	Decision Tree	8.5	13.6	82.7	NO	F=18
London – Athens	TF-0	MLP	6.5	31.9	178.0	NO	N=2 ; H=(128, 128) ; T=300
London – Athens	TF-0	Random Forest	4.9	5.1	5.2	YES	I = 150
London – Malta	TF-15	MLP	6.1	25.6	117.4	NO	N=2 ; H=(128, 128) ; T=350
London – Malta	TF-15	Decision Tree	6.6	9.9	50.8	NO	F=16
London – Malta	TF-15	Random Forest	3.9	4.1	4.5	YES	I = 150

Table 14 - Results for the cruise flight duration prediction

Annex III: Modelling results for flight plan prediction

Lateral flight plan prediction

Route London-Athens

Long-Term (15 days before EOBT)

Confusion matrix				Performance metrics				
		Prediction			Flight plan	Precision	Recall	F-Score
		1	2	3				
True Class	1	4699	661	1780	1	0.396	0.658	0.494
	2	3408	814	1648	2	0.365	0.139	0.201
	3	3764	753	1993	3	0.368	0.306	0.334
					Weighted	0.377	0.385	0.353
					Overall accuracy			
								0.385

Model: Induction Decision Tree

Table 15 – Long-term 10-fold cross validation results.

Confusion matrix				Performance metrics				
		Prediction			Flight plan	Precision	Recall	F-Score
		1	2	3				
True Class	1	1201	132	445	1	0.402	0.675	0.504
	2	822	212	438	2	0.408	0.144	0.213
	3	961	175	494	3	0.359	0.303	0.329
					Weighted	0.390	0.391	0.358
					Overall accuracy			
								0.391

Model: Induction Decision Tree

Table 16 – Long-term hold-out validation results.

Mid-term (5 days before EOBT)

Confusion matrix				Performance metrics				
		Prediction			Flight plan	Precision	Recall	F-Score
		1	2	3				
True Class	1	4200	1300	1640	1	0.547	0.588	0.567
	2	1747	2728	1395	2	0.514	0.465	0.488
	3	1727	1282	3501	3	0.536	0.538	0.537
					Weighted	0.533	0.534	0.533
					Overall accuracy			
								0.534

Model: Induction Decision Tree

Table 17 - Mid-term 10-fold cross validation results.

Confusion matrix				Performance metrics				
		Prediction			Flight plan	Precision	Recall	F-Score
		1	2	3				
True Class	1	999	376	403	1	0.525	0.562	0.543
	2	461	710	301	2	0.489	0.482	0.486
	3	443	366	821	3	0.538	0.504	0.520
Weighted						0.519	0.518	0.518
Overall accuracy								
0.518								

Model: Induction Decision Tree

Table 18 - Mid-term hold-out validation results.

Short-term (1 day before EOBT)

Confusion matrix				Performance metrics				
		Prediction			Flight plan	Precision	Recall	F-Score
		1	2	3				
True Class	1	7140	0	0	1	0.999	1.000	0.999
	2	5	5864	1	2	1.000	0.999	0.999
	3	3	0	6507	3	1.000	1.000	1.000
Weighted						1.000	1.000	1.000
Overall accuracy								
0.999								

Model: Random Forest l=250

Table 19 - Short-term 10-fold cross validation results.

Confusion matrix				Performance metrics				
		Prediction			Flight plan	Precision	Recall	F-Score
		1	2	3				
True Class	1	1778	0	0	1	0.998	1.000	0.999
	2	2	1470	0	2	1.000	0.999	0.999
	3	2	0	1628	3	1.000	0.999	0.999
Weighted						0.999	0.999	0.999
Overall accuracy								
0.999								

Model: Random Forest l=250

Table 20 - Short-term hold-out validation results.

Route London-Malta

Long-Term (15 days before EOBT)

Confusion matrix					Performance metrics			
		Prediction			Flight plan	Precision	Recall	F-Score
		1	2	3				
True Class	1	829	1047	4314	1	0.371	0.134	0.197
	2	681	1213	4456	2	0.358	0.191	0.249
	3	725	1132	4833	3	0.355	0.722	0.476
	Weighted						0.361	0.358
Overall accuracy								
0.358								

Model: Induction Decision Tree

Table 21 - Long-term 10-fold cross validation results.

Confusion matrix					Performance metrics			
		Prediction			Flight plan	Precision	Recall	F-Score
		1	2	3				
True Class	1	172	231	1127	1	0.360	0.112	0.171
	2	140	281	1193	2	0.364	0.174	0.236
	3	166	260	1237	3	0.348	0.744	0.474
	Weighted						0.357	0.352
Overall accuracy								
0.352								

Model: Induction Decision Tree

Table 22 - Long-term hold-out validation results.

Mid-term (5 days before EOBT)

Confusion matrix					Performance metrics			
		Prediction			Flight plan	Precision	Recall	F-Score
		1	2	3				
True Class	1	5089	579	522	1	0.791	0.822	0.806
	2	640	5185	525	2	0.808	0.817	0.812
	3	701	657	5332	3	0.836	0.797	0.816
	Weighted						0.812	0.812
Overall accuracy								
0.812								

Model: Random Forest I=20

Table 23 - Mid-term 10-fold cross validation results.

Confusion matrix				Performance metrics				
		Prediction			Flight plan	Precision	Recall	F-Score
		1	2	3				
True Class	1	1214	151	165	1	0.768	0.793	0.780
	2	174	1286	154	2	0.789	0.797	0.793
	3	193	193	1277	3	0.800	0.768	0.784
Weighted						0.786	0.786	0.786
Overall accuracy								
0.786								

Model: Random Forest l=20

Table 24 - Mid-term hold-out validation results.

Short-term (1 day before EOBT)

Confusion matrix				Performance metrics				
		Prediction			Flight plan	Precision	Recall	F-Score
		1	2	3				
True Class	1	6184	2	4	1	0.999	0.999	0.999
	2	3	6342	5	2	0.999	0.999	0.999
	3	4	4	6682	3	0.999	0.999	0.999
Weighted						0.999	0.999	0.999
Overall accuracy								
0.999								

Model: Random Forest l=250

Table 25 - Short-term 10-fold cross validation results.

Confusion matrix				Performance metrics				
		Prediction			Flight plan	Precision	Recall	F-Score
		1	2	3				
True Class	1	1528	2	0	1	0.997	0.999	0.998
	2	3	1609	2	2	0.998	0.997	0.998
	3	1	1	1661	3	0.999	0.999	0.999
Weighted						0.998	0.998	0.998
Overall accuracy								
0.998								

Model: Random Forest l=250

Table 26 - Short-term hold-out validation results.

Flight level

Route London-Athens

Long-term (15 days before EOBT)

Confusion matrix						Performance metrics			
		Prediction				Flight Level	Precision	Recall	F-Score
		330	350	370	390				
True Class	330	276	150	29	4	330	0.564	0.601	0.582
	350	171	288	119	40	350	0.411	0.466	0.437
	370	38	199	164	128	370	0.403	0.310	0.350
	390	4	64	95	183	390	0.515	0.529	0.522
							Weighted	0.390	0.391
Overall accuracy									
0.467									

Model: Inductive decision tree

Table 27 - Long-term 10-fold cross validation results.

Confusion matrix						Performance metrics			
		Prediction				Flight Level	Precision	Recall	F-Score
		330	350	370	390				
True Class	330	75	26	6	0	330	0.500	0.701	0.584
	350	62	60	39	16	350	0.405	0.339	0.369
	370	12	50	46	14	370	0.357	0.377	0.367
	390	1	12	38	31	390	0.508	0.378	0.434
							Weighted	0.431	0.434
Overall accuracy									
0.434									

Model: Inductive decision tree

Table 28 - Long-term hold-out validation results.

Mid-term (5 days before EOBT)

Confusion matrix						Performance metrics			
		Prediction				Flight Level	Precision	Recall	F-Score
		330	350	370	390				
True Class	330	358	99	2	0	330	0.810	0.780	0.795
	350	80	444	85	9	350	0.644	0.718	0.679
	370	4	133	322	70	370	0.643	0.609	0.625
	390	0	13	92	241	390	0.753	0.697	0.724
							Weighted	0.702	0.699
Overall accuracy									
0.699									

Model: Inductive decision tree

Table 29 - Mid-term 10-fold cross validation results.

Confusion matrix						Performance metrics			
		Prediction				Flight Level	Precision	Recall	F-Score
		330	350	370	390				
True Class	330	78	28	1	0	330	0.765	0.729	0.746
	350	22	113	41	1	350	0.693	0.638	0.665
	370	2	21	83	16	370	0.542	0.680	0.604
	390	0	1	28	53	390	0.757	0.646	0.697
			0	1	28	53	Weighted	0.682	0.670
Overall accuracy									
0.670									

Model: Inductive decision tree

Table 30 - Mid-term hold-out validation results.

Short-term (1 day before EOBT)

Confusion matrix						Performance metrics			
		Prediction				Flight Level	Precision	Recall	F-Score
		330	350	370	390				
True Class	330	415	44	0	0	330	0.950	0.904	0.926
	350	22	551	44	1	350	0.854	0.892	0.873
	370	0	50	447	32	370	0.863	0.845	0.854
	390	0	0	27	319	390	0.906	0.922	0.914
			0	0	27	319	Weighted	0.888	0.887
Overall accuracy									
0.887									

Model: Inductive decision tree

Table 31 - Short-term 10-fold validation results.

Confusion matrix						Performance metrics			
		Prediction				Flight Level	Precision	Recall	F-Score
		330	350	370	390				
True Class	330	94	13	0	0	330	0.940	0.879	0.908
	350	6	153	18	0	350	0.895	0.864	0.879
	370	0	5	109	8	370	0.838	0.893	0.865
	390	0	0	3	79	390	0.908	0.963	0.935
			0	0	3	79	Weighted	0.893	0.891
Overall accuracy									
0.891									

Model: Inductive decision tree

Table 32 - Short-term hold-out validation results.

Route London-Malta

Long-term (15 days before EOBT)

Confusion matrix						Performance metrics			
		Prediction				Flight Level	Precision	Recall	F-Score
		330	350	370	390				
True Class	330	328	101	47	5	330	0.565	0.682	0.618
	350	191	174	146	58	350	0.371	0.306	0.335
	370	56	147	170	159	370	0.359	0.320	0.338
	390	6	47	110	178	390	0.445	0.522	0.480
			6	47	110	178	Weighted	0.429	0.442
Overall accuracy									
0.442									

Model: Inductive decision tree

Table 33 - Long-term 10-fold cross validation results.

Confusion matrix						Performance metrics			
		Prediction				Flight Level	Precision	Recall	F-Score
		330	350	370	390				
True Class	330	79	22	11	2	330	0.513	0.693	0.590
	350	49	45	26	18	350	0.388	0.326	0.354
	370	21	35	39	47	370	0.424	0.275	0.333
	390	5	14	16	52	390	0.437	0.598	0.505
			5	14	16	52	Weighted	0.437	0.447
Overall accuracy									
0.447									

Model: Inductive decision tree

Table 34 - Long-term hold-out validation results.

Mid-term (5 days before EOBT)

Confusion matrix						Performance metrics			
		Prediction				Flight Level	Precision	Recall	F-Score
		330	350	370	390				
True Class	330	405	70	6	0	330	0.802	0.842	0.822
	350	91	371	100	7	350	0.681	0.652	0.666
	370	9	93	349	81	370	0.672	0.656	0.664
	390	0	11	64	266	390	0.751	0.780	0.765
			0	11	64	266	Weighted	0.721	0.723
Overall accuracy									
0.723									

Model: Inductive decision tree

Table 35 – Mid-term 10-fold validation results.

Confusion matrix						Performance metrics			
		Prediction				Flight Level	Precision	Recall	F-Score
		330	350	370	390				
True Class	330	104	10	0	0	330	0.765	0.912	0.832
	350	28	78	29	3	350	0.690	0.565	0.622
	370	4	25	84	29	370	0.651	0.592	0.620
	390	0	0	16	71	390	0.689	0.816	0.747
						Weighted	0.696	0.701	0.694
						Overall accuracy			
						0.701			

Model: Inductive decision tree

Table 36 - Mid-term hold-out validation results.

Short-term (1 day before EOBT)

Confusion matrix						Performance metrics			
		Prediction				Flight Level	Precision	Recall	F-Score
		330	350	370	390				
True Class	330	444	37	0	0	330	0.945	0.923	0.934
	350	26	511	31	1	350	0.886	0.898	0.892
	370	0	28	468	36	370	0.891	0.880	0.886
	390	0	1	26	314	390	0.895	0.921	0.908
						Weighted	0.904	0.903	0.903
						Overall accuracy			
						0.903			

Model: Inductive decision tree

Table 37 - Short-term 10-fold cross validation results.

Confusion matrix						Performance metrics			
		Prediction				Flight Level	Precision	Recall	F-Score
		330	350	370	390				
True Class	330	104	10	0	0	330	0.897	0.912	0.904
	350	12	117	9	0	350	0.860	0.848	0.854
	370	0	9	124	9	370	0.879	0.873	0.876
	390	0	0	8	79	390	0.898	0.908	0.903
						Weighted	0.881	0.881	0.881
						Overall accuracy			
						0.881			

Model: Inductive decision tree

Table 38 - Short-term hold-out validation results.