D5.2 Verification and Validation report

Deliverable 5.2

Pilot3

Grant:	**863802**
Call:	**H2020-CS2-CFP09-2018-02**
Topic:	**JTI-CS2-2018-CfP09-SYS-01-12**
Consortium coordinator:	**University of Westminster**
Dissemination level:	**Public**
Edition date:	**28 March 2022**
Edition:	**01.00**

EUROPEAN UNION

Authoring & Approval

Authors of the document

Name/Beneficiary	Position/Title	Date
Jovana Kuljanin / Universitat Politècnica de Catalunya	Project member	28 March 2022
Xavier Prats / Universitat Politècnica de Catalunya	Project member	28 March 2022
Júlia De Homdedeu / Universitat Politècnica de Catalunya	Project member	28 March 2022
Adeline Montlaur / Universitat Politècnica de Catalunya	Project member	28 March 2022
Luis Delgado / University of Westminster	Project member	28 March 2022
Paolino De Falco / University of Westminster	Project member	28 March 2022
Clara Argerich / Innaxis	Project member	28 March 2022
Damir Valput / Innaxis	Project member	28 March 2022
Ralph Schultz / PACE	Project member	28 March 2022

Reviewers internal to the project

Name/Beneficiary	Position/Title	Date
Xavier Prats / Universitat Politècnica de Catalunya	Project member	28 March 2022
Luis Delgado / University of Westminster	Project member	28 March 2022

Approved for submission to the CSJU

Beneficiary	Position	Date
University of Westminster	Project Coordinator	28 March 2022

Document History

Edition	Date	Status	Author	Justification
01.00	28 March 2022	Release	Pilot3 Consortium	New document for review by Topic Manager

Pilot3

A SOFTWARE ENGINE FOR MULTI-CRITERIA DECISION SUPPORT IN FLIGHT MANAGEMENT

This deliverable is part of a project that has received funding from the Clean Sky Joint Undertaking under grant agreement No 863802 under European Union's Horizon 2020 research and innovation programme.

Abstract

The deliverable provides the outcomes from the verification and validation activities carried during the course of work package 5 of the Pilot3 project, and according to the verification and validation plan defined in deliverable D5.1 (Pilot3 Consortium, 2020c). Firstly, it presents the main results of the verification activities performed during the development and testing of the different software versions. Then, this deliverable reports on the results of internal and external validation activities, which aimed to demonstrate the operational benefit of the Pilot3 tool, assessing the research questions and hypothesis that were defined at the beginning of the project.

The Agile principle adopted in the project accompanying with the five five-level hierarchy approach on the definition of scenarios and case studies enabled the flexibility and tractability in the selection of experiments through different versions of prototype development. As a result of this iterative development of the tool, some of the research questions initially defined have been revisited to better reflect the validation results.

The deliverable also reports the feedback received from the experts during the internal and external meetings, workshops and dedicated (on-line) site visits. During the validation campaign, both subjective qualitative information and objective quantitative data were collected and analysed to assess the Pilot3 tool. The document also summarises the results of the survey that were distributed to the external experts to assess the human-machine interface (HMI) mock-up developed in the project.

EUROPEAN UNION

Table of Contents

EUROPEAN UNION

List of figures

EUROPEAN UNION

EUROPEAN UNION

List of tables

Executive summary

This verification and validation report describes the results obtained through the project verification and validation activities execution, which aim to provide a thorough evidence of the status of the project prototype. Different prototype versions of Pilot3 tool have been released during the project development following the Agile principle and considering high-level functionalities. On each iteration, when a new functionality was implemented and before the release of the new version, the verification of the code was performed. Classical **verification** activities (both **static** and **dynamic**) were performed through the development of the prototype with software design technical reviews, code walk-through reviews, unit and interfaces testing, integration testing and functional testing. **System** testing, to verify that the requirements defined for Pilot3 are satisfied, was conducted prior to both software releases.

The activities performed and reported in this deliverable are:

1- Verification activities

- Pilot3 prototype **requirements test** - the test cases results defined for the final set of requirements are presented, identifying for each requirement a final qualitative classification. Out of the 33 final requirements, 26 have been deemed as passed, 5 as not passed and 2 as partially passed. Considering that the 5 not-passed test case results derive from optional requirements, the overall verification activity of the system requirements is considered to be successfully performed.
- **HMI requirements test** - new HMI requirements are presented in this deliverable in order to justify the verification status of the HMI design. A total of 18 requirements are presented, of which 14 have been deemed as passed, and 4 partially passed. Therefore, the overall HMI verification is also considered successful.

2- Internal validation activities (IVA):

Among the seven IVAs initially defined in deliverable D5.1 (Pilot3 Consortium, 2020c), four of them were finally executed. The main results are summarised as follow:

- **IVA1 -** Validation of Pilot3 optimised trajectory plans with PACE FPO trajectory plans: the results of this action demonstrated that the trajectories generated by Pilot3 are realistic and of similar expected performance compared to state-of-the-art Pacelab Flight Profile Optimiser (FPO) tool.
- **IVA2 -** Validation of indicators and estimators' prediction: the validation activities were performed independently of the Pilot3 framework and they showed that Performance Indicators Estimator (PIE) and Operational ATM Estimator (OAE) modules of Pilot3 are able to improve the estimation of parameters. At this validation phase, for each specific machine learning model developed, standard validation activities in a machine learning pipeline were performed to ensure the model performs well (accuracy) and according to the expectations. The machine learning and heuristics models were subsequently assessed by the experts within consortium to ensure that the development was moving in the "right direction".
- **IVA3 -** Assessment of the optimisation framework: by using a theoretical example of flight (but using realistic values and aircraft performance models), this validation action demonstrated how different choices of priorities for the airlines would lead to different results (rankings) from the Performance Assessment Module. The results of this action have been used to

analyse the sensitivity of the AHP-VIKOR algorithm implemented in the Performance Assessment Module.

- **IVA4 -** Pilot3 performance at generation of optimised trajectories plans: the results of the Pilot3 optimised trajectory plans have been analysed and compared against the operational flight plan (OFP) by using several metrics. The particular benefit of Pilot3 has been demonstrated for two types of flights: short/medium-haul (within Europe) and long-haul (with an oceanic segment). In addition to the metrics used for the purpose of the validation, the results of the Pilot3 optimised trajectory plan and the OFP were also provided in a graphical form depicting the vertical and speed trajectory profiles of the respective trajectories. In this context, the expected total cost function is also displayed, as a function of the arrival time and clearly identifying the incurred cost for both plans.
- **IVA7 -** HMI validation: the internal validation of the HMI was performed iteratively during the course of the project with the active participation of all consortium members. The partners eventually agreed that the final version of HMI met their individual requirements and expectations.

3- External validation activities (EVA):

Two EVAs were performed from the three initially planned. Namely:

- **EVA1 -** Pseudo-live demonstration of the HMI prototype and overall capabilities: the HMI prototype was presented in the form of mock-ups rather than an interactive dashboard including a number of screenshots for different HMI functionalities. The general feedback obtained during the Final Advisory Board meeting was that HMI prototype contains all important aspects relevant for the operations and may ease the decision-making process of the aircraft crew. A similar feedback was received by the survey sent to the Advisory Board members after the meeting.
- **EVA2 -** Presentation of the results obtained with stand-alone simulations at trajectory level: the external experts acknowledged that results of Pilot3 are, in general, meaningful and in line with current operational strategies/practice. Consequently, Pilot3 has a considerable potential to support the pilot in making a proper decision during the flight execution.

As already mentioned above, verification has been conducted in parallel to the model development and the internal validation actions have been executed after each software release, once the certain level of maturity of the prototype was ensured. EVA1 has been executed independently to the core prototype, since it has been a not integrated prototype design. Finally, EVA2 aimed at evaluating the benefits of the prototype and therefore has been executed with the external expert's panel with fully working versions of the code. At the end of the validation campaign, with all activities that were performed meanwhile, we succeeded to **successfully validate 10 research questions aimed for the internal validation** and **6 research questions defined for external validation**, out of 14 and 10 initially planned in D5.1 (Pilot3 Consortium, 2020c), respectively. It is worth emphasising that the research questions that have not been validated mainly stem from IVA5 and IVA6, which require additional development of the tool. Nevertheless, the successfully validated research questions proves that the results obtained by the Pilot3 tool met the expectations defined at the beginning of the project.

1 Introduction

1.1 Pilot3 background

Pilot3 aimed at developing a **software engine model** for supporting crew decisions for civil aircraft in the execution phase of flight. By triggering the tool, this software provides the crew with at least, **two trajectory options** along with information on different KPIs to aid the crew to select the most suitable one. The selection is performed considering the **multi-criteria business objectives** of the airline, including the **impact on the network** of flights of the airline of those decisions. With the current version of the tool, Pilot3 is capable of providing support in several domains:

- Gain understanding on the **types of performance** that airlines are seeking in order to translate these into high level objectives that can be formulated in **measurable parameters** which are relevant for airlines.
- Understand different **airlines policies and flight management policies** in order to identify the best **multi-criteria decision-making technique** to provide the crew with the different options and their trade-offs.
- Provide a tool which allows airlines define their preferences, **enriching their flight policies**.
- Develop an enhanced system to **estimate the different indicators** for the different trajectories alternatives. From using only available airborne information to the use of advanced machine learning trained ground predictors.
- Incorporate flexibility to **select which approach to use to estimate the indicators**, trading accuracy and complexity with efficiency, and considering the costs associated to develop enhanced indicators predictors by the airline and to operate the required data-links.
- Estimate the **overall impact of each trajectory option** not only for the current flight, but considering follow-up rotations of the same aircraft.
- Create a **software engine model** which can be used by crews to produce alternatives.
- Provide the **design of a possible HMI** for such a tool and a software interface for the system.

When a flight is disrupted, the crew faces different options and, nowadays, it could be difficult to understand their impact on the overall airline business policy. This is due to the fact that there are different parameters that should be considered at the same time, which can represent trade-offs such as total operating cost, adherence to a given flight schedule or the environmental impact of the flight. Moreover, understanding the full value of these indicators can be challenging as their overall value does not depend solely on the disrupted flight but on the whole network. For example, connecting passengers missing their connections might have a significant impact on the overall cost of a given flight, but these potential missed connections depend on the performance of other flights (e.g. if outbound connecting flights are delayed on their own); or uncertainty in the system means that suboptimal decisions can be selected, for example, speeding up a flight to encounter congestion at arrival airport. Pilot3 endeavoured to mitigate some of these problems by allowing the estimation of the performances of each alternative, not only based on the information available within a particular flight, but considering trained machine learning predictors.

One of the main objectives of Pilot3 is therefore, to **provide a comprehensive selection of options** with their associated trade-offs, considering the airline's business objectives, and to **maximise the likelihood that estimated values of those parameters are accurate**.

1.1.1 Trajectory optimisation scope

As presented in Figure 1, the following flight phases are considered:

- **Departure:** Departure procedure at the origin airport from take-off runway.
- **En-route**: From last point of the departure up to fix of the destination airport TMA (terminal manoeuvring area).
- **Arrival:** From the destination airport TMA entry fix to a given holding fix. Typically, this phase of flight will correspond with a STAR (standard terminal arrival route) or a transition.
- **Holding** (if any) at a given holding fix and altitude. Typically, the Initial approach fix (IAF) or just the end of the transition.
- **Sequencing and merging**: from the holding fix to the landing runway. It is assumed the aircraft might be subject to some tactical path stretching due to ATC instructions in this phase. Depending on the concept of operations in place at the destination airport, this phase could be in the form of standard approach procedure, tromboning, point merge or just conventional radar vectoring. In all these cases, this will be translated into a given distance to be flown as part of this final stage of the flight.
- **Taxi-in**: from the landing runway to the arrival gate in the terminal.

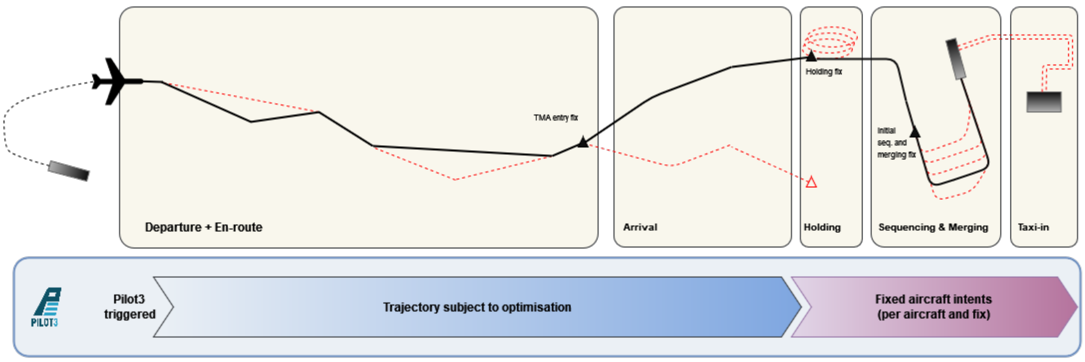


Figure 1 Trajectory optimisation scope

It should be noted that the departure phase does not correspond to the climb phase. Likewise, the en-route phase is not the cruise phase. Climb and cruise depend on aircraft performance (and weather), while departure and en-route procedures are set for ATM purposes. Thus, the top of climb (TOC) - the transition from climb to cruise - could either fall in the departure or en-route phase. Similarly, the top of descent (TOD) - transition from cruise to descent - could fall in the en-route or arrival phases.

Although the previous model is generic enough to consider a wide set of possible scenarios (airports, TMAs, procedures, etc.), in some busy TMAs more than one holding pattern is published along the

EUROPEAN UNION

same arrival. Yet, their usage is rare and only in extremely disrupted operations (temporary runway closure, for instance). Only the most frequently used holding fix (typically at the IAF) is modelled in Pilot3.

In the general case, Pilot3 could be **triggered** at any point of the flight (from the departure procedure to the en-route phase). Pilot3 **optimises** the aircraft trajectory from the current aircraft state (i.e., the moment Pilot3 is triggered) down to reaching FL100 at the proximity of the destination. After consultation with the Advisory Board, it is understood and assumed that from that altitude to the runway, the actions of the pilot are limited and standardised. Therefore, the trajectory plan will be computed assuming standard operations for this final segment (from FL100 to the runway) of the flight (i.e., a fixed sequence of aircraft intents). The reason of not optimising the trajectory after this altitude is twofold:

1. the optimisation control space is significantly reduced since the aircraft is flying near the limits of the flight envelope (i.e., min/max speeds), the standard operating procedures constraint significantly the trajectory (e.g. approach speeds, glide path on the instrument landing system - ILS) and ATM strategic constraints might also be in place (e.g. speed and altitude limitations for certain legs);
2. the aircraft trajectory is likely to be modified several times by tactical ATC intervention, thus forcing the pilot to no longer follow the Pilot3 plan.

Besides the appropriate models to define the optimisation cost functions, the trajectory optimisation engine uses weather forecast and an estimation (or assumptions) on operational ATM constraints that might affect the planning of the trajectory ahead. Weather itself is subject to uncertainty, but the operational ATM operations too, since they might depend on ATC tactical interventions, traffic conditions, airport operations and, also, weather conditions.

Pilot3 uses the most up to date weather mean forecast available, i.e., no uncertainty is explicitly modelled on the weather forecast. For the ATM uncertainty in the operations, as indicated in Figure 1, a significant part is experienced during the final part of the flight. Pilot3 does not models uncertainties which affect the lateral route length from the triggering point to reaching FL100 in the descent (e.g. shortcuts). Therefore, all operational uncertainty is concentrated on the following operational phases: holding time, if holding is experienced, distance to be flown during the sequencing and merging (understood as the distance from FL100 to the runway), and taxi-in time. As all these uncertainties are experienced after FL100 (even if the holding could be before, it is just a temporal displacement). These operational ATM uncertainties are directly integrated in the expected cost function used by the optimiser as presented in Section 1.1.2 and the optimiser is in that manner deterministic minimising the expected total cost of the operations.

1.1.2 Consideration on cost function modelling

As sown in Figure 2 the cost as a function of arrival time depends on the arrival time at the gate as this will be translated into potential reactionary delay, passenger compensations or missed connections, etc. This cost is non-linear and discontinuous. The cost function can be seen as a step-wise function, as increments are produced linked to events, e.g. passenger missing connections, reaching the curfew for having to compensate passengers due to Regulation 261, or breaching a curfew at the end of the day due to reactionary delay. Most of these parameters, however, have some degree of uncertainty. For example, if passengers miss their connection does not only depend on the arrival time of the flight but also on the actual time taken by passengers to do the connection at the airport and on the status of the remaining flights in the fleet (if the connecting flight is delayed on its own some extra-buffer is generated for these connecting passengers); or breaching a curfew will depend on the propagation of delay through the day, which is uncertain.

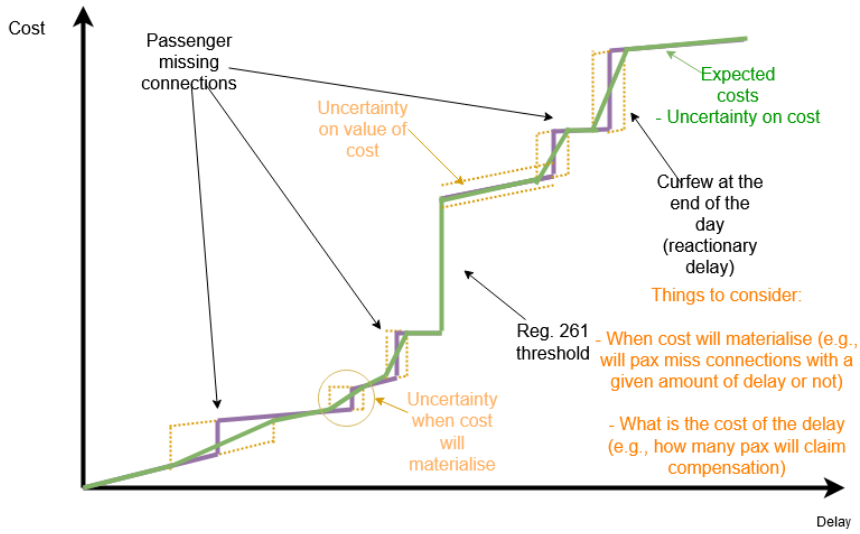
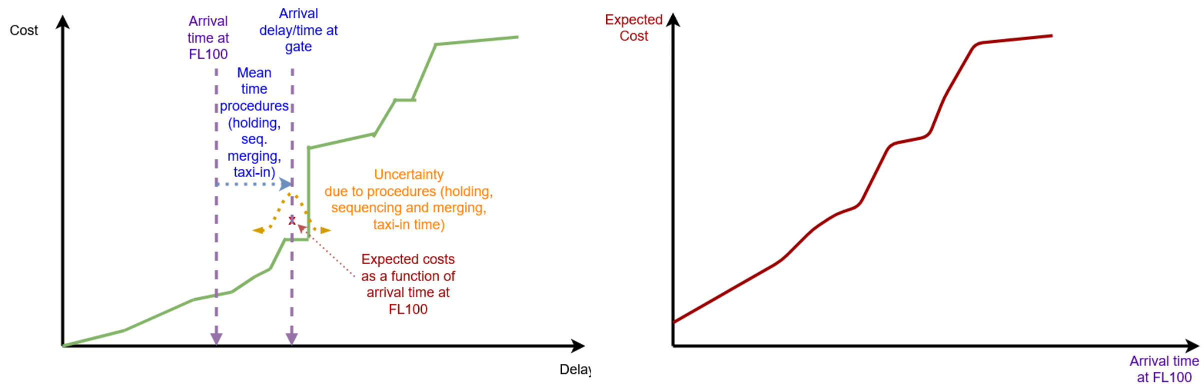


Figure 2 Cost of delay as a function of arrival delay at gate

After considering all these uncertainties, the total expected cost function as the one represented in Figure 3 (a) can be built. This cost function takes into account these internal uncertainties and it is computed with respect to the arrival time at the gate. This process will be done by the Performance Indicator Estimator as explained in Section 1.1.3.



(a) Integration from FL100 to runway

(b) Expected cost as a function of arrival time at FL100

Figure 3 Integration of uncertainty from FL100 to runway on cost function

As shown in Figure 3 (a) given an arrival time to FL100, the actual time of arrival at the gate will be in average that time plus the expected time at holding, the expected time of sequencing and merging and the expected taxi-in time. If the whole distributions are considered, their stochastic processes will be added, i.e., convolved, to produce the distribution of times when the flight will arrive to the gate as a function of the arrival time to FL100. Then as presented in the Figure, the expected cost can be computed considering the probability of arriving at the different times at the gate as a function of the arrival time at FL100 computing the expected value with a sliding-window.

EUROPEAN UNION

Therefore, after integrating in a sliding-window the stochastic process of the holding, sequencing and merging and taxi-in times, the results is an expected cost function (as shown in Figure 3 (b)), which will be smoother and shifted in average the expected time of the arrival processes, that will be experienced by the flight as a function of the arrival time at FL100. In this way, all the arrival uncertainty has been directly considered in this cost function. As presented in Section 1.1.3 Section, this integration of uncertainty will be carried out by the Objective Function Generator with the uncertainties estimated by the Operational ATM Estimator.

1.1.3 Architecture and components

Figure 4 presents the high-level view of the different components of Pilot3. The main characteristics and responsibilities of the different modules are as follow:

- **Performance Indicators Estimator**: Provides to the objective function estimator the cost of fuel and the expected costs at the arrival gate incorporating their intrinsic uncertainties. Some uncertainty on the materialisation of costs might exist (e.g. costs associated with potential reactionary delay or passenger related costs to miss connections which might or not occur). These uncertainties are estimated an integrated in the cost of delay functions by the Performance Indicator Estimator. Three main estimators are implemented: **cost fuel**, **IROP costs** and **other costs** estimators. These three estimators, however, require a set of different costs and operational estimators to individually compute the different components of the cost function (e.g. models to estimate the reactionary delay and associated costs). Therefore, not only the aggregated cost functions are produced but also information on their components, which might be useful for human-machine interface (HMI) purposes. Moreover, during the configuration of Pilot3, the user can indicate which estimator should be used for each component (e.g. heuristic or machine learning).
- **Operational ATM Estimator**: Which estimates uncertainties associated with operational aspects which might affect the trajectory and the cost such as taxi time, sequencing and merging distance or holding time. Like with previous module, the user can configure different ways to perform these estimations (e.g. heuristic or machine learning).
- **Alternative Generator**: The alternatives generator is in charge of the optimisation of the trajectory from the triggering point. It is composed, in turn, of different elements:
    - **Objective function estimator**: which, using the outcome of the Performance Indicators Estimator and of the Operational ATM Estimator, integrates the uncertainty in the cost of delay function producing the expected cost as a function of arrival time at FL100 in the descent towards the destination airport.
    - **Trajectory optimiser,** which using an optimisation framework and produces trajectories which minimise the total expected cost (delay - IROP and other- and fuel) by modifying the vertical and speed profile of the flight plan. This can be done by optimising the Cost Index or with a full altitude/speed grid-search.
    - **Trajectory predictor:** is used to estimate the fuel and time from FL100 to the runway, to translate the uncertainty from the Operational ATM Estimator into uncertainty on time and fuel.
- **Performance assessment module**: Filters and ranks the different available trajectories considering if they meet or not OTP and their costs following the airline policy on prioritisation of cost components.

- **Human machine interface**: Mock-up designed in Pilot3 to provide the crew with information on the different alternatives generated and their impact on the airline objectives.
- **Data manager**: in charge of provide with the required data to the different components of Pilot3. The architecture allows for the distinction of air and ground data, and system status data

All these components have been created during the project with a certain technology readiness level (TRL) which is mainly 4. More details on the approach followed in Pilot3 are presented D4.3 – Crew Assistant Decision model description (final release) (Pilot3 Consortium, 2022a).

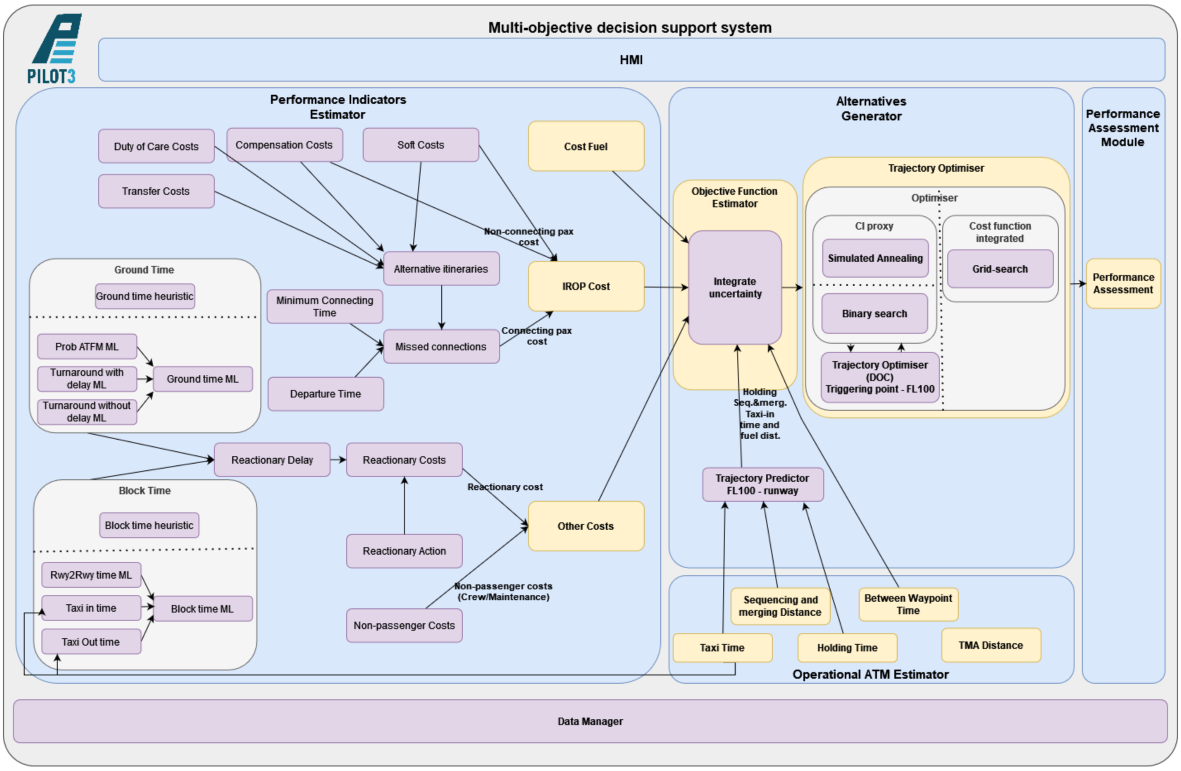


Figure 4 Different components of Pilot3

1.2 Verification and validation plan

The verification and validation plan initially defined in deliverable D5.1 (Pilot3 Consortium, 2020c), is a complex document built on the principle of the Agile methodology. The Agile methodology proposes breaking up the project into several parts in order to manage it in a more efficient way. In addition, it also involves constant collaboration within stakeholders and continuous improvement at every stage. Following the Agile principles, during the course of the project different prototype versions have been developed by gradually upgrading each subsequent version until we finally reached the fully functional prototype. However, D5.1 already acknowledged that the line between verification and validation is very often blur, indicating that some verification and validation activities may overlap. In this vein, we attempted to follow the guidelines defined in the verification and validation plan in order to minimise the potential deviation from the plan and enable the seamless workflow.

EUROPEAN UNION

1.3 Verification and validation activities and limitations

According to the definitions provided in D5.1, the software developed in Pilot3 was first subject to a verification process, which was carried as part of the developed following the Agile methodology explained above (Pilot3 Consortium, 2020c). On each iteration, when a new functionality was implemented and before the release of the new version, the verification of the code was performed. Classical **verification** activities (both **static** and **dynamic**) were performed through the development of the prototype with software design technical reviews, code walk-through reviews, unit and interfaces testing, integration testing and functional testing. **System** testing, to verify that the requirements defined for Pilot3 are satisfied, was conducted prior to both software releases.

In addition, the validation actions performed during the project aimed at quantifying the performance of the fully functional prototype (with internal validation actions) and evaluating the acceptance of the solution with external experts (external validation actions). These actions targeted the functionalities of Pilot3 (considering the optimised trajectories plan) and the **Human Machine Interface (HMI)** designed for the tool. Among the seven different internal validation actions that were envisioned to be performed by the plan, the five of them were eventually executed during the course of the project. Namely:

- **Actions aiming at validating the different components of the model**
    - **IVA1 - Validation of Pilot3 optimised trajectory plans with PACE FPO trajectory plans** - the aim of this action is to compare the result of Pilot3 with state-of-the-art FPO tool, ensuring that the trajectories generated by Pilot3 are realistic and with similar (or better) expected performance. These actions focus on evaluating the Trajectory Generator of Pilot3.
    - **IVA2 - Validation of indicators and estimators' prediction** - the aim of these actions is to validate the capabilities of the performance indicators (from the Performance Indicators Estimator module of Pilot3), and of the ATM uncertainties estimations (from the Operational ATM Estimators module).
    - **IVA3 - Assessment of the optimisation framework** - the objective of these actions is to assess how Pilot3 is able to generate different alternative trajectories and trade-offs.
- **Actions aiming at assessment the benefit of Pilot3**
    - **IVA4 - Pilot3 performance at generation of optimised trajectories plans** - the objective of this step is to assess the benefits of Pilot3 optimised trajectories plans against several baseline plans at the moment of considering the decision by the pilot. I.e., comparison of Pilot3 alternatives suggested to pilot with respect to baselines (original flight plan, or basic pilot trajectory behaviour).
- **Actions aiming at the validation of the HMI**
    - **IVA7 - HMI** -these action aims to ensure that HMI prototype is well designed with respect to the information and mechanism available to the pilot.

EUROPEAN UNION

For the purpose of internal validation actions and to address the benefits of Pilot3, a set of experiments were employed accompanied by a set of research questions (RQs) and their corresponding hypotheses (HPs). As some validation actions (i.e., IVA5 and IVA6) have not been performed during the project, it is worth mentioning that some RQs and HPs related to these actions, were eventually discarded. Instead of organising two internal workshops that were scheduled in D5.1, the internal validation activities were rather conducted in a more iterative way encompassing a large number of bi- and three lateral meetings among the partners within the consortium (Pilot3 Consortium, 2020c). In this way, we succeeded to refine the experiments and discuss the results obtained by different prototype versions in order to bring them closer to real operations and thus, maximising the benefits of Pilot3 tool.

The external validation was conducted using fully functional versions of the prototype and based on the results of experiments studies performed in the internal validation. Two external workshops were organised, as well as a continuous interaction with the Advisory Board was maintained in order to provide input into the project. The first external workshop aimed at collecting the feedback on the PIlot3 architecture and general capabilities. In addition, at the second external workshop, some results obtained with fully-matured version of the software were presented and discussed together with the experts from the Advisory Board. The external validation was performed through two main types of actions:

1. **EVA1 - Live or pseudo-live demonstration of the HMI prototype and overall capabilities** - the objective of this external validation action is to validate the interface, how the information is presented to and gathered from the crew, and to show the overall capabilities of Pilot3.

2. **EVA2 - Presentation of results obtained with stand-alone simulations at trajectory level** - in this case, the results from the experiments executed in the internal validation IVA4 were used. The objective was to validate the relevance of the findings.

However, it is worth mentioning that the verification and validation activities performed during the course of the project slightly deviate from what was planned in D5.1 (Pilot3 Consortium, 2020c). This is anticipated due to the complex nature of the project with a backlog of tasks emerged during the preparation of different datasets and developments of different software versions. For instance, in the absence of the simulator which would enable to perform **several simulations** considering different realisations of uncertain parameters, IVA5 were eventually exempted from the report. In similar manner, IVA6 that aimed at evaluating the impact of operating a fleet of aircraft equipped with Pilot3 through a day of operations, was not performed during validation campaign. Namely, this action required the substantial effort to be allocated to the configuration of the fast time simulations, which was not critical for Pilot3.

Bering in mind that EVA3 fully relies on the results of IVA6, this is the main reason why this validation action is not reported in this deliverable either.

1.4 Deliverable structure

This document is organised in seven sections and four appendices:

- Section 1 introduces the context of Pilot3 decision support tool for crew support on trajectory management. Then it briefly presents the verification and validation activities performed during the Pilot3 prototype development.

- Section 2 provides an overview of the different scenarios and case studies that were considered in the project.
- Section 3 lays out the verification activities performed during the development of different software versions. The final list of functional and non-functional requirements is also described in this section.
- Section 4 reports the validation activities performed during the project. It separately details validation activities performed internally within the consortium and those conducted in a close interaction with the experts from the Advisory Board.
- The document closes with some conclusions (in Section 5).
- References and acronyms are provided in Sections 6 and 7 respectively.
- Finally, the document contains four Appendices:
  - Appendix A provides the weather analysis for scenario 201 (Madrid - Frankfurt) to identify weather forecast which are representative of specific operational days.
  - Appendix B presents the HMI for Pilot3 and the results of the survey distributed to the Advisory Board members as a part of external validation of HMI.
  - Appendix C provides a literature review on the modelling of uncertainty with machine learning models.
  - Appendix D contains a detailed example of the computation of reactionary cost as performed in Pilot3 step by step for the consideration of pre-tactical/strategic actions.

2 Definition of scenarios

The Agile methodology adopted in the project required a special consideration for definition of the experiments that will be eventually run for verification and validation purposes. As already explained in detail in D5.1 (Pilot3 Consortium, 2020c), the validation of Pilot3 will be based on the simulation of specific flights in given conditions taking into account operational aspects (e.g. airline type, cargo, number of passengers and connections), environmental/external considerations (e.g. weather, ATFM (Air Traffic Flow Management) conditions), the event which triggers the use of Pilot3 (e.g. late arrival at TOC (top of climb) with respect to planned), or the configuration of Pilot3 tool itself.

In order to efficiently track the progress of different verification and validation activities and their associated experiments, we followed a five-level hierarchy that has been defined in D5.1. For the sake of traceability of D5.2, we will recall the definition of each of the five components that constitutes experiments:

1. **Scenario** is high-level item linked to **specificities of the routes and operations** that are modelled. A scenario specifies operational variables such as origin-destination pair, airline characteristics, baseline flight used to define this scenario.
2. **Sub-scenario** further particularises the **operational environment** (i.e., "external" factors), such as, type of weather ATM characteristics.
3. **Case study** is related to the different **events that may trigger Pilot3**.
4. **Sub-case study** is related to the different possible **configurations of Pilot3** (e.g. different ways to estimate the performance indicators).
5. **Parametrisation** refers to changing **parameters** that define a (sub)scenario or (sub)case-study to allow sensitivity studies.

The combination and particularisation of these five components provide a specific condition into where to test Pilot3 and this is considered an **experiment**.

2.1 Scenarios used for the verification and validation

The scenario is placed at the first hierarchy level specifying the most basic variables which help to particularise the specific flight (see the columns of Table 1). Among the nine scenarios initially identified in D5.1, four of them are executed as listed in Table 1. Note that since the scenarios were highly data-driven, their creation was typically time and effort consuming as it required different activities involving data collection, data preparation, development and verification and validation. Nevertheless, the four scenarios cover different operational contexts and were identified as relevant in close interaction with the experts from the Advisory Board that was continuously maintained along the whole project. Following the methodology for the creation of experiments defined in D5.1., a batch of experiments encountering different instantiation have been run to identify those that can reflect the benefit of Pilot3 in its most magnitude. Table 1 summarises the four different scenarios eventually

EUROPEAN UNION

executed and assessed internally among the consortium teams. As shown in Table 1, the particular benefit of Pilot3 have been demonstrated for two types of flights:

- short/medium-haul (within Europe). For this type of flight, the following scenario have been considered:
  - Athens (LGAV) - London Heathrow (EGLL) (scenario ID 100)
  - Madrid (LEMD) - Frankfurt (EDDF) (scenario ID 201)
- long-haul (with oceanic segment). The following long-haul flights were investigated:
  - New York (KJFK) - Frankfurt (EDDF) (scenario ID 600)
  - New York (KJFK) - London Heathrow (EGLL) (scenario ID 800)

Table 1: Description of scenarios used in the verification and validation campaign

Scenario ID	Example Airline OD pair	Type of route	En-route uncertainty	Destination characteristics	Airline type	Destination Type	Time of the day
P3-SCN-100	BAW: LGAV (ATH)- EGLL (LHR)	Intra-ECAC	Normal	Holding	FSC	Hub	Morning
P3-SCN-201	DLH: LEMD (MAD) - EDDF (FRA)	Intra-ECAC	Normal	Tromboning	FSC	Hub	Morning
P3-SCN-600	DLH: KJFK (JFK) - EDDF (FRA)	North - Atlantic	High	Tromboning	FSC	Hub	Morning
P3-SCN-800	BAW: KJFK (JFK) - EGLL (LHR)	North - Atlantic	High	Holding	FSC	Hub	Morning

The distinctive characteristics of two arrival airports considered in the experiments (i.e., London Heathrow and Frankfurt) with respect to their arrival procedures in terminal manoeuvring area (TMA) allows us to observe the benefits of Pilot3 in the context of the very dense TMA currently found at the ECAC area and operating with two different main methodologies for sequencing and merging arrival traffic: tromboning (Frankfurt) and holdings (London Heathrow).

Next subsections further particularise each of the four scenarios identified above providing the information on their respective operational flight plans (OFP) and cost function specification. The OFP description contains the information split into three categories:

- **Flight schedule** -this section provides the information on:
  - Aircraft type operated,
  - Origin and destination airports,
  - Schedule off-block time (SOBT) - the time that an aircraft is scheduled to depart from the parking position,
  - Schedule in-block time (SIBT) - the time that an aircraft is scheduled to arrive at its first parking position,
  - Scheduled block time - the time duration between the scheduled departure time and the scheduled arrival time for a given flight.

- **Flight dispatch** - the information provided within section are as follows:
    - Nominal cost index (CI) - the CI is the ratio of the time-related cost of an airplane operation and the cost of fuel. Nominal CI for the scenarios with short/medium-haul routes (LGAV-EGLL and LEMD-EDDF) was set to 10 kg/min, while for long-haul scenarios (KJFK-EDDF and KJFK-EGLL) the CI was specified to 20 kg/min.
    - Payload - total amount of payload in the flight divided into cargo and passenger load. According to the EU-OPS 1.620 (Official Journal of European Union, 2008) for the flight "within the European region", the standard mass of an adult passenger accounts for 84kg in addition to 13kg of luggage. In the case of "Intercontinental" flight, the regulation specifies the same mass for an adult passengers (i.e., 84kg) but with 15 kg of luggage mass.
    - Number of connecting passengers.
    - Weather forecast - the nominal weather forecast for LEMD - EDDF route is based on the statistical analysis of a large datasets containing the information on weather ensembles of the whole 2018 year. The day eventually selected was 02-07-2018 based on two criteria: average wind and ISA temp and small forecasting error (see additional information on weather analysis in Appendix A). The weather forecast for the three other scenarios (LGAV-EGLL, KJFK-EDDF and KJFK-EGLL) considered the available weather forecast for 28-07-2016.
- **Other operational information** further details the following aspects:
    - The published standard instrumental departures, arrivals and approaches in the concerned TMAs.
    - Taxi-in time and buffer: the OFP computed provides the information on the expected landing time (ELDT), which in combination with SIBT allows us to calculate taxi-in time and buffer (i.e., difference between SIBT and ELDT)
    - Taxi-out time: the time spent by a flight between its actual off-block time (AOBT) and actual take-off time (ATOT).

The OFP trajectory has been computed using the same trajectory optimisation engine that is embedded in the Pilot3 software prototype, based on the UPC in-house tool Dynamo (Dalmau et al., 2018). The lateral trajectory (i.e., sequence of waypoints) was directly taken from EUROCONTROL's Demand Data Repository 2 (DDR2). Then, the vertical trajectory was optimised with Dynamo taking into account the flight schedule, dispatch and operational considerations enumerated above. The optimisation criterion for this vertical optimisation was the standard direct operating cost function (i.e., *Fuel + CI·Time*) and it is assumed that (for each cruise flight level) the cruise Mach is kept constant. In this optimisation, aircraft performance data (and models) were taken from EUROCONTROL's BADA V4.2.

Modelling the cost function that will be then used when triggering Pilot3, is a very complex task as it considers passenger related cost (connecting and non-connecting) and other costs (including reactionary delay, curfew, etc). In this regard, it substantially differs from the most widely used approach, which is based on the **cost index** definition (such the cost function used to compute the OFP). For each specific scenario that has been used during the validation campaign, Pilot3 has modelled the expected costs at gate and, taking into account uncertainty for taxi-in and sequencing and merging operations, the expected costs at FL100 are derived. Below FL100, the trajectory is no longer optimised and a predefined sequence of aircraft intents is used to compute the remaining trajectory down to the runway threshold, which take into account operational restrictions in the TMA and the operating particularities for each aircraft type. The detailed description of each specific cost functions will be provided in Section 4.4

Next sections provide, for each scenario, the details on the OPF trajectory that will follow for the different validation experiments.

2.1.1 OFP description for SCN 100: LGAV (ATH) to EGLL (LHR)

The main characteristics of the operational flight plan (OFP) for the flight between Athens and London Heathrow are summarised in Table 2 below.

Table 2: Main characteristics of SCN 100 OFP

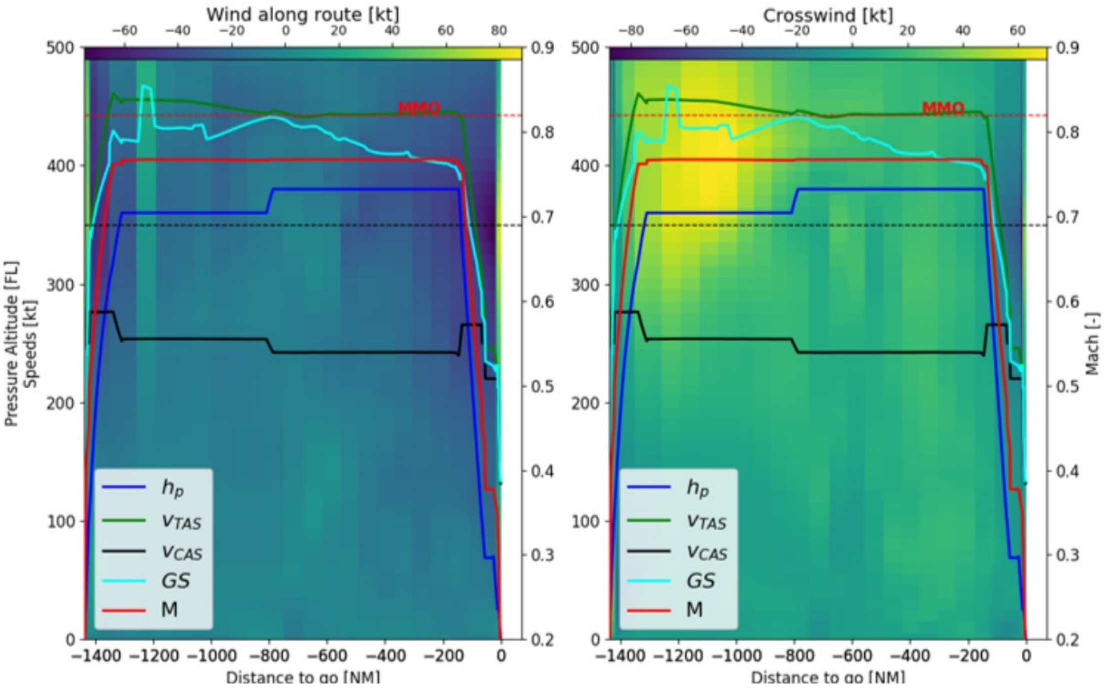
Flight schedule	Flight dispatch	Other operational information

• Airline: British Airways (BAW)
• Aircraft type: A320-231
• LGAV - EGLL
• Scheduled off-block time (SOBT): 5h 15min UTC
• Scheduled in block time (SIBT): 9h 10min UTC
• Scheduled block time: 235’
	• Nominal Cost Index: 10 kg/min
	• Payload:
	  ○ Passengers 144
	  ○ Cargo: 1,000 kg
	• Number of connecting pax: 124
	• Weather forecast: nominal, issued 2016-07-28
	• Passengers entitled to Regulation 261 compensation if delay thresholds meet
		• Taxi-out: 10’
		• OFP trip time: 216’
		• Taxi-in + padding at arrival: 9’
		• LOGAN 2H arrival
		• Holding at LAM
		• ILS approach to RWY 09R

Figure 5 below depicts the horizontal and vertical trajectory profiles of the OFP. The horizontal trajectory (see Figure 5 (a)) is given as a sequence of waypoints defining the OFP route. Climb, cruise and descent phases are represented, respectively, by green, blue and red trajectories in this figure. Figure 5 (b), in turn, shows the vertical and speed profiles of the OFP trajectory, along with the along track and cross-wind components at different altitudes (coloured background). In these plots, pressure altitude (hp) for the whole trajectory is depicted together with Mach number (M), calibrated airspeed (CAS), true airspeed (TAS) and ground speed (GS). It is worth mentioning that the (apparently) sudden changes in ground speed of these figures (such as observed at around 1300 NM from the destination airport) are due to track changes in the lateral route, which change the relative wind direction along and cross-track and therefore the resulting ground speed. Finally, these plots also depict the maximum operational speeds for that aircraft type: MMO (maximum Mach in operation) and VMO (maximum CAS in operation).

As observed in Figure 5 (b) the OFP for this scenario consists on an initial cruise at FL360 followed by a step-climb to FL380 at around 850 NM from the destination airport. The optimal cruise speed resulting for this OFP is M0.77. As observed in the figures, the first half of the cruise is mainly affected by a relative strong crosswind component (around 60 kt), while a relative mild headwind and crosswind components dominate the remaining cruise.

a) Horizontal trajectory profile

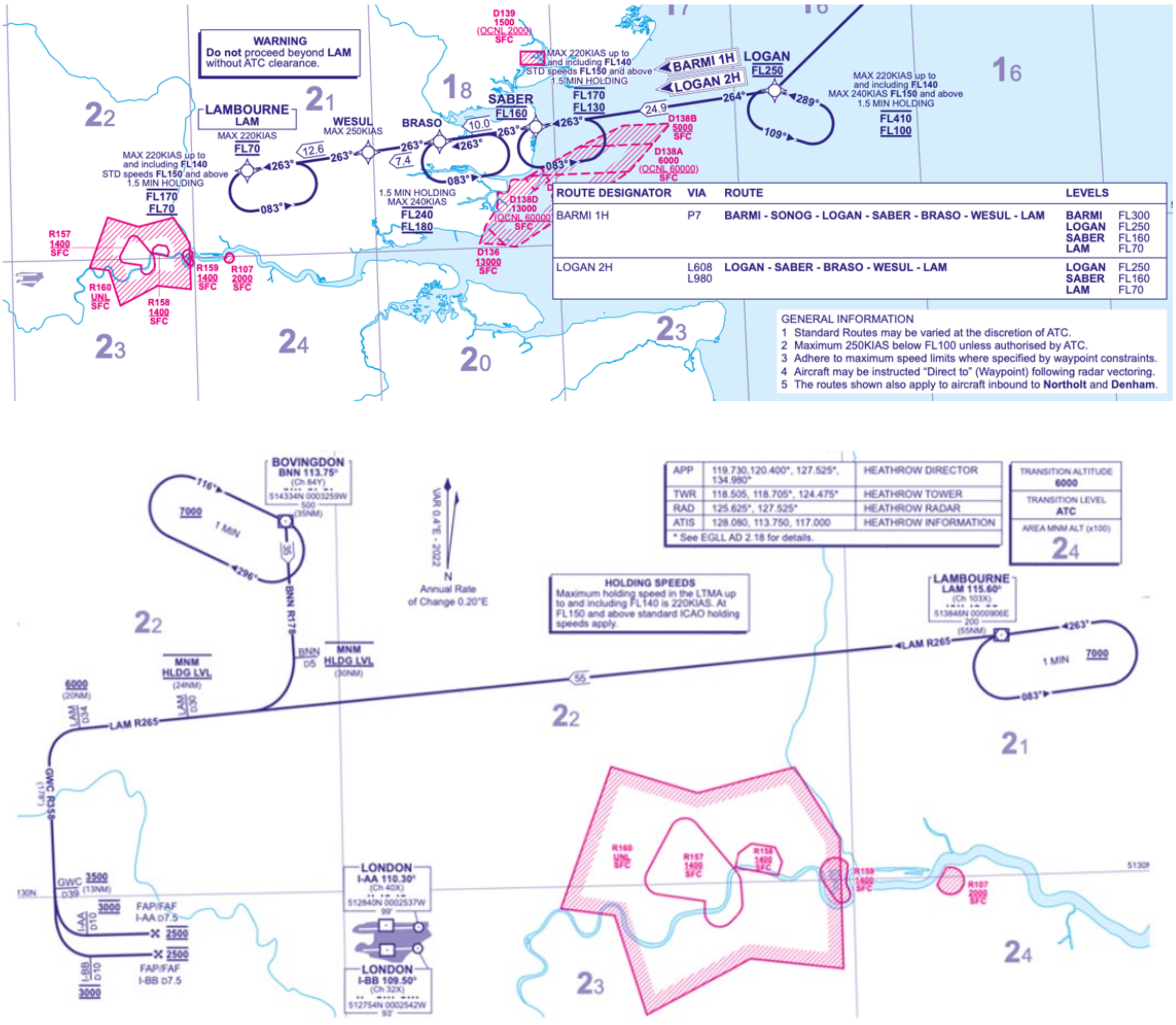


b) Vertical and speed trajectory profiles (with wind components)

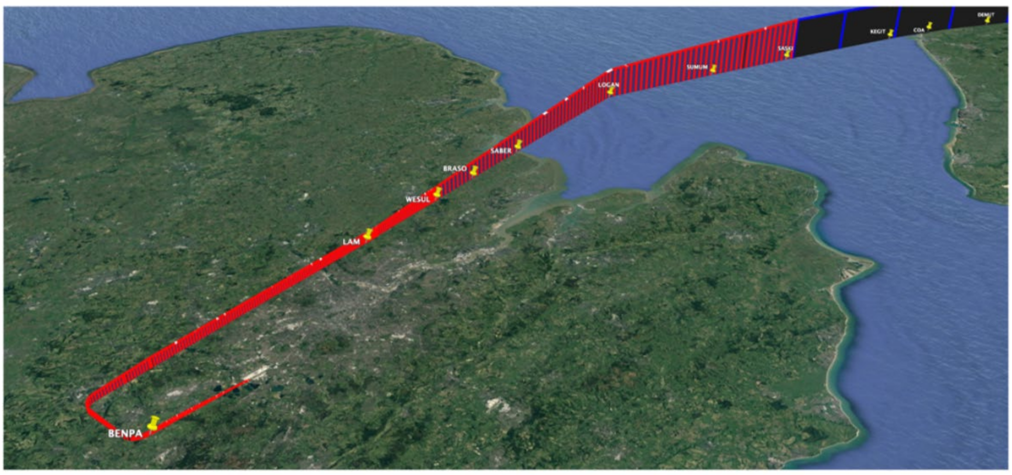
Figure 5 Resulting trajectory for SCN-100 operational flight plan (OFP)

For this scenario, the STAR procedure LOGAN 2H is used in London Heathrow TMA, which ends at Lambourne fix (LAM), where the holding pattern is located. Then, for flight and fuel planning purposes (i.e., to compute the OFP), the approach to runway 9R is chosen, since it is the longest possible. The AIRAC 2111 (issued on Nov 4th 2021) was taken from the UK AIP (NATS, 2021) and these procedures are depicted in Figure 6 (a), while Figure 6 (b) provides a screenshot of the decent of the OFP generated by Dynamo to RWY 09R and following the STAR and approach mentioned before.

EUROPEAN UNION



a) Arrival and initial approach procedure charts published in the AIP



b) Computed descent trajectory

Figure 6 Descent and arrival for SCN-100 operational flight plan (OFP)

2.1.2 OFP description for SCN 201: LEMD (MAD) to EDDF (FRA)

The main characteristics of the operational flight plan (OFP) for the flight between Madrid and Frankfurt are summarised in Table 3 below.

Table 3: Main characteristics of SCN 201 OFP

Flight schedule	Flight dispatch	Other operational information

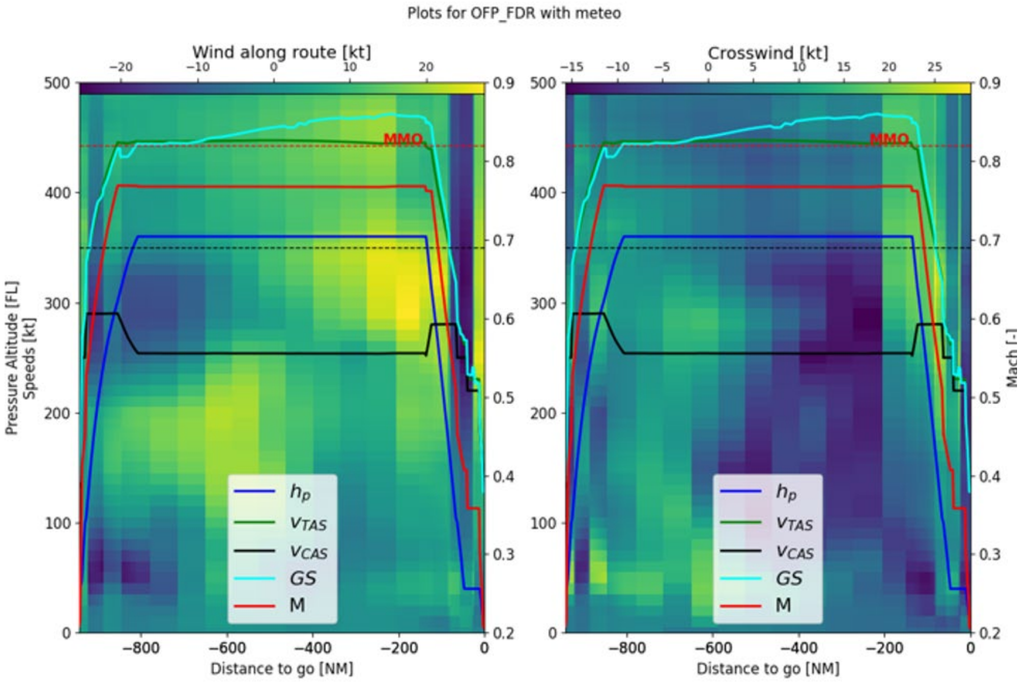
• Airline: Lufthansa (DLH)
• Aircraft type: A320
• LEMD-EDDF
• Scheduled off-block time (SOBT): 6h 35min UTC
• Scheduled in block time (SIBT): 9h 10min UTC
• Scheduled block time: 155’
	• Nominal Cost Index: 10 kg/min
	• Payload:
	  ○ Passengers: 171
	  ○ Cargo: 1,000 kg
	• Number of connecting pax: 65
	• Weather forecast: nominal, issued 2018-02-07
	• Passengers entitled to Regulation 261 compensation if delay thresholds meet
		• Taxi-out: 10’
		• OFP trip time: 138’
		• Taxi-in + padding at arrival: 7’
		• EMPAX 1C arrival
		• ILS approach to RWY 07C

Figure 7 below depicts the horizontal and vertical trajectory profiles of the OFP. For this OFP, we have a single cruise altitude at FL360 at M.77. Winds are relatively calm (around 10-20kt of headwind in the last part of the cruise, with a relatively small crosswind component.

To compute the OFP for this scenario, the AIRAC 2013 (issued on Dec 3rd 2020) was taken from the German AIP (DFS, 2020). For flight and fuel planning purposes (i.e., to compute the OFP), EMPAX 1C is the longest STAR in Frankfurt (taking into account that EMPAX 1D and EMPAX1A are labelled "by ATC only" and therefore, cannot be used for flight planning). Furthermore, since Frankfurt operates with a tromboning philosophy, the actual distance that will be flown in the tromboning is not known when planning the flight. Hence, the German AIP asks the operators to consider 83 NM from SPESA to the landing runway as average flight distance for fuel planning purposes. All these considerations have been taken into account when computing the OFP for this scenario.

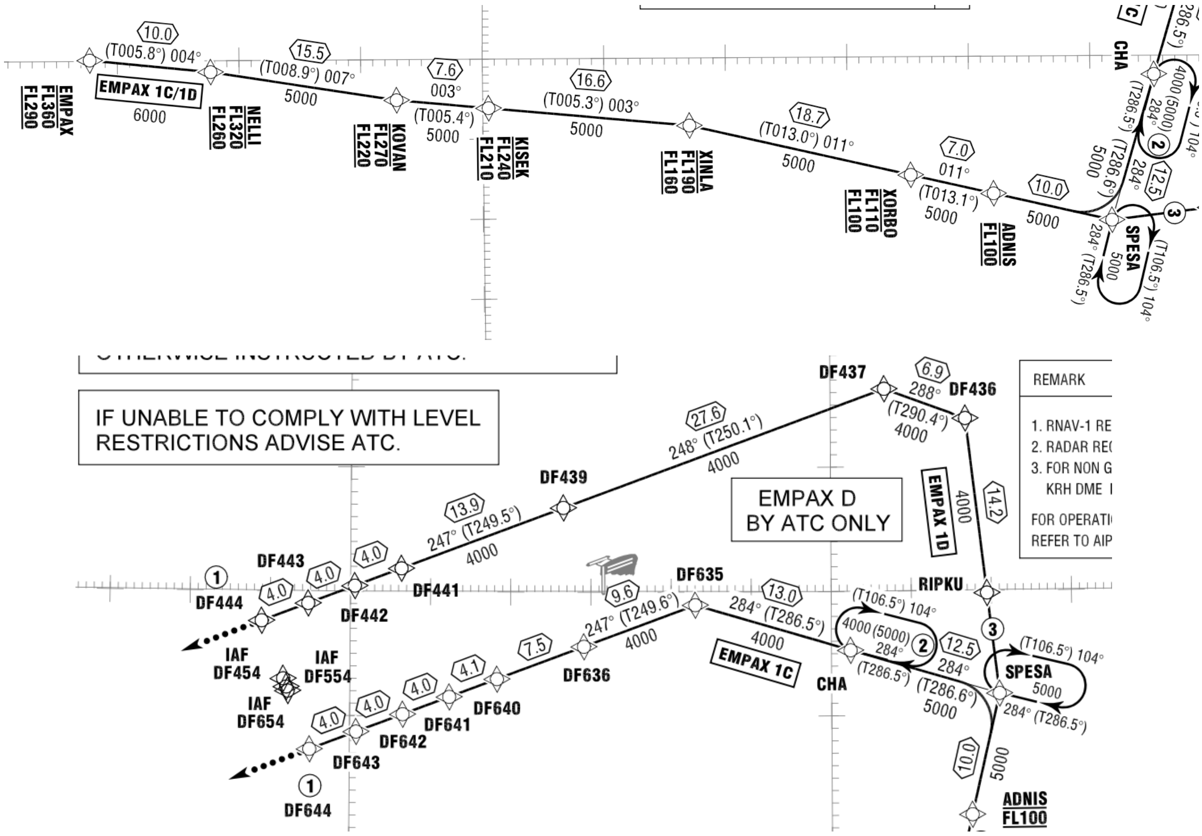
a) Horizontal trajectory profile

EUROPEAN UNION

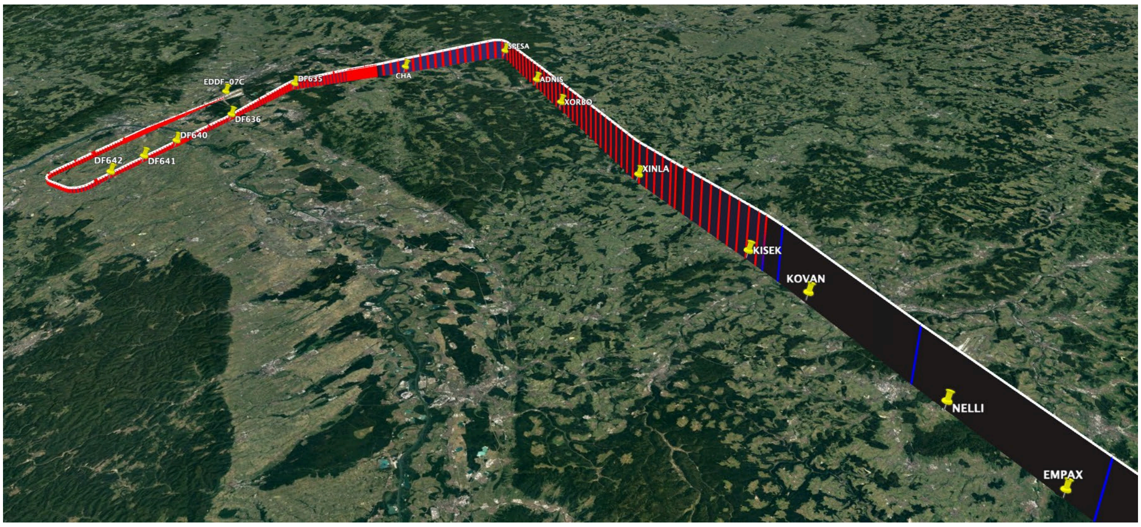


b) Vertical and speed trajectory profiles (with wind components)

Figure 7 Resulting trajectory for SCN-201 operational flight plan (OFP)



a) Arrival and initial approach procedure charts published in the AIP



b) Computed descent trajectory

Figure 8 Descent and arrival for SCN-201 operational flight plan (OFP)

Figure 8 (a) shows the AIP charts for the arrival and initial approach considered, while Figure 8 (b) provides a screenshot of the OFP decent generated by Dynamo to RWY 07C.

2.1.3 OFP Description SCN 600: KJFK (JFK) - EDDF (FRA)

The main characteristics of the operational flight plan (OFP) for the flight between Madrid and Frankfurt are summarised in Table 4 below.

Table 4: Main characteristics of SCN 600 OFP

Flight schedule	Flight dispatch	Other operational information

• Airline: Lufthansa (DLH)
• Aircraft type: B747-400
• KJFK-EDDF
• Scheduled off-block time (SOBT): 1h 45min UTC
• Scheduled in-block time (SIBT): 9h 20min UTC
• Scheduled block time: 455'
	• Nominal Cost Index: 20 kg/min
	• Payload:
	  ○ Passengers: 303
	  ○ Cargo: 20,000 kg
	• Number of connecting pax: 246
	• Weather forecast: nominal, issued 2016-07-28
		• Taxi-out: 30 min
		• Taxi-in + padding at arrival: 22'
		• OFP trip time: 403'
		• UNOKO 1B arrival
		• Tromboning procedure
		• ILS approach to RWY 25C

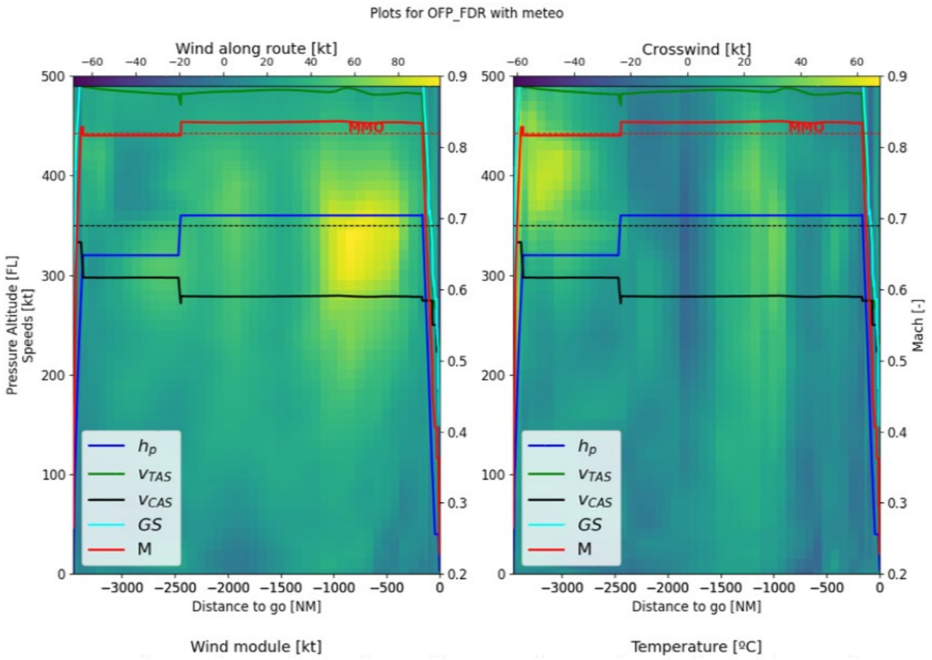
Figure 9 depicts the horizontal and vertical trajectory profiles of the OFP. For this OFP, we observe an initial cruise at FL320 and M.81 followed by a step-climb to FL360 and M.83. All flight benefits from tailwind conditions that increase when approaching to the European continent, while the cross-wind component is relative moderate along the flight.

To compute the OFP for this scenario, the AIRAC 2013 (issued on Dec 3rd 2020) was taken from the German AIP (DFS, 2020). For flight and fuel planning purposes (i.e., to compute the OFP), UNOKO 1B is the longest STAR in Frankfurt. Furthermore, since Frankfurt operates with a tromboning philosophy, the actual distance that will be flown in the tromboning is not known when planning the flight. Hence, the German AIP asks the operators to consider 113 NM from UNOKO to the landing runway as average flight distance for fuel planning purposes. All these considerations have been taken into account when computing the OFP for this scenario. Figure 10 (a) shows the AIP charts for the arrival and approach

EUROPEAN UNION

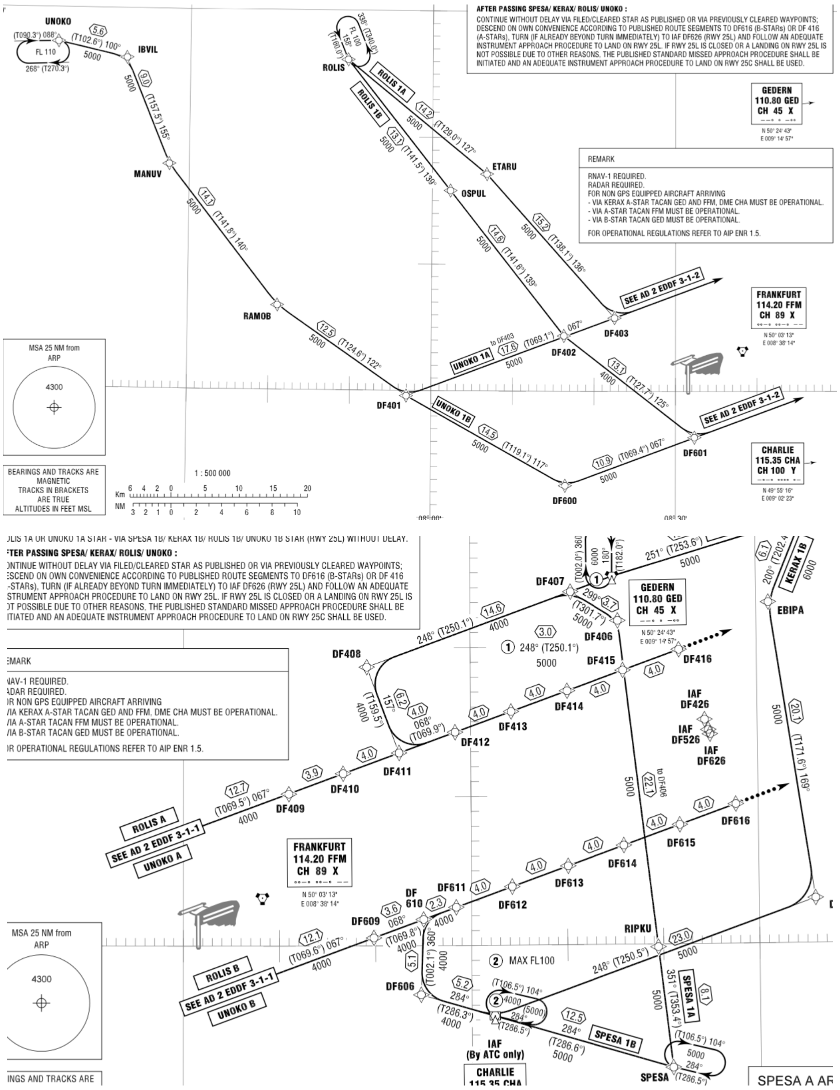
considered, while Figure 10 (b) provides a screenshot of the OFP decent generated by Dynamo to RWY 25C.

b) Horizontal trajectory profile

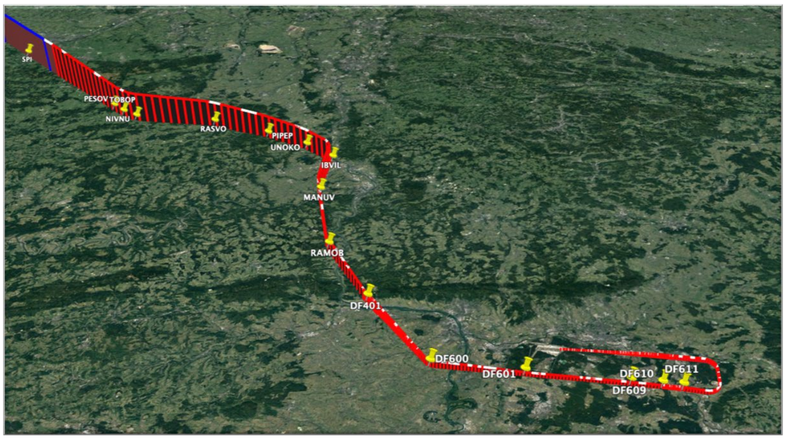


b) Vertical and speed trajectory profiles (with wind components)

Figure 9 Resulting trajectory for SCN-600 operational flight plan (OFP)



a) Arrival and initial approach procedure charts published in the AIP



b) Computed descent trajectory

Figure 10 Descent and arrival for SCN-600 operational flight plan (OFP)

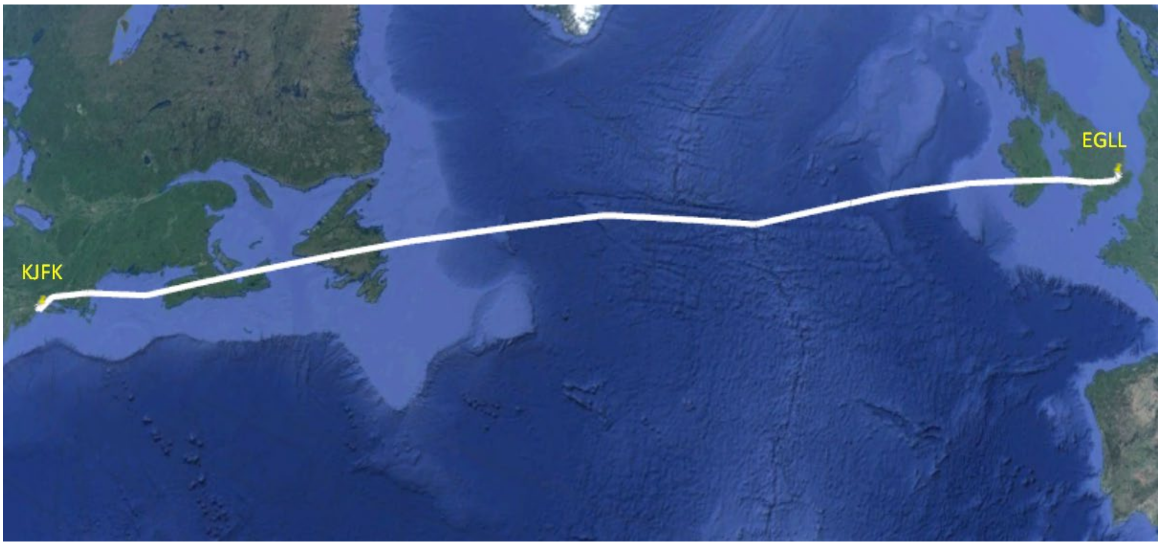
2.1.4 OFP description SCN 800: KJFK (JFK) - EGLL (LHR)

The main characteristics of the operational flight plan (OFP) for the flight between New York and London Heathrow are summarised in Table 5 below. Figure 11 below depict the horizontal and vertical trajectory profiles of the OFP.

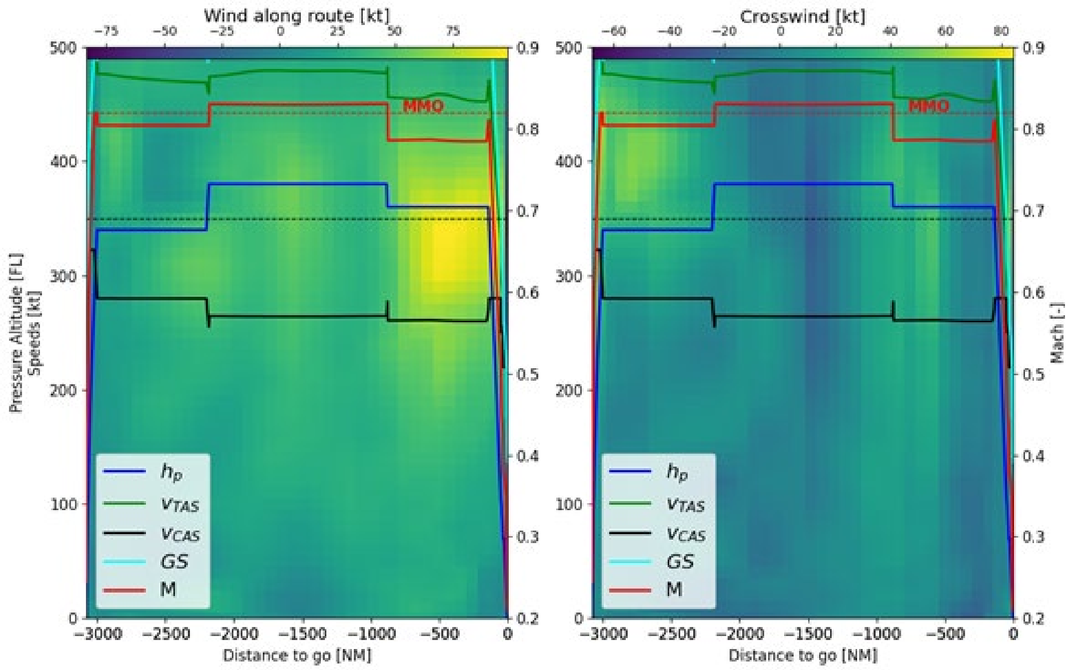
Table 5: Main characteristics of SCN 800 OFP

Flight schedule	Flight dispatch	Other operational information
• Airline: British Airways (BAW) • Aircraft type: B747-400 • KJFK/EGLL • Scheduled off-block time (SOBT): 0h 40min UTC • Scheduled in block time (SIBT): 7h 35min UTC • Scheduled block time: 410'	• Nominal Cost Index: 20 kg/min • Payload: ○ Passengers: 234 ○ Cargo: 0 kg (no cargo) • Number of connecting pax: 79 • Weather forecast: nominal, issued 2016-07-28	• Taxi-out: 30' • OFP trip time: 365' • Taxi-in + padding at arrival: 15' • BEDEK 1H arrival • Holding at OCK • ILS approach to RWY 27R

For this OFP, we observe an initial cruise at FL340 and M.80 followed by a step-climb to FL380 and M.83 at approximately 2200 NM from the destination airport. Then, the optimal trajectory includes a step-descent to FL360 at approximately 800 NM from destination with a reduction of the cruise speed to M.078. These altitude and speed changes can be explained with the wind profile forecasted for that route. As observed in the same figure, all flight benefits from tailwind conditions and some moderate cross-wind. Yet, at the beginning of the cruise (around 2900 to 2000 NM from destination) there is a region with small tailwind components (or even some headwind) at altitudes above the chosen cruise altitude. This fact, together with a heavier aircraft at the beginning of the flight could explain the optimal altitude chosen for the first cruise flight level. Then, between 800 to 100 NM from destination stronger tailwind is found at lower altitudes. This explains the step-descent of the OFP and the selection of a rather low Mach number for that portion of the cruise: the trajectory is taking advantage of the stronger tailwind found at FL360 and although cruise Mach is reduced the ground speed still increases at this altitude, saving in this way some fuel and time.



a) Horizontal trajectory profile



b) Vertical and speed trajectory profiles (with wind components)

Figure 11 Resulting trajectory for SCN-800 operational flight plan (OFP)

For this scenario, the STAR procedure BEDEK 1H is used in London Heathrow TMA, which ends at Ockham fix (OCK), where the holding pattern is located. Then, for flight and fuel planning purposes (i.e., to compute the OFP), the approach to runway 27R is chosen, since it is the longest possible. The AIRAC 2111 (issued on Nov 4th 2021) was taken from the UK AIP (NATS, 2021) and these procedures are depicted in Figure 12 (a), while Figure 12 (b) provides a screenshot of the decent of the OFP generated by Dynamo to RWY 27R and following the STAR and approach mentioned before.

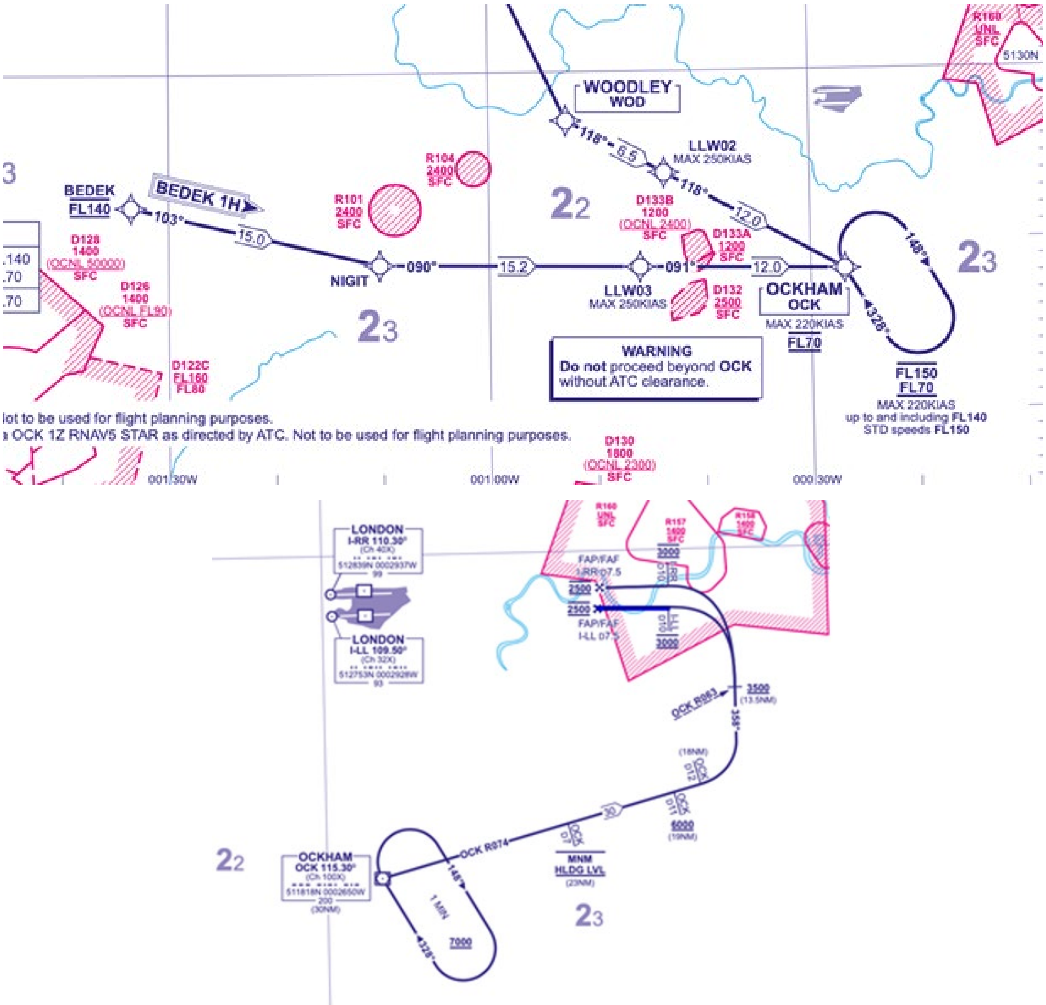
2.2 Sub-scenarios used for the verification and validation

The sub-scenario aims to further particularise the given scenario in terms of operational environment and weather. Among the five different variables which particularise the sub-scenario, we eventually used the four of them, each of which is further defined by different values (see Table 6 below). At the beginning of the project, TTA (Target Time of Arrival) and ATFM parametrisation of the sub-scenarios was envisioned in order to further explore Pilot3 usability for an extended contextualisation of each scenario. Nevertheless, the grade of complexity has been developed only including the sub-scenarios defined in Table 6.

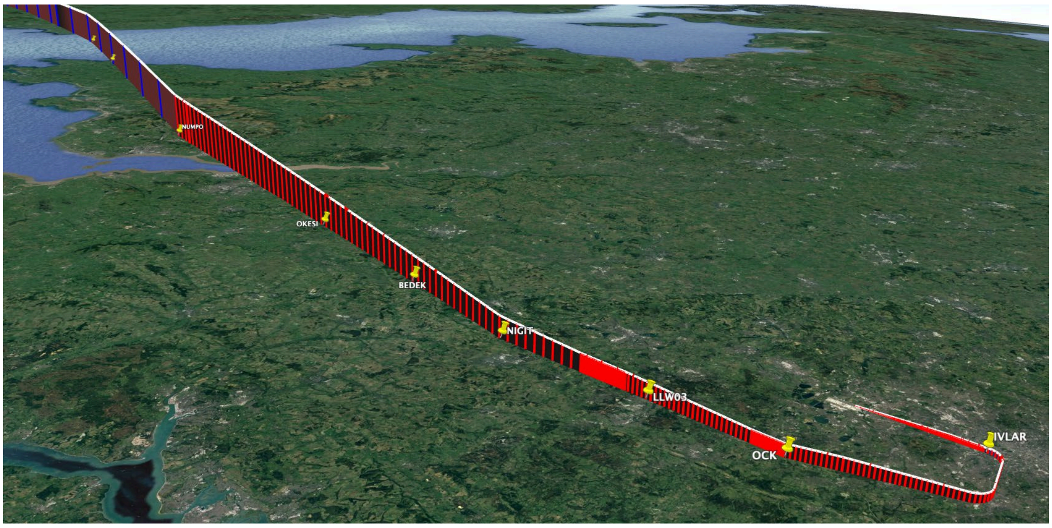
A default setup for the sub-scenarios used in the different validation activities has been defined as follows:

- Curfew; .destination airport default
- Regulation 261: YES

Hence, if not stated otherwise these are the values used in all experiments.



a) Arrival and initial approach procedure charts published in the AIP



b) Computed descent trajectory

Figure 12 Descent and arrival for SCN-800 operational flight plan (OFP)

Table 6: The description of sub-scenario identified for the validation campaign

Weather	Curfew	Entitlement to compensation (Reg. 261)
• Nominal • Strong beneficial wind • Strong decremental wind	• No • Yes, at arrival • Yes, at the end of the day	• Yes • No

2.3 Case studies used for the verification and validation

As already explained above, the case study refers to different events that may trigger Pilot3. Table 7 identifies the final four case studies, among the nine initially identified in D5.1, that were considered in the specification of different experiments.

Table 7: The description of case study identified for the validation campaign

Case study ID	Case study – Pilot3 triggering events	Possible parametrisation for this case-study
CS-10	Early/Late take-off	• Amount of departure delay
CS-40	Delay at destination TMA updated in cruise	• Amount of expected delay at TMA (holding and/or sequencing and merging time at arrival)
CS-60	Updated weather forecast	• Weather characteristics
CS-80	Turbulence in current FL	• Turbulence volume (i.e., restricted airspace) characteristics.

2.4 Sub-case studies used for the verification and validation

The sub-case study particularises how the performance indicators and the operational ATM parameters are estimated, and the airline flight policy with respect to the prioritisation of different airline costs (see Table 8). At the beginning of the project, it was envisioned an additional parametrisation of the sub-case studies, where ground or air information could be used for the estimation. Nevertheless, the grade of complexity has been reduced to only contemplate a single source of data.

Unless otherwise stated, the default setup for sub-case study is as follows:

- PIE: Uncertainty modelled. All estimators modelled with heuristics except for reactionary delay which is modelled with machine learning models (i.e., block time and ground time estimated with machine learning models).
- OAE: Uncertainty modelled. All estimators modelled with heuristics.
- Optimisation ranking: IROPS, other and fuel.

Table 8: The description of sub-case study identified for the validation campaign

Performance Indicator Estimator	ATM Operational Estimator	Optimisation ranking (airline policies to configure Pilot3)
• Heuristic • Machine Learning	• Distance sequence and merging -heuristic • Distance sequence and merging -machine learning • Holdings – heuristic • Holding - machine learning • TMA distance to FL100 – heuristic • Taxi-in - heuristic	• Cost of fuel • Cost of IROPs • Other cost

3 Verification

In early stages of the project life-cycle, a set of preliminary requirements were defined for the aimed Pilot3 prototype. Two requirement sets were defined, one for the system definition of the Pilot3 software prototype, and one for the functional definition of the Pilot3 HMI prototype. The former, was aimed to be developed separately from the software development cycle and therefore integration between the two was deemed as out of the scope of the Pilot3 project.

In this section, a results report will be presented against each requirement, showing the evidences that have been used to deem correct verification of the requirements in a separate form, one dedicated section for the software and another one for the HMI. Additionally, a results and rationale section is presented to embody the numerical results into verbal conclusion of the activity.

3.1 System Verification

As defined in D5.1 (Pilot3 Consortium, 2020c), the system verification consisted on the testing of requirements. Therefore, for each requirement a Verification test-case has been developed as evidence of the fulfilment of the functionality. There exist three types of requirements: Domain, Functional and non-functional. As it can be expected, the Domain requirements will be tested by means of a developed rationale.

3.1.1 Experiments executed for the verification campaign

As already foreseen in D5.1 the experiments developed to verify Pilot3 software prototype have been created ad-hoc alongside the Test-Case definition. Differently from the validation activities experiments, the experiments designed for the verification activities do not have a five-level hierarchy definition associated in order to ease the process and focus on the pure verification of the requirement. As explained in D5.1, verification aims at ensuring that the system does the "right things", not the "things right". Same default parameters will be used as defined in section 2: Definition of scenarios.

Table 9 summarises the experiments designed and execute for the verification of the Pilot3 prototype. Note, the sub-scenario ID will identify in an unequivocal way the experiment number for a set of experiments using the same Scenario (i.e., route). Additionally, the case-study ID will be set to 1, meaning that no behavioural Pilot3 triggering is being analysed, but rather the execution of the prototype modules.

3.1.2 Functional verification test report

This section aims at purely reporting the results of the verification test-cases. For this purpose, Table 10 presents the relevant data traced to the requirement has been generated with the following fields:

- Requirements ID: ID of the requirements
- Mandatory: With Y/N indicating whether it is a hard of soft requirements
- SW rel.: SW version used when the test-case was executed
- Date: date when the test-case was executed
- Description: description of the requirement
- Verification test-case: set of steps followed to test the requirement
- Experiments: experiments involved on the test-case
- Test Result: color-coded result of the test-case:

-	NOT TESTED
-	NOK
-	OK
-	PARTIAL

Table 9: The description of experiments executed for the verification campaign

Experiment ID	Scenario ID	Sub Scenario		Case Study			Sub case Study		Purpose
		Weather	other	ID	Triggering point	Other	PIE/OAE configuration	Other	
901	100	Nominal		1	TOC	Taxi-out: 10' Dep. Delay: -20'	- PIE: default, manual holding of 20' - AOE: default		Test of requirements: P3-DR-SYS-010, P3-DR-SYS-020, P3-FR-SYS-010, P3-FR-SYS-020, P3-FR-SYS-121.
902	100	Nominal	Reg261: "N"	1	TOC	Taxi-out: 10' Dep. Delay: 10'	- PIE: default, manual holding of 20' - AOE: default		Test of requirements: P3-DR-SYS-010 .
903	100	Nominal		1	TOC	Taxi-out: 10' Dep. Delay: 10'	- PIE: default - AOE: ground_time_estimator and g2g_time_estimator: machine_learning		Test of requirements: P3-FR-SYS-010.
904	100	Nominal		1	FL100	Taxi-out: 30' Dep. Delay: 145'	- PIE: default - AOE: default		Test of requirements: P3-FR-SYS-030.
905	100	Nominal		1	TOD	Taxi-out: 30' Dep. Delay: 20'	- PIE: default - AOE: default		Test of requirements: P3-FR-SYS-030.
906	201	Nominal	Restricted airspace	1	TOC	Taxi-out: 10' Dep. Delay: 10'	- PIE: default - AOE: default	Sub-objectives ranking: IROPS > Other > Fuel	Test of requirements: P3-FR-SYS-040, P3-FR-SYS-050.
907	201	Nominal	Restricted airspace	1	TOC	Taxi-out: 10' Dep. Delay: 10'	- PIE: default - AOE: default	Sub-objectives ranking: Fuel > IROPs > Other	Test of requirements: P3-FR-SYS-050.
908	201	Nominal		1	TOC	Taxi-out: 10' Dep. Delay: 10' Weather: Heavy_head wind	- PIE: default -AOE: default		Test of requirements: P3-FR-SYS-110.
909	600	Nominal		1	TOC	Taxi-out: 30' Dep. Delay: 145'	- PIE: default - AOE: default		Test of requirements: P3-FR-SYS-182.
910	600	Nominal		1	TOC	Taxi-out: 30' Dep. Delay: 145'	- PIE: default - AOE: default		Test of requirements: P3-FR-SYS-182.

EUROPEAN UNION

Table 10: Verification test results report

Requirement ID	Mandatory requirement	SW release	Date	Description	Verification test-case	Experiments	Test result
P3-DR-SYS-010	Y	V1.3	19/01/22	Pilot3 prototype shall model EC reg. 261/2004 in the estimation of the appropriate indicators.	- Inputs/Pre-conditions: - 901 experiment *reg_261* parameter set to 'Y' - 902 experiment *reg_261* parameter set to 'N' - Expected results: - Cost function of experiment 901 should have greater Irops_costs than experiment 902 - Obtained results . Experiment 901 has greater Irops_costs than experiment 902	901, 902	
P3-DR-SYS-020	Y	V1.3	19/01/22	Pilot3 prototype shall take into consideration the SESAR 2020 Transition ConOps.	- Inputs/Pre-conditions: None, domain requirement Rationale: Given an RBT, Pilot3 optimises the downstream 4D trajectory by means of several estimators, resulting of an updated RBT.	901	
P3-FR-SYS-010	Y	V1.3	19/01/22	The Pilot3 prototype shall be manually configured by a human user (airline operator).	Inputs/Pre-conditions: - Configuring heuristics and machine learning for the OAE/PIE estimators - 901: Default - 903: block_time_estimator and g2g_time_estimator of OAE module configured as machine learning. holding_time_estimator configured as manual of 40'. - Expected results: Estimation outputs for each single estimator differs between 901 and 903 experiments. - Obtained results Estimation differs between 901 and 903 experiments.	901, 903	
P3-FR-SYS-020	Y	V1.3	19/01/22	The Pilot3 prototype shall be manually triggered by a human user (pilot).	Inputs/Pre-conditions: None, domain requirement Rationale: The prototype can be triggered at any point of the trajectory.	901	
P3-FR-SYS-030	Y	V1.3	19/01/22	The Pilot3 prototype shall be triggered at any moment of flight from FL100 in climb down to TOD.	Inputs/Pre-conditions: - 904: Trigger Pilot3 at FL100 on the climb phase - 905: Trigger Pilot3 at TOD Expected results: - Pilot3 should provide an output for both cases Obtained results	904, 905	

Requirement ID	Mandatory requirement	SW release	Date	Description	Verification test-case	Experiments	Test result
					- Pilot3 provides an output for both cases		
P3-FR-SYS-040	Y	V1.3	19/01/22	Pilot3 prototype shall automatically provide a set of alternative 4D trajectories down to the runway threshold.	Inputs/Pre-conditions: Not applicable Expected results: hp[ft] value for the last point of the Pilot3 trajectory should correspond to a distance of 52ft in altitude. Obtained results hp[ft] value for the last point of the Pilot3 trajectory corresponds a distance of 52ft in altitude.	906	
P3-FR-SYS-050	Y		19/01/22	The Pilot3 prototype shall rank the set of trajectory alternatives according to the airline policies.	Inputs/Pre-conditions: Execute Vikor with two alternatives of the same total cost but different in sub-cost ranking. Execute Vikor configuring {'sub_obj_rank':{'fuel':3,'irops':1,'other':2},'otp':{'threshold':0.8}} Execute Vikor configuring {'sub_obj_rank':{'fuel':3,'irops':1,'other':2},'otp':{'threshold':0.8}} Expected results: Ranking for 901 and 907 is different	906, 907	
P3-FR-SYS-060	Y		19/01/22	The Pilot3 prototype shall interact with the pilot to select among the alternatives generated.	Rationale: The VIKOR module has been verified and validated within IVA3 scope.	N/A	
P3-FR-SYS-070	Y		19/01/22	For each 4D trajectory, the prototype shall quantify its impact on airline performance by means of several PIs.	Inputs/Pre-conditions: N/A Expected results: Pilot3 should be able to provide information with regards to OTP Cost of fuel IROPs cost Total Cost Obtained results:	901	

Requirement ID	Mandatory requirement	SW release	Date	Description	Verification test-case	Experiments	Test result
					Pilot3 provides the pertinent information.		
P3-FR-SYS-080	Y	V1.3	19/01/22	The Pilot3 prototype shall be able to produce up to two different estimators for the same PI when applicable.	Inputs/Pre-conditions: Configuring heuristics and machine learning for the OAE/PIE estimators 901: Default 903: block_time_estimator and g2g_time_estimator of OAE module configured as machine learning. holding_time_estimator configured as manual of 40'. Expected results: Estimation outputs for each single estimator differs between 901 and 903 experiments. Obtained results Estimation differs between 901 and 903 experiments.	901, 903	
P3-FR-SYS-090	Y	V1.3	19/01/22	The Pilot3 prototype shall allow the user (airline operator) to specify which estimator(s) should be used.	Inputs/Pre-conditions: Configuring heuristics and machine learning for the OAE/PIE estimators 901: Default 903: block_time_estimator and g2g_time_estimator of OAE module configured as machine learning. holding_time_estimator configured as manual of 40'. Expected results: Estimation outputs for each single estimator differs between 901 and 903 experiments. Obtained results Estimation differs between 901 and 903 experiments.	901, 903	
P3-FR-SYS-100	N	V1.3	19/01/22	The Pilot3 prototype may produce an estimated accuracy/confidence level for each PI estimator.	Rationale: Software development not deployed to satisfy this requirement.	N/A	
P3-FR-SYS-110	Y	V1.3	20/01/22	The Pilot3 prototype shall take into account the influence of the wind when computing the alternative trajectories.	Inputs/Pre-conditions: Having heavy wind Pilot3 weather input. Run experiment 909 and compare the optimised trajectory and weather with 901. Expected results:	908	

Requirement ID	Mandatory requirement	SW release	Date	Description	Verification test-case	Experiments	Test result
					Resulting 908 Pilot3 trajectory should get greater delay than 901 due to the effect of the head wind. Obtained results: Resulting 909 trajectory gets greater delay than 901.		
P3-FR-SYS-121	N	V1.3	20/01/22	The Pilot3 prototype may take into account real forecast air temperature conditions when computing trajectories.	Inputs/Pre-conditions: Real grib weather used: 2016-07-28 Mean, Nominal conditions Expected results: Pilot3 optimiser should provide the temperature along the 4D trajectory Obtained results: Temp[ºC] output parameter is provided along the 4D trajectory	901	
P3-FR-SYS-122	N	V1.3	20/01/22	The Pilot3 prototype may take into account real forecast air pressure conditions when computing trajectories.	Inputs/Pre-conditions: Real grib weather used: 2016-07-28 Mean, Nominal conditions Expected results: Pilot3 optimiser should provide the pressure along the 4D trajectory Obtained results: Press[hPa] output parameter is provided along the 4D trajectory	901	
P3-FR-SYS-130	N	V1.3	20/01/22	The Pilot3 prototype may take into account uncertainties related to weather forecasts.	Rationale: Weather uncertainties have not been implemented in Pilot3 prototype framework. Nevertheless, mean ensembles have been considered as weather input data.	N/A	
P3-FR-SYS-140	Y	V1.3	20/01/22	The Pilot3 prototype shall implement solutions for an Airbus A320.	Inputs/Pre-conditions: A/C modelled for the experiment should be A320 Expected results: Pilot3 optimiser should provide a 4D trajectory Obtained results:	901	

Requirement ID	Mandatory requirement	SW release	Date	Description	Verification test-case	Experiments	Test result
					Pilot3 optimiser provides a 4D trajectory		
P3-FR-SYS-150	Y	V1.3	20/01/22	The Pilot3 prototype shall implement solutions for an Boeing B747.	Inputs/Pre-conditions: A/C modelled for the experiment should be A747 Expected results: Pilot3 optimiser should provide a 4D trajectory Obtained results: Pilot3 optimiser provides a 4D trajectory	909	
P3-FR-SYS-160	N	V1.3	20/01/22	The Pilot3 prototype may be able to implement other turbojet aircraft (different from an A320 and a B777).	Rationale: Software development not deployed to satisfy this requirement.	N/A	
P3-FR-SYS-170	N	V1.3	20/01/22	The Pilot3 prototype may be able to implement turboprop aircraft.	Rationale: Software development not deployed to satisfy this requirement.	N/A	
P3-FR-SYS-181	Y	V1.3	20/01/22	When computing trajectories, the Pilot3 prototype shall consider aircraft operation (flight envelope) constraints.	Inputs/Pre-conditions: Expected results: Obtained results:	--	
P3-FR-SYS-182	Y	V1.4	20/01/22	When computing trajectories, the Pilot3 prototype shall consider constraints set by the pilot.	Inputs/Pre-conditions: Pilot3 version set to 1.4 or 2.0 Turbulence field active, with a path to a volume restricted csv. Restricted airspace overlapping with Pilot3 trajectory, experiment 909. Expected results: Pilot3 optimised trajectory should avoid the defined restricted airspace. Obtained results: Pilot3 optimised trajectory avoids the defined restricted airspace. Rationale: Speed constraints have not been implemented.	909, 910	
P3-FR-SYS-183	Y	V1.3	20/01/22	When computing trajectories, the Pilot3 prototype shall	Inputs/Pre-conditions: Pilot3 version set to 1.4 or 2.0 Airlines policies captured as indication of ranking for sub-objectives (Fuel, IROPSs and Other costs).	905; 906; 907; and dedicated analysis of Performance	

Requirement ID	Mandatory requirement	SW release	Date	Description	Verification test-case	Experiments	Test result
				consider constraints driven by airline specific policies.	Obtained results: The filtering and ranking phase of Pilot3 (Performance Assessment Module) produces different ranking for alternatives as a function of the configuration of the airlines' priorities.	Assessment Module	
P3-FR-SYS-184	Y	V1.3	20/01/22	When computing trajectories, the Pilot3 prototype shall consider static ATM related constraints.	Inputs/Pre-conditions: Pilot3 airport chart configuration set for the arrival airport of the experiment. Expected results: Pilot3 optimised trajectory should follow LHR STAR and approach procedures. Obtained results: Pilot3 optimised trajectory follows LHR STAR and approach. procedures	901	
P3-FR-SYS-191	N	V1.3	20/01/22	The Pilot3 prototype may estimate the following ATC tactical intervention: extra flown distance in en-route.	Rationale: Software development not deployed to satisfy this requirement.	N/A	
P3-FR-SYS-192	N	V1.3	20/01/22	The Pilot3 prototype may estimate the following ATC tactical intervention: extra flown distance TMA.	Inputs/Pre-conditions: Run an experiment with tromboning procedures at destination airport. ofg_config set to 'sequencing':'dynamo'. Expected results: Pilot3 should output an estimation for the sequencing and merging distance Obtained results: Pilot3 provides an estimation for the sequencing and merging distance. Rationale: Results obtained only for TMA and below FL100	906	
P3-FR-SYS-193	N	V1.3	20/01/22	The Pilot3 prototype may estimate the following ATC	Inputs/Pre-conditions: Run an experiment with holding procedures at destination airport ofg_config set to 'holding':'dynamo'.	901	

EUROPEAN UNION

Requirement ID	Mandatory requirement	SW release	Date	Description	Verification test-case	Experiments	Test result
				tactical intervention: air holding in published patterns.	Expected results: Pilot3 should output an estimation for the holding time. Obtained results: Pilot3 provides an estimation for the holding time.		
P3-FR-SYS-194	N	V1.3	20/01/22	The Pilot3 prototype may estimate the following ATC tactical intervention: taxi-in time	Inputs/Pre-conditions: ofg_config set to 'taxi':'dynamo'. Expected results: Pilot3 should output an estimation for the taxi_in time. Obtained results: Pilot3 provides an estimation for the taxi_in time.	901	
P3-FR-SYS-200	Y	V1.4	28/01/22	All alternatives, metadata and the interaction with the pilot shall be stored in DataBeacon.	Inputs/Pre-conditions: Run an experiment with no data for the particular folder on Databeacon's platform. Expected results: Dedicated folder containing all the results should be created in Databeacon. Obtained results: Dedicated folder containing all the results should be created in Databeacon.	901	
P3-NFR-SYS-010	Y	V1.4	28/01/22	Pilot3 prototype shall be a standalone software.	Rationale: Running Pilot3 script pilot3_main.py and obtaining a Pilot3 optimised trajectory output, indicates a correct function of the inner data exchange between modules.	901	
P3-NFR-SYS-020	Y	V1.4	28/01/22	Pilot3 prototype shall run in a conventional PC platform under Linux.	Inputs/Pre-conditions: Run Pilot3 in a conventional PC platform with Linux iOS. Expected results: Results should be obtained with no fatal error reported. Obtained results: Results are obtained with no fatal error reported.	901	
P3-NFR-SYS-030	N	V1.4	28/01/22	Pilot3 prototype may run in DataBeacon.	Inputs/Pre-conditions: Hold access to DataBeacon platform. Run Pilot3 in DataBeacon platform. Expected results: Results should be obtained with no fatal error reported.	901	

Requirement ID	Mandatory requirement	SW release	Date	Description	Verification test-case	Experiments	Test result
					Obtained results: Results are obtained with no fatal error reported.		
P3-NFR-SYS-040	Y	V1.4	28/01/22	Pilot3 prototype shall store all input and output files in the DataBeacon platform.	Inputs/Pre-conditions: Run an experiment with no data for the particular folder on Databeacon's platform. Expected results: Dedicated folder containing all the results should be created in Databeacon. Obtained results: Dedicated folder containing all the results should be created in Databeacon.	901	

3.1.3 Update and revision of requirements

The aim of the Requirements Verification not only lies on the Software verification itself but the requirements definition. The first release of requirements was defined in deliverable D1.1 (Pilot3 Consortium, 2020a), where the final prototype of Pilot3 was envisioned for the first time. Throughout the development of WP4 and WP5, the prototype, and therefore the requirements, have suffered some adaptations with respect to the initial plan. Therefore, and following the Agile methodology, the requirements have been reviews and corrected according to the new definition needs.

In this section, the updated requirements have been listed and the reason for modification have been presented.

Table 11: Requirements update

Req. ID	version	Old description	version	New description	Rationale
P3-FR-SYS-030	1.0	The Pilot3 prototype shall be triggered at any moment of flight from FL100 in climb down to FL100 in descent.	2.0	The Pilot3 prototype shall be triggered at any moment of flight from FL100 in climb to TOD.	The Pilot3 prototype has been designed to be triggered from FL100 in climb to Top of Descend.
P3-FR-SYS-150	1.0	The Pilot3 prototype shall implement solutions for a Boeing B777.	2.0	The Pilot3 prototype shall implement solutions for a Boeing B747.	Due to data availability with regards to the aircraft performance model, aircraft type B747-400 have been implemented-
P3-FR-SYS-121	1.0	The Pilot3 prototype may take into account real temperature and/or pressure atmospheric conditions when computing the alternative trajectories.	2.0	The Pilot3 prototype may take into account real temperature and/or pressure atmospheric conditions forecast when computing the alternative trajectories.	Correction performed in order to clearly specify that realistic forecast data will be used rather than real data.
P3-FR-SYS-122	1.0	The Pilot3 prototype may take into account real air pressure conditions when computing trajectories.	2.0	The Pilot3 prototype may take into account real air pressure conditions forecast when computing trajectories.	Correction performed in order to clearly specify that realistic forecast data will be used rather than real data.
P3-FR-SYS-080	1.0	The Pilot3 prototype shall be able to produce up to four different estimators for the same PI when applicable.	2.0	The Pilot3 prototype shall be able to produce up to two different estimators for the same PI when applicable.	Considering the re-scope of Pilot3 project ambition, two estimators will be aimed for each PI, one for Heuristics and one for Machine learning.

3.1.4 Summary of software verification results

Table 12 summarises the results obtained through the last cycle of verification.

Table 12 Summary of software verification results

Result	Total	Comments
	5	All these 5 requirements are non-mandatory: P3-FR-SYS-100, P3-FR-SYS-130, P3-FR-SYS-160, P3-FR-SYS-170 and P3-FR-SYS-191. These requirements are related to either: • Uncertainty: en-route distance, weather forecast uncertainty, provide confidence level for each performance indicator estimator. • New aircraft models to be considered: other turbojets and turboprop aircraft.

EUROPEAN UNION

	26	Includes all requirements defined as mandatory and all non-functional requirements specified for the prototype.
	2	Two requirements are partially verified, and further work is required to completely implement the capabilities needed: • P3-FR-SYS-192, considers the use of uncertainty due to ATC in the TMA. This is considered for some type of operations (tromboning, radar vectoring and holdings) and dedicated uncertainty estimator is developed in heuristic form but not integrated in Pilot3 architecture. • P3-FR-SYS-182 requires that the pilot can set constraints for the optimisation. Some of these capabilities are incorporated in Pilot3 optimisation engine V2.0, e.g. flight level restrictions. However, further experiments and the integration of these constraints in the data manager are still pending.

3.2 Verification of the human-machine interface

Differently from the software requirements, the HMI prototype ones have been tested in a qualitative manner, since the prototype consists of a set of mock-ups presented in a linear form (i.e., the features are not interactive)[1].

3.2.1 HMI verification test report

The set of requirements that have been presented in this section, were developed in earlier stages of the project and have been finally presented in this deliverable D5.2.Table 13Table 13 presents the results of the verification of the HMI requirements.

Table 13: HMI Verification requirements report

Activity	Module description	ID	Requirement short description	Fulfilled
Alternatives interaction	The pilot shall interact through the HMI with the alternatives provided by the system	P3-FR-HMI-010	The HMI shall display (**output**) a list of alternatives (4D trajectories)	
		P3-FR-HMI-020	The HMI shall allow the pilot to rank (**input**) the presented alternatives according to the airline policies criteria.	
		P3-FR-HMI-030	The HMI shall allow the pilot to inspect (**input**) the alternative details in depth.	
		P3-FR-HMI-040	The HMI shall allow the pilot to re-compute (**input**) the alternatives considering the activated constraints	
Alternative details	The pilot shall be able to inspect all	P3-FR-HMI-050	The HMI shall display (**output**) the inspected alternative, plotting the horizontal and vertical profiles in a graph	

[1] A set of mock-ups have been made available in the following link (Accessed March 2022): https://xd.adobe.com/view/8e3baa9d-e838-4e26-b2ca-a27822b10a12-db68/?fullscreen.

	the alternatives in depth.	P3-FR-HMI-060	The HMI shall display (**output**) the Performance Indicators Estimators assessment for the given alternative.	
		P3-FR-HMI-070	The HMI shall display (**output**) a confidence level for each estimator.	
Alternative comparison	The pilot shall be able to compare the alternatives each other	P3-FR-HMI-080	The HMI shall allow the pilot to compare (**in/out**) the 4D trajectories of alternatives each other	
		P3-FR-HMI-090	The HMI shall allow the pilot to compare (**in/out**) alternatives from previous computations, to analyse the impact of the enabled constraints	
		P3-FR-HMI-100	The HMI shall allow the pilot to compare (**in/out**) alternatives with the planned one.	
Constraints configuration	The HMI shall provide a module for the pilot to set up the constraints.	P3-FR-HMI-110	The HMI shall provide an editor (**input**) to modify the constraints.	
		P3-FR-HMI-120	The HMI shall allow the pilot to enable/disable (**input**) one or multiple constraints.	
		P3-FR-HMI-130	The HMI shall display (**output**) a list of constraints activated by the pilot	
Operational ATC Estimators	The HMI shall provide a set of operational ATC estimators generated by Pilot3 system	P3-FR-HMI-140	The HMI shall display (**output**) a set of airborne and ground estimators used to compute the alternatives.	
		P3-FR-HMI-150	The HMI shall display (**output**) an estimated extra flown distance in en-route phase.	
		P3-FR-HMI-160	The HMI shall display (**output**) an estimated extra flown distance inside Terminal Airspace.	
		P3-FR-HMI-170	The HMI shall display (**output**) an estimated number of holdings plus the holding time	
		P3-FR-HMI-180	The HMI shall display (**output**) an estimated taxi-in time.	

3.2.2 Summary of HMI verification results

Table 14 summarises the results obtained through the last cycle of verification on the HMI.

Table 14: Summary of HMI verification results

Result	Total	Comments
	0	All requirements have been considered and incorporated in the design of the HMI.
	14	Most requirements have been fully incorporated in the HMI and feedback gathered from the Advisory Board for their validation.
	4	Four requirements have been only partially incorporated into the HMI. This is due to the limitation of some of the prototype capabilities which impact the capabilities of what can be presented and requested to/from the crew.

EUROPEAN UNION

4 Validation

The internal validation has two objectives: **to validate the functionalities of the components** of Pilot3, and to **evaluate the operational benefits** of the prototype against the set of **research questions** and corresponding hypotheses defined for this purpose.

The internal validation campaign is based on the interaction within the members of the consortium, and with the Topic Manager. The results for a set of scenarios and case studies will be presented to the internal experts in order to understand if the objectives and goals specified in the hypotheses have been successfully achieved. In addition, the internal validation provides a set of quantifiable metrics to facilitate the assessment of the tool.

The internal validation is carried out through seven different internal validation actions (IVA), which can be grouped between:

- **Actions aiming at validating the different components of the model**
    - **IVA1 - Validation of Pilot3 optimised trajectory plans:** the aim of this action is to compare the result of Pilot3 with state-of-the-art FPO tool, ensuring that the trajectories generated by Pilot3 are realistic and with similar (or better) expected performance. These actions focus on evaluating the Trajectory Generator of Pilot3.
    - **IVA2 - Validation of indicators and estimators' prediction:** the aim of these actions is to validate the capabilities of the performance indicators (from the Performance Indicators Estimator module of Pilot3), and of the ATM uncertainties estimations (from the Operational ATM Estimators module).
    - **IVA3 - Assessment of the optimisation framework:** the objective of these actions is to assess how Pilot3 is able to generate different alternative trajectories and trade-offs.
- **Actions aiming at assessment the benefit of Pilot3**
    - **IVA4 - Pilot3 performance at generation of optimised trajectories plans:** the objective of this step is to assess the benefits of Pilot3 optimised trajectories plans against several baseline plans at the moment of considering the decision by the pilot. I.e., comparison of Pilot3 alternatives suggested to pilot with respect to baselines (original flight plan, or basic pilot trajectory behaviour).
    - **IVA5 - Pilot3 performance at trajectory realisation:** the aim of this action is to consider the impact of uncertainty in the execution of the optimised trajectory plans, and to assess the real benefits of Pilot3 against several baseline plans by simulating the trajectory to its arrival at the destination gate.
    - **IVA6 - Pilot3 performance full day of operations:** the aim of this action is to assess the benefit of Pilot3 at network-wide level in a full day of operations

- **Actions aiming at the validation of the HMI**
  - **IVA7 – HMI:** these action aims to ensure that HMI prototype is well designed with respect to the information and mechanism available to the pilot.

The external validation will be conducted using fully functional versions of the prototype and based on the results of experiments studies performed in the internal validation will be used as an input for the external validation. Dedicated activities (e.g. workshop) will be organised, but also a continuous interaction with the Advisory Board will be seek in order to provide input into the project, therefore, some overlap between internal and external validation might occur. For example, once results for relevant scenarios are produced, these can be used to do a targeted interaction with some members of the Advisory Board. The external validation will be performed through three main types of actions:

- **EVA1 - Live or pseudo-live demonstration of the HMI prototype and overall capabilities** - the objective of this external validation action is to validate the interface, how the information is presented to and gathered from the crew, and to show the overall capabilities of Pilot3.
- **EVA2 - Presentation of results obtained with stand-alone simulations at trajectory level** - in this case, the results from the experiments executed in the internal validation IVA4 and IVA5 will be used. The objective is to validate the relevance of the findings.
- **EVA3 - Presentation of results obtained with network-wide simulations** - if EVA6 is implemented and results are obtained at network level for a full day of operations, providing insight on the potential benefit of Pilot3 for airlines, these will be validated as part of this external validation action.

This section presents the different internal validation actions with detail on the methodology and metrics that will be generated for the assessment of the research questions presented in Section 6.

4.1 IVA1 – Validation of Pilot3 optimised trajectory plans

As described in D5.1 (Pilot3 Consortium, 2020c), the scope of this validation activity was to demonstrate that Pilot3 is able to provide meaningful trajectories by comparing them with the trajectories generated by Pacelab Flight Profile Optimiser (FPO). Therefore, the FPO was placed as a benchmark tool.

When D5.1 was defined, the initial objective was to compare the executed PACE FPO planned trajectory to the Pilot3 optimised planned trajectory. During the execution of Verification and Validation plan, it was realised that for the sake of this activity, the Pilot3 OFP module (DYNAMO) would largely fulfil the aim of this activity since it considers a DCO (Direct Cost Operations) trajectory optimisation and it is exactly the same trajectory optimisation engine that is embedded within the Pilot3 software.

4.1.1 Approach

In order to properly perform this validation activity, the consortium partners, namely UPC and PACE, conducted a set of bilateral meetings during October and November, 2021. The process was performed in an iterative way in which they exchanged a batch of emails. In general, the iterative approach allowed the partners to easily identify the potential bottlenecks and corresponding corrective actions during the execution of the validation activities. In order to ensure the seamless flow and accomplish this validation activity, the following steps were initially traced, namely:

EUROPEAN UNION

1. **Validation of the trajectory input file required for the FPO**
   1.1. **Agree/understand the file format and content :** The task consisted on understanding the interface requirements of the FPO module in order to set and define the required interface updates for the Pilot3 OFP. For this, specification documents for the ARINC files were provided to the UPC and bilateral meetings and emails were exchanged. Two main files were aimed to be provided:
   - flight plan: initial optimised trajectory in ARINC 633 format.
   - weather: UAD weather file in ARINC 633 format.

   1.2. Generate the AIRNC input files required by FPO: The task consisted on the implementation of required changes on the Pilot3 OFP interface to produce readable output files for the FPO module, automating export from DYNAMO and import in FPO Cloud in ARINC 633 format.

   1.3. **Validate the input files:** PACE reviewed and validated the aforementioned input files provided by the UPC in order to execute the FPO module.
2. **Execution of Trajectories for several relevant experiments**
   2.1. **Agree on relevant experiments to be executed:** Two meaningful experiments were selected in order to perform the trajectories comparison. Section 4.1.2.defines the specific particularities of each experiment.

   2.2. **Validate Pilot3 outputs against FPO outputs**: For each experiment, the defined parameters were being used to compare the outputted trajectories. It is important to remark that while executing the experiments, both modules separately (FPO and DYNAMO) used:
   - the same weather information
   - the same objective function *C* defined as *C = Fuel + CI· TIme*
   - the same operational constraints in the trajectory optimisation process (except for the initial climb and approach phases).

This allowed the comparison to be the most reliable (and fair) possible.

4.1.2 Experiments

As aforementioned two experiments were used to perform the Pilot3 trajectories, Table 15 describes the main particularities of the experiments used:

Table 15: IVA1 experiments

ID	Origin	Destination	Weather	Aircraft type	Cost Index	Payload	Trip fuel
101	LGAV	EGLL	Nominal Date: 2016-07-28.	A320-232	20 kg/min	11837 kg (121 pax, 97 kg/pax, 100 kg cargo)	8052 kg
102	LEMD	EDDF	Nominal Date: 2016-07-28.	A320-232	20 kg/min	17490 kg (170 pax, 97 kg/pax, 1000 kg cargo)	5498 kg

As already foreseen in D5.1, Pilot3 and FPO differ in terms of the aircraft performance model used in their respective trajectory generation engines (i.e., Pilot3 uses EUROCONTROL BADA4 while FPO uses Original Equipment Manufacturer - OEM - data).

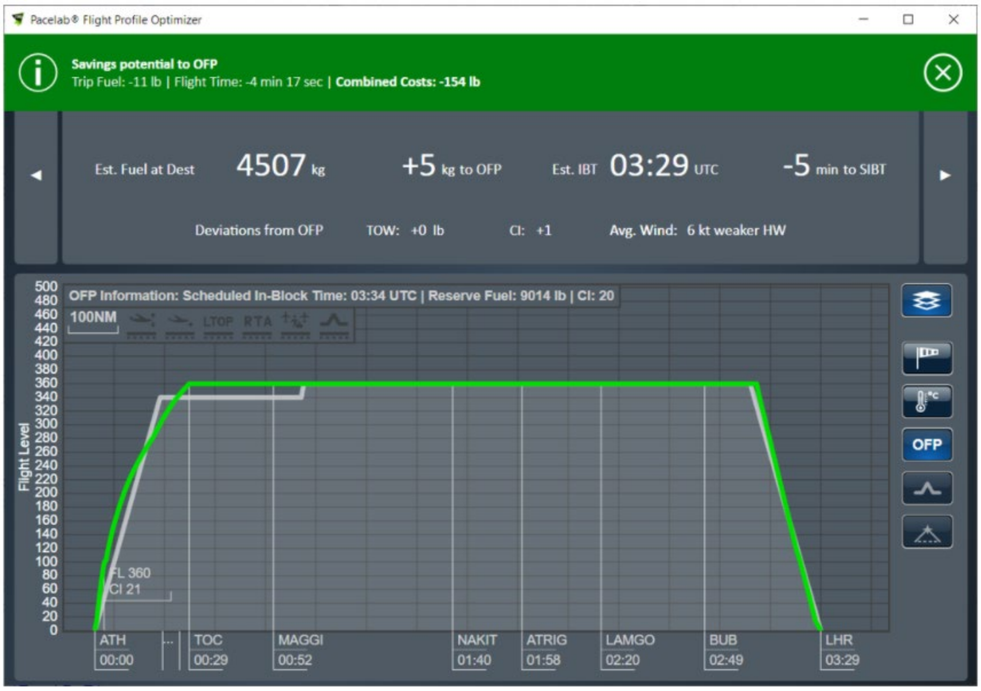
4.1.3 Results

Following the D5.1 plan, a set of metrics were used to validate the Pilot3 optimiser module trajectories, namely:

- Fuel consumption: The difference in total fuel consumption computed from the executed PACE FPO trajectory plan and executed Pilot3 trajectory plan.
- Flight Level: Difference in the number of speed level between the executed PACE FPO trajectory plan and executed Pilot3 trajectory plan.
- Number of steps: Difference in the number of flight level changes (as the optimiser will not consider lateral deviations) between the executed PACE FPO trajectory plan and executed Pilot3 trajectory plan.
- Time: The difference in total flight time computed from the executed PACE FPO trajectory plan and executed Pilot3 trajectory plan.

Figure 13 shows the vertical profile for both trajectories, corresponding to experiment 101 (LGAV-EGLL) and the comparison made by the Pacelab FPO tool. Table 16 contains the values for the validation metrics of this experiment.

As observed in the Figure, the vertical profiles are very similar. Dynamo proposes an initial cruise flight level at FL340 followed by a step climb to FL360 after 1h of flight, approximately. The FPO trajectory, conversely propose FL360 as initial (and unique) cruise altitude. These differences in the vertical profile results in a difference of 5 kg and 5 minutes: the FPO solution uses 5 kg less of fuel (0.06% of the trip fuel difference) and is 4’17’’ faster than the Dynamo solution (2.0% faster than the trip time).



PACElab Dynamo

Figure 13 FPO vs Dynamo trajectories for Experiment 101 (LGAV-EGLL)

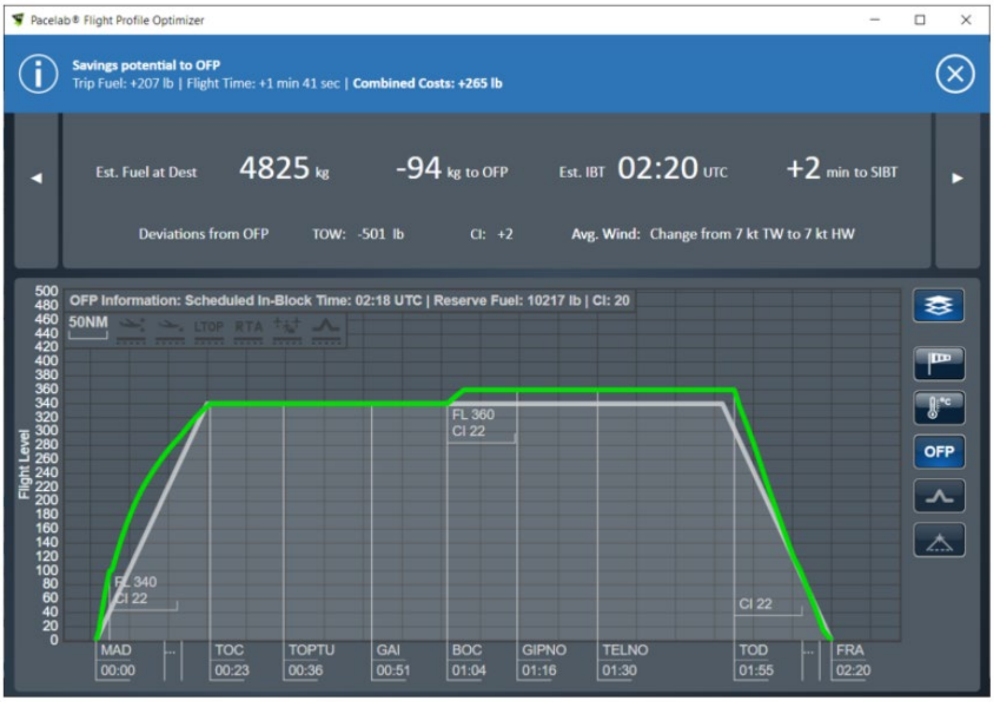
EUROPEAN UNION

Table 16: IVA1 results. Comparison of PACElab vs. DYNAMO (Pilot3 trajectory optimisation engine)

Indicator	Experiment 101: PACElab vs DYNAMO	Experiment 102: PACElab vs DYNAMO
Fuel consumption	- 5kg (-0.06% of. trip fuel)	+ 94 kg (+0.54% of. trip fuel)
Flight Level	360 vs 340/360	340/360 vs 340
Number of steps	0 vs 1	1 vs 0
Time	- 4' 17'' (-2.0% of. trip time)	+ 1' 41'' (+1.2% w.r.t trip time)

Figure 14 shows the vertical profile for both trajectories, corresponding to experiment 101 (LGAV-EGLL) and the comparison made by the Pacelab FPO tool. Table 16 contains the values for the validation metrics of this experiment.

As observed in the Figure, the vertical profiles are very similar. Both solutions propose an initial cruise at FL340. Dynamo keeps this cruising altitude for the whole flight, while the FPO trajectory proposes a step climb to FL360 at approximately 1h after take-off. These differences in the vertical profile results in a difference of 94 kg and 2 minutes: the FPO solution uses 94 kg more of fuel (0.54% of the trip fuel) and is 1'41'' slower than the Dynamo solution (1.2% of the trip time).



PACElab Dynamo

Figure 14. FPO vs Dynamo trajectories for Experiment 102 (LEMD-EDDF)

4.1.4 Summary of Research Questions and Hypothesis

The validation of the trajectories of Pilot3 was deemed as successful given the great similarities between the trajectories in terms of metrics and graphical comparison. For the tested scenario there is less than 1% discrepancy for fuel with respect to the trip fuel of the operational flight plan and 2% discrepancy for the trip time.

Table 17: Summary of research questions (RQ) and hypotheses addressed in IVA1

RQ ID	Rationale	Research question	Hypothesis	Success criteria	Status
P3-RQ-IV-010	Validate that Pilot3 is able to create trajectories which are realistic and representative.	Are trajectories computed by the trajectory generator of Pilot3 realistic enough?	It is expected to obtain similar trajectories than those obtained with state-of-the-art trajectory planning applications running in EFBs under similar execution conditions. Yet, discrepancies might be found due to mismatches in aircraft performance models.	• Pilot3 vs. FPO fuel and time discrepancies will not differ more than 4% and 6% respectively. • Discrepancies in number of speed/altitude changes along the trajectory can be explained by discrepancies in aircraft performance models.	**Validated**

4.2 IVA2 – Validation of indicators and estimators' prediction

Performance Indicators Estimator (PIE) and Operational ATM Estimator (OAE) are the two modules of Pilot3: performance indicators are used to compute the cost function, while the operational estimators aim at predicting operational uncertainties. As presented in Section 1.1.2 and Section 1.1.3, the Objective Function Estimator integrates the estimated costs at gate computed using the estimators of the PIE with the uncertainty provided by the OAE.

The validation activities can be broadly divided into two categories:

- **Validation activities as a part of model development** - A set of machine learning models have been developed for some of these estimators. In standard machine learning model development, some validation activities on the models are performed as part of their development: for each machine learning model a reference level of accuracy or a benchmark model is established. A set of standard metrics commonly used in machine learning will be used for each specific machine learning model to evaluate the model results.
- **Validation with experts within the consortium** - The models (heuristics and machine learning) are presented and reviewed with experts within the consortium to ensure that they produce adequate results. This was done with support of metrics and visualisations. For this purpose, four internal meetings were organised during July 2021 in which the teams involved in the estimator development presented the main results to other partners and experts within consortium. Table 18 summarises the main details on the technical meetings conducted as a part of IVA2.

Table 18 The internal meetings held as a part of IVA2

Date	Institution	Topic
12JUL21 - 11h-14h	Innaxis	Arrival procedure and sequencing and merging distance estimation (heuristics and ML)
19JUL21 - 11h-14h	UoW	Uncertainty modelling with machine learning: block time and turnaround models
20JUL21 - 11h-14h	UoW	Reactionary delay and costs with machine learning

Different models for the same indicator (e.g. heuristic and machine learning) should produce compatible but not fully equivalent predictions. For example, a model is to predict the rotation time of subsequent flights is used as part of the estimation of the reactionary delay. The heuristic version of this rotation time model does not consider the impact of features which might be related with the possibility of ATFM delay. This means that the machine learning version will provide predictions which have more operational factors into consideration which, even if considering relevant operational aspects, might render the predictions more uncertain. These alternatives to estimate the same indicator will be translated into different expected costs functions as a function of the configuration of Pilot3; and therefore, they could impact the result of the optimisation. For this reason, for each estimator the comparison between heuristic and machine learning alternative is provided in Sections 4.2.2 and 4.2.3 and the impact of these different alternatives on the cost function presented in 4.2.4.

4.2.1 Estimation approach

4.2.1.1 Estimation approach in Pilot3

As mentioned in the previous section, Pilot3 allows the user to select different implementations for the same estimator. Pilot3 software architecture allows the user to define, during the configuration of the tool, for each indicator a chain of estimators. In some cases, advance estimators might require data which might not be available during the flight or they might be valid only for specific operational conditions (e.g. a block time estimator might require the METAR at arrival which might not be loaded in the system or a machine learning model might have been trained (and be valid) only for a specific arrival airport). In these cases, Pilot3 will automatically revert to a lower estimator. This lower estimator, on its turn, might have defined other lower estimators creating this chain of alternatives. The main principle is that the system can try to use the most specific and advance estimator as possible. For example, a chain of estimators of the holding at arrival could be compose of five different estimators:

1. First, try to use a specific advance machine learning estimator which has been trained for arrivals at EGLL only;
2. if the previous estimator is not usable (e.g. the flight is going to an airport different than EGLL), then try to use a machine learning model trained with data for hub airports in Europe;
3. if that is not possible then use a model which requires the METAR weather at arrival;
4. if METAR is not available then a model with only static features (which don't change over time) might be used;
5. and finally, if this is not possible it will revert to an heuristic model which considers historical holdings per airport type.

For this reason, several approaches and models (heuristic and machine learning) have been implemented for some estimators.

4.2.1.2 Prediction horizons

Estimations are performed using different data (and features) as input into the models. These inputs can be static, i.e., which do not change over time such as if the airport at arrival is or not a hub airport, or dynamic, i.e., which might evolve over time, such as the weather at the arrival airport as reported in the METAR. Therefore, when using estimators (and machine learning models in particular), it is paramount to ensure that the data used for the prediction is consistent with the data used for the training of the models. This data characteristics (and in particular their temporal availability) is what is considered as the prediction horizon of the machine learning model. For example, the rotation time (ground time) of a flight can use dynamic data (e.g. METAR) which are available right before that flight departure, 3 hours, 6h, 15h prior departure, etc. Each of these horizons might require an individual model to be trained.

Pilot3 architecture allows the user to provide several estimators (trained at different prediction horizons) for the same indicator. Pilot3 will automatically use the estimators which are more relevant as a function of the data availability and, if necessary, interpolate the results across estimators. This capability is however not used in the results presented in this deliverable although preliminary analyses to highlight the importance of using different prediction horizons are presented in this section.

The need for having different models trained at different prediction horizons will vary as a function of the importance of the dynamic features on the performance of such models. In this section an analysis is presented to provide an initial understanding of the impact that input data have on the model estimations for the different time horizons; specifically, for the estimation of minimum turnaround time.

An initial subdivision between static and dynamic features (as described in Table 22) for the four-time horizons. The relevance of the feature is computed as the (normalized) total reduction of the criterion (that is used to measure the quality of a split) brought by that feature. In Figure 15 it is possible to observe that the relevance of static and dynamic features remains constant over the 4-time horizons. Specifically, static features contribute mainly to the explainability of the model with 80% of computed relevance.

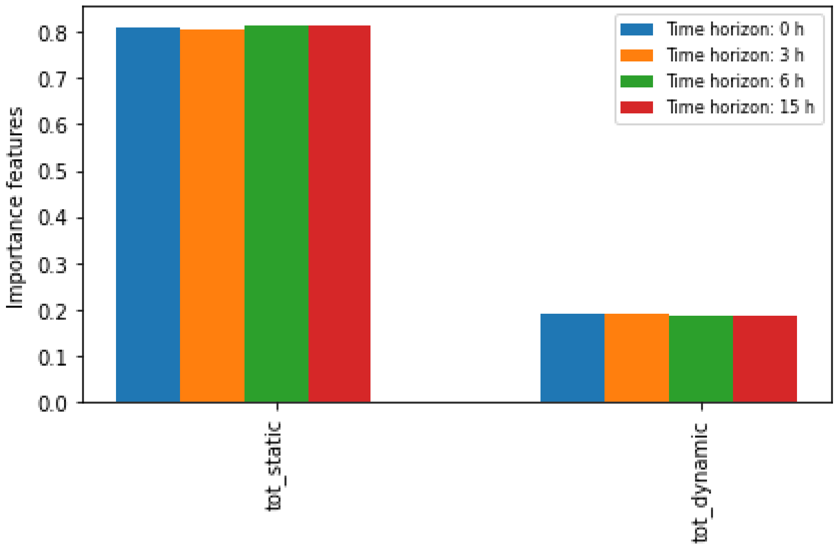


Figure 15 Features importance of static and dynamic features over the four time horizons of the model estimating the minimum turnaround time.

In a more detailed analysis we considered the contribution of the grouped features to the model for minimum turnaround time. As a result (see Figure 16), we can observe that the main features are the wake turbulence category of the aircraft and the four features for congestion at departure and arrival computed as shown in Table 22.

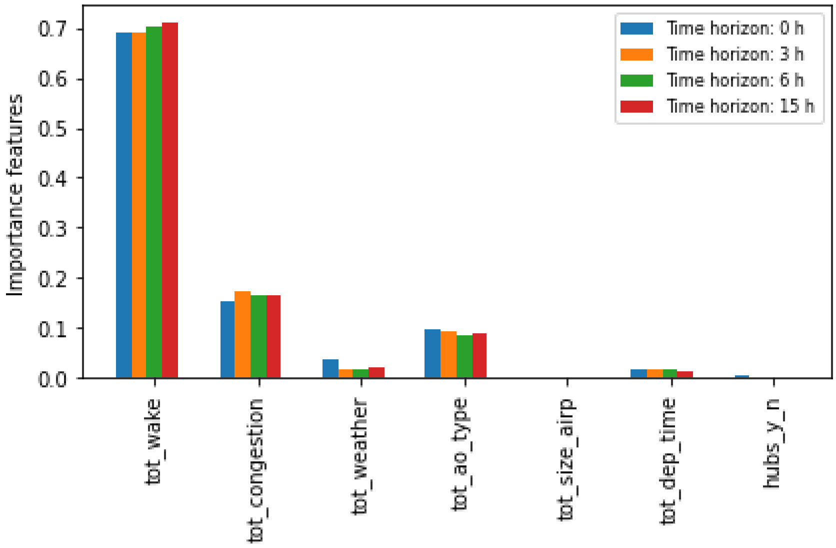


Figure 16 Features importance of grouped static and dynamic features over the four time horizons for the model estimating the minimum turnaround time.

This analysis confirms that overall the dynamic features play quite an important role in the minimum turnaround model (20% of model explainability), however, their contribution remains almost constant over the four time horizons. Further analysis will be required to quantify how much each of these dynamic features changes over time for a given instance (flight).

4.2.1.3 Modelling of uncertainty

Machine Learning (ML) models can be used to predict air traffic operational parameters, which, for example, are required to estimate how delays and costs will be generated and propagated in the air transport network. These costs typically grow non-linearly as the delay increases and can present sharp increments after certain thresholds (Cook and Tanner, 2015), e.g. breaching a curfew at the end of the day (Boeing, 2019; Gurtner et al., 2021) or having to compensate passengers due to delay as indicated in Regulation 261 (European Commission, 2004). Therefore, the expected cost associated with small probabilities of high delay can be very relevant and sometime dominate the expected cost generated, instead, by high probabilities of delay. Being able to capture not only the possible delay that will be propagated through the day, but the distribution of that delay is therefore paramount to estimate the expected cost of these operations. This is particularly relevant for Pilot3 where the estimators are used to compute the expected cost of delay as a function of the uncertain arrival time.

These ML models, which aim to describe an uncertain environment are necessarily complex and present some level of inaccuracy. This inaccuracy is generally measured with metrics that are representative of the overall quality of the models but do not provide information about the level of inaccuracy and uncertainty of each single prediction. Once a model has been trained, some error is expected between predicted and actual realisation of the target variable. This error accounts for both aleatory uncertainty in the phenomena being modelled and epistemic uncertainty in the capability of the model to represent the relationship between features and target variables. In most cases, while the quantification of the uncertainty would be crucial to the comprehension of the problem, the differentiation between its two possible sources is highly challenging or even not possible. The homoscedasticity of the error on the predictions by the model cannot be assumed for several reasons: the training set could be more or less disperse on different regions of the feature space; the underlying processes and relationships being modelled could present aleatory uncertainty; and the machine learning model might have limitations which could produce more accurate predictions on different regions of the feature space. For this reason, averaged statistics and the distribution of the error on the predictions for the entire validation set cannot generally be used as an estimation of the uncertainty of a single prediction. The local uncertainty of the model could be different than the average dispersion of the error and even present some skewness.

Different approaches have been suggested in the literature to overcome these limitations and to estimate the uncertainty and reliability of the individual predictions (please see Appendix C). Most of these methods provide either an estimation of the variance of the error or an interval of reliability but are not able to describe the distribution of possible values. In Pilot3, we propose the use of a probabilistic classifier to characterise the distribution of the error of a prediction relying on the estimation of this error on the training set, obtaining the discrete distribution of the possible expected values of the prediction.

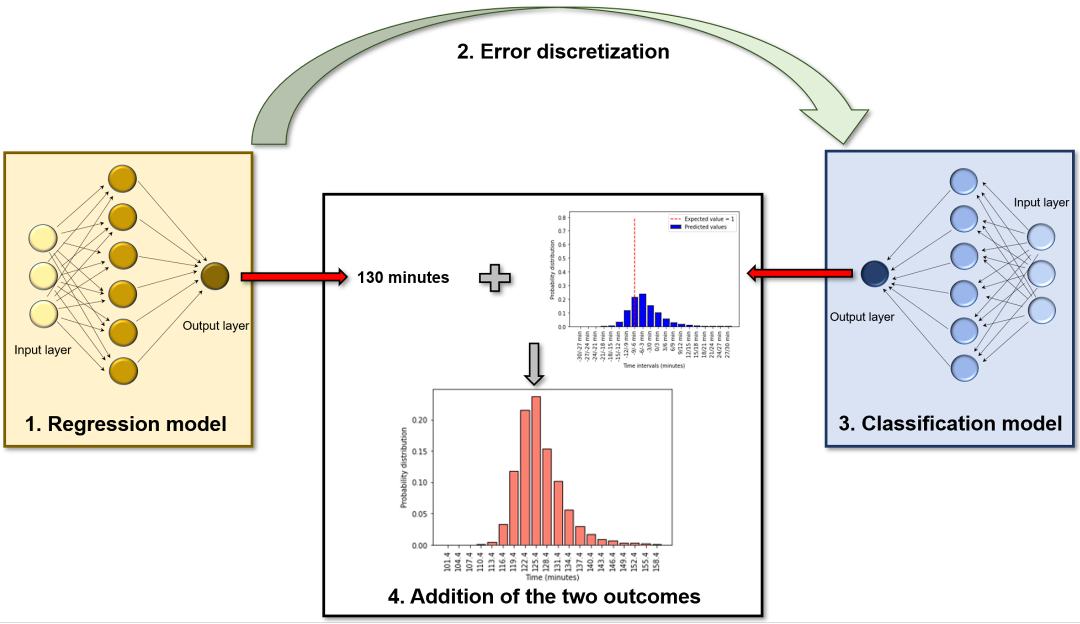


Figure 17 Steps to compute a ML outcome as a probabilistic distribution using regression and classification models

The approach can be summarised by Figure 17. Specifically, a regression model is designed to predict a desired continuous target variable (step 1). Secondly, the error of the model for each prediction is computed and discretised (step 2). Then, a classifier predicts the discretised error (step 3). This classifier will use the categorical cross-entropy as loss function and it will therefore predict a distribution of the error as a probability distribution for each individual prediction. The outcomes of the regression and classification models are finally combined producing a discrete probability distribution of the initial target variable (step 4). More in detail, the error distribution predicted by the classifier is centred on and added to the value predicted by the regression model leading to the probability distribution of the target variable. Note that the first and second steps could be avoided and a classifier directly used to predict the target variable. However, the use of the regression model

EUROPEAN UNION

reduces the range of possible values to be modelled by the classifier which only focuses on the error of this first regression model.

It is certainly desirable to quantify the level of uncertainty associated with these discrete distributions. Some concepts can be borrowed from information theory in an attempt to describe the uncertainty embedded by probabilistic distributions (Brownlee, 2019). As an example, entropy could be used as a metric to quantify the level of uncertainty of discrete probability distributions and cross-entropy can be a measure of the quality of predictions. Entropy provides a measure of the average amount of information needed to represent an event as probabilistic distribution. The lowest entropy for a random variable is 0 and occurs when there is a single event with a probability of 1, i.e., a certainty. The largest possible entropy represents a random variable for which all events are equally likely. Cross-entropy is a measure of the difference between two probability distributions. Cross-entropy is the calculation of the total entropy of the distributions. More accurately, when a target probability distribution P and an approximation of it, Q, are considered, the cross-entropy of Q from P is the number of bits to represent an event using Q instead of P.

Although, these metrics follow a formal and precise definition, their interpretation might be not intuitive and might require further knowledge of the topic. Therefore, in Pilot3 we introduced more intuitive metrics translating the uncertainty to more operational parameters. For this, the range of values (time in our models) that is captured by a given percentile (95%) is used. A larger time range will correspond to higher uncertainty and vice-versa. To assess the accuracy of the models, instead, the expected value of the individual predicted distributions is computed (after combining the output of the regression and classifier models - step 4 in Figure 17) and these values are used to calculate the mean absolute error with respect to the target values of the regression model.

An interesting metric that could support the assessment of the quality of the models could be the 'reliability' interpreted as the model capability to provide predictions with low uncertainty when the accuracy is high and with a higher uncertainty when the accuracy is low. However, this metric has been neither defined nor used in Pilot3 and could be included as part of possible future developments.

4.2.2 Performance Indicator Estimator

4.2.2.1 Architecture and main validation results

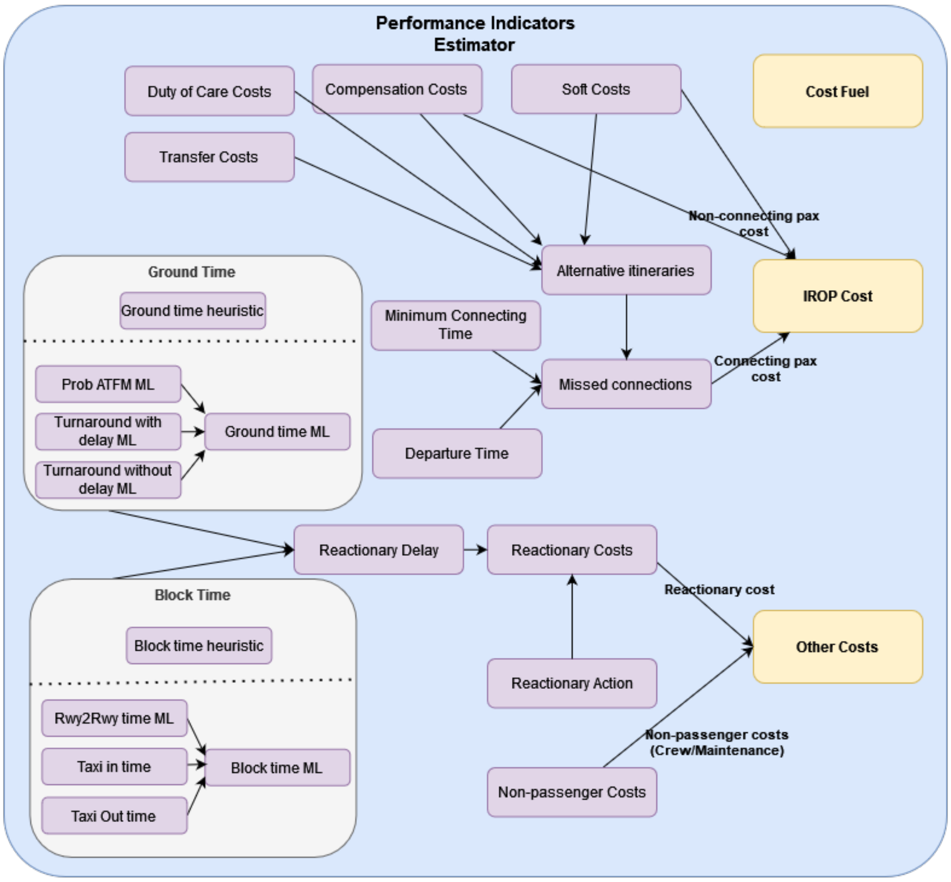


Figure 18 Performance Indicator Estimator components

Figure 18 presents the different estimators that are provided by the Performance Indicator Estimators module. These are used to estimate the cost of delay related to IROPS (due to passenger) and other costs (considering crew, maintenance and reactionary costs).

These costs are computed as follow:

- IROP Costs consisting on:
  - Non-connecting passenger costs, which require modelling of compensation costs (due to Regulation 261) and soft costs, and
  - Connecting passengers, which will experience costs if they miss their connection and as a function of when the following alternatives are available to reach their destination. Therefore to estimate these costs Pilot3 requires:
    - Estimation of the different alternatives to reach their final destination and their associated costs: compensation, soft costs, duty of care and transfer costs.

        - Estimation of the possibility of missing a connection and of making a given alternative as a function of the arrival time at the gate of the flight. This needs two estimators:
            - Estimation of the minimum connecting time, and
            - Estimation of the departing time of subsequent flights.
- Other costs consisting on:
    - Non-passenger costs for the current flight modelling crew and maintenance related cost, and
    - Reactionary delay costs which will depend on the estimated amount of reactionary delay propagated to subsequent flights and on the possibility (and cost) of performing a pre-tactical (strategic) action such as cancelling or swapping a flight downstream. This means that the reactionary cost will depend on:
        - Reactionary costs computed as a function of the estimated reactionary delay, which is estimated modelling the rotation of subsequent flights with (see Section 4.2.2.2.2)):
            - Ground time (rotation) time estimation and
            - Block time estimation.
        - Reactionary action estimator to capture the possibility and cost of an action being performed by the AOCC to limit the propagation of delay (see Section 4.2.2.2.4)

Therefore, besides the models used to estimate the costs, which are based on the European cost of delay report (Cook and Tanner, 2015) the following estimators are required, for which in this section information on the validation and results of the heuristics and machine learning models are provided:

- Passenger connections estimation (Section 4.2.2.2.1.3), which requires:
    - minimum connecting time estimator, and
    - departure time estimator.
- Reactionary cost estimation (Section 4.2.2.2.2), which requires for subsequent flights:
    - block time estimation (Section 4.2.2.2.2.1)
    - ground time (rotation time) estimation (Section 4.2.2.2.2.2), and
    - reactionary strategic action being performed (Section 4.2.2.2.2.4)

Table 19 presents a summary of the different estimators and how they have been validated.

Table 19: Summary of Performance Indicator Estimators and their validation

Component	Estimator	Estimator approach	Validation approach	Result	Mode details in Section
Minimum connecting time	Heuristic	Based on reported historical data (considering airport and type of connection). Uncertainty can be added to the prediction.	Provide values from historical dataset (Cook et al., 2012).	Values used in previous research projects (Gurtner et al., 2021)	4.2.2.2.1.1
Departure time estimation	Heuristic	Provide SOBT as estimation of departure time for other flights in the network. Data could be updated in Pilot3 during the flight. Uncertainty can be added to the prediction.	-	-	4.2.2.2.1.2
Block time estimation	Heuristic	Block time planned by airline (SIBT - SOBT)	Schedules from historical dataset (Gurtner et al, 2021)	Values used in previous research projects (Gurtner et al, 2021)	4.2.2.2.2.1
Block time estimation	Machine learning	Two model approach (prediction + classifier) to provide prediction with uncertainty. ANN for prediction, ANN for classification	Label dataset based on DDR historical block times.	Accuracy ~5.6 minutes Average uncertainty ~27 minutes	4.2.2.2.2.1
Ground time (Rotation time)	Heuristic	Only minimum turnaround time is estimated	-	-	4.2.2.2.2.2
Ground time (Rotation time)	Machine learning	Combination of three models: probability of ATFM delay, minimum turnaround time if no ATFM delay is assigned, and ground time if ATFM is assigned.	-	-	4.2.2.2.2.2
ATFM delay probability estimation	Machine learning	Binary classifier with ANN.	Label dataset based on DDR historical information of flights being regulated due to ATFM.	Confusion matrix (predicted/actual): • delayed/delayed: 47% • non-delayed/delayed: 53% • delayed/non-delayed: 28% non-delayed/non-delayed: 72%	4.2.2.2.2.2
Minimum turnaround time	Heuristic	Provide minimum turnaround time based on analysis of rotation times from historical dataset (DDR) considering 2 percentile grouping the data based on WTC, airport size and airline type.	-	Values used in previous research projects (Gurtner et al, 2021).	4.2.2.2.2.2

Component	Estimator	Estimator approach	Validation approach	Result	Mode details in Section
Minimum turnaround time	Machine learning	Two model approach (prediction + classifier) is used to estimate the minimum turnaround time.	Labelling performed with a regression tree. This is used to group similar rotation times per WTC category aircraft. The 2 percentile of each leave node is used as the estimated minimum turnaround time of those flights.	Accuracy ~ 2.6 minutes Average uncertainty ~ 4.8 minutes	4.2.2.2.2.2
Rotation time if flight regulated	Machine learning	Two model approach (prediction + classifier) is used to estimate the rotation time if flights have ATFM delay assigned.	Label dataset based on DDR historical information considering only flights which have ATFM delay issued.	Accuracy ~ 18 minutes Average uncertainty ~ 72.6 minutes	4.2.2.2.2.2
Reactionary delay strategic action	Heuristic	Strategic action considered that will be performed as an outcome of two models: possibility of doing an action (dependent on how many legs downstream the flight to which the action will be performed is), and the expected cost of delay if no action is done.	Parameters manually tuned as proof of concept and expected reactionary delay cost compared with estimates obtained from analysis of European cost of delay report (Cook and Tanner, 2015)	Values significantly lower than do-nothing approach and aligned with European cost of delay report.	4.2.2.2.2.4

4.2.2.2 Components

4.2.2.2.1 Passenger connections estimation

4.2.2.2.1.1 Minimum connecting time

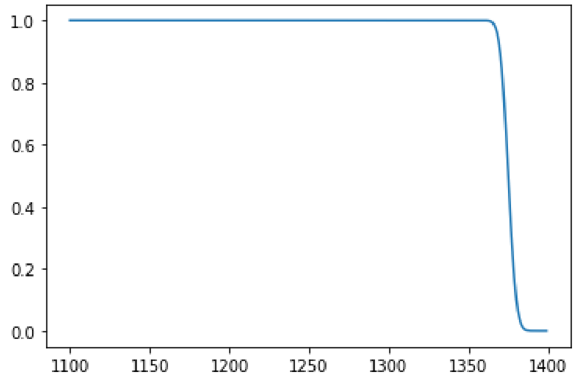
The minimum connecting time is computed by using heuristics which consider the airport where the connection is performed and the type of connection: domestic, international or standard. These values are obtained from the analysis of minimum connecting times at ECAC airports originally performed in POEM project (Cook et al., 2012) and used in previous research projects such as Domino (Gurtner et al., 2021).

Pilot3 gives the possibility to provide some uncertainty. This uncertainty will be used to create a normal distribution centered at the minimum connecting time with a sigma as the provided by the user.

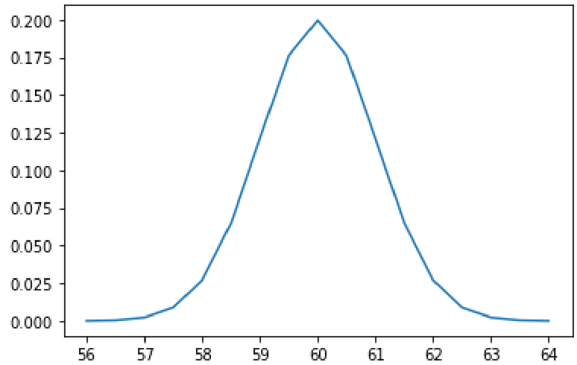
4.2.2.2.1.2 Departure time

For the departure time of flights which are not the current flight, Pilot3 relies on using the SOBT as loaded in the flight prior departure. The system allows for this information to be updated during the flight. As with the minimum connecting time a parameter of uncertainty could be provided.

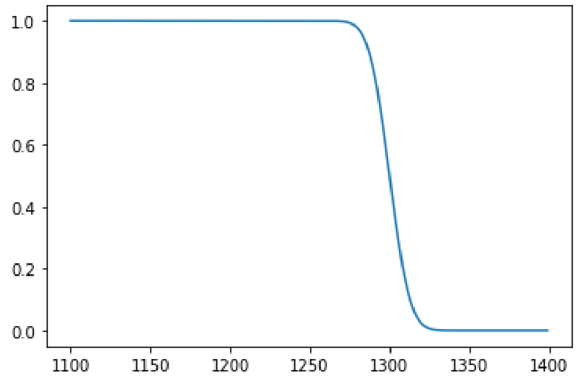
4.2.2.2.1.3 Passenger connections estimations



(a) Probability connecting flight will still be at the hub



(b) Minimum connecting time of passenger



(c) Probability of making the connection

Figure 19 Modelling of probability of missing connections

As presented in Figure 19, Pilot3 estimates for each alternative available to a passenger to arrive to their destination the possibility that the passenger could make the connection as a function of the

EUROPEAN UNION

arrival time of the inbound flight by combining the probability that the connecting flight is there and the minimum connecting time required by the passenger.

4.2.2.2.2 Reactionary estimation

4.2.2.2.2.1 Block time estimation

Heuristics

The heuristic version of the block time of a flight is estimated as the time between SOBT and SIBT (i.e., SIBT-SOBT). There is the possibility to add as a parameter some uncertainty (sigma). Then, instead of estimating a certainty of SIBT-SOBT, the heuristics estimates a normal distribution with parameters mu = SIBT-SBOT and sigma = sigma.

Machine Learning

In this section a description of both the regression and classification ML models that have been used in Pilot3 to predict the duration of flights (runway-to-runway) is provided. The datasets used for these models are listed below:

- Traffic data from EUROCONTROL's Demand data repository 2 (DDR2)(EUROCONTROL, 2015a). This includes information on the flight plan and the actual realization of flight operations for flights in September 2018. For each flight, among others, there is information on their off-block and landing times. These data and an estimation of the taxi-in times provide the required parameters to compute the labels of the datasets: block time as time from off-block at departure to in-block at arrival. This dataset is also used to estimate some input features such as congestion at airports or arrival direction.
- Airport weather reports. METeorological Aerodrome Reports (FAA, 2016) records describing the weather present at airports from which information such as wind, temperature from which the ATM Airport Performance (ATMAP) weather score can be estimated (Schultz et al., 2018).
- En-route weather extracted from ECMWF Re-Analysis 5 (ERA5) which provides hourly estimates of a large number of atmospheric variables notably wind aloft.
- Aircraft characteristics, such as Wake-turbulence category (WTC) extracted from EUROCONTROL's Base of Aircraft Data (BADA) (EUROCONTROL, 2015b).
- Airport and airlines characteristics, such as status of airport (e.g. hub/no-hub) or type of airline (i.e., full-service, low-cost, regional and charter).

The input features that were extracted and/or computed from these datasets are described in Table 20.

Table 20 Input features for the block-time ML model

Input features	Data sources	Description	Static (S) vs Dynamic
Time departure	DDR2	Categorical (morning, afternoon, evening)	S
Airline type	DDR2	Categorical (type of airline, e.g. regional)	S
Congestion at departure during the day of operations	DDR2	Numerical (number of flights departing from the airport during the hour of interest)	S

Input features	Data sources	Description	Static (S) vs Dynamic
Congestion at arrival during the day of operations	DDR2	Numerical (number of flights arriving at the airport during the hour of interest)	S
Landing direction	DDR2	Categorical (e.g. North-West)	D
ATFM regulations	DDR2	Categorical (presence of imposed ATFM regulations)	S
Great Circle Distance (GCD)	DDR2	Numerical (distance between departing and arrival airports)	S
Direction of flight	DDR2	Categorical (e.g. North-West)	S
ATMAP weather score at departure airport	METAR	Numerical	D
ATMAP weather score at arrival airport	METAR	Numerical	D
Wind speed at departure airport	METAR	Numerical	D
Wind speed at arrival airport	METAR	Numerical	D
Temperature at departure airport	METAR	Numerical	D
Temperature at arrival airport	METAR	Numerical	D
Airport hub	Airport data	Categorical (yes/no)	S
Average wind along trajectory	ERA5	Numerical	D
Size aircraft	BADA	Categorical (low, medium, high, jumbo)	S
Size airport departure	Airport data	Categorical (small, medium, big)	S

The selection of the input features for the model initially relies on the experience and expertise of the modeller. However, there are well established techniques that allow to assess the contribution to the model explainability of the features. One of these techniques is called SHAP (Shapley additive explanations) (Shapley, 1953) analysis which quantifies the relevance of the input features used by a ML model. We performed this analysis for the regression model to rate but also to verify our initial intuition about the more explanatory features. This analysis, indeed, confirmed our initial hypotheses that the phenomenon we are here modelling is mainly 'static' (in other words it is affected mainly by the static features that we included in the model) and its accuracy is mainly driven by the Great Circle Distance with approximately 33% of SHAP relevance (See Figure 20).

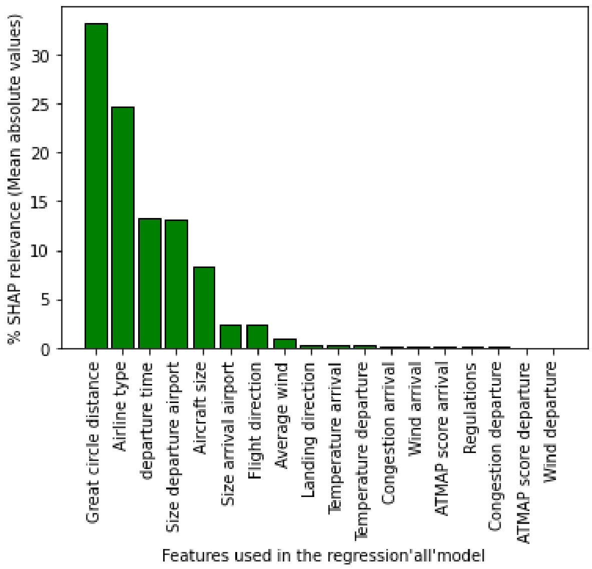


Figure 20 SHAP analysis to assess the impact of the features on the regression problem for the block time estimation

An analysis of the possible models' hyper-parameters was implemented via Grid Search which allows to define a search space as a grid of hyper parameters values and evaluate the combinations of parameters. This analysis led to the hyper-parameters described in Table 21.

Table 21 Hyper-parameters used for the regression and classification ML models.

	Regression model	Classification model
Learning rate	1e-4	1e-4
Number of hidden layers	2	1
Number of neurons in the hidden layers	20-8	30
Optimizer	Adam	Adam
Mini-batch size	64	32
Activation functions	Relu/Linear	Relu/Softmax
Weights initialization	Glorot uniform	Glorot uniform
Number of epochs	25	25
Loss function	Mean absolute error	Categorial Cross-entropy
% of data for training/validating/ testing	60/20/20	60/20/20

In order to assess the quality of the entire two-model approach we firstly characterised the accuracy of the single regression and classification models and later we introduced specific metrics to characterise the accuracy, uncertainty and reliability of the entire two-model approach.

The mean absolute error and its standard deviation are used to characterise the accuracy of the regression model, which are respectively of 6.9 and 5.2 minutes. For the classification models, however, we are not interested on the accuracy as usually defined for these types of problems, i.e., class with highest probability being the one of the labelled sets. Instead, the quality of the distribution of uncertainty across classes should be captured. This is due to the fact that the different classes in this problem are ordered and their probability will be used to describe the uncertainty of the prediction. Therefore, specific metrics need to be considered. Cross-entropy (Brownlee, 2019) can be a measure of the level of accuracy of the predictions delivered by ML classifier models. However, the cross-entropy function does not have a maximum limit making its normalization and comparison across different distributions challenging. As it is a normalised measure of diversity when comparing probability distributions, Jensen-Shannon divergence (Brownlee, 2019) is preferred a measure of the level of accuracy of the ML classifier models. Another metric that can be considered for quantifying the performance of a classifier is an average of the absolute error values. In this case, the error is computed as the difference between the expected value from the predicted probability distributions and the target probability distributions. The Jensen-Shannon value for the classifier we used is ≈0.83 and the average of the absolute error values is ≈5.7 minutes.

The accuracy of the full model is computed combining the output of the regression and classifier models (i.e., adding the error computed by the classifier to the value estimated by the regression model) and using the expected value of the individual predicted distributions. According to this, the accuracy of the model is overall of ≈5.6 minutes. For the assessment of the uncertainty embedded with the prediction, a normalized entropy function could be used. However, the temporal range within which 95% of the probability values for each distribution falls is used instead as it provides a more operational metric. As a result, the uncertainty of the model is estimated at ≈27 minutes (in average).

In order to obtain the gate-to-gate block time of flights the duration of the taxi-in and taxi-out are added to the predictions of the runway-to-runway model via convolution.

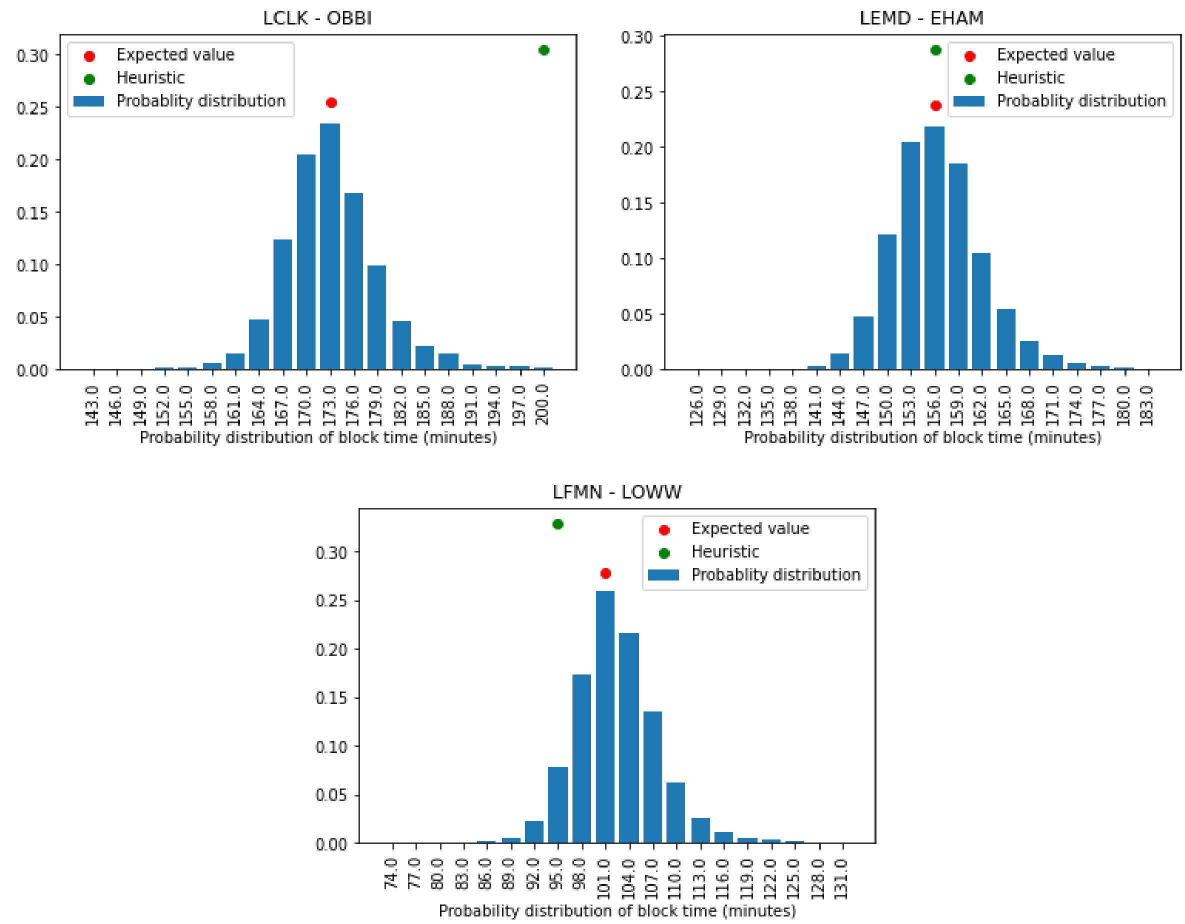


Figure 21 Comparison between the estimations of gate-to-gate block time obtained with a heuristic approach (from schedule times) and with our ML model

Figure 21 presents the comparison of three example of estimations of block-times performed by the heuristic approach (i.e., planned time between SIBT and SOBT) and the machine learning approach. In

EUROPEAN UNION

these three examples, one can observe how in some instances the ML model will in average estimate a block time smaller than the one planned by the airline, i.e., some padding is present, be aligned with it, or estimate that the block time will be larger than the planned time.

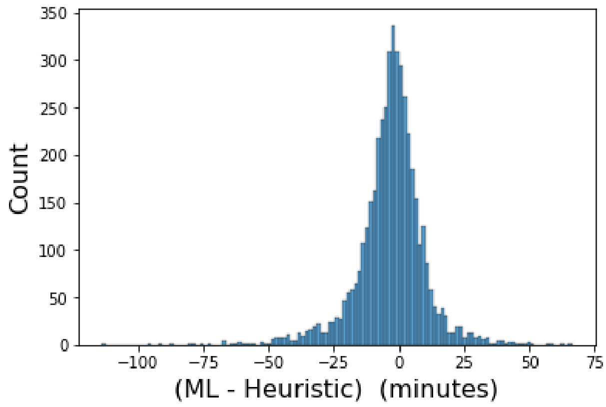


Figure 22 Histogram showing the difference between the predictions of the ML and the heuristic models over a set of ≈5000 testing samples

In general, the ML gate-to-gate block time model produces shorter estimations than the heuristic ones (approximately 3.5 minutes less over a testing set of ≈5000 samples). This is consistent with the general practice of having padding on the schedules. Figure 22, shows the distribution of difference between the expected prediction of the ML model and the estimation by the heuristic using only the planned block time by the airline.

4.2.2.2.2.2 Rotation time (ground time)

In order to estimate the reactionary delay, Pilot3 uses an estimation on when the flight will be ready for the next rotation. It was not possible to directly estimate the time that the aircraft would be doing the rotation as this will require a dataset which includes planned schedules and realised rotations. The only dataset available is the EUROCONTROL DDR dataset from which it is possible to estimate when an aircraft arrives at an airport and when it leaves again (EUROCONTROL, 2015a). However, it is not possible to estimate from that if the aircraft was performing actions while on ground or if some buffer was being used. For this reason, ground time models tend to focus on the estimation of minimum turnaround time which is then combined with a sequence of schedules to estimate the propagation of delay, i.e., if the flight is ready after its schedule it is assumed it will depart then, if it's ready before it will wait until its planned SOBT. This will be the approach followed on the heuristic modelling of ground time.

For the machine learning model, a combination of minimum turnaround time, which is estimated using dynamic features, and delay due to ATFM regulations is considered. The assumption is that if a flight has been assigned ATFM delay the observed time on the ground will not count with the above-mentioned buffer, and therefore it can be assumed it's the minimum time the flight needs to be on the ground if ATFM delay is issued. Therefore, the machine learning approach will feature a combination of the outcome of three models: one to determine if a flight is being affected by ATFM delay, another one for the estimation of the minimum turnaround time if the flight does not have ATFM delay assigned, and a final model to estimate the total rotation time if the flight is affected by the regulation.

For these reasons the individual models (heuristic to estimate minimum turnaround time, probability ATFM delay, machine learning minimum turnaround time, and machine learning rotation time if ATFM issue) will be validated among their dataset, i.e., among their labelled data. But the integrated rotation/ground time, cannot be validated against an historic dataset of rotations.

Heuristic

The heuristic approach to estimate the minimum rotation time is based on the use of an historical analysis of ground times observed in data. Data on arriving and departing of subsequent flights has been collected from ALL_FT+ and analysed grouping the information considering aircraft wake turbulence, airport size and airline type. For each of these categories the 2 percentile of the rotation times is used as an approximation for the minimum rotation time required by a flight with those characteristics (wake turbulence, airport size and airline type). This is the same approach as followed in previous European SESAR research projects such as Domino (Gurtner et al., 2021).

Machine Learning

The ground time is estimated as a probabilistic distribution of the time that an aircraft will spend at the gate between each rotation. For the estimation of the ground time three ML models are implemented. Figure 23 shows the working principles of the full ground time model.

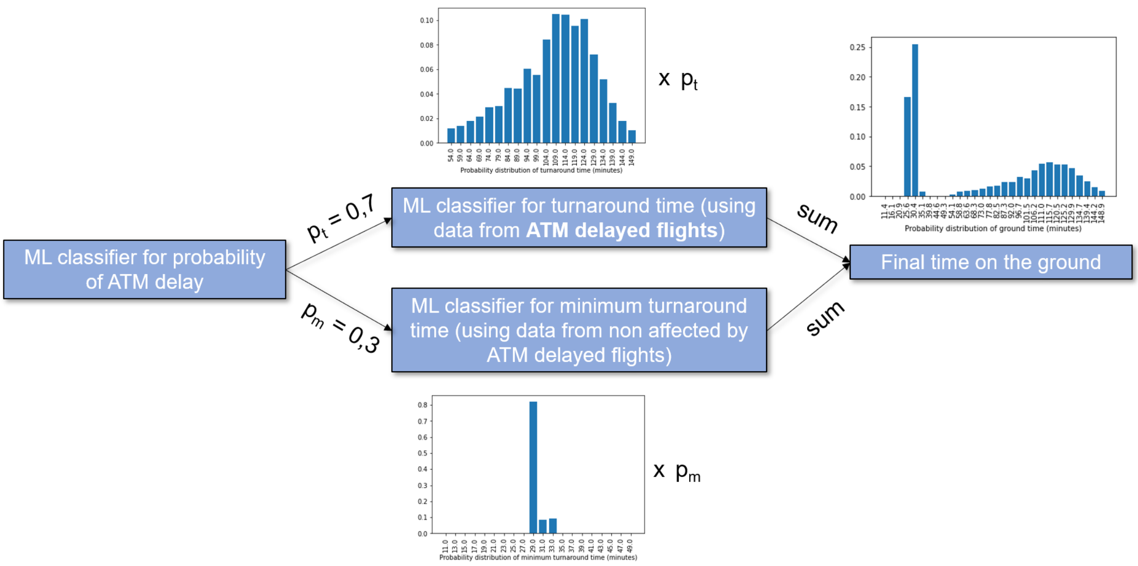


Figure 23 Schematic of the working principles of the three models implemented to predict the ground time of flights

First, a classifier which outputs two coefficients, one predicting the probability that ATFM delay will be assigned to the flight and another describing the possibility for the aircraft to operate on the ground as fast as possible without any imposed delay is used. Then, the two further models predict the time distribution for turnaround operations when ATFM delay is imposed ('Turnaround with delay model'), and another the time distribution for similar operations when no delay is assigned to the flight ('Minimum turnaround model'). These two predicted distributions are then combined together considering the probability of their occurrence as estimated by the first model (weighted sum).

The three models have been trained using the input features listed in Table 22.

EUROPEAN UNION

Table 22 Input features for ground-time ML models

Input features	Data sources	Description	Static (S) vs Dynamic
Time departure	DDR2	Categorical (morning, afternoon, evening)	S
Airline type	DDR2	Categorical (type of airline, e.g. regional)	S
Congestion at departure during the day of operations	DDR2	Numerical (number of flights departing from the airport during the hour of interest)	S
Size of arrival airport	Airport data	Categorical (small, medium, big)	S
ATFM regulations	DDR2	Categorical (presence of imposed ATFM regulations)	S
Congestion at departure during the hour of operations	DDR2	Categorical based on threshold value (yes/no)	D
Congestion at arrival during the hour of operations	DDR2	Categorical based on threshold value (yes/no)	D
ATMAP weather score at departure airport	METAR	Numerical	D
Wind speed at departure airport	METAR	Numerical	D
Temperature at departure airport	METAR	Numerical	D
Airport hub	Airport data	Categorical (yes/no)	S
Size aircraft (except for the 'Minimum turnaround model')	BADA	Categorical (low, medium, high, jumbo)	S
Size airport departure	Airport data	Categorical (small, medium, big)	S

Probability ATFM delay

This first model (an ANN (Artificial Neural Network) classifier) predicts the probability for a flight to be assigned ATFM delay. The target vector during the training process was obtained from for the ALLFT+ DDR data by labelling as delayed flights the ones having a non-empty COBT (computed off-block time) field. The input features are listed in the table above (Table 22) while the model parameters are listed in Table 23.

Table 23 Hyper-parameters for probability ATFM delay

	Classification model
Learning rate	1e-4
Number of hidden layers	1
Number of neurons in the hidden layers	15
Optimizer	Adam
Mini-batch size	32
Activation functions	Relu/Sigmoid
Weights initialization	Glorot uniform
Number of epochs	15
Loss function	Binary Cross-entropy
% of data for training/validating/ testing	60/20/20

EUROPEAN UNION

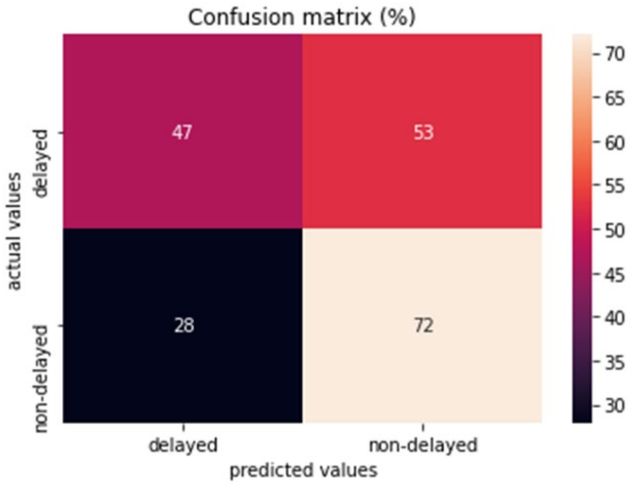


Figure 24 Confusion matrix describing the accuracy of the ATFM delay model

Having an initial unbalanced distribution of target classes (≈66% non-delayed sample data *vs* ≈34% with-delay sample data), the input data has been balanced by up-sampling the input data of the less represented class. The accuracy of the model can be described by the confusion matrix reported in Figure 24. The confusion matrix is achieved by imposing a threshold value of 0.5 when considering the predictions. Therefore, we labelled the predictions (probability coefficients) that are higher than the threshold as 'delayed' and the predictions that are lower than the threshold as 'non delayed'. As shown by Figure 24, 72% of non-delayed and 47% of delayed flights are expected to be predicted as such.

Turnaround estimation with delay

This model predicts probabilistic distributions of ground time (turnaround) under the assumption that ATFM delay is assigned to a certain flight. The input features for this model are listed in Table 22. The target vector for the regression model was extracted using ALLFT+ DDR data by subtracting the actual landing time and the average Taxi-in time (provided by EUROCONTROL statistics) from the actual off-block time (AOBT) of the next leg in the rotations. The approach of regression and classifier is used to estimate the distribution of possible turnaround times (see Section 4.2.1.3).

The model parameters for the regression and classification models used to estimate the turnaround time with delay are shown in Table 24.

Table 24 Hyper-parameters used for the regression and classification ML models used for the turnaround estimation of ATFM delayed flights.

	Regression model	Classification model
Learning rate	1e-5	1e-4
Number of hidden layers	2	1
Number of neurons in the hidden layers	16-8	21
Optimizer	Adam	Adam
Mini-batch size	64	32
Activation functions	Relu/Linear	Relu/Softmax
Weights initialization	Glorot uniform	Glorot uniform
Number of epochs	25	10
Loss function	Mean squared error	Categorial Cross-entropy
% of data for training/validating/ testing	60/20/20	60/20/20

EUROPEAN UNION

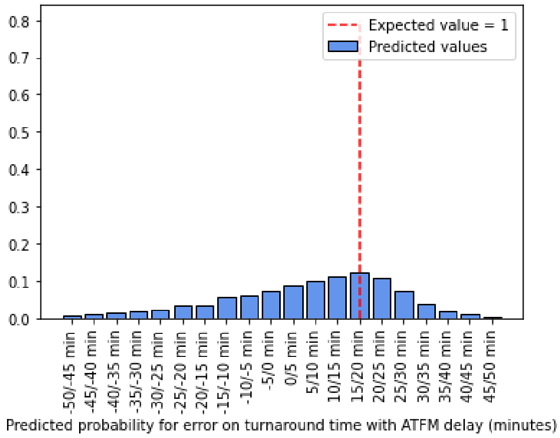


Figure 25 Probability distribution of the error estimation by the classifier model used for the estimation of the turnaround time when ATFM delay is applied

A typical error estimation computed by the classification model is shown in the Figure 25. An initial visual assessment of the quality of the model is possible by comparing the expected value (dashed red line) with the peak area of the probability distribution.

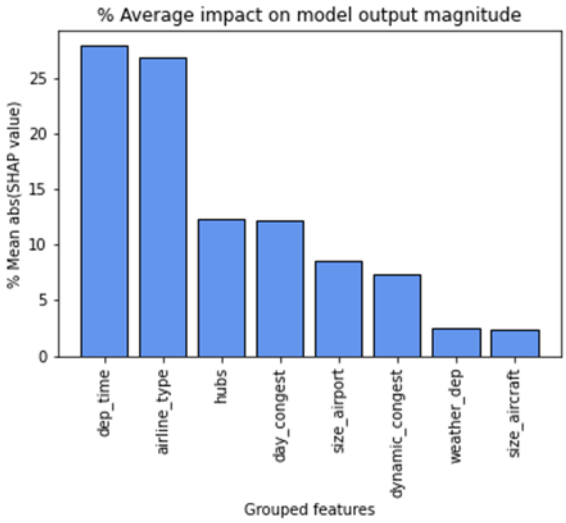


Figure 26 SHAP analysis showing the % mean absolute value of the relevance values

In order to improve the transparency and understanding of this model a SHAP analysis was implemented, as shown in Figure 26. The most relevant features for this model are the time of departure and the airline type.

Table 25 Accuracy performance of turnaround with delay model

Model	Metric	Sample testing size	Result
Regression model to predict the turnaround time	Mean difference (absolute value) between predicted and target value	~40,000 flights	~22 minutes
Classification model to estimate the error of the regression model	Mean difference (absolute value) between expected value from probabilistic distribution and target value	~40,000 flights	~15 minutes
Turnaround with delay model (regression + classification)	Mean difference (absolute value) between expected value from the probabilistic error distribution (predicted by classifier) centred around values predicted by regression model and target value	~40,000 flights	~ 18 minutes

Table 25 presents the performance of the model and the components of the turnaround with delay estimation model. The mean difference (absolute value) between the target and the predicted values is used as a metric for the regression model. This metric results in ≈22 minutes of mean error over a testing sample size of ≈40000 flights.

The mean difference (absolute value) between the expected is used to quantify the performance of the probabilistic distributions and the target values. This metric results in ≈15 minutes of mean error over a testing sample size of ≈40000 flights.

Finally, for the combined two-models the mean difference (absolute value) between the target values and the expected value from the probabilistic error distributions (predicted by the classifier) and centered around the values predicted by the regression model is computed. This metric results in ≈18 minutes of mean error over a testing sample size of ≈40000 flights. The uncertainty of the model, defined as usual in Pilot3 (the range of values that is captured by a given 95% percentile), is ~ 72.6 minutes.

Minimum turnaround time

This model predicts probabilistic distributions of ground time (minimum turnaround) under the assumption that ATFM delay is not assigned to a certain flight. The first problem when trying to estimate the minimum turnaround time is that a dataset with this information is not available. Therefore, a modelling approach is followed to estimate this minimum turnaround time and to build the dataset required to train a model able to predict this value considering dynamic features.

EUROPEAN UNION

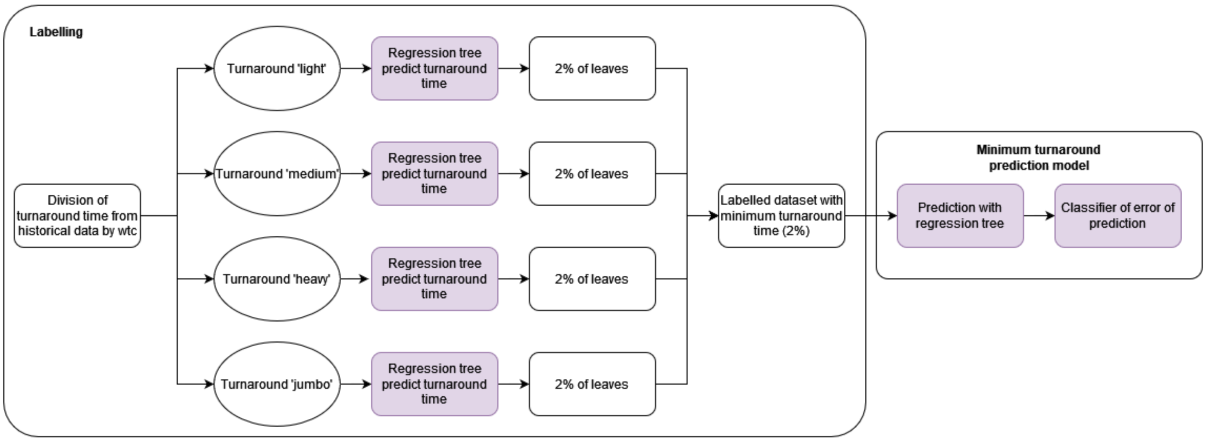


Figure 27 Minimum turnaround estimation approach

Figure 27 presents the overall approach followed:

1. Labelling of the dataset to estimate the minimum turnaround time of a given flight:

    a) The turnaround time for each flight is computed considering historical data from ALL_FT+ base on the landing time of the previous flight and the actual off block time of the subsequent flight subtracting the estimated taxi-in time, as performed in Turnaround estimation with delay presented above.

    b) The size of the aircraft has an important factor on this minimum turnaround time, therefore first the dataset is divided as a function of this parameter.

    c) For each cluster a regression tree is computed to predict the actual rotation time as a function of different features (see Table 20). Then following the approach of the heuristic modelling, the 2 percentile of the values of each final-leave from the tree are computed. This information will build the labelled dataset. This means that even for the same WTC different minimum turnaround times will be expected considering static and dynamic features.

2. The two-model approach presented in Section 4.2.1.3 is used to predict the minimum turnaround time of a given flight while providing information on the distribution of possible values. This is done by training two models:

    a) A regression model to predict the minimum turnaround time of the flight (done using a regression tree).

    b) A discrete classifier ANN to predict the error of the previous regression model. As explained in Section 4.2.1.3. the prediction provided by the estimator will be the combination of these two models.

Table 26 Hyper-parameters for regression tree model for predicting turnaround time used for the labelling phase

	Regression tree model used for labelling of dataset
Pruning (Yes/No)	No
Criterion	Mean squared error
Splitter	Feature with the highest importance
Max depth	6
Minimum samples split	2
Minimum sample leaf	1
CCP alpha	0

Table 26 presents the hyper-parameters of the regression tree used in the labelling phase of the dataset.

Table 27 Hyper-parameters for regression models for minimum turnaround time.

	Regression tree model used to predict minimum turnaround time
Pruning (Yes/No)	Yes
Criterion	Mean squared error
Splitter	Feature with the highest importance
Max depth	6
Minimum samples split	2
Minimum sample leaf	1
CCP alpha	0.03

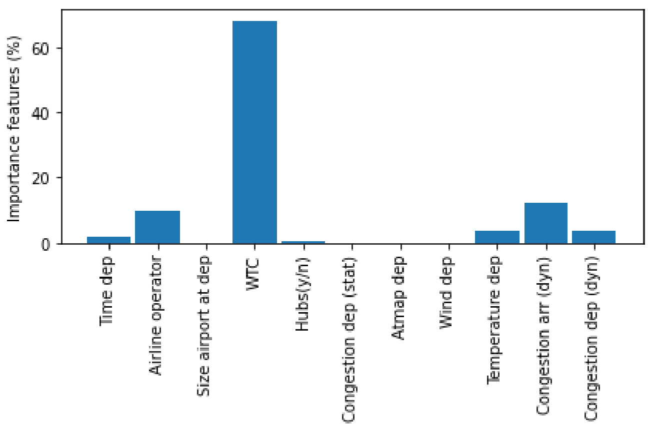


Figure 28 Importance of the input features of the regression model for minimum turnaround data

Table 27 presents the hyper-parameters of the regression tree used to predict the minimum turnaround time (regression model of the two-model approach).

The feature importance for the regression model was computed as the (normalized) total reduction of the criterion brought by that feature. As a result of this analysis (see Figure 28) we can observe that the wake turbulence category (WTC) is the most relevant feature (more than 60% of the feature importance over the entire set of input features) followed by a combination of static and dynamic features: airline type and congestion at arrival and departure.

Table 28 Hyper-parameters for prediction minimum turnaround time classification model.

	Classification model -predict minimum turnaround time
Learning rate	1e-4
Number of hidden layers	1
Number of neurons in the hidden layers	21
Optimizer	Adam
Mini-batch size	32
Activation functions	Relu/Softmax
Weights initialization	Glorot uniform
Number of epochs	35
Loss function	Categorial cross-entropy
% of data for training/validating/ testing	60/20/20

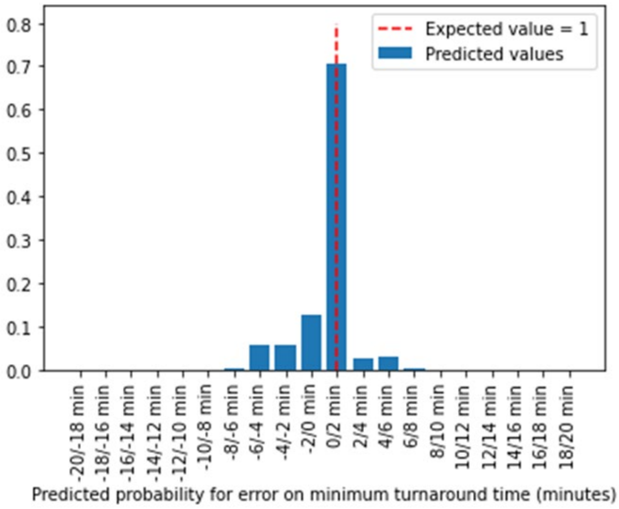


Figure 29 Probability distribution of the error estimation by the classifier model used to estimate the turnaround time when no ATFM is applied

Finally, Table 28 presents the hyper-parameters of the classifier used to estimate the error of the regression tree which predicts the minimum turnaround time. A typical error estimation computed by the classification model is shown in the Figure 29. An initial visual assessment of the quality of the model is possible by comparing the expected value (dashed red line) with the peak area of the probability distribution.

Table 29 The performance of the model and the components of the turnaround without delay estimation model

Model	Metric	Sample testing size	Result
Regression model to estimate the minimum turnaround time	Mean difference (absolute value) between predicted and target value	~78,000 flights	~2 minutes
Classification model to estimate the error of the regression model	Mean difference (absolute value) between expected value from probabilistic distribution and target value	~78,000 flights	~1.5 minutes
Turnaround with delay model (regression + classification)	Mean difference (absolute value) between expected value from the probabilistic error distribution (predicted by classifier) centred around values predicted by regression model and target value	~78,000 flights	~ 2.6 minutes

The different performance observed by the models are provided in Table 29. The mean difference (absolute value) between the target and the predicted values is used as performance metric of the regression model. This metric results in ≈2 minutes of mean error over a testing sample size of ≈78000 flights.

The mean difference (absolute value) between the expected value from the probabilistic distributions and the target values is used as a metric for the classifier. This metric results in ≈1.5 minutes of mean error over a testing sample size of ≈78000 flights.

The mean difference (absolute value) between the target values and the expected value from the probabilistic error distributions (predicted by the classifier) and centred around the values predicted by the regression model is used for the two-model approach for the final prediction. This metric results in ≈2.6 minutes of mean error over a testing sample size of ≈78000 flights. The uncertainty of the model, defined as usual in Pilot3 (the range of values that is captured by a given 95% percentile), is ~ 4.8 minutes.

<u>**Comparison heuristic and machine learning for rotation time estimation**</u>

Figure 30 show sample examples where the estimations from the ML models for the ground time are compared with the heuristic predictions.

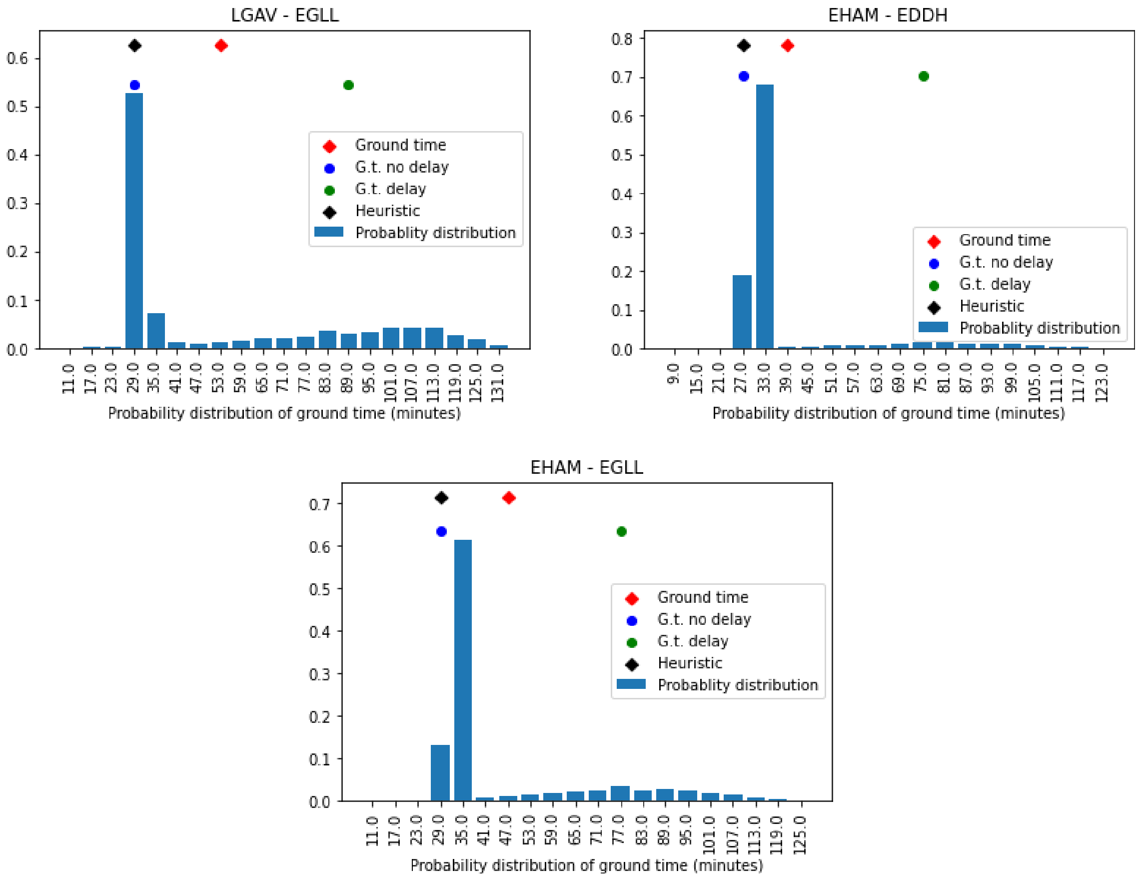


Figure 30 Probability distributions for the ground time predicted using ML models. The figures also allow a visual comparison between the heuristic and the ML estimators

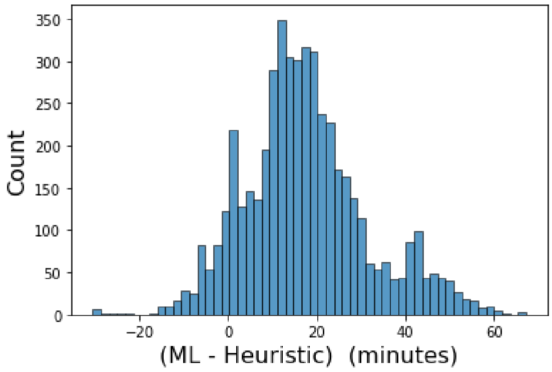


Figure 31 Histogram showing the difference between the predictions of the ML and the heuristic models over a set of ≈5000 testing samples

Although the ML model estimating the probability for a flight of being assigned with ATFM delay will tend to assign an overall higher probability to non-delayed ground time operations we can observe that our ML ground time estimator overall overestimates the predictions of the heuristic model (see Figure 31). Indeed, when comparing the predictions of two estimators, the ML approach overall overestimates the ground time of approximately 17.5 minutes. This is to be expected and shows how considering the possibility of ATFM delay on the turnaround time will lead to estimation of higher probability of propagating delay. Note also how in Figure 30 the heuristic estimation of the turnaround time is aligned with the expected value of the estimation of the ML models if no ATFM delay is considered.

4.2.2.2.2.3 Reactionary delay

Integration of reactionary delay approach and example

The delay that a certain (delayed) flight propagates, affecting the ground and flight operations of the subsequent flights in the rotations. As mentioned in Section 4.2.2.1 the reactionary delay of each subsequent planned rotation is computed by combining the outcome of rotation and block time estimation of the different flights (see Figure 32).

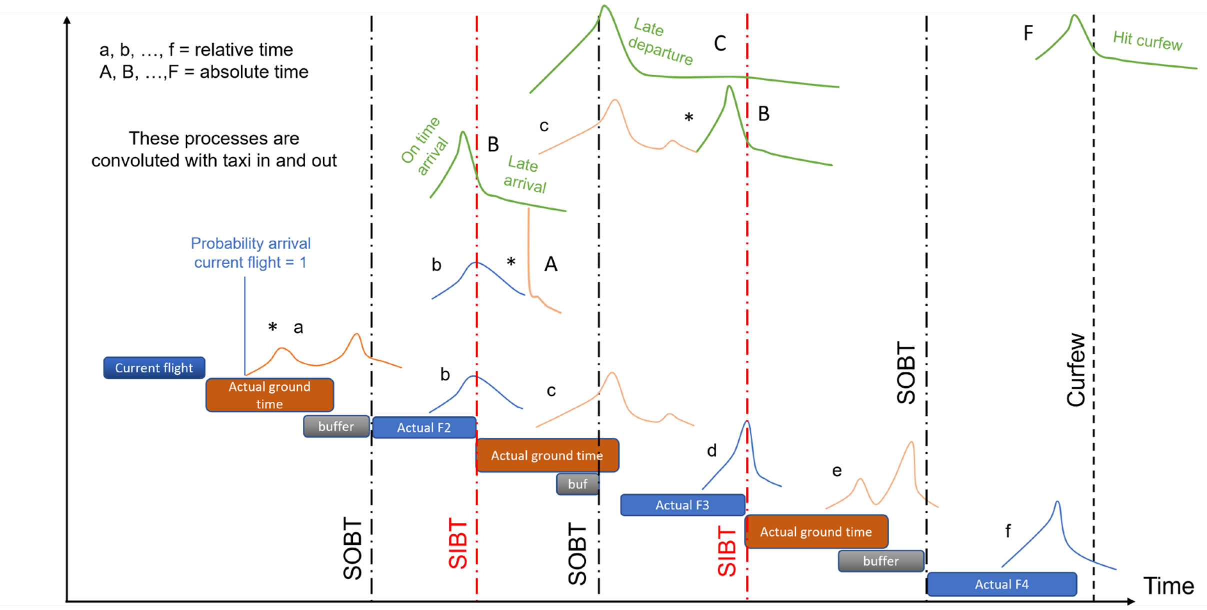


Figure 32 Example of estimation of ground and block times and their convolution to estimate reactionary delay

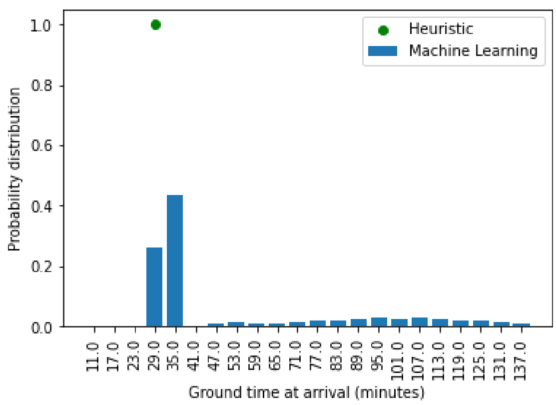
Each model (ground and block time) produces as an outcome a probability distribution of the duration of these processes. The convolution of these times produces allows Pilot3 to estimate the arrival and departure time distributions for each flight. The reactionary delay model will compute for each flight the time when the aircraft is ready to depart as the distribution of arrival time of previous rotation and the ground time estimated. If this time is earlier than the planned SOBT of the flight, the model assumes that the aircraft will wait at the departure airport and leave at the SOBT, otherwise it will leave immediately. This approach allows Pilot3 to implicitly estimate, and to consider, possible padding that the airline has planned on their schedules.

EUROPEAN UNION

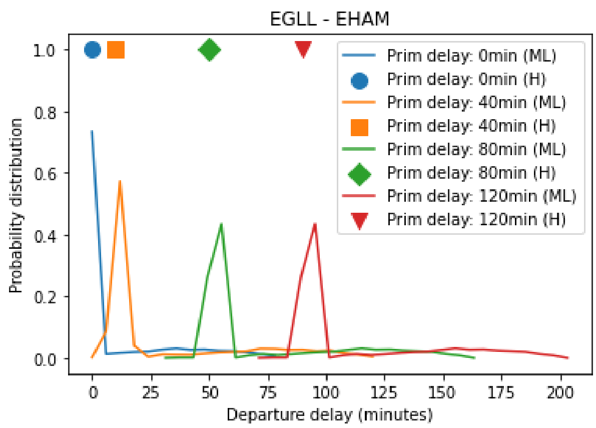
Table 30 Flight information for a set of rotations of a specific aircraft. In red is highlighted the flight provoking primary arrival delay

Origin	Destination	SOBT	SIBT	Planned time for rotation from previous flight
EHAM	EGLL	05:15	06:40	-
EGLL	EHAM	07:40	09:00	1h00
EHAM	LGAV	10:05	13:20	1h05
LGAV	EHAM	14:15	17:45	0h55
EHAM	UUEE	18:50	22:05	1h05

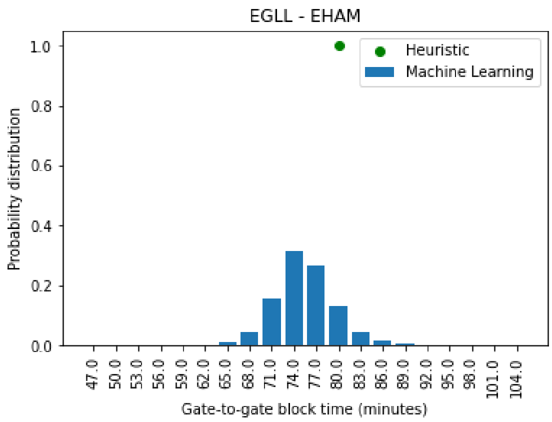
The rotations of a particular flight (an B737-900) for a given day in September 2014, planned as shown in Table 30, are reported as an example. Note that it's assumed that the first flight (EHAM-EGLL) is the *current* flight in Pilot3. Therefore, the propagation of delay will be estimated as a function of the different arrival times of this flight to EGLL.



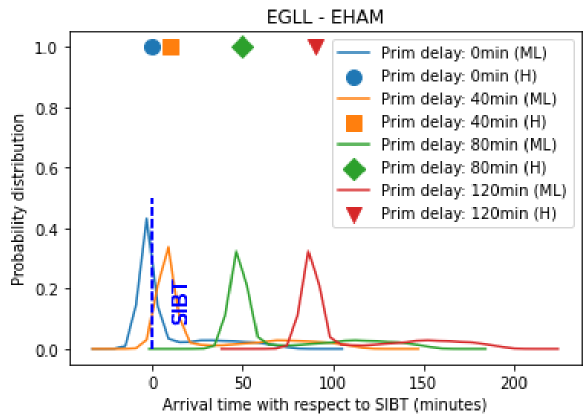
(a) Ground time distribution of the first leg



(b) The convolution of the primary arrival delay with the ground time distribution of the second leg produces the departure delay distribution of the second leg



(c) Gate-to-gate block time distribution

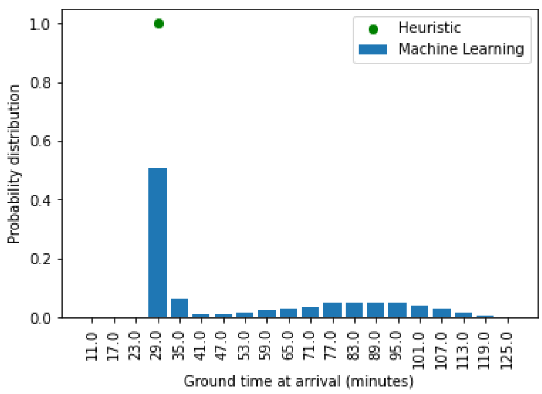


(d) The convolution of the distributions in (b) with the distribution in (c) produces the arrival time distribution of the second leg as function of primary arrival delay

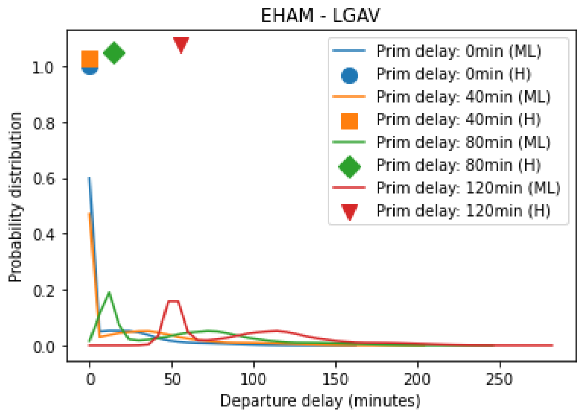
Figure 33 Estimation of propagation of delay for EGLL to EHAM flight (first rotation) (includes heuristic predictions)

All the plots report also the corresponding values obtained with the heuristic model (please note that while the probability of the heuristic values is 1 a slight shift of their values was needed for visual convenience in case of overlaps).

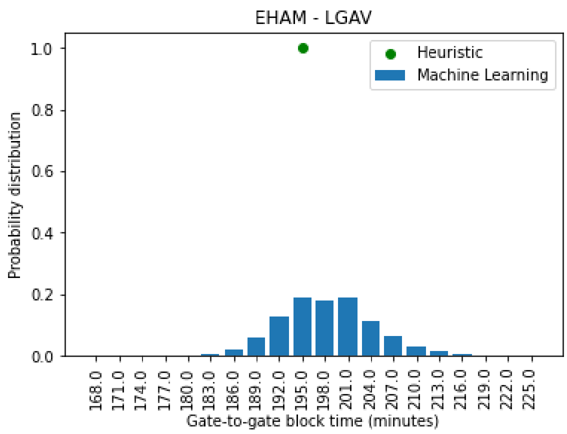
First, Figure 33 presents how once the flight EHAM-EGLL arrives to EGLL there is an estimated rotation (ground) time that will be required before the return flight to EHAM (Figure 33 (a)). As a function of the primary delay experience by the first flight the aircraft will be ready to depart at different moments as shown in Figure 33 (b). Note how if the aircraft is ready before the SOBT for the EGLL-EHAM flight then the flight will wait until that time increasing the probability of departing without delay. Figure 33 (c) presents the estimation of the block time between EGLL and EHAM. Convolving this block time distribution with the different departing time Figure 33 (d) is produced which estimates the arrival time at EHAM of this rotation. It is then possible to compute the probability of the flight to arrive before or after its planned SIBT. All plots Figure 33 also report the corresponding values obtained with the heuristic models (please, note that probability of the heuristic values is 1 a slight shift of their values was needed for visual convenience in case of overlaps).



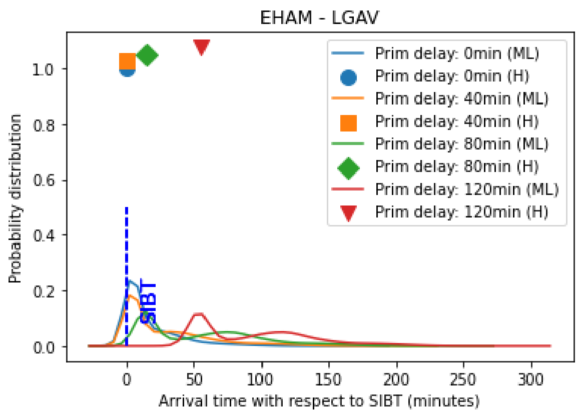
(a) Ground time distribution of the second leg



(b) The convolution of the primary arrival delay with the ground time distribution of the third leg produces the departure delay distribution of the third leg



(c) Gate-to-gate block time distribution

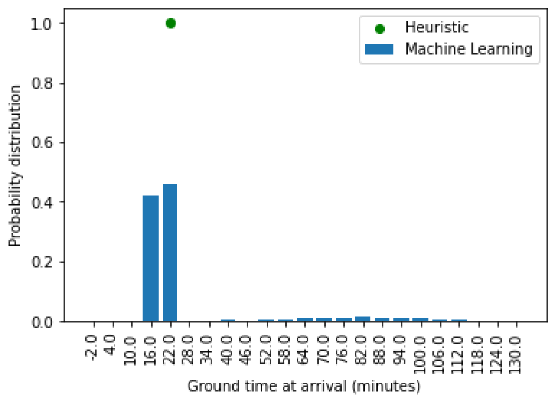


(d) The convolution of the distributions in (b) with the distribution in (c) produces the arrival time distribution of the third leg as function of primary arrival delay

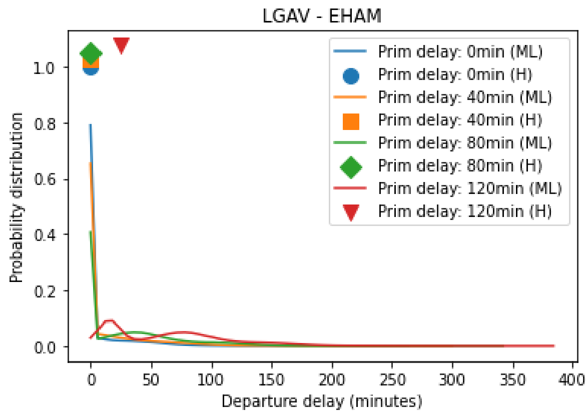
Figure 34 Estimation of propagation of delay for EHAM to LGAV flight (second rotation) (includes heuristic predictions)

EUROPEAN UNION

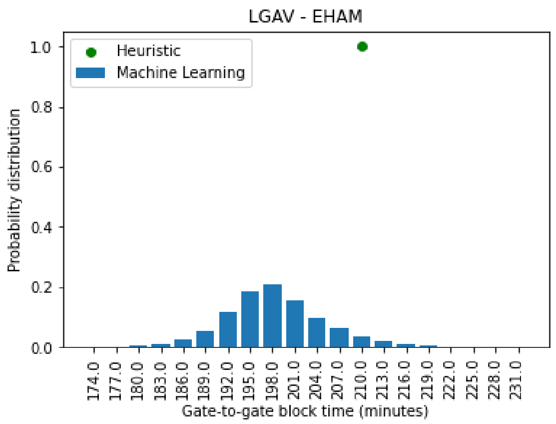
The same process is performed for the following rotation: computation of distribution of ground time for the rotation between EHAM and LGAV at EHAM (Figure 34 (a)), computation of expected departure time from EHAM by convolving the rotation time with the arrival times (Figure 34 (b)). Note how if the primary delay is 40 minutes or less, the probability of departing on time is high (>0.4) as delay will be absorbed by the padding on the schedules.



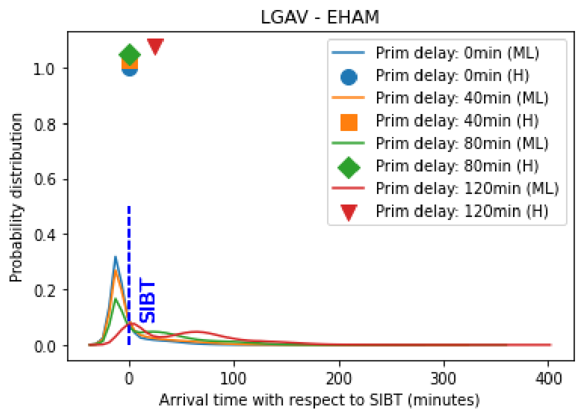
(a) Ground time distribution of the third leg



(b) The convolution of the primary arrival delay with the ground time distribution of the fourth leg produces the departure delay distribution of the fourth leg

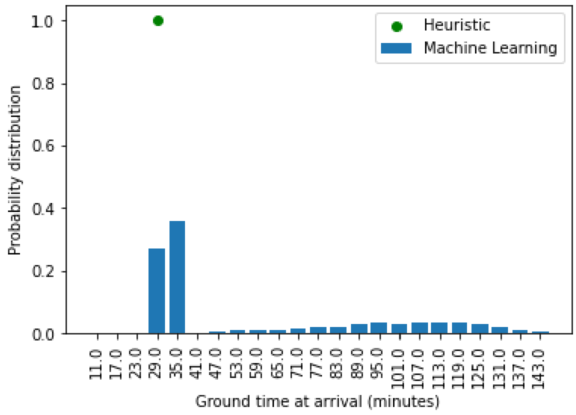


(c) Gate-to-gate block time distribution

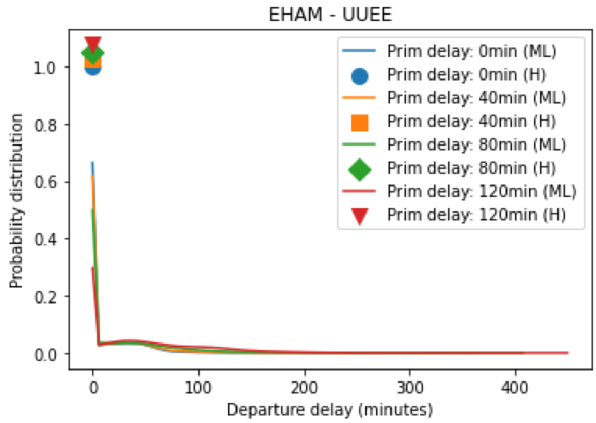


(d) The convolution of the distributions in (b) with the distribution in (c) produces the arrival time distribution of the fourth leg as function of primary arrival delay)

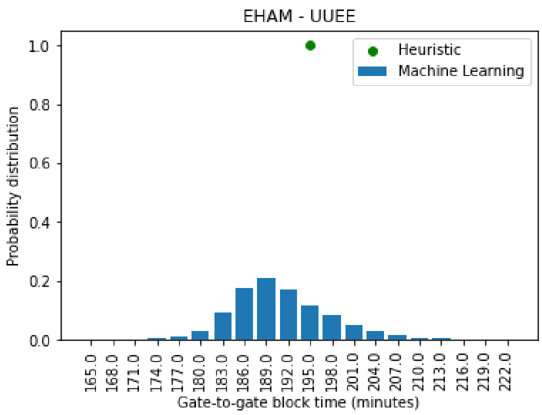
Figure 35 Estimation of propagation of delay for LGAV to EHAM flight (third rotation) (includes heuristic predictions)



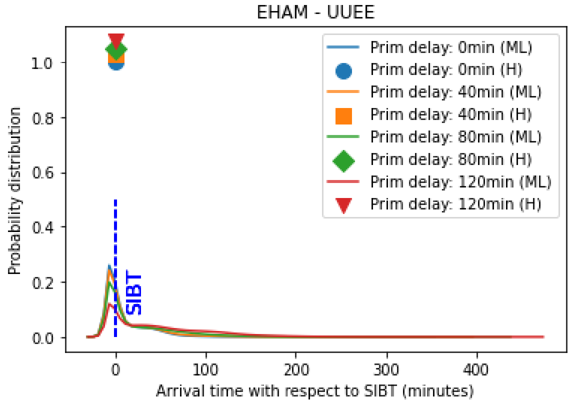
(a) Ground time distribution of the fourth leg



(b) The convolution of the primary arrival delay with the ground time distribution of the fifth leg produces the departure delay distribution of the fifth leg



(c) Gate-to-gate block time distribution



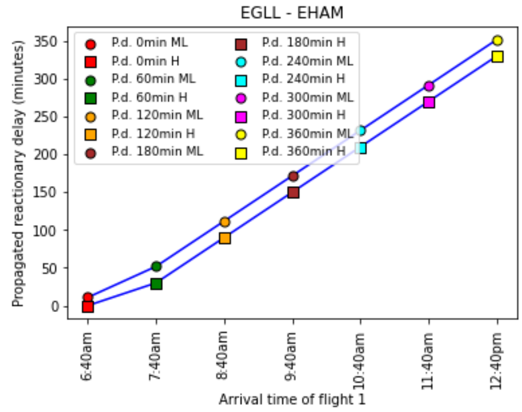
(d) The convolution of the distributions in (b) with the distribution in (c) produces the arrival time distribution of the fifth leg as function of primary arrival delay

Figure 36 Estimation of propagation of delay for EHAM to UUEE flight (forth rotation) (includes heuristic predictions)

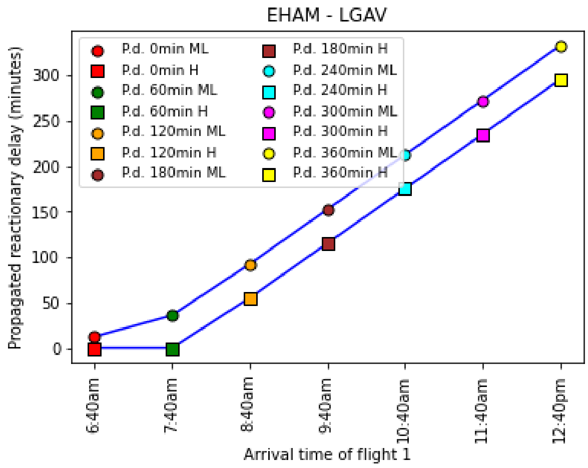
The same process if followed by the remaining flights in the sequence as presented in Figure 35 and Figure 36. Note how in some cases even if the primary delay is high the departure delay for the last rotation (EHAM to UUEE flight) is low. This is due to the padding in the schedules and the abortion of some reactionary delay. On the other hand, even if the first flight is on-time, some delay might be experienced by subsequent flights as some ATFM delay or longer rotations might also occur as captured by the models.

EUROPEAN UNION

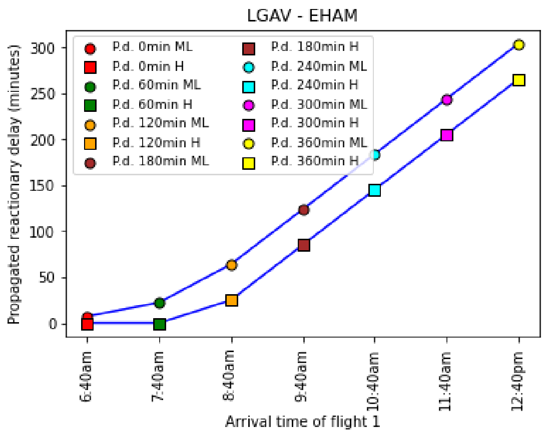
Comparison heuristic and machine learning reactionary delay estimation



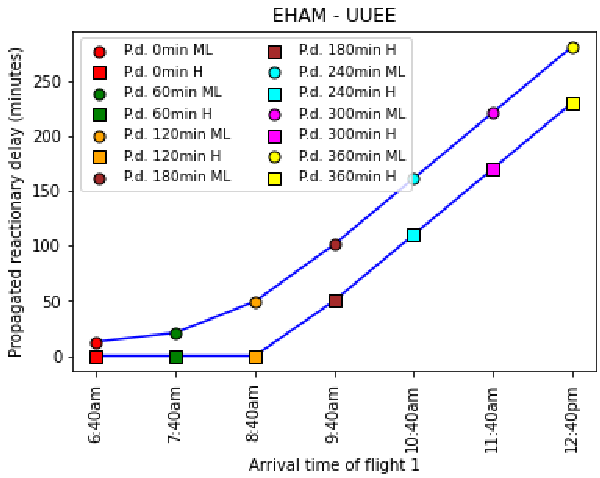
(a) Second leg



(b) Third leg



(c) Fourth leg



(d) Fifth leg

Figure 37 Propagated reactionary delay as a function of primary delay. A comparison between heuristic and ML.

As presented the reactionary delay is computed as the convolution of rotation and block time processes. These processes can be estimated with heuristics or with machine learning models. Pilot3 even allows the combination of these, e.g. rotation time estimated with heuristics and block time with machine learning models. As presented on the different components, machine learning models tend to produce larger ground times and shorter block times estimations. This will have an impact on the expected average delay (and cost) to be propagated. This is particularly relevant as the machine learning models can capture small probabilities of large delays (which will have a significant impact on the average expected cost of delay due to the non-linearity of cost of delay).

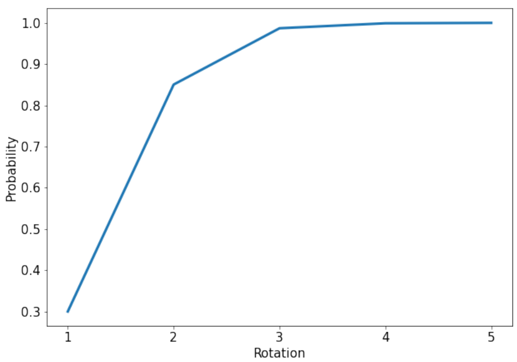
Figure 37 presents the comparison of the expected departing delay of the subsequent legs of the flight described in Table 30 as a function of the primary delay (i.e., as a function of the arrival time of the first flight) computing the reactionary delay with machine learning models and with heuristics. As observed the differences can be significant (e.g. 5th leg is expected to have an average departing delay of 50 minutes if the first flight has 120 minutes of delay according to the ML models, while that primary delay would be absorbed between the 4th and 5th leg according to the heuristic model).

Finally, it is worth noticing that not all the subsequent delay (and cost) experienced by the remaining rotations of the flight should be attributed to the primary delay of the first flight. For example, according to the machine learning models all legs (and noticeably in the second and fifth) will experience some departure delay (greater than zero) even if the first flight does not have any primary delay (initial arrival delay set at zero). This is due to the estimation of potential rotations and block times which are longer than the allocated times by the airline in the schedules (for example due to ATFM delay). However, an earlier than schedule arrival for the first flight might help to alleviate this subsequent delay and cost.

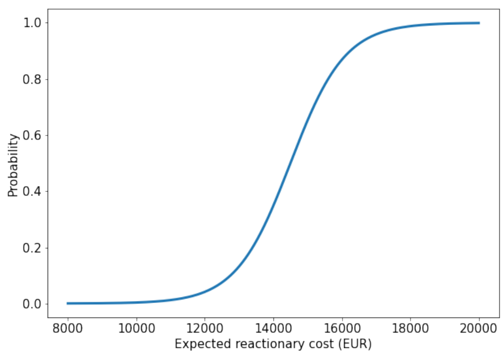
4.2.2.2.2.4 Reactionary delay strategic action

The modelling approach followed to estimate the reactionary delay (convolution of expected ground and block times) has the inconvenience that all processes are assumed to be independent and therefore delay propagated without any further consideration. For example, if the primary delay is 2 hours, these are just propagated on the subsequent flights leading to a unrealistically high total propagation of delay. The model lacks the consideration that some strategic actions (pre-tactically) might be performed by the AOCC to prevent this propagation such as cancelling or swapping flights downstream. These actions will generally be performed by the duty manager and their explicit modelling is difficult as it is difficult to obtain a dataset with historical information on this, and the actions might vary due to a high number of factors.

For this reason, in Pilot3 a process 'strategic action' has been modelled. This model captures the probability that an action, with a pre-defined cost, will be performed by the AOCC if the total expected reactionary cost is considered to be too high.



(a) possibility to perform strategic action as a function of rotation number



(b) Probability of performing strategic action to avoid reactionary delay propagation as a function of expected total cost

Figure 38 Modelling functions for strategic action

The strategic action model computes the probability of performing such action as combination of two factors:

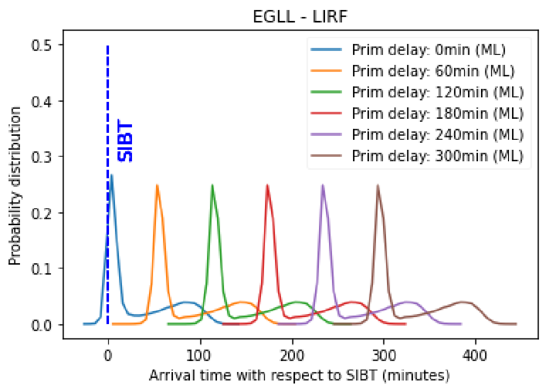
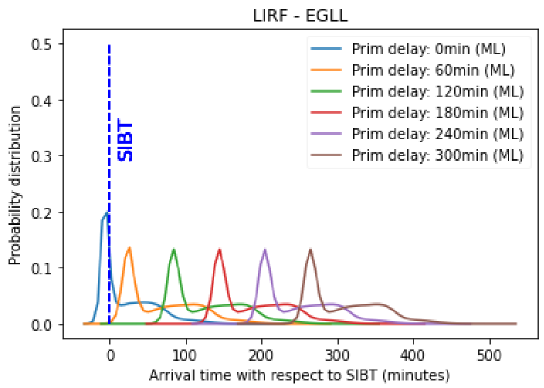
- first a model which captures the possibility of performing this action. It is assumed that the further in number of leg with respect to the current flight more time will be allowed for the AOCC to have the possibility to perform an action. The current model follows a sigmoid with a probability of 0.3 if the flight is the first rotation and a probability of 0.99 if it is the third rotation from the current flight (as shown in Figure 38 (a)). This model could be replaced by advanced trained ML models which can estimate what is the possibility for an airline to perform an action (such a swap of aircraft) as a function of operational parameters (e.g. if the flight is departing from the airline hub, time of the day).

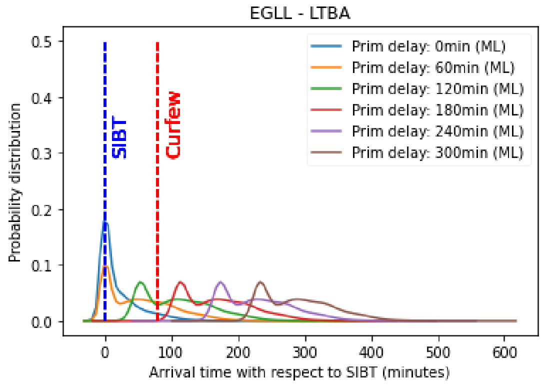
- the second model captures the willingness of the airline of doing this action. This probability is also modelled with a sigmoid. In this case, it is assumed a given cost for the strategic action (10.000 EUR) and the airline will start to consider to perform this action if the expected cost is higher than a given threshold (9.000 EUR) with a probability of 0.99 if the cost is higher than a given value (20.000 EUR). The use of the sigmoid captures the no-rational aspects associated with this decision-making process (as shown in Figure 38 (b)).

Note that both models could be replaced by advanced heuristics or models provided by the AOCC.

An example is provided to describe how these models are used in Pilot3 and the validation of the manually adjusted parameters of the possibility and probability models. For this Table 31 presents a flight with four legs. A curfew has been manually defined in the final arrival airport (LTBA) at 22h to present this impact on the reactionary delay cost model. In the same table, following the reactionary delay estimation presented in Section 4.2.2.2.2.3, the arrival reactionary delay of the subsequent legs is presented as a function of the primary delay of the first flight.

Table 31 Rotations and arrival delay as a function of primary delay for reactionary strategic action example

Leg	SOBT	SIBT	Arrival delay
LGAV-EGLL[1]	5h15	9h10	
EGLL-LIRF	9h50	12h20	
LIRF-EGLL	13h15	15h50	

Leg	SOBT	SIBT	Arrival delay
EGLL-LTBA[2]	16h50	20h40	

[1] Current inbound flight. Reactionary delay and cost computed with respect time of arrival of this flight to gate.
[2] Curfew set *manually* at LTBA at 22h00

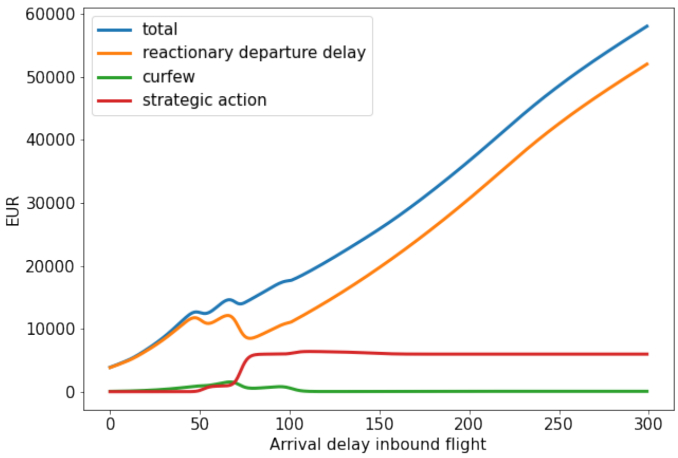


Figure 39 Total expected reactionary costs

Appendix D presents how the costs are computed in a step-by-step manner. Here only the final results are presented. Figure 39 shows the expected cost for the different components of reactionary delay (departing delay cost, strategic action cost and curfew cost).

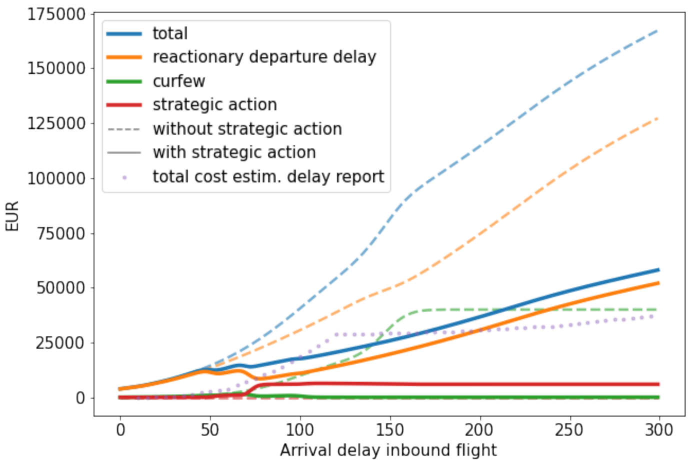


Figure 40 Comparison total expected reactionary costs with and without strategic action and with European cost of delay reference values

A comparison with the expected strategic costs if no strategic action is modelled is presented Figure 40. As shown, with the strategic action the expected total reactionary delay cost grows in a much lesser intensity. The total expected reactionary cost with a primary delay of 300 minute is higher than 170.000 EUR as this delay is propagated in subsequent flights eventually breaching the curfew and with very high departing delay costs in subsequent legs. With the modelling of the strategic action the costs are kept under 60.000 EUR. Finally, the European cost of delay reference provides an estimation of the cost of delay per minute of primary delay considering the full cost (including reactionary costs) and just with the primary cost (without reactionary costs) (Cook and Tanner, 2015). Doing the difference between these estimations, it is possible to get an indication of the order of magnitude of the reactionary delay costs of this cost reference model. Note that these values are valid in average and therefore do not consider the particularities of this specific rotation. In any case, it is possible to observe how the values of the estimated cost of reactionary delay by the cost of delay report are aligned and closer to the values obtained by Pilot3 model with the strategic action. Future work should be done to better calibrate the parameters of the strategic action model.

4.2.3 Operational ATM Estimator

4.2.3.1 Architecture

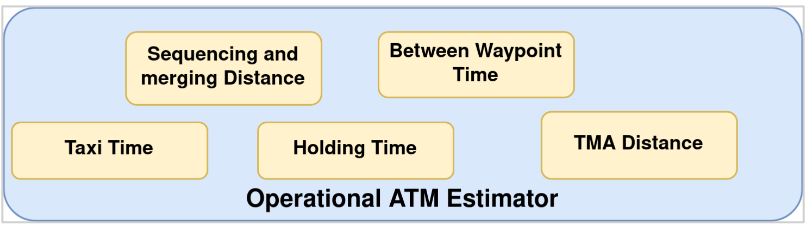


Figure 41 Operational ATM Estimator components

Figure 41 presents the different estimators that are provided by the Operational ATM Estimator module. As indicated in Section 1.1.2 the uncertainty modelled in Pilot3 focuses on the arrival and is used by the Objective function estimator to provide the estimation of the expected cost as a function of arrival time at FL100 in the descent. This is done by computing the expected costs of delay as a function of arrival time at gate (as provided by the Performance Indicator Estimator) and then integrating the different uncertainties between FL100 to the gate. Notably, holding, sequencing and merging and taxi-in times. For this reason it is particularly important that the Operational ATM Estimator does not provides only a value as prediction of the different components but a distribution which captures the uncertainty of the operations.

The flight might experience some discrepancies between planned and realised lateral trajectory. These variations would affect the optimisation (as the distance remaining in the flight might vary) and have not been considered in the prototype yet. However, a first estimation of the actual distance within the Terminal Manoeuvring Area (TMA) (specifically, from entering TMA to reaching FL100) has been computed as a first step toward integrating these type of uncertainties in Pilot3.

Finally, the Operational ATM Estimator provides an estimator to predict the time between two given waypoints in the flight plan. This is used by the Objective function estimator to estimate the expected cost function at any waypoint along the flight plan to be used as heuristic by the Trajectory optimiser with the full grid-search approach.

Table 32 summarises the estimators developed in the Operational ATM Estimator. Note that most of the predictions are route dependent as the datasets specific to the destination airport were used to train the particular machine learning models or/and to perform heuristics models.

Table 32: Types of predictors used for several ATM operational indicators

Estimator	Type	Routes estimator computed/trained	Data sources	Technical specifications
Distance sequencing and merging	Heuristic	• LEMD - EDDF • ENGM - LEBL • ELLX - EHAM • LGAV - EGLL • JKFK – EGLL	Radar data from ADS-B Opensky	Historical analysis of distance from FL100 to runway.
	Machine learning	xxx – EDDF	DDR data and METAR data	Two implementations: • Multi classifier • Two-model approach (regression and classification)
Probability of holdings	Heuristic	• LEMD - EDDF • ENGM - LEBL • ELLX - EHAM • LGAV – EGLL • JKFK – EGLL	Radar data from ADS-B Opensky	Historical analysis of holdings..
	Machine learning	xxx – EGLL	Radar data from ADS-B Opensky DDR and METAR data	Random Forest Classifier
TMA distance to FL100	Heuristic	xxx – EDDF	ADS-B Opensky	Historical analysis of distance flown from a point defined as the intersection of the flown trajectory with a circle centred at the airport with radius the start point of the STAR the TMA STAR (EMPAX) and reaching FL100.
Taxi time	Heuristic	All IATA airports	CODA files	Logarithmic-normal distribution.

EUROPEAN UNION

Estimator	Type	Routes estimator computed/trained	Data sources	Technical specifications
Between waypoint time	Heuristic	–	Flight plan	Directly estimating the time between points as indicated in the flight plan operated in Pilot3. Uncertainty is added to this value as configured in Pilot3.

The internal validation of these models is driven by the experts from the consortium who are directly involved in the development of the machine learning and heuristics models. As previously mentioned, the results of the OAE were presented during the internal meeting which was organised as an on-line meeting in July, 2021.

4.2.3.2 Components

4.2.3.2.1 Taxi time

The taxi time is the time an aircraft spends taxiing form the gate to the runway (taxi-out) and from the runway to the gate (taxi-in). For assessing this value we have used the CODA files. EUROCONTROL/CODA has published airport taxi times for many years, divided in seasonality with summer and winter times. These times are calculated from flight-by-flight data provided to CODA by airlines and airports. This includes airports for which CODA receives data on more than 100 flights, thus covering small and large airports, mainly in Europe, but also some non-European airports with direct flights to Europe. The values within CODA files times are calculated using the airline reported actual off-block time, actual take-off time, actual landing time and actual in-block time. The files provides a mean value for taxi in each airport, its standard deviation, the 10th percentile, median and 90th percentile. For the purpose of assessing uncertainty on taxi time a normal logarithmic distribution is calculated with the mean and standard values for each airport, resulting into a probabilistic distribution as shown in Figure 42. The use of these files as data source and the probability distribution obtained from them was validated by the Advisory board.

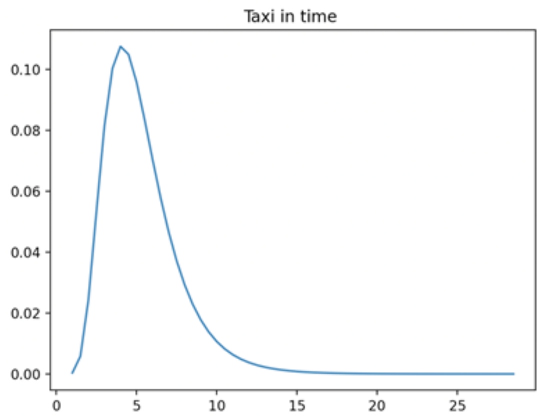


Figure 42 Taxi-in in Frankfurt Airport as estimated from CODA files

As observed in Figure 42, the distribution of taxi time is skewed to the left, indicating that most taxi in time values will be around 6 minutes, with a probability of 12%. This outcome will be the same for all flights landing at Frankfurt airport.

4.2.3.2.2 Sequencing and merging distance

4.2.3.2.2.1 Heuristic

The arrival at destination distance is defined as the distance left to fly by an aircraft when it hits FL100 until it reaches the runway. To compute this distance we used the great circle distance point by point along the trajectory from FL100 to RWY and add it. In order to validate this calculus we performed the same analysis with two different datasets: ADS-B data from Opensky and R&D EUROCONTROL data, with data from September 2018.

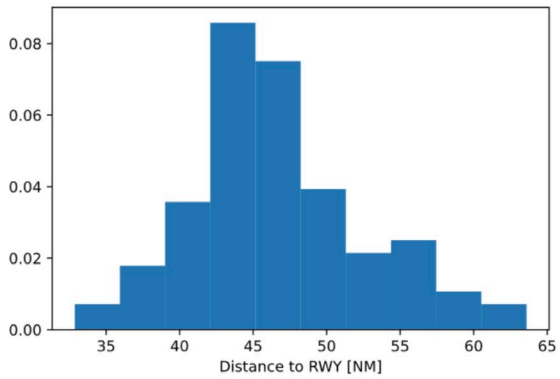


Figure 43 Histogram for sequence merging distance for LEMD-EDDF arrivals

Having estimated the sequence and merging distance for all aircraft landing at a given airport during the selected period (September 2018) a histogram is produced as outcome of the estimator as shown in Figure 43.

4.2.3.2.2.2 Machine learning – regressor

To estimate the sequencing and merging distance with machine learning models different datasets are used to build the input features: EUROCONTROL R&D Archive and METAR data. Two feature analysis techniques are used to identify the importance of the different features modelled: PCA (Principal Component Analysis) and feature regression which rank features in the same order if all the features are positively correlated with the target. Table 33presents the features used for the models.

Table 33: Features used for the sequencing and merging models

Component	Feature	Detailed information
Trajectory	Altitude at FIR entry (geo and baro)	Implicit route and date (time of the day) information.
	Latitude at FIR entry	The coordinates at the FIR entry are estimated using the flight plan
	Longitude at FIR entry	
	Day of the week	Time of the day and Hour of the day are obtained from the scheduled arrival time
	Hour of the day	
Performance	Velocity at FIR entry	
	Vertical rate at FIR entry	
Weather at arrival airport	Precipitation	From METAR
	Wind speed	
	Air temperature	

Component	Feature	Detailed information
	Dew Point temperature	
	Visibility	
	Altimeter hpa	
Network status	Occupancy at arrival time	Network disruptions. The occupancy is defined as all flights landing in Heathrow airport during a 40 minutes window centred either on the scheduled arrival time or the scheduled departure time
	Occupancy at departure time	

PCA shows how with 10 features approximately 90% of the dataset is explained (Figure 44). Features regression (Figure 45) shows that the highest correlated feature with the sequence and merging distance is the latitude at which the aircraft is planned to enter the FIR, followed by the hour of the day at which it is expected to land and the occupancy of the airport at that time.

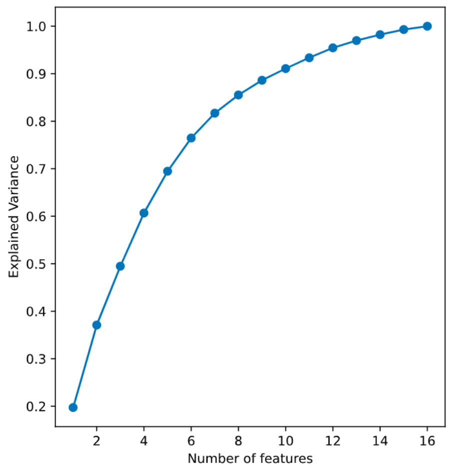


Figure 44 PCA results for sequence-merging distance dataset

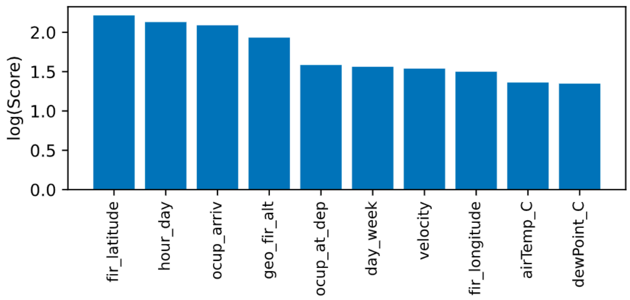


Figure 45 Feature regression results for sequence-merging distance dataset

Two different approaches are developed to predict not only the expected sequencing and merging distance but the uncertainty associated with these estimations:

1. a classifier directly on the target variable (distance of sequencing and merging). A clustering (k-means) algorithm is used to define the different classes, i.e., possible sequencing and merging distances. This produced an accuracy of 0.5.
2. the two-model approach of a regression and a classification of the error of the regression as presented in Section 4.2.1.3 and used by the Performance Indicator Estimator models (see Figure 17). In this case a random forest regressor is used to estimate the sequencing and merging distance followed by a classifier.

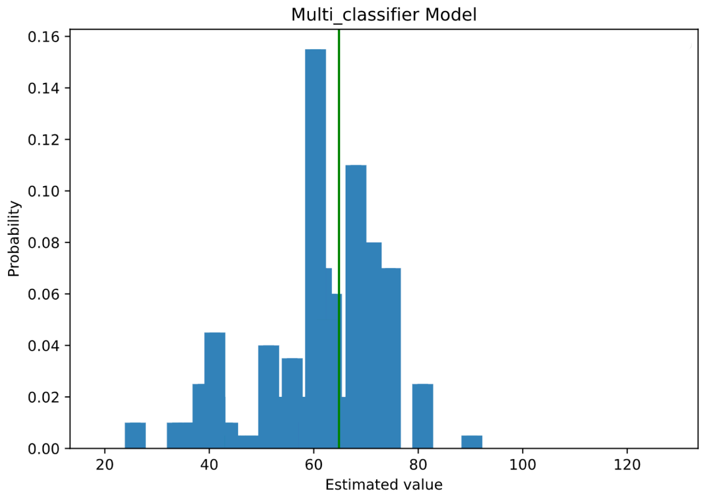


Figure 46 Prediction with classifier on the target variable (sequencing and merging distance)

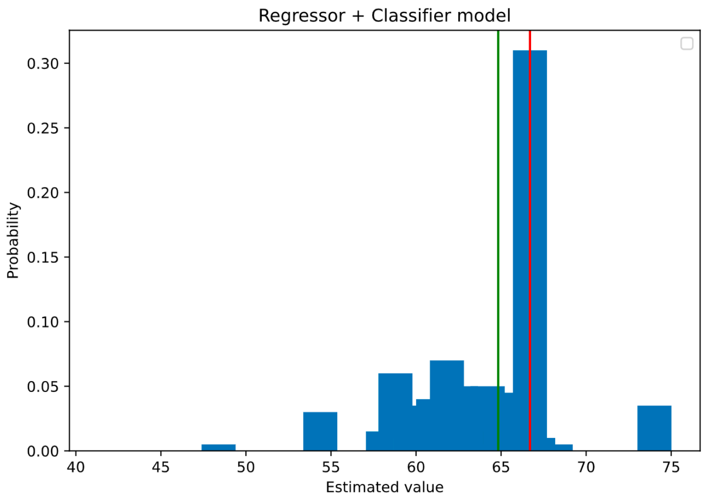


Figure 47 Prediction with two-model approach for sequencing and merging distance

4.2.3.2.2.3 Comparison heuristic and machine learning

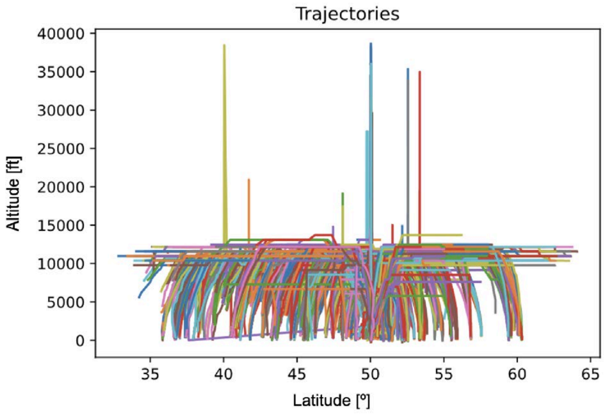
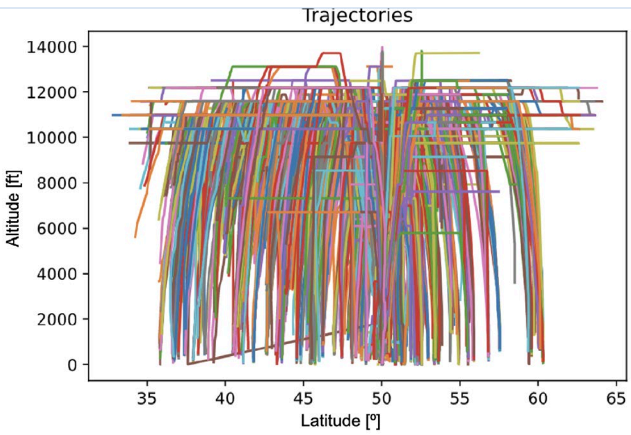
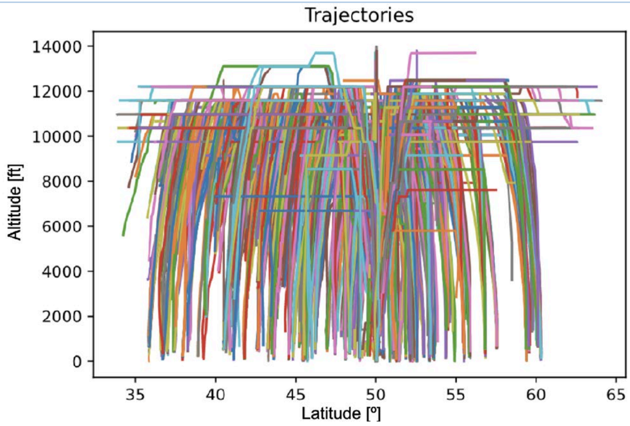
Note that in this case, the heuristic approach provides the same distribution independent of the flight. The machine learning models adjust their prediction as a function of some flight characteristics.

4.2.3.2.3 Holding time

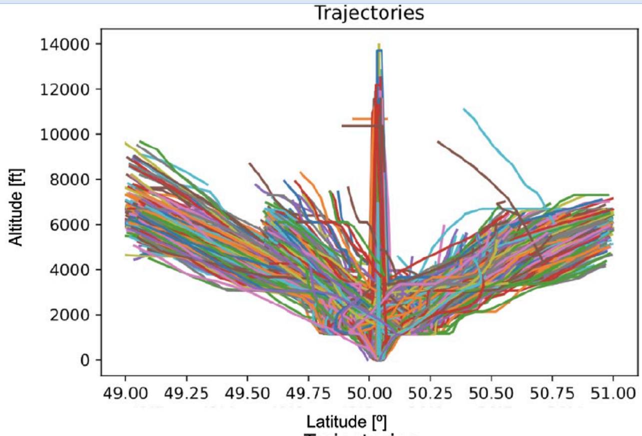
4.2.3.2.3.1 Heuristic

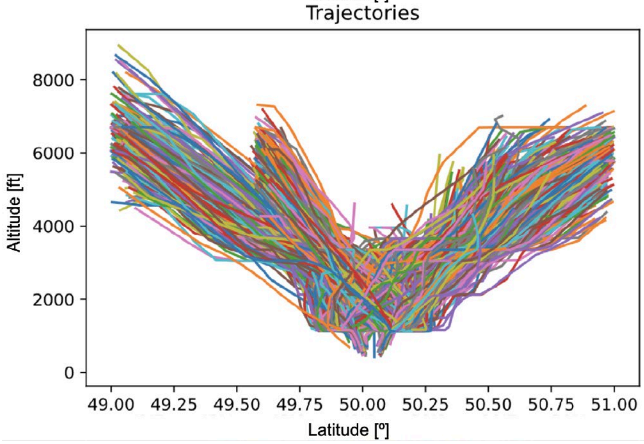
For this case-study ADS-B data from Opensky data is used. ADS-B provides higher resolution than other datasets such as EUROCONTROL R&D Archive. However, ADS-B from Opensky is usually noisy and it requires a cleaning pre-process in order to validate trajectories and work with them. Table 34 presents the different cleaning processes that were applied to ADS-B data from Opensky are described and illustrated with an example of trajectories.

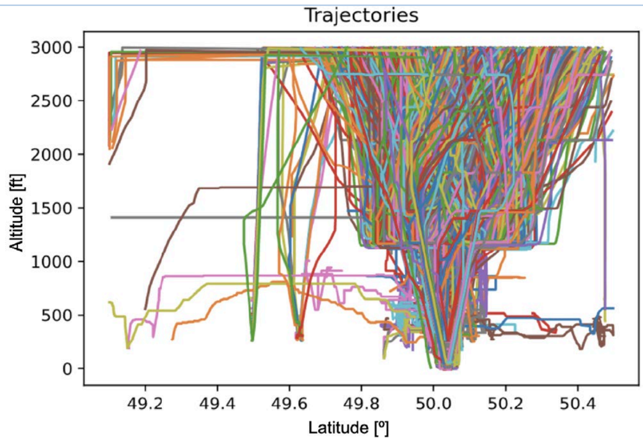
Table 34: Cleaning processes applied to ADS-B data for holding estimation

Action	Trajectories result
Raw data	
Removing too high and too low altitudes. In this stage we remove non-realistic altitudes NaN removal	
Removal of flights presenting peaky outliers There are flights which trajectory is not smooth which means radar detection issues.	

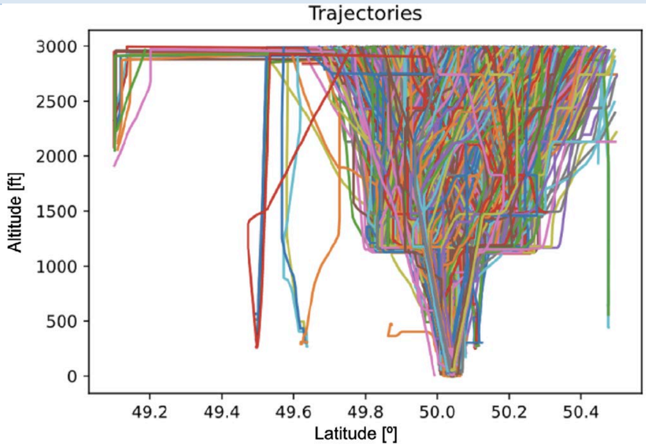
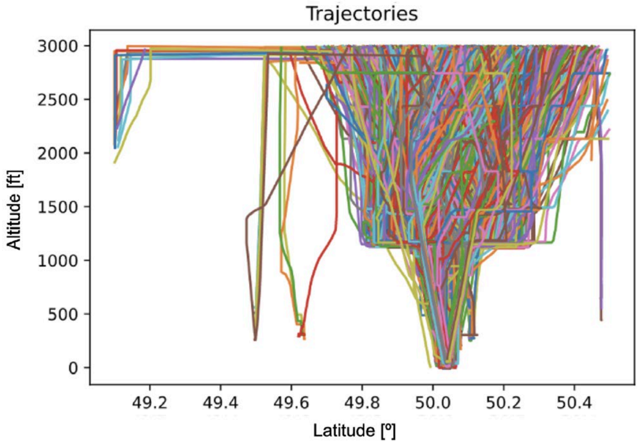
<table>
<tr><th>Action</th><th>Trajectories result</th></tr>
<tr><td>Removal of corrupted data at arrival.

For some flights, once the aircraft touches ground the recorder keeps adding points to the trajectory that are corrupted. These need to be removed.</td><td>


</td></tr>
<tr><td>Removal of flights flying at low altitude.

For the case studies within Pilot3 we do not require to analyse this type of planes.</td><td>
</td></tr>
</table>

EUROPEAN UNION

Action	Trajectories result
	
Removal of very short haul flights These flights are usually test flights or flights with missing data. In both cases they are not object of study in pilot3	
Resampling (10 seconds)	Finally, to obtain smoother trajectories and remove noise a re-sampling with a sampling period of 10 seconds was performed on all trajectories. This procedure will also reduce the size of data and thus the computational cost associated to dealing with these data

The holdings are estimated for EGLL using data from September 2018. Figure 48 presents all the flights considered.

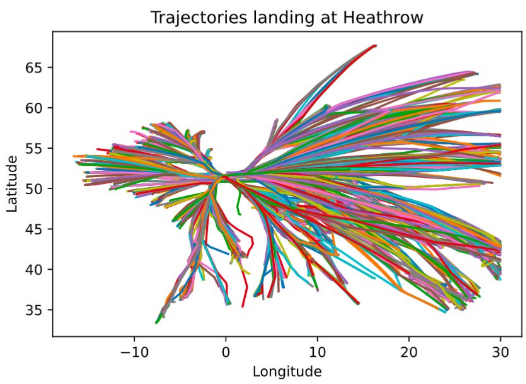


Figure 48 All trajectories landing at Heathrow during September 2018

The goal was to label the holdings, that means being able to obtain from a trajectory if a holding has occurred. For that purpose the LineString method from Python was used. This method is able to determine whether a 2D array is complex or simple. A simple array is that in which no crossing occurs, on the other hand a complex array is a line that crosses itself. In aircraft trajectories, if a trajectory crosses itself on the plane, a holding has happened. Figure 49 presents several trajectories where holdings are present.

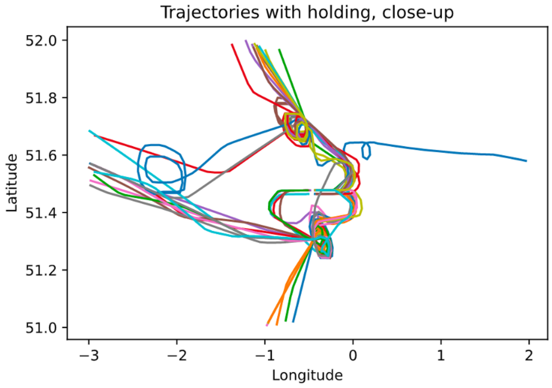


Figure 49 Close-up of all flights presenting holdings during the first week of September 2018

The probability of a holding occurring at a given airport is then estimated using data of one month. For instance, in September 2018, all flights flying the route LGAV-EGLL were analysed, turning out in 55% of them having a holding. Therefore, in Pilot3 every flight having this route will be associated with a probability of holding equal to 55%.

4.2.3.2.3.2 Machine learning

Table 35 presents the different features that are considered by the machine learning models.

Table 35: Features used for the sequencing and merging models

Component	Feature	Detailed information
Trajectory	Altitude at FIR entry (geo and baro)	Time of the day and Hour of the day are obtained from the scheduled arrival time Entry FIR coordinates are extracted from R&D Archive
	Latitude at FIR entry	
	Longitude at FIR entry	
	Day of the week	
	Hour of the day	
Performance	Velocity at FIR entry	
	Vertical rate at FIR entry	
Weather at arrival airport	Precipitation	Holdings may be weather sensitive as bad weather could imply network disruptions A 'complex' feature is proposed: Dew point temperature minus temperature at the airport. As the difference of these values is closely related to fog creation and thus, visibility.
	Wind speed	
	Air temperature	
	Dew Point temperature	
	Visibility	
	Altimeter hpa	
Network status	Occupancy at arrival time	Network disruptions
	Occupancy at departure time	

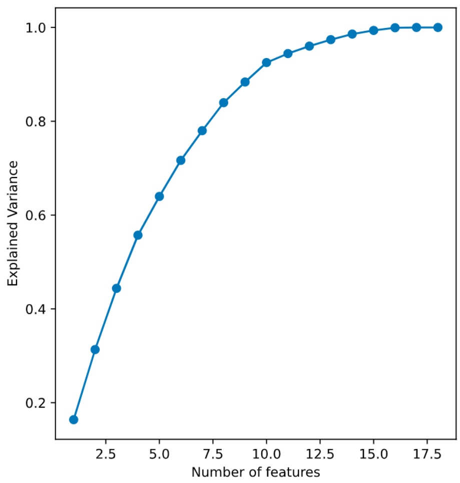


Figure 50 PCA Analysis for the holdings dataset

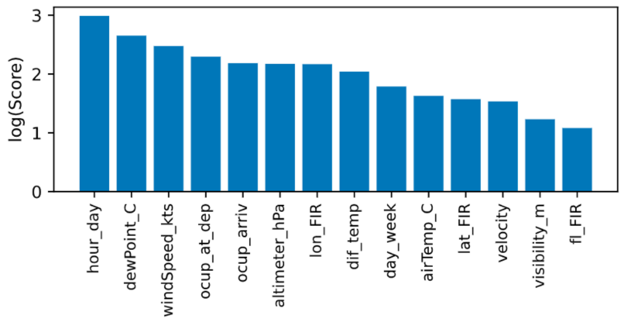
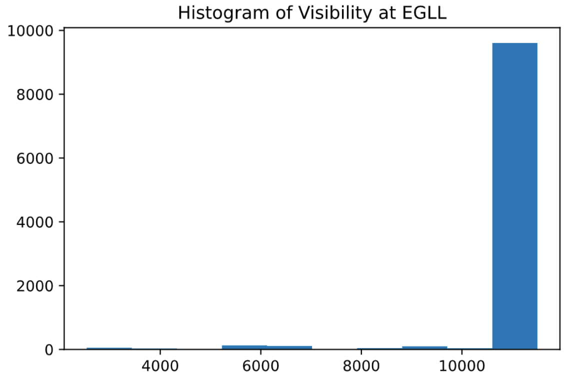
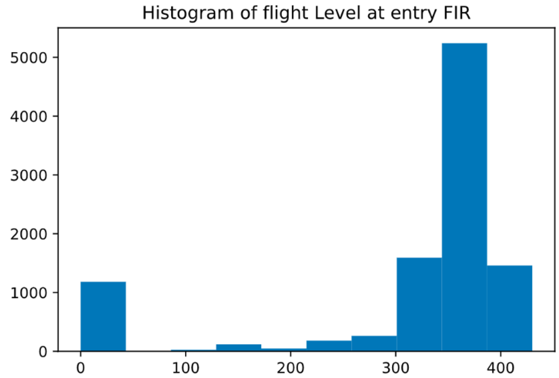


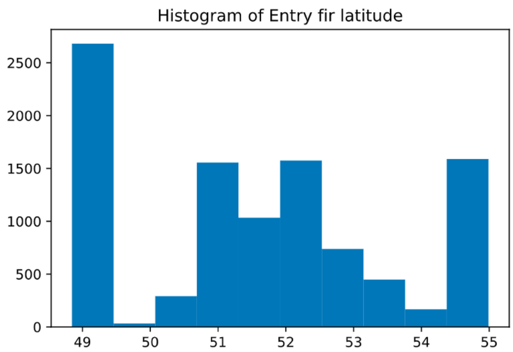
Figure 51 Feature regression analysis for the holdings dataset



(a) Visibility



(b) Flight level entry



(c) Entry FIR latitude

Figure 52 Significant features for holdings analysis

The PCA analysis indicates that 9 features are enough to retain 0.96 of variance, hence letting us know that the selected input features are enough to explain holdings (see Figure 50). The feature analysis indicates that apparently, dew point and wind speed at the arrival airport are the most relevant features, followed by the network features. Surprisingly visibility ap-pears as a features having medium impact on holdings, however, it is directly linked with the dew point (see Figure 51). The dew point is the point at which the water in the air condensates, forming fog, the closer this temperature is to the ambient temperature, the more likely fog will form, thus leading to a loss of visibility. If we take a deeper look into the features we can see how visibility (see Figure 52 (a)) is almost always the same value, which will explain its low im-pact on the feature analysis, as it adds no information. This low variance on the feature may be explained by poor data quality from the data source used, METAR. On the other hand the air temperature is already considered in the *difference temperature* feature, and taking a look into the probability histograms of flight level entry and latitude entry (Figure 52 (b) and Figure 52 (c)) we can see how the values are very similar for the flight level and the variation range very low for the fir latitude, so these features do not add a lot information to our model, they can even difficult the model comprehension. For that reason, 9 features from the total proposed features have been selected for the model uptake, their impact on holdings is displayed in Figure 53.

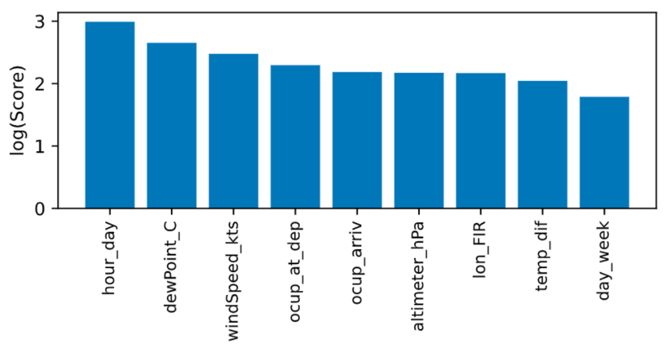


Figure 53 Most important features for holdings

Several machine learning algorithms were tested: logistic regression, classification tree, random forest classifier, k-nearest neighbours and AdaBoost. These were evaluated as a function of their accuracy (ratio of correctly predicted observation to the total observations), precision (ratio of correctly predicted positive observations to the total predicted positive observations) and recall (ratio of sum of true positive and true negatives out of all the predictions made).

Table 36 presents the preliminary results on the performance of the different methods considered.

Table 36: Algorithms tested for Machine Learning models on holdings and their performance

Algorithm	Accuracy	Precision	Recall
Logistic Regression	0.69	0.78	0.72
Classification Trees	0.79	0.85	0.81
Random Forest classifier	0.86	0.89	0.88
K-nearest Neighbour	0.80	0.84	0.83
AdaBoost	0.79	0.85	0.80

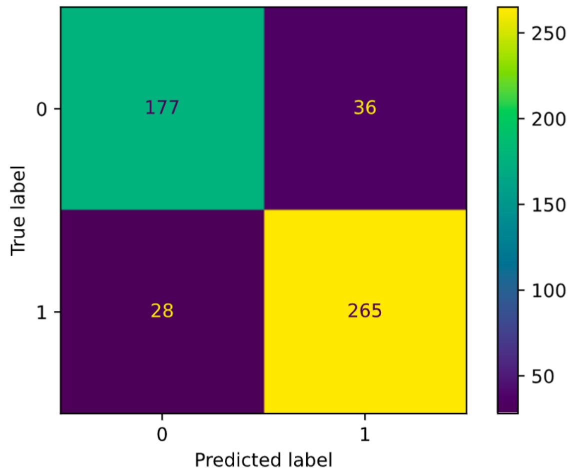


Figure 54 Confusion matrix for random forest algorithm to detect holdings at EGLL

The hyper-parameters of the random forest classifier are further tuned as this is the method which provides best performance obtaining a final accuracy of 0.87, precision of 0.9 and re-call of 0.88. Figure 54 presents the confusion matrix for these predictions:

- True Positive(TP): Flights which have a holding and are predicted to have a holding (bot-tom right square).
- True Negative(TN): Flights which do not have a holding and are predicted to not have a holding (top left square).
- False Positive(FP): Flights which do not have a holding and are predicted to not have a holding (top right square)
- False Negative(FN): Flights which have a holding and are predicted to not have a holding (bottom left square).

4.2.3.2.4 Comparison heuristic and machine learning

In the case of holdings for a given flight the current heuristic version returns the probability of having a holding in that airport based on historical data without any further consideration. So this probability is the same for every flight. The machine learning model returns the probability of having a holding based on the accuracy of the model, which is 87%.

4.2.3.2.5 TMA distance estimation

In order to optimise the last segments of the route a distance estimator that computes the distance from the TMA entry point to FL100, which the point at which Pilot3 will optimise.

To do so we select the EMPAX point, which indicates the entrance to the TMA for the flights entering the space using the southern route. Considering the Great Circle Distance between EMPAX (N48 27 43, E 008 58 53) and EDDF airport (N50 1.96 E 008 32 08), which is 97NM a circumference is drawn with that radius with the airport as centre. We define that the moment a flight crosses that imaginary line it has entered the TMA space. Thus, following the approach previously defined, the distance from the TMA entry point until the aircraft reaches FL100 is computed. This rationale is depicted in Figure 55. As it can be appreciated in the picture, the segment of the trajectories within the circumference is the one presenting larger dispersion, and thus the one with greater uncertainty and the most relevant to predict.

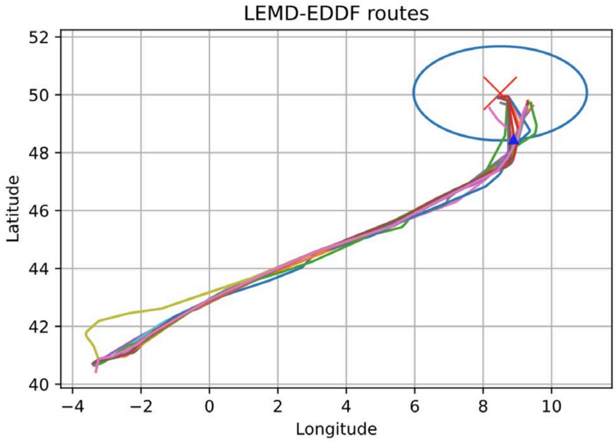


Figure 55 TMA distance rationale, LEMD - EDDF southern trajectories, EMPAX point (blue triangle), EDDF airport (red cross), and circumference delimiting the entry at the TMA

In the case of the TMA distance the results are also given in a shape of a probability distribution histogram, as depicted in Figure 56. This distribution will represent all incoming flights at Frankfurt airport.

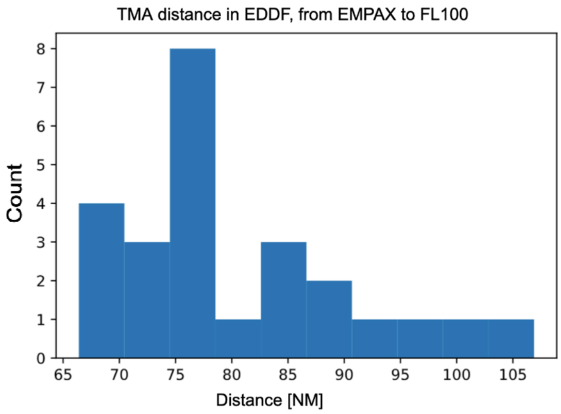


Figure 56 Histogram depicting the TMA distance for LEMD-EDDF southern routes

4.2.4 Cost function and alternatives

As shown in the previous sections, the different configuration of the Performance Indicator Estimator and the Operational ATM Estimator will have a direct impact on the computed expected cost of delay function used in Pilot3. However, from an optimisation perspective what is relevant is not necessarily the value of the amount of expected cost but how this cost varies as a function of arrival time in order to consider if a trade-off between fuel (and extra cost of fuel) and cost of delay is worth it. For this reason the derivative of the cost over time will be provided.

In this section different configurations of the Performance Indicators Estimator and the Operational ATM Estimator are used in three flights from the scenarios defined in Section 2:

- SCN 100 — LGAV – EGLL (see Section 2.1.1)
- SCN 600 — KJFK – EDDF (see Section 2.1.3)
- SCN 800 — KJFK – EGLL (see Section 2.1.4)

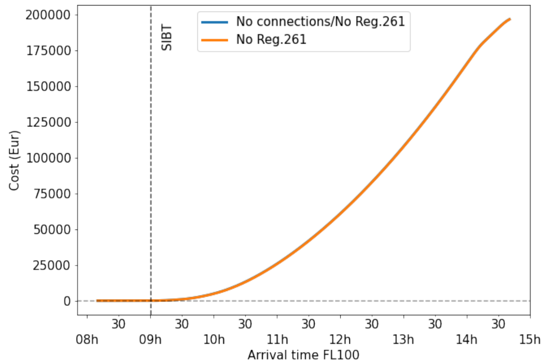
EUROPEAN UNION

The shape and order of magnitude of the total expected costs for these flights when using the default configuration of Pilot3 have been validated with the Advisory Board. In this section, different configurations of Pilot3 will be used to analyse how this impact the shape and possible optimisation. All expected cost of delay functions are computed by the PIE as a function of arrival time at the gate and then integrated by the Objective Function Generator up to FL100 using the outcome of the OAE as presented in Section 1.1.2. Therefore, the results of the expected cost at FL100 are shown in this section, as these will be the values used by the optimiser (see Section 1.1.3).

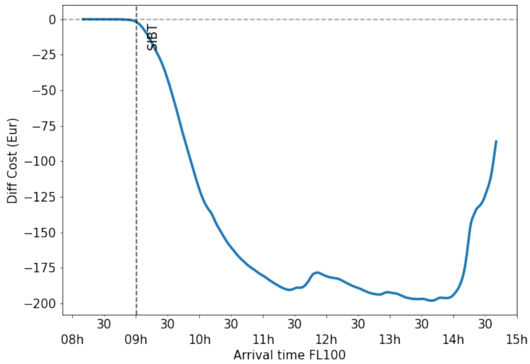
For each comparison four different representations are shown:

- Expected cost when reaching FL100.
- Difference on expected cost when reaching FL100 between two configurations being analysed.
- Derivative of cost of delay over time for each alternative. Note that this is an indication of the variation of cost of delay (EUR) over time (min). As Cost Index is defined as kg fuel / minutes, there is a relationship between the derivative of cost over time and the cost of fuel, being able to estimate an operational cost index as the derivative of cost of delay over time divided by cost of fuel. This is something that should be explored further.
- Difference between derivative of cost over time for each alternative configuration.

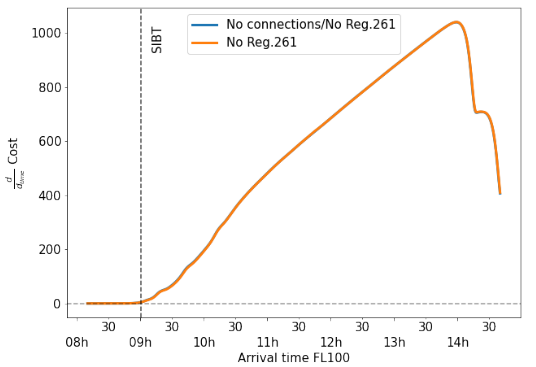
4.2.4.1 SCN100 – LGAV – EGLL cost function alternatives analysis



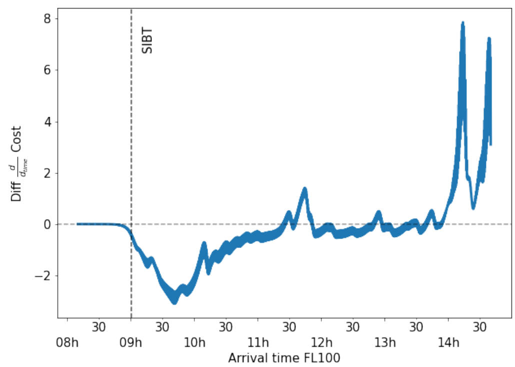
(a) Expected cost



(b) Difference of cost

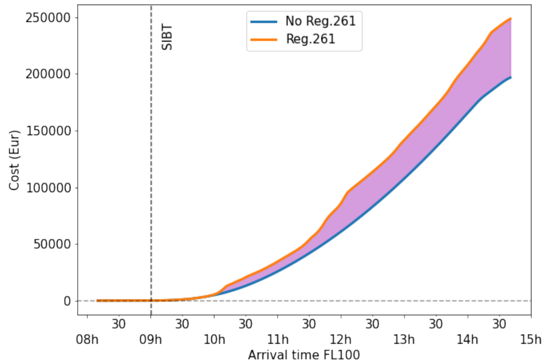


(c)Derivative of cost over arrival time

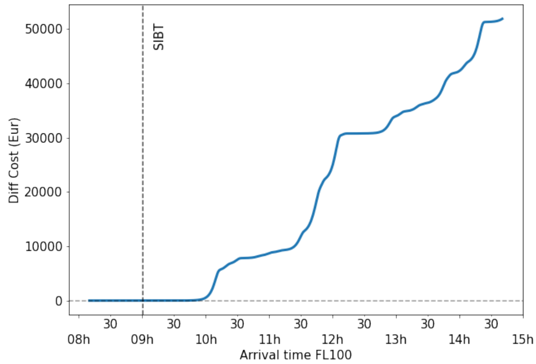


(d) Difference on derivative cost

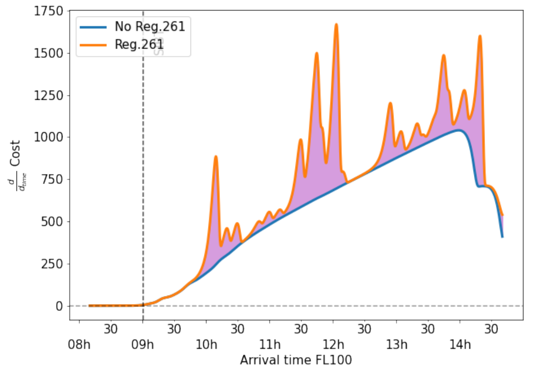
Figure 57 No entitlement to Regulation 261 consideration of connecting passengers



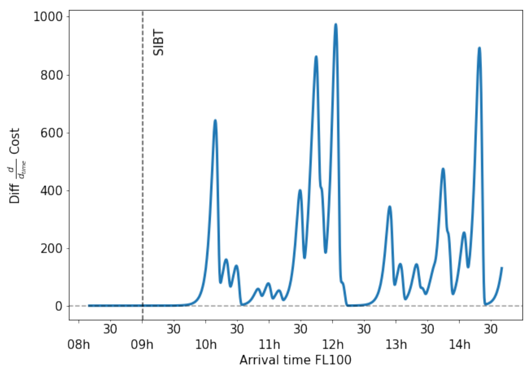
(a) Expected cost



(b) Difference of cost



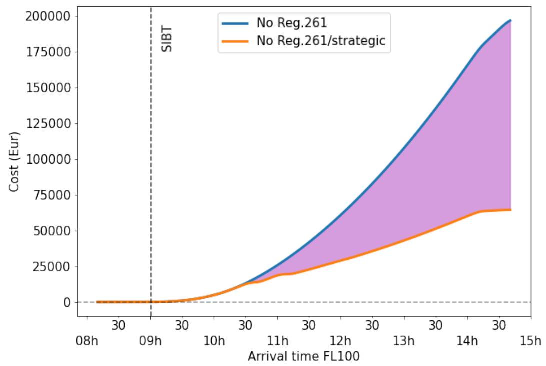
(c)Derivative of cost over arrival time



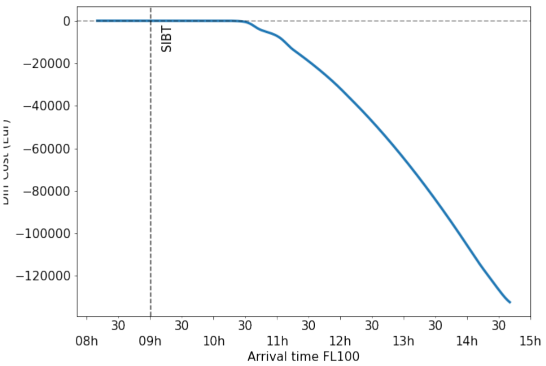
(d) Difference on derivative cost

Figure 58 Regulation 261 vs non Regulation 261 cost with connecting passengers

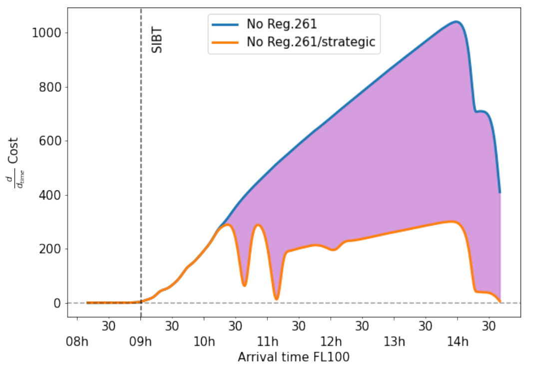
The first analysis performed is assess the impact of considering connecting passengers in the cost function. As shown in Figure 57, if passengers are not entitled to compensation due to Regulation 261 the expected costs considering or not connecting passengers are very similar. Not including the connecting passengers has overall a lower cost as when passenger miss a connection some extra delay might be experienced which, even if not entitled to compensation (Regulation 261), will have an impact on the total IROP costs for example on the soft cost component. However, when analysing how the cost varies over time and the comparison between both alternatives, (Figure 57 (c) and Figure 57 (d)), it is clear that the impact is low. When passengers are entitled to Regulation 261, connecting passengers have a larger impact on the total experienced cost as shown in Figure 58. It is interesting to observe how when passengers miss their connections some non-linear increment on the cost function are produced (see Figure 58 (b)). These abrupt changes lead to variations of cost as a function of arrival time with sharp increments and descends. However, as shown in Figure 58 (d), for this particular flight, the variation of cost over time is always the same or higher that if passengers are not entitled to Regulation 261. So, passenger entitlement to Regulation 261 pro-duces an optimisation cost function with discontinuities and overall higher benefits from recovering delay for all possible arrival times analysed.



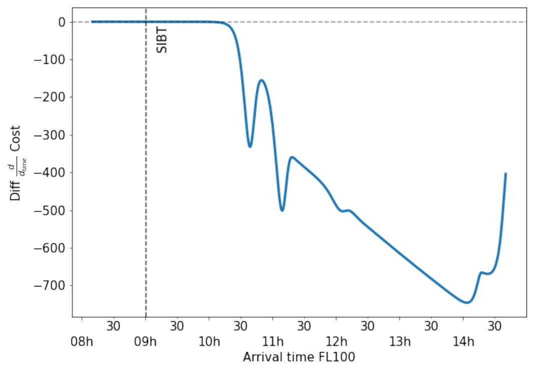
(a) Expected cost



(b) Difference of cost

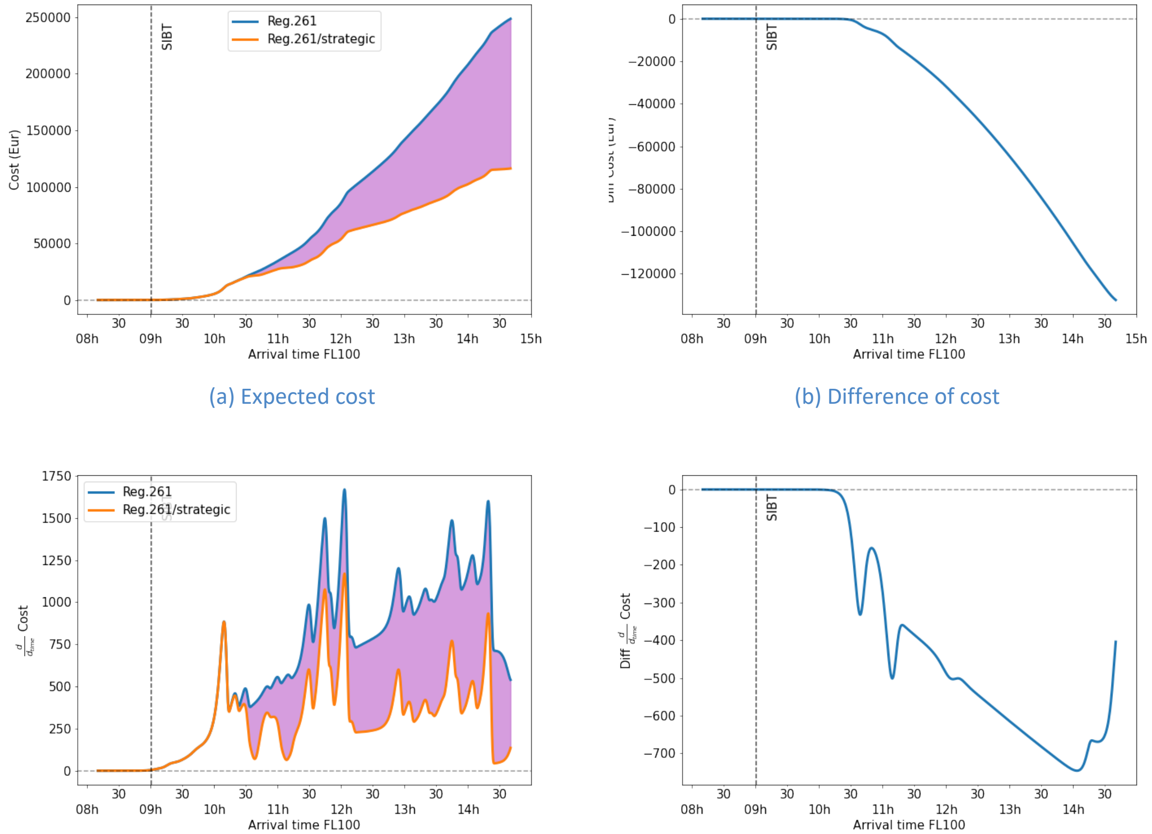


(c)Derivative of cost over arrival time



(d) Difference on derivative cost

Figure 59 Strategic action for reactionary delay consideration no compensation (Reg. 261) entitlement



(a) Expected cost

(b) Difference of cost

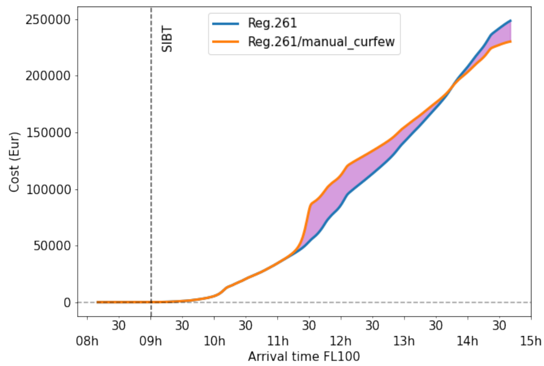
(c)Derivative of cost over arrival time

(d) Difference on derivative cost

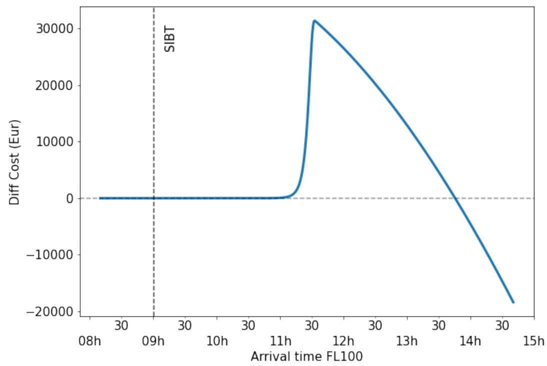
Figure 60 Strategic action for reactionary delay consideration with compensation (Reg. 261) entitlement

Figure 59 and Figure 60 present the impact of including the strategic modelling action for reactionary delay cost with passengers are considered to be entitled to compensation or not. As observed, the impact of the strategic action is to reduce the overall cost, but not only that, the variation of cost over time also reduces. This means that when these pre-tactical actions are modelled the optimised solutions might in general produce lower recovery actions (or of lower intensity) that if this action is not modelled. This highlight the importance of modelling the reactionary expected costs in a realistic manner, not only for the absolute amount of expected cost modelled but for its impact on the variation of this cost over arrival time.

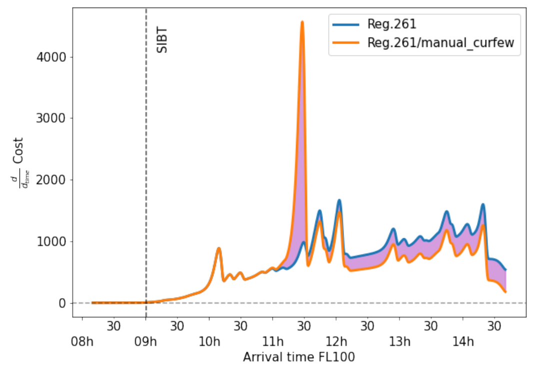
EUROPEAN UNION



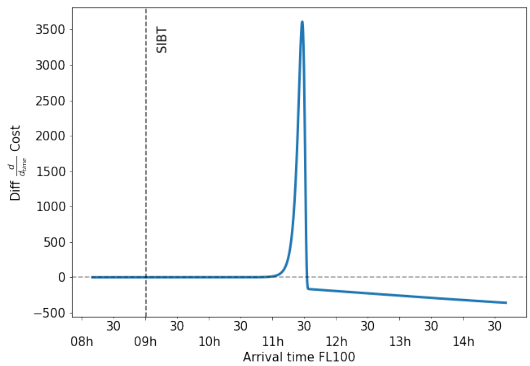
(a) Expected cost



(b) Difference of cost

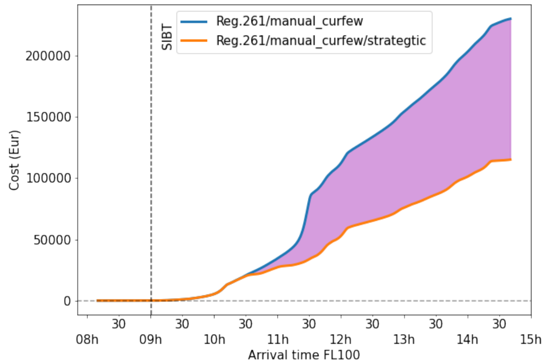


(c)Derivative of cost over arrival time

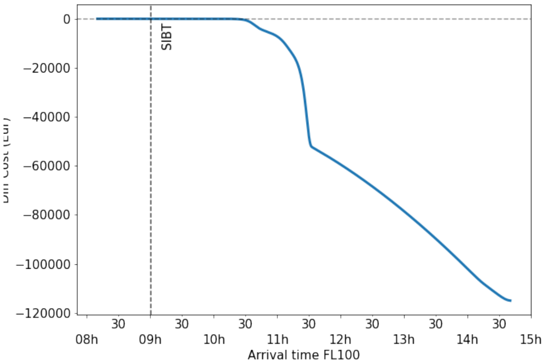


(d) Difference on derivative cost

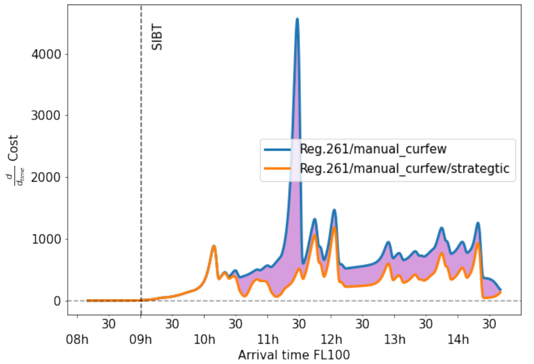
Figure 61 Impact of manually defining a curfew



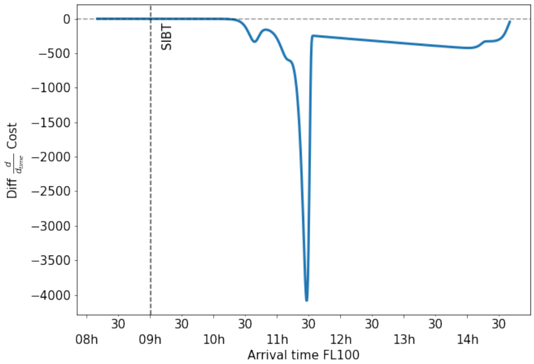
(a) Expected cost



(b) Difference of cost



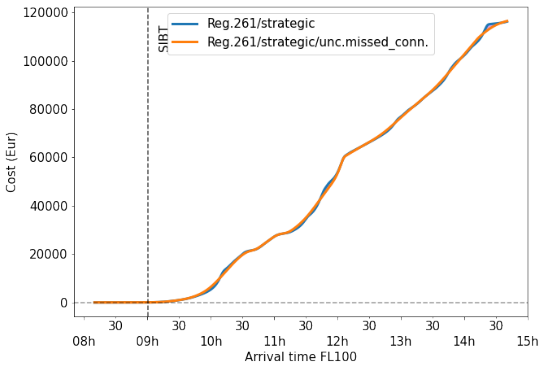
(c)Derivative of cost over arrival time



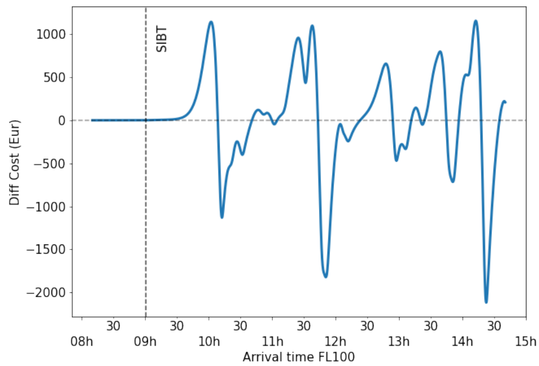
(d) Difference on derivative cost

Figure 62 Impact of manually defining a curfew and considering a potential strategic action

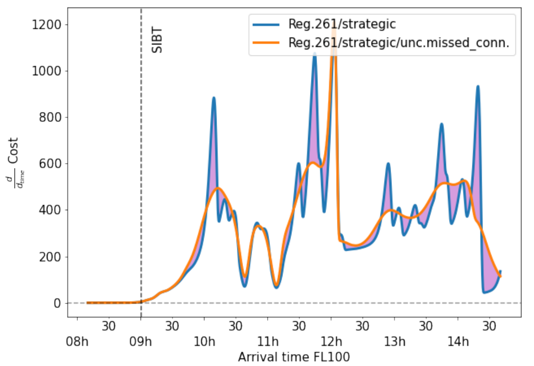
The introduction of a curfew at the end of the day of operation and its modelling translate into a sharp increment on the cost function (see Figure 61). It is worth noticing how after this discontinuity, even if the total expected cost is higher (Figure 61 (a)), the variation of cost over time is actually lower that if curfew is not modelled (Figure 61 (d)). This means that if the trajectory is arriving at the airport right when the cost is potentially materialised, the variation of cost over time is very large and this will push the optimiser toward solutions which try to recover as much delay as possible to limit the expected cost of delay. However, once that cost has been materialised, or cannot be avoided by reducing the delay enough, the incentive to recover delay is lower that if the curfew is not modelled (Figure 61 (d)). Note however that the expected cost over time is still high (due to other type of costs of delay) and the optimiser might still try to recover some delay (see Figure 61 (c)). Finally, if a strategic action is modelled (Figure 62), it can be observed how the high increment due to breaching the curfew is smooth out with the strategic action. Therefore as shown in Figure 62 (c) and Figure 62 (d) the expected variation of cost over time remains similar but without the large step experienced around 11h30. Once again, the use of this strategic action might lead to more stable solutions.



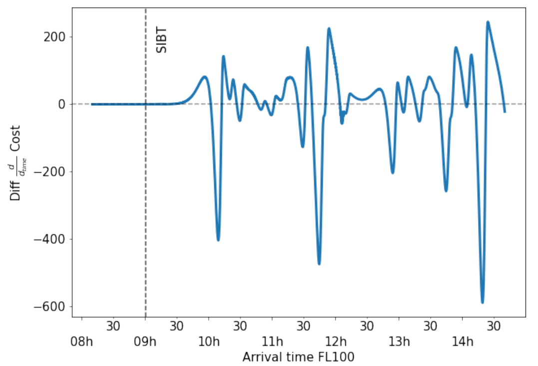
(a) Expected cost



(b) Difference of cost



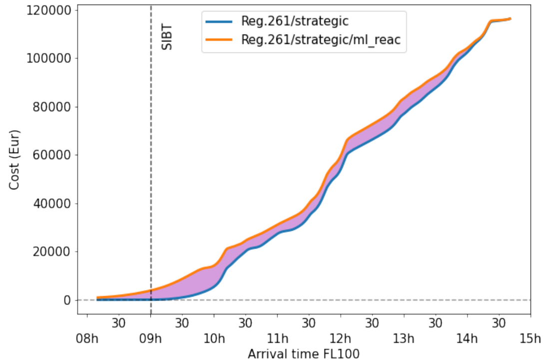
(c)Derivative of cost over arrival time



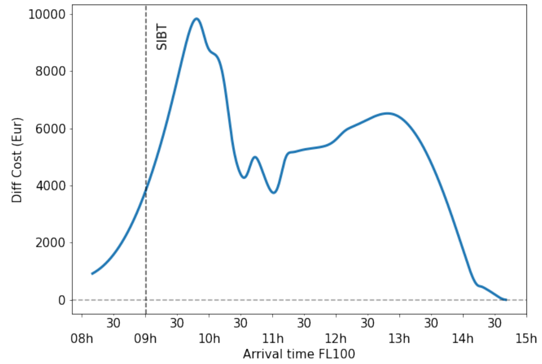
(d) Difference on derivative cost

Figure 63 Considering uncertainty on passenger connections (minimum connecting time and departure estimation)

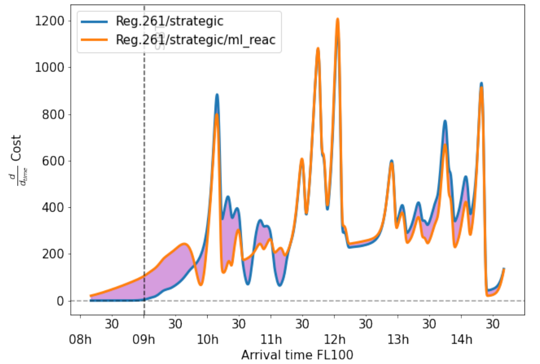
As shown in Figure 63 the consideration of uncertainty on when cost will materialised, e.g. due to uncertainties on the connecting time of passengers, lead to smoother cost functions. Therefore the total expected cost is similar for both alternatives (with uncertainty on passenger connection or without), but the uncertainty reduces the non-linear jumps.



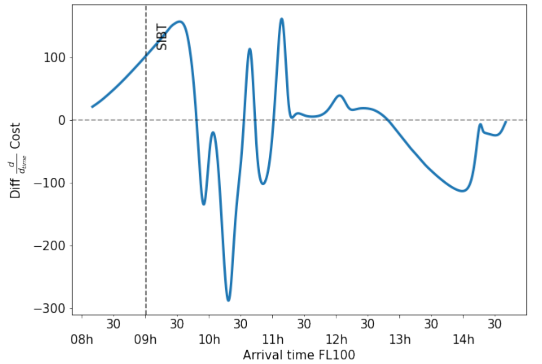
(a) Expected cost



(b) Difference of cost



(c)Derivative of cost over arrival time

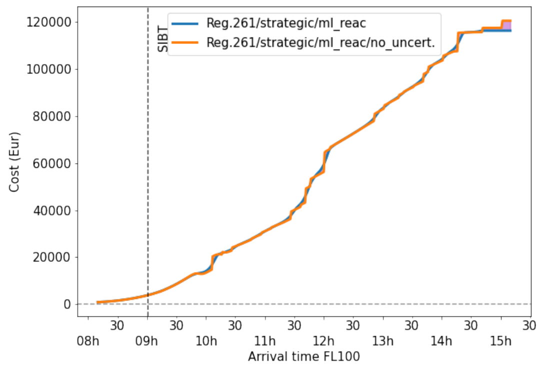


(d) Difference on derivative cost

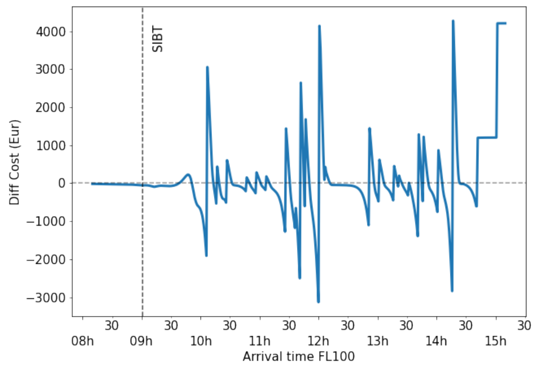
Figure 64 Effect of machine learning model to estimate reactionary delay on cost function

As presented in Section 4.2.3.2.2.3, the use of machine learning models to estimate the reactionary delay leads to overall higher estimation of delay. This is since the heuristic version focuses on the estimation of solely the minimum turnaround time while the machine learning model approach incorporates the estimation of delays due to ATFM congestion. As shown in Figure 64 (a) this leads to, in some instances, having a non-zero cost even if the flight arrives at its intended SIBT. This, however, does not necessarily mean that the optimisation will always be biased towards recovering delay. See for instance in Figure 64 (d), how the expected cost as a function of time variation is in some cases lower for the estimation of reactionary delay and cost with ML models with respect to the heuristic approach. As shown in Figure 64 (d), if flight arrives between 10h0 and 10h30 to FL100 the increase in cost is lower in the ML approach. In general, the use of ML models for reactionary delay will, how-ever, increase the slope of the cost of delay function close to SIBT which might lead, as shown in this case, to potentially some recovery of delay even if the planned arrival time is at SIBT.

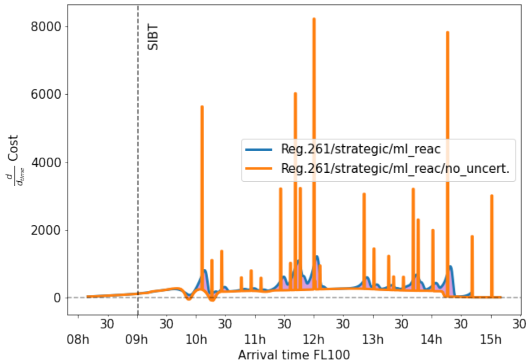
EUROPEAN UNION



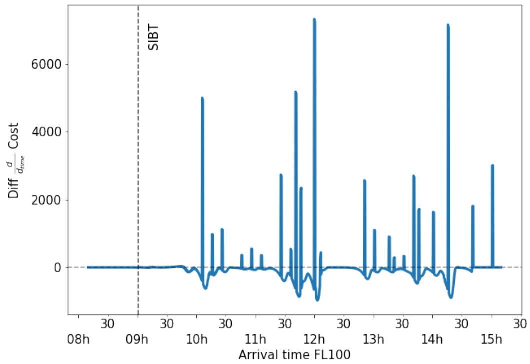
(a) Expected cost



(b) Difference of cost

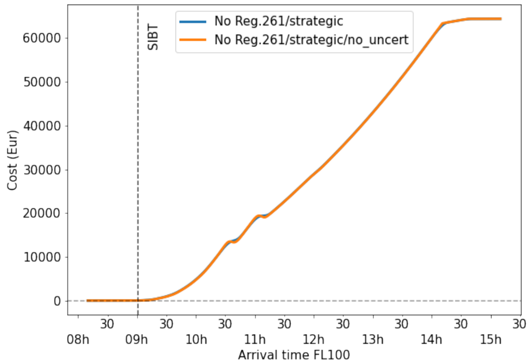


(c)Derivative of cost over arrival time

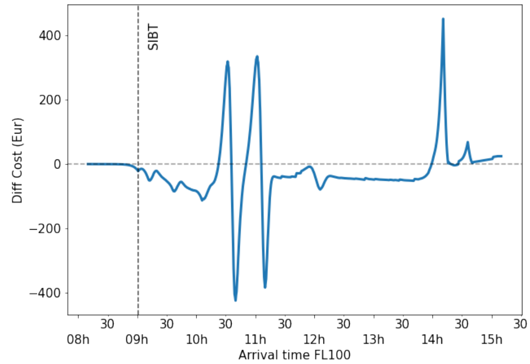


(d) Difference on derivative cost

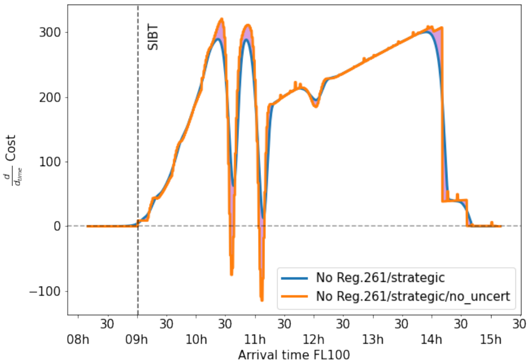
Figure 65 Not modelling uncertainty on Operational ATM Estimators



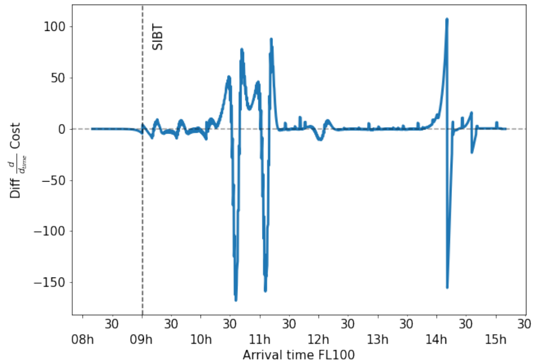
(a) Expected cost



(b) Difference of cost



(c)Derivative of cost over arrival time

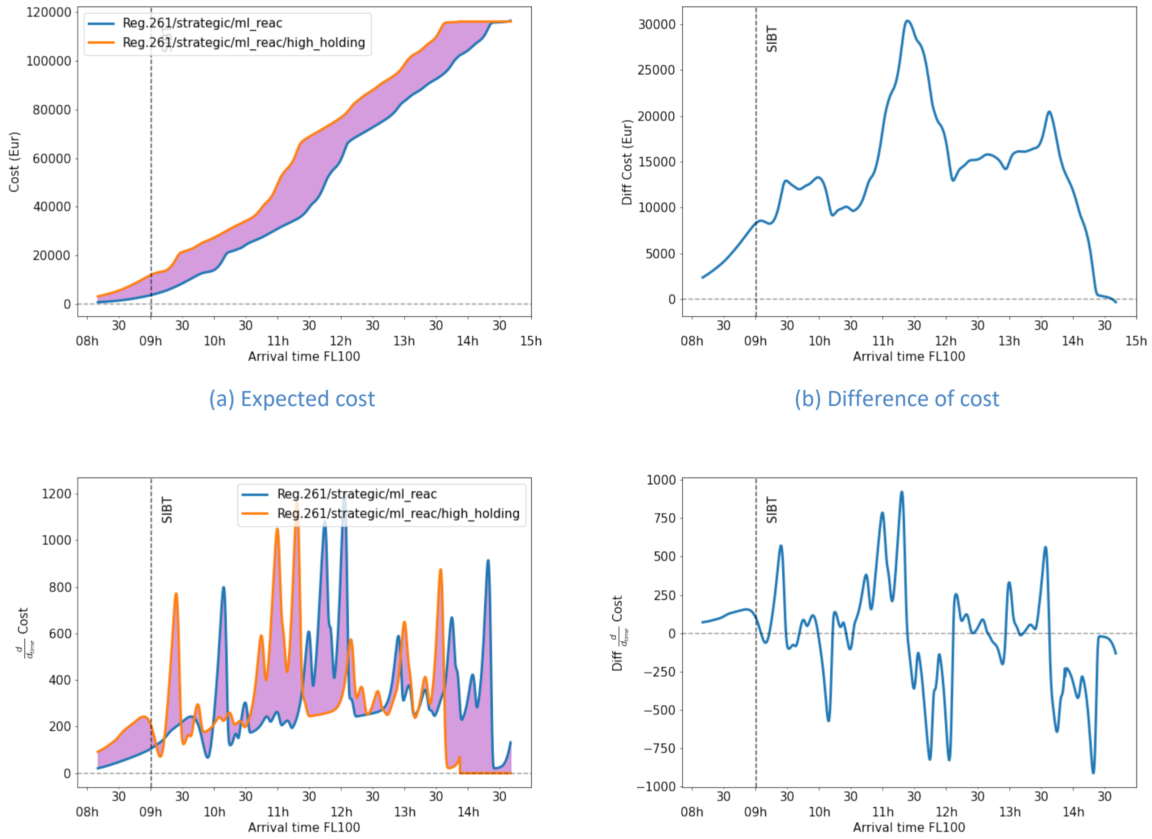


(d) Difference on derivative cost

Figure 66 Not modelling uncertainty on Operational ATM Estimator and no entitlement to Regulation 261

As with the modelling of uncertainty for passenger connections, the modelling of uncertainty on the operational ATM estimators (holding, sequencing and merging and taxi-in times) leads to smoother cost of delay function. As shown in Figure 65 (a) and Figure 65 (b) the expected cost is similar, however, the not use of uncertainty leads to abrupt changes on the cost function variation. This will mean that without the uncertainty the optimiser will consider that costs are or not materialised in a binary approach leading to solutions close to the cost of delay jump. However, the approach with uncertainty will produce smoother optimisations. A similar behaviour is observed if passengers are not entitled to Regulation 261 (see Figure 66). However, in this case as the cost function is smoother (no abrupt changes on cost due to passenger missing connection and entitled to the compensation), the impact is also reduced with lower discrepancies between the two approaches.

EUROPEAN UNION



(a) Expected cost

(b) Difference of cost

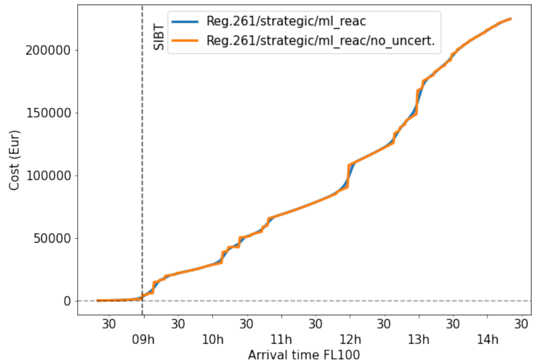
(c)Derivative of cost over arrival time

(d) Difference on derivative cost

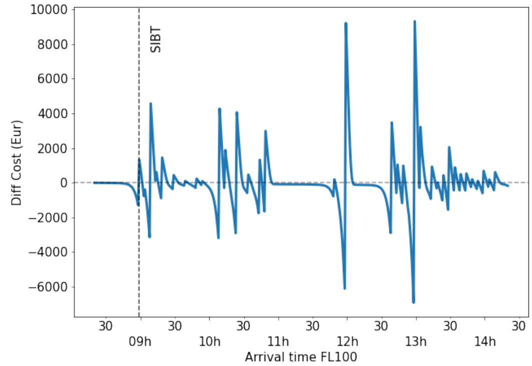
Figure 67 Effect of high delay at arrival

Finally, the effect of a high delay at arrival, e.g. due to a predicted long holding, is analysed an presented in Figure 67. The delay is translated into a shift on the cost function. Therefore, if for a given amount of initial delay the optimiser would or not decide to recover more or less delay would depend on the shape of the cost of delay as a function of time at different points in the curve. For example, the extra holding delay might mean that some costs are al-ready materialised (or not recoverable, or) and therefore a lower delay recovery suggested, or on the contrary higher costs variations expected and therefore higher delay recovered. Note that as in the machine learning modelling of reactionary delay (see Figure 64), in general this delay at arrival translated into the fact that even if arriving at the planned time at FL100 with respect to the schedule some costs might be present, and delay might be recovered prior the holding.

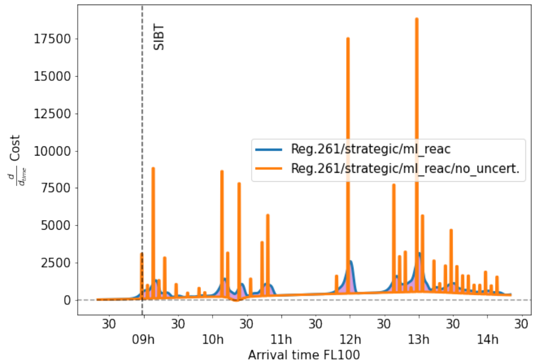
4.2.4.2 SCN 600 – KJFK – EDDF cost function alternatives analysis



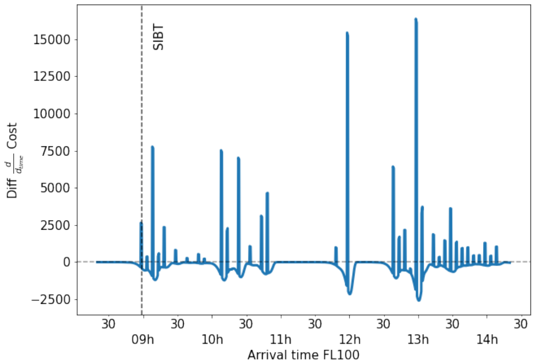
(a) Expected cost



(b) Difference of cost



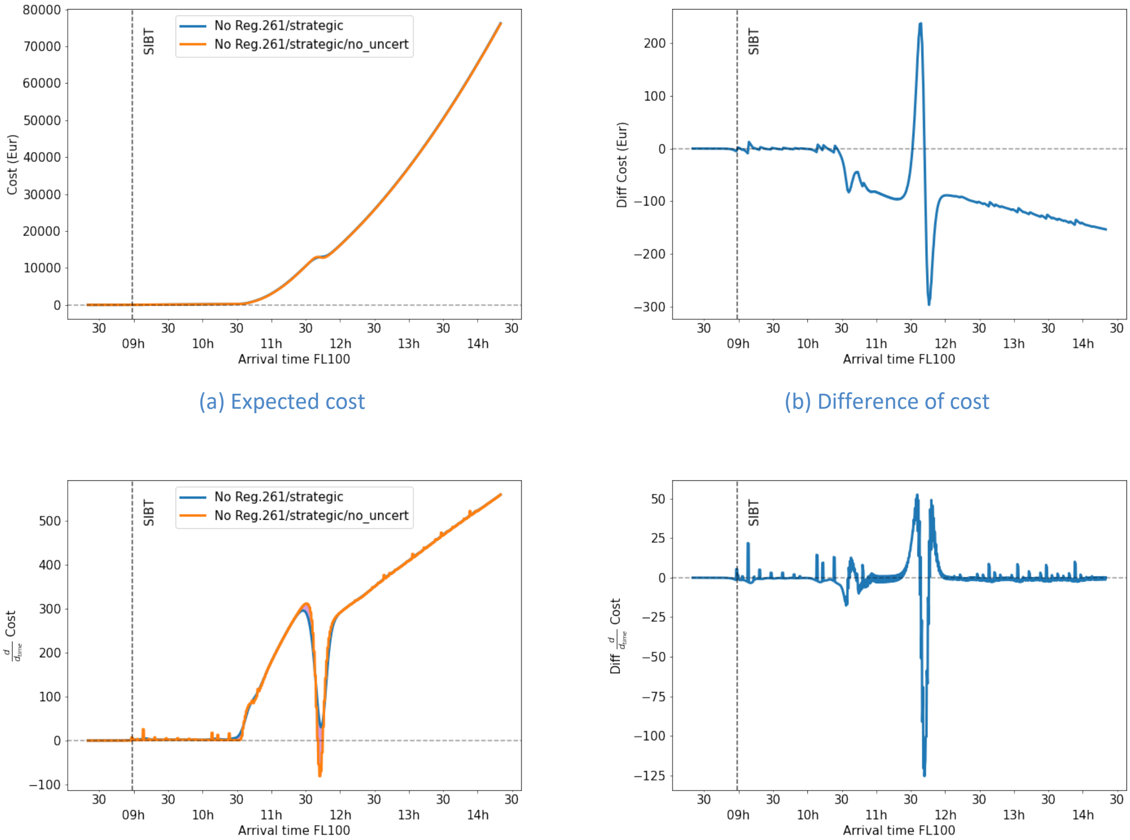
(c)Derivative of cost over arrival time



(d) Difference on derivative cost

Figure 68 Modelling or not uncertainty with passengers entitled to Regulation 261 compensation KJFK – EDDF

EUROPEAN UNION



(a) Expected cost

(b) Difference of cost

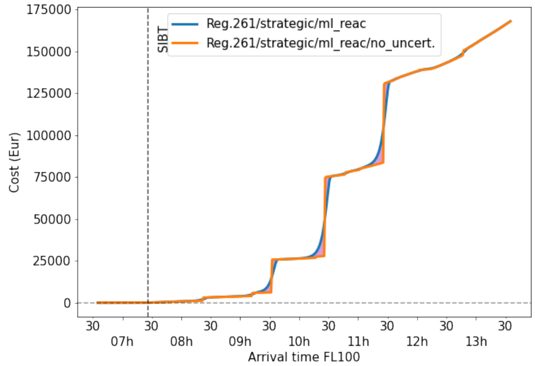
(c)Derivative of cost over arrival time

(d) Difference on derivative cost

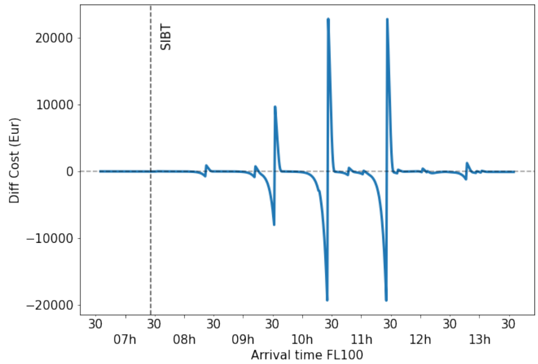
Figure 69 Modelling or not uncertainty with passengers not entitled to Regulation 261 compensation KJFK – EDDF

Figure 68 and Figure 69 analyse the impact of uncertainty on the OAE for the KJFK – EDDF flight when passengers are or not entitled to Regulation 261 compensation. As previously seen, the use of uncertainty leads to smoother cost functions which reduce the discontinuities. This is particularly relevant for the case when passengers are entitled to compensation. Overall, however, the difference in terms of variation of cost over time is not very high be-tween the different options (see Figure 68 (d) and Figure 69(d)) except for the above-mentioned discontinuities.

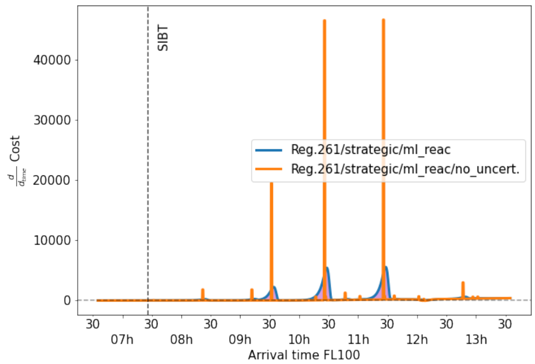
4.2.4.3 S CN 800 – KJFK – EGLL cost function alternatives analysis



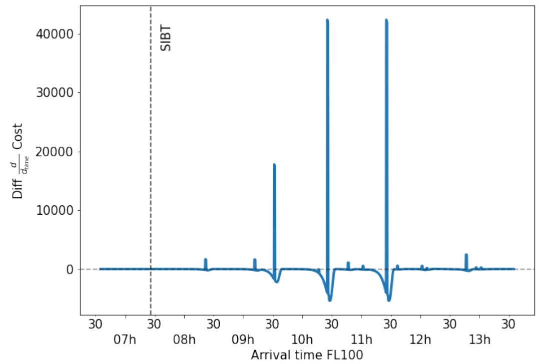
(a) Expected cost



(b) Difference of cost



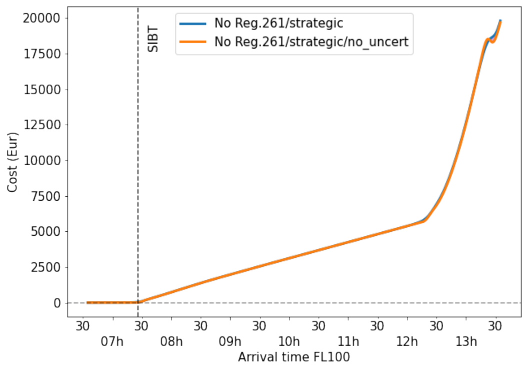
(c)Derivative of cost over arrival time



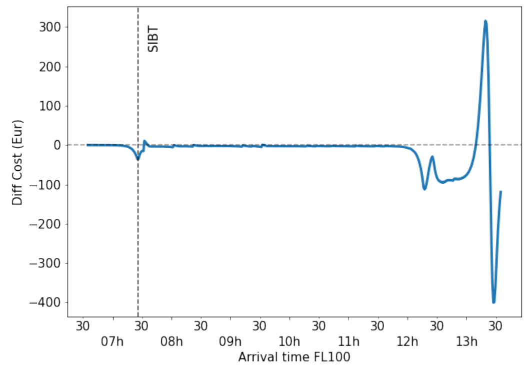
(d) Difference on derivative cost

Figure 70 . Modelling or not uncertainty with passengers entitled to Regulation 261 compensation KJFK – EGLL

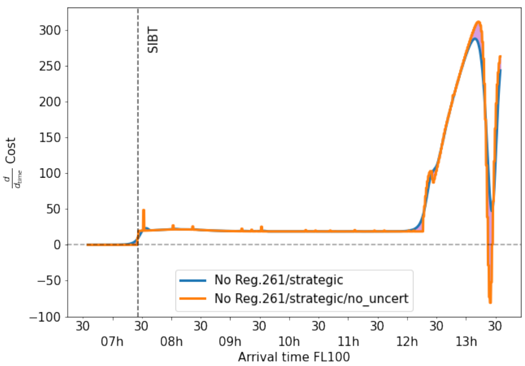
EUROPEAN UNION



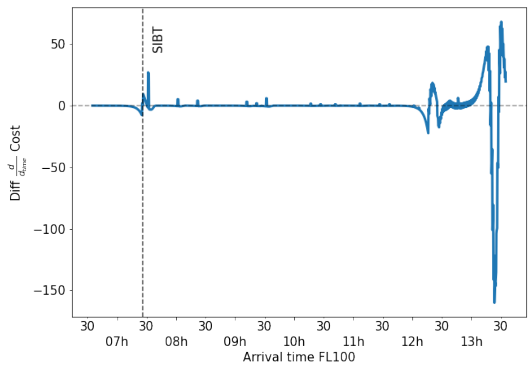
(a) Expected cost



(b) Difference of cost



(c)Derivative of cost over arrival time



(d) Difference on derivative cost

Figure 71 Modelling or not uncertainty with passengers not entitled to Regulation 261 compensation KJFK – EGLL

In this flight it is worth observing how if passengers are entitled to compensation due to Regulation 261 different plateaus are generated in the cost function (see Figure 70 (a)). When cost is materialised, i.e., passengers miss their connection the flight might consider it is not worth it to recover delay (see Figure 70 (c) how the variation of cost over time can be practically zero), in some cases even incurring in cheap extra delay by slowing down could be beneficial if fuel is saved. This behaviour, and recovering up to just before a cost jump, will be particularly prevalent when uncertainty is not considered. As shown in Figure 70 (a) and Figure 70 (c) when uncertainty at arrival is modelled these sharp cost increments are smoothed and variations on cost of delay over time reduced.

For this flight, the passengers' compensation drives the cost function. Therefore if passengers are not entitled, as shown in Figure 71, the increment of cost of delay is rather gradual, leading to a constant derivative of cost of delay over time which would produce very consistent optimisation performance. Note that in this case, due to the shape of the cost function, the temporal shift experienced by the modelling of uncertainty at arrival is practically non-relevant.

4.2.4.4 Conclusions on configuration of Pilot3 cost modelling components

As presented in this section, the most important factor when optimising the trajectory is not necessarily the absolute expected cost of delay but how this cost varies over time. Therefore different flight with different drivers for the cost function (passenger, reactionary, compensation) might benefit

EUROPEAN UNION

in different degree from the modelling of uncertainties and cost factors with more or less detail. The same amount of expected arrival delay might produce very di-verse optimisation results as a function of which factors are considered in the modelling of the cost function. Further research should try to identify the characteristics of flights which might benefit from different levels of quality (and complexity) on the estimation of cost and uncertainty components.

4.3 IVA3 – Assessment of the optimisation framework

IVA3 aims at validating the Performance Assessment Module. To that end, a theoretical example of flight (but using realistic values) is used.

4.3.1 Approach

In this validation activity, we first show how different choices of priorities for the airlines would lead to different results (rankings) from the Performance Assessment Module. Then we study the sensibility of the AHP-VIKOR algorithm. See deliverables D2.1 -- Trade-off report on multi-criteria decision making techniques (Pilot3 Consortium, 2020b) for more details. This focuses on the validation of P3-RQ-IV-060.

Note that the current framework of Pilot3 can handle multiple alternatives but does not provide a set of equivalent optimised alternatives automatically. These limitations are described in D6.1 – System evolution and uptake (Pilot3 Consortium, 2022b).

4.3.2 Flight characteristics

The flight considered goes between Madrid (LEMD) and Frankfurt (EDDF), with an Airbus A320. It is an early flight scheduled to arrive to Frankfurt at 9h10 UTC with 38% of passengers with further on-going connections. Note that, in Frankfurt the domestic minimum connecting time is 45 minutes, and 60 minutes for international passengers. This means that passengers need at least that time to ensure their connection.

For this flight, the variation in fuel and time that can be achieved optimising the trajectory by selecting a different CI with respect to the planned one (i.e., CI of 10 kg/min) is obtained. These fuel and delay trade-offs are computed using the trajectory optimiser DYNAMO. In this particular case, the flight can recover up to 6.4 minutes by selecting a higher CI (100 kg/min), using in this case 154 EUR extra cost of fuel (308 kg of fuel at a fuel cost of 0.5 EUR per kg of fuel). The flight could also consider slowing down, increasing its delay by up to 2.2 minutes with a saving of 20 EUR of fuel. This means that if the flight is expected to arrive later than 6.4 minutes after OTP it will not be possible to meet OTP, and if the flight is expected to arrive earlier than 2.2 minutes before OTP then it will always meet OTP.

Four different case studies are defined to represent different operational situations of interest:

- Case 1 (OTP): Flight arriving at scheduled in-block time (SIBT). In this case, OTP would always be reached but the trajectory could still be optimised considering trade-offs between fuel and cost of delay.
- Case 2 (OTP and no-OTP): Small expected arrival delay of 17 minutes, which would present the opportunity to recover enough delay to meet OTP; here OTP and no-OTP cases are available and ranked independently.

- Case 3 (no-OTP): High expected arrival delay of 55 minutes, with low variability of delay cost around expected arrival time.

4.3.3 Combination of possible configurations

For these four cases, we study here three combinations of airline priorities:

- **IROPs priority**: IROPs is the most important cost, followed by fuel, followed by other costs
- **Fuel priority**: fuel is the most important cost, followed by IROPs, followed by other costs
- **Other costs priority**: other costs is the most important cost, followed by IROPs, followed by fuel

to validate how different choices of proprieties lead to different optimization results. These combinations lead to different weights used in the AHP-VIKOR algorithm.

The expected total cost of a given trajectory is obtained by adding the expected costs of delay and fuel of this trajectory. Therefore, given an expected arrival time at the gate, the possible available alternatives that the system will consider are obtained based on the variations of time and fuel. If only total cost is minimised, due to the characteristics of the cost of delay curve, only one alternative is generally found. Therefore, to be able to consider more than one alternative, some buffer (extra cost) should be used.
With small buffers, 10 EUR, several alternatives can already be obtained, as a function of the expected arrival time to the gate. For example, if the flight is expected to arrive at its SIBT (Case study 1), with 10 EUR buffer, a range of 5 minutes can be considered, for which all solutions lie within a maximum of 10 EUR of extra total cost with respect to the minimum total cost alternative. In this example, to increase the number of potential *equivalent* trajectories, a buffer of 50e is used in all cases.

Table 37: Ranked solutions (Extra delay [min]) for different cases as a function of the order of priorities

	IROPs priority	Fuel priority	Other costs priority
Case 1 (OTP)	-3, -2, -4	-2, -1, -3	-3, -4, -5, -2
Case 2 (OTP)	-5, -6	-4, -3, -5	-6, -5
Case 2 (no-OTP)	-1, -2	1, 0, 2	-2, -1
Case 3 (no-OTP)	-4, -5, -3	-3, -2, -1	-6, -5

Table 37 shows the results obtained for each case study when applying the ranking AHP-VIKOR algorithm. It presents the available solutions, indicating the difference on arrival time with respect to the expected arrival time of the flight.

In case 1, all options reach OTP, as the flight was expected to arrive at its SIBT. Therefore, it should be expected that fuel cost is the only cost and thus that slowing down to save fuel should be the only possible option. However, the cost function considers uncertainties linked to operations, such as holdings, missed connections. This means that IROPs costs, even if low, are not null. It can be seen that if IROPs costs are considered as the most relevant cost, the ranked solutions consist in recovering 3, 2 and 4 minutes, while if fuel is the most important cost, the solution would be recovering 2, 1 and 3 minutes. One more minute is considered when fuel is considered more important, in order to lower fuel costs. In case 2 (no-OTP), when giving priority to fuel instead of IROPs, solutions by two minutes. Case 3, which has a delay of 55 minutes, sees IROPs costs increase more rapidly than in the previous cases. As a consequence, the solutions tend to recover as much delay as possible (4-5 minutes) if IROPs cost is the most important one and less if fuel is more important.

As expected, we observe in all cases that different configurations of priorities lead to different results, in particular, we can see that giving more priority to fuel leads to results tending to recover less delay (in order to save more on fuel).

4.3.4 AHP-VIKOR algorithm sensibility analysis

One tuning parameter of the VIKOR method (see D4.3 (Pilot3 Consortium, 2022a)), corresponds of the weight of the strategy of maximum group utility versus of the individual regret, was set to a "neutral" value of 0.5. We have proven that changing this value does not affect the results, unless extreme values (<0.1 or >0.9) are chosen. It is thus considered that maintaining the commonly used value of 0.5 is adequate. The effect of the choice of this tuning parameter is thus rather limited, proving the robustness of the optimisation framework proposed.

4.3.5 Summary of Research Questions and Hypothesis

After revision of the initially defined RQs and HPs aimed for validation of IVA3, with the previously discussed results, we were able to successfully validate the RQ-IV-040, RQ-IV-050 and RQ-IV-060 (see Table 38).

Table 38: Summary of research questions (RQ) and hypotheses addressed in IVA3

RQ ID	Rationale	Research question	Hypothesis	Success criteria	Status
P3-RQ-IV-040	Validate that the Pareto can be computed by Pilot3 (e.g. if trade-offs between OTP and cost exist they can be computed with Pilot3).	For a given triggering event, will Pilot3 generate a meaningful set of alternative 4D trajectories when trade-off between objectives are present?	It is expected to obtain trade-off 4D trajectories between Total Cost and OTP (i.e., Pareto efficient solutions). Moreover, it is expected to obtain different 4D trajectories with same cost objective but different sub-cost components (KPIs).	• When the trade-off between Total Cost and OTP exists, the **two** Pareto optimal trajectories are generated. • When the trade-off between cost KPIs exists (fuel, IROPS and other), **different trajectories** are generated.	**Partially Validated**
P3-RQ-IV-050	Validate that if more than one alternative produce equivalent results Pilot3 can compute them.	For a given triggering event, will Pilot3 generate a meaningful set of alternative equivalent 4D trajectories?	It is expected to obtain different 4D trajectories that lead to the same (and/or statistically equivalent) objective functions (i.e., Total Cost or OTP).	• **At least two** trajectories that lead to the same (and/or statistically equivalent) objective function (i.e., at least two trajectories for Total Cost and at least two trajectories for OTP) when trade-off between cost KPI are possible	**Not validated**

EUROPEAN UNION

P3-RQ-IV-060	Validate that the airlines' policies captured as preferences in the configuration are considered adequately by the trajectory optimisation.	For a given triggering event, will Pilot3 show different 4D trajectories for different airline policies configured in the tool?	Pilot3 will provide its full potential to the airline industry as it will capture different airline policies (as reflected in the Pilot3 configuration) that will lead to different solutions (i.e., trajectories) to the same problem (i.e., triggering event).	• KPIs and PIs have **different values** for different Pilot3 configurations • **Different ranking** of alternative trajectories for different Pilot3 configurations	**Validated**

4.4 IVA4 – Pilot3 performance at optimised trajectories plan

As defined in D5.1, the purpose of the internal validation activity 4 (IVA4) was to capture the benefits of Pilot3 optimised trajectory plans when compared it against several baseline trajectory plans.

4.4.1 Approach

The comparison of planned trajectories (e.g. the baseline trajectory plan and the optimised trajectory plan) is useful to evaluate expected benefits, but should be done with caution as misinterpretations could arise (e.g. if stronger headwind than originally planned is forecasted with a weather update, the optimised trajectory will likely use more fuel and/or have a longer duration than the original baseline which was estimated prior the weather update). For this reason, an integration of the remaining trajectory by the different baseline trajectories is required to provide more meaningful comparisons. In the D5.1, two groups of baseline trajectory were proposed, namely:

- the **OFP** being executed regardless of the different Pilot3 triggering events that might arise in flight; and
- some **new plan(s)** assuming some typical pilot's reactions based on their experience and in the absence of Pilot3.

During the project validation, the results of the Pilot3 optimised trajectory plan have been analysed and compared against OFP by using several metrics. Following the OFP could be seen as "do nothing", as basic pilot reaction. No other pilot reaction was finally modelled and simulated in the execution of the project. More details can be found in the subsection below.

4.4.2 Summary of the Experimental Scenario

The benefits of the Pilot3 optimised trajectory plan have been validated on four different scenarios among the nine initially identified in D5.1. As already introduced in Section 2, for each experiment considered, the characteristics of the different components of the experiment (scenario, sub-scenario, case studies, sub-case study and parametrisation) were defined. These experiments are reported in Table 39 below.

The benefit of Pilot3 optimised trajectory with respect to baseline (i.e., keep flying OFP) will be discussed using the following metrics:

- difference in total fuel consumption;
- difference in total trip time;
- difference in IROPs cost;
- difference in other cost;
- difference in total cost.

Table 39: IVA4 experiments

EXP	Scenario	Sub Scenario	Case Study (CS)				Sub case Study	Purpose
ID	ID	OFP weather	ID	Triggering point	CS main feature	Other features	PIE/OAE configuration	
401	100	Nominal	40	TOC	Expected holding at arrival TMA: +40'	Departure delay: -20' (i.e., early departure)	- PIE: default, manual holding of 40’ - AOE: default	- Tight buffer at arrival (SIBT-ETA: 9'). - Although we arrive 20 earlier at the TOC, we speed-up because some significative holding is expected at destination TMA and a significative reactionary delay is predicted by the PIE. .
414	100	Nominal	40	TOC	Expected holding at arrival TMA: +40'	Departure delay: -20' (i.e., early departure)	- PIE: default except for reactionary delay with heuristics, manual holding of 40' - AOE: default	- Reactionary delay is estimated with heuristics and arriving before the SIBT implies (almost) no cost. - Similar purpose as experiment 401. Still speeding-up, but less than in experiment 401.
403	201	Nominal	60	15 NM after TOC	Heavy head wind weather Date: 2018-04-17	Departure delay: -20' (i.e., early departure)	- PIE: default - AOE: default	- Although we arrive 20 earlier at the TOC, we speed-up because the weather update indicates more headwind than expected.
404	201	Nominal	60	15 NM after TOC	Heavy head wind weather Date: 2018-04-17	Departure delay: 0' (i.e., on-time departure)	- PIE: default - AOE: default	- Similar purpose as experiment 403, but we expect to speed up even more.
412	201	Nominal	60	TOC	Heavy tail wind weather Date: 2018-04-17	Departure delay: +20'	- PIE: default - AOE: default	Although the unexpected tailwind component is beneficial, we still speed up due to the reactionary delay estimated with the ML model of the PIE
413	201	Nominal	60	TOC	Heavy tail wind weather Date: 2018-04-17	Departure delay: 0' (i.e., on-time departure)	- PIE: default - AOE: default	Same purpose as in experiment 412 with different departure delay.
415	201	Nominal	60	TOC	Heavy tail wind weather Date: 2018-04-17	Departure delay: 0' (i.e., on-time departure)	PIE: default except for reactionary delay with heuristics. AOE: default	- Reactionary delay is estimated with heuristics and arriving before the SIBT implies (almost) no cost. - Due to beneficial tailwind conditions we slow-down to maximum range cruise (zero cost index) to save fuel (and still arrive on time).

EXP	Scenario	Sub Scenario	Case Study (CS)			Sub case Study		Purpose
ID	ID	OFP weather	ID	Triggering point	CS main feature	Other features	PIE/OAE configuration	
405	600	Nominal	10	TOC	Departure delay: +30'	-	- PIE: default - AOE: default	- Normal buffer at arrival (SIBT-ETA: 22') - Nominal arrival time is within OTP, but because the first group of connecting PAX there is a high probability to miss these connections. We speed-up to lower the expected total cost.
406	600	Nominal	80	600 NM before the turbulence area	Turbulence[2]	Departure delay: 0' (i.e., on-time departure)	- PIE: default - AOE: default	- Demonstrate the effect of triggering Pilot3 in case of turbulence ahead. The new trajectory avoids the turbulence volume. - Demonstrate the pilot can put constraints.
407	600	Nominal	80	600 NM before the turbulence area	Turbulence	Departure delay: +30'	- PIE: default - AOE: default	- Demonstrate the effect of triggering Pilot3 in case of turbulence ahead. The new trajectory avoids the turbulence volume. - Demonstrate the pilot can put constraints. - Demonstrate that Pilot3 takes into account the departure delay and the avoidance trajectory is different than previous one.
408	600	Nominal	80	600 NM before the turbulence area	Turbulence	Departure delay: +60'	- PIE: default - AOE: default	- Demonstrate the effect of triggering Pilot3 in case of turbulence ahead. The new trajectory avoids the turbulence volume. - Demonstrate the pilot can put constraints. - Demonstrate that Pilot3 takes into account the departure delay and the avoidance trajectory is different than previous one.
409	600	Nominal	80	600 NM before the turbulence area	Turbulence	Departure delay: +90'	- PIE: default - AOE: default	- Demonstrate the effect of triggering Pilot3 in case of turbulence ahead. The new trajectory avoids the turbulence volume. - Demonstrate the pilot can put constraints. - Demonstrate that Pilot3 takes into account the departure delay and the avoidance trajectory is different than previous one.

[2] See Figure 21 for further details.

EXP	Scenario	Sub Scenario			Case Study (CS)		Sub case Study	Purpose
ID	ID	OFP weather	ID	Triggering point	CS main feature	Other features	PIE/OAE configuration	
410	800	Nominal	10	250 NM after TOC	Departure delay: +153'	-	- PIE: no uncertainty modelled - AOE: default	- Tight buffer at arrival (SIBT-ETA: 15') - Demonstrate that Pilot3 can optimise with a deterministic cost function. - Demonstrate the sensitivity of the solution to cost function modelling.
411	800	Nominal	10	250 NM after TOC	Departure delay: +153'	-	- PIE: default - AOE: default	- Tight buffer at arrival (SIBT-ETA: 15') - Demonstrate that Pilot3 can optimise with a probabilistic cost function. - If uncertainty is considered, there is a high probability to miss these connections. The optimised arrival time will be earlier than in case of not considering uncertainty.

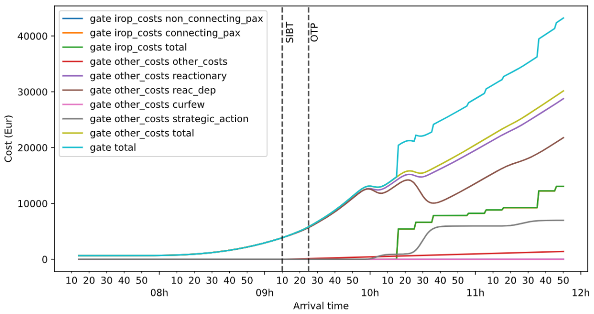
4.4.3 Results

4.4.3.1 Scenario 100: Athens (LGAV) to London Heathrow (EGLL)

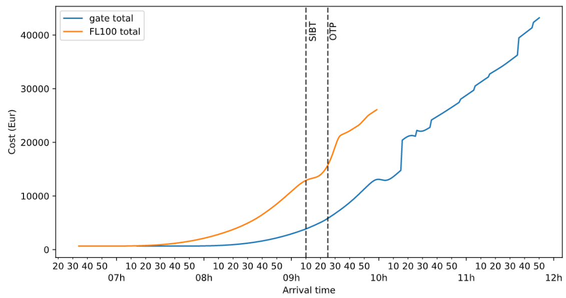
The description of the experiments carried out for the scenario 100 is given in Table 39 above. As seen from Table 39, two very similar experiments are run with only difference in the estimation of reactionary delay – while experiment ID 401 was run by applying machine learning for the estimation of the reactionary delay, the experiment ID 414 aims to demonstrate the benefit of Pilot3 when the reactionary delay is estimated by applying heuristics. In this way, one can observe the flexibility of Pilot3 to configure the PIE in different ways.

4.4.3.1.1 Cost function

Figure 72 below depicts the outlook of the cost function for the Athens - London route. Figure 72 (a) shows the expected costs as a function of arrival time at the gate. As observed, it consists of different components involving the IROPs costs (e.g. passenger compensation costs (it is assumed that passengers are entitled to compensation due to Regulation 261 if delay thresholds are passed)) and other costs (i.e., reactionary delay, crew costs, maintenance cost). After considering the uncertainties at arrival (i.e., holding, sequencing and merging procedure, taxi-in), the expected costs at arrival is translated into expected costs at FL100 (orange line in Figure 72 (b)).



a) Total expected costs as a function of arrival time at the gate



b) Total expected costs at FL100

Figure 72 Cost function for Athens – London Heathrow route

As shown in Figure 72 a) for this particular flight, costs are dominated by other costs. Reactionary costs (propagated in subsequent rotations) are the main driver of the total costs expected. As seen in Table 40, the time allowed for rotations is relatively tight. The aircraft has 40 minutes for the rotation at EGLL before departing to LIRF. Therefore, if the arrival to EGLL is delayed, the probability of delay being propagated to LIRF (and to subsequent flights) is high. In this particular example, the distribution of reactionary delay is computed using machine learning models. This means that not only the minimum turnaround time is considered but the probability of longer rotations and ATFM delay are also modelled. Even if the flight arrives to the gate at schedule, it can be observed how, some reactionary cost (and delay) is expected. This might mean that even if departing on-time from LGAV, Pilot3 might consider that recovering time is beneficial to reduce the total expected cost as it might be able to reduce the delay expected to be propagated to LIRF.

See the difference with Figure 73 where the same cost function is computed only using heuristics for the reactionary delay estimation for comparison. In this case, reactionary delay and cost do not occur after 30 minutes after SIBT. Hence, machine learning models (if accurate (properly trained)) would provide information on the impact of external factors which might affect the cost function leading to

EUROPEAN UNION

more optimal solutions (e.g. it is worth it to recover delay because subsequent flights have a high probability of propagating delay. This won't have been noticed without these advanced models).

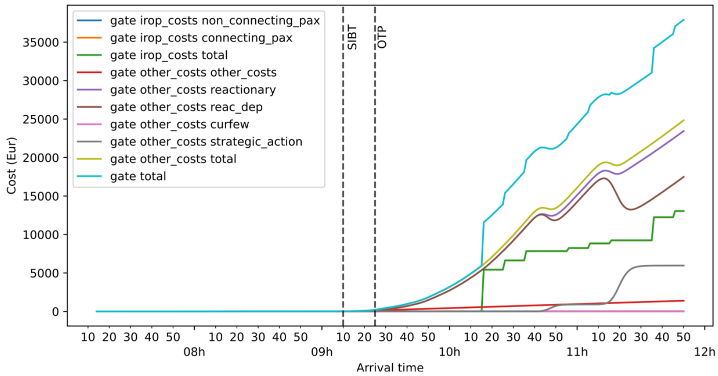


Figure 73 Cost function for Athens - London Heathrow route if reactionary delay modelled with heuristics

Table 40: Rotation planned for Athens – London Heathrow flight

Origin	Destination	SOBT	SIBT	Rotation time available from previous flight
LGAV	EGLL	05h15	09h10	-
EGLL	LIRF	9h50	12h20	0h40
LIRF	EGLL	13h15	15h50	0h50
EGLL	LTBA	16h50	20h40	1h

It is worth noticing how when the flight arrives after 10h there is a possibility that the airline might do an action to reduce the propagation of delay (pre-tactical (strategic action) cost such as cancelling or swapping a flight originally planned as a follow up rotation), this probability (and associated cost) increases over time (see cost at 10h30). This has the overall impact that the expected reactionary delay cost is reduced even if still dominating the total expected cost of the flight.

For the passenger related costs, Figure 74 shows how the passengers connecting in EGLL form this flight have a relatively large connecting time at the hub. The first passenger group with a connection has an ongoing flight scheduled close to 11h (note that the SIBT of the LGAV-EGLL flight is 9h10). This means that passengers will not miss connections until an arrival delay greater of one hour. As observed in the cost function, the first significant increment of passenger cost is produced when the expected arrival time at the gate is 10h15 which would correspond to this group of passengers missing their connection and being entitled to Regulation 261.

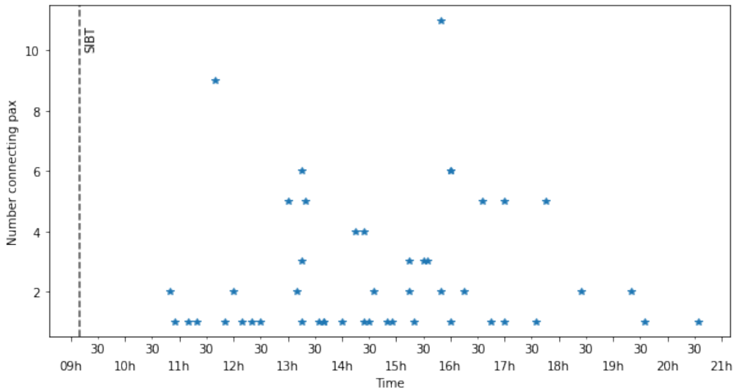
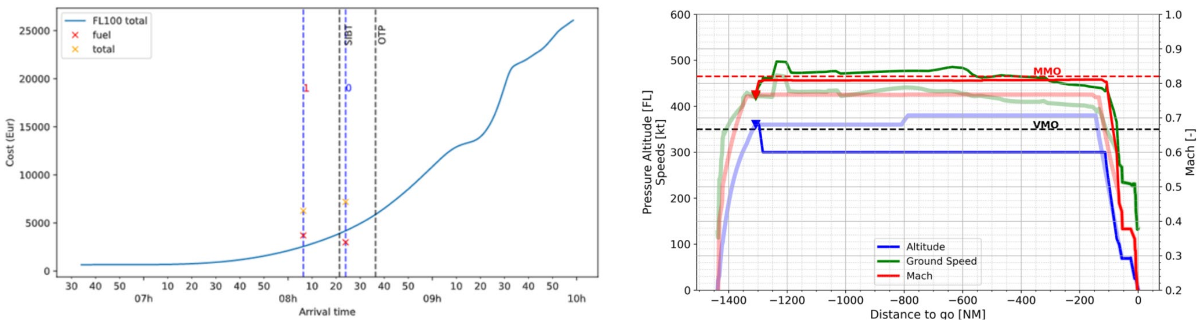


Figure 74 Passenger groups connecting at EGLL into follow up flights on the LGAV — EGLL flight

The experiment assumed no changes in weather condition with respect to the OFP and the triggering point is at TOC (top of climb). The reason for triggering Pilot3 is that the aircraft reaches the cruise altitude (at TOC) earlier than planned. The aircraft crew queries Pilot3. Although the aircraft arrives 20 minutes earlier at the TOC with respect to the OFP, Pilot3 will speed-up due to the holding is expected at destination.

4.4.3.1.2 Results for CS40 – Departure delay with holding at arrival

The resulting trajectory of the Pilot3 optimised trajectory plan is presented in Figure 75 (b) in addition to the total expected costs as a function of arrival time at FL100 (Figure 75 (a)). As observed from Figure 75 (a), the line denoted as "1" indicates the arrival time of the Pilot3 optimised trajectory plan, whereas the line "0" corresponds to the time of arrival of the OFP.



a) Total expected costs as a function of arrival time at FL100

b) vertical profile with speed profile of Pilot3 optimised trajectory

Figure 75 Results of Pilot3 optimisation for Experiment 401 (**20' earlier** at TOC and expected arrival **holding** of **40'**)

Table 41: Different KPIs for Pilot3 optimised trajectory plan and OFP (from triggering point to FL100)

	Total fuel	Total trip time	Fuel costs	IROPs cost	Other costs
Pilot3 optimised trajectory	7,398 kg	159 min	3,699 EUR	0.01 EUR	2,566 EUR
Keep flying OFP	5,986 kg	177 min	2,993 EUR	0.49 EUR	4,201 EUR
Difference between Pilot3 and OFP	+ 1.412 kg	-18 min	+ 706 EUR	- 0.48 EUR	- 1,634 EUR

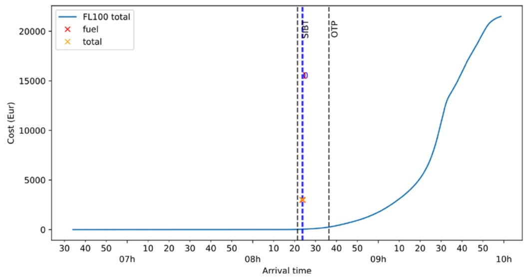
As observed from Figure 75 (a), Pilot3 optimised trajectory plan outperforms the baseline trajectory plan (i.e., keep flying OFP) in terms of total cost (indicated by the red cross) mainly stemming from the savings obtained through "other costs". Total trip time of Pilot3 optimised trajectory plan accounts for 2h and 39 minutes which is around 18 minutes shorter in comparison to the OFP trajectory. This allows the flight to arrive at its destination before SIBT. As commented before the machine learning model of the PIE still estimates some non-negligible expected cost even if the flight arrives before the SIBT. For this reason, the speed recovery of the Pilot3 solution is significant. This shorter travel time undoubtedly comes at the expense of higher fuel consumption of around 1,4 tonnes which corresponds to extra fuel cost of approximately EUR 700. However, the total savings expected, considering the whole objective function are 928.5 Eur (see Table 41).

As seen in Figure 75 (b) this important delay recovery (18 minutes) can only be achieved by increasing the speed to the maximum Mach in operations (MMO) -minus a safety/operational buffer- and performing a significant descent to FL300 to increase the true airspeed (and therefore ground speed) for that Mach.

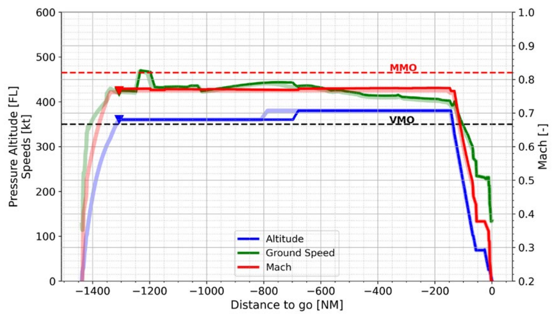
EUROPEAN UNION

As discussed above, the experiment ID 414 reproduces the same conditions as in the case of ID 401, but estimating, in the PIE, the reactionary delay costs with heuristics instead of machine learning (see Figure 73). As observed from Figure 76, arriving before the SIBT makes no difference in the cost function and for this reason the Pilot3 trajectory is speeding up just to arrive slightly before the SIBT. In this case, the speeding-up is of lower magnitude than in the case of experiment ID 401.

Concerning the trajectory profile of Pilot3 (see Figure 76 b)), one can observe a slight speed increase in the Mach number with respect to the OFP, while the cruise altitudes are not changed (although the position of the step climb is delayed by around 100NM). This new trajectory profile enables to recover the 1 minutes needed to arrive on-time at the destination gate, by accruing negligibly fuel costs in comparison to the OFP (see Table 42).



a) Total expected costs as a function of arrival time at FL100



b) vertical profile with speed profile of Pilot3 optimised trajectory

Figure 76 Results of Pilot3 optimisation for Experiment 414 (**PIE: heuristics for reactionary delay**)

Table 42: Different KPIs for Pilot3 optimised trajectory plan and OFP (from triggering point to FL100)

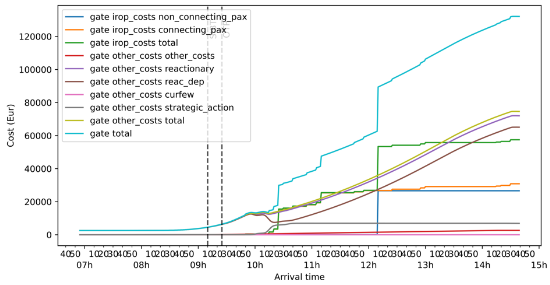
	Total fuel	Total trip time	Fuel costs	IROPs cost	Other costs
Pilot3 optimised trajectory	6,016 kg	176 min	3,008 EUR	0.44 EUR	34.5 EUR
Keep flying OFP	5,986 kg	177 min	2,993 EUR	0.49 EUR	37.3 EUR
Difference between Pilot3 and OFP	+30 kg	1 min	+15 EUR	-0.05 EUR	-2.8 EUR

4.4.3.2 Scenario 201: Madrid (LEMD) to Frankfurt (EDDF)

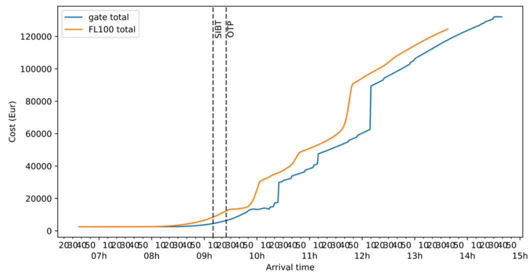
Five experiments were eventually run for the LEMD - EDDF scenario to show the benefit of Pilot3 optimised trajectory plan in different operational context encountering different weather conditions. Namely, the rationale behind the first two experiments (ID 403 and 404) is to explore the capabilities of Pilot3 prototype in the case of an updated weather forecast indicating more headwind and crosswind received shortly after the aircraft reaches TOC. The experiment ID 404 addresses the context in which the departure is performed on time. This was also one of the case studies emphasised as relevant by the experts from the Advisory Board. The experiment ID 403 aimed to show the benefit of Pilot3 in the case of similar operational context (i.e., weather update at TOC), but the aircraft reaches TOC 20 minutes earlier than OFP ETA (Estimated Time of Arrival). The third and four experiments (ID 413 and 412) aimed to explore the benefit of Pilot3 when facing a weather forecast update with heaving tail-wind shortly after reaching TOC under similar conditions as in the two previous experiments – when the flight is on-time and with 20 minutes delay at TOC (with respect to OFP ETA) respectively. It is worth mentioning that all these four experiments were run by using the reactionary delay estimated by machine learning model. Finally, in order to observe the behaviour of Pilot3 in the case when the reactionary delay is modelled by heuristics, the experiment ID 415 was run.

4.4.3.2.1 Cost function

Following the same rational as in the case of Athens - London Heathrow route, Figure 77 provides the expected costs at gate (Figure a) and expected cost at FL100 (figure b) for Madrid-Frankfurt route.



a) Total expected costs as a function of arrival time at the gate

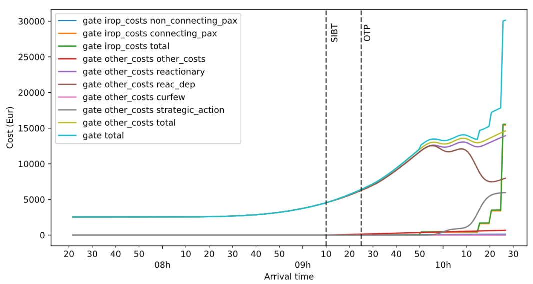


b) Total expected costs at FL100 considering uncertainty

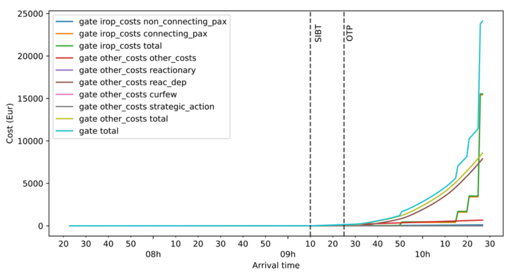
Figure 77 Cost function for Madrid – Frankfurt route

As shown in the cost function (Figure 77) costs are generally dominated by reactionary related costs. Similar to the LGAV to EGLL flight, the use of machine learning models for the estimation of reactionary delay produces an estimation on some cost even if flight arriving at its SIBT (see Figure 78 for a comparison between using machine learning models or not for this estimation). Rotation times allocated for subsequent flights (as shown in Table 43) are generally 50 minutes. Being one of the first flights in the morning, the propagation of delay can be significant along the subsequent 5 flights. That is one of the reasons of the early costs associated with pre-tactical actions to limit the propagation of reactionary delay and cost (see expected costs due to these strategic actions increasing after 10h30 (1h20 minutes of arrival delay).

For passengers' costs, Figure 79 presents the number of passengers which are connecting into forward flights as a function of the departing time of their first desired connection. It can be observed that a significant number of passengers (greater than 40) have a flight departing earlier than 12h. We can observe this first increment on passenger related costs around 10h20 which would correspond to the group of 20 passengers missing their connection (considering the minimum connecting time, if the flight arrives after 10h20 they won't be able to reach their connecting flight). Finally, the largest increment on passenger related costs appears when the flight arrives after 12h10 (as shown in Figure 77 (a)) which correspond to the threshold of 180 minutes of delay and therefore the entitlement of non-connecting passengers to compensation due to Regulation 261 for this flight distance.



a) Reactionary delay with machine learning models



b) Reactionary delay with heuristics

Figure 78 Detail cost at gate around SIBT for Madrid-Frankfurt route comparison reactionary delay with and without machine learning models

EUROPEAN UNION

Table 43: Rotation planned for Madrid – Frankfurt flight

Origin	Destination	SOBT	SIBT	Rotation time available from previous flight
LEMD	EDDF	06h35	09h10	-
EDDF	EGLL	10h00	11h40	0h50
EGLL	EDDF	12h30	14h10	0h50
EDDF	EGBB	15h00	16h35	0h50
EGBB	EDDF	17h25	19h00	0h50
EDDF	EGCC	19h55	21h40	0h55

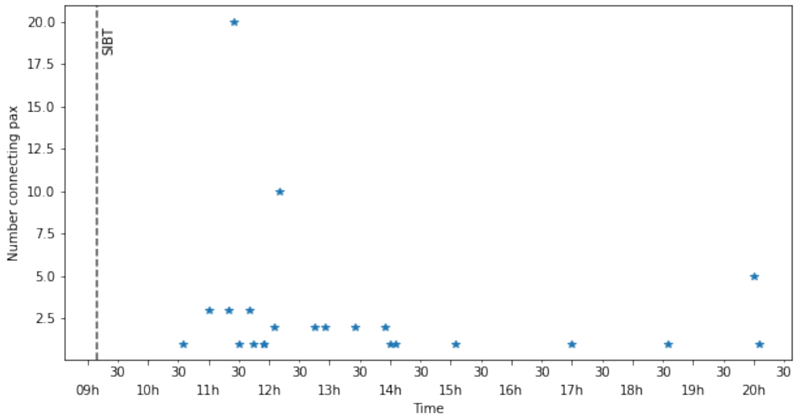
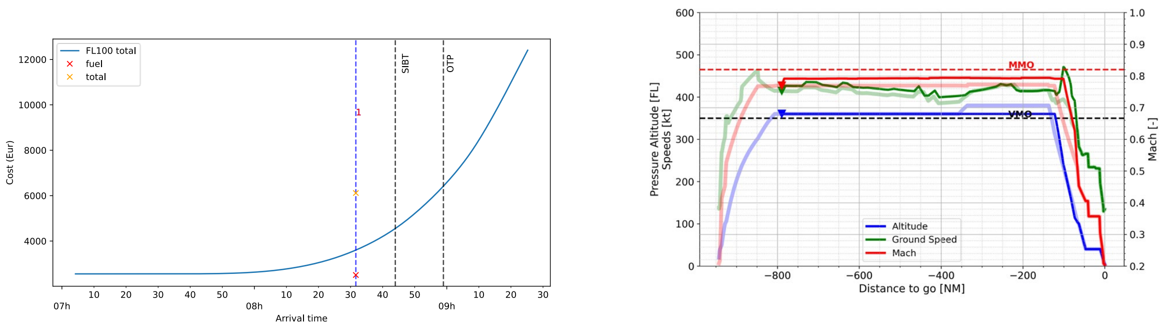


Figure 79 Passenger groups connecting at EDDF into follow up flights on the LEMD — EDDF flight

4.4.3.2.2 Results for CS60 - Heavy head wind weather forecast update

Figure 80 and Figure 81 show, respectively, the results for experiments 403 and 404. Subfigures (a) show the comparison between the OFP and the Pilot3 total expected costs as function of the arrival time at FL100, while the comparisons of the vertical and speed profiles of these two trajectories are given in sub figures (b).

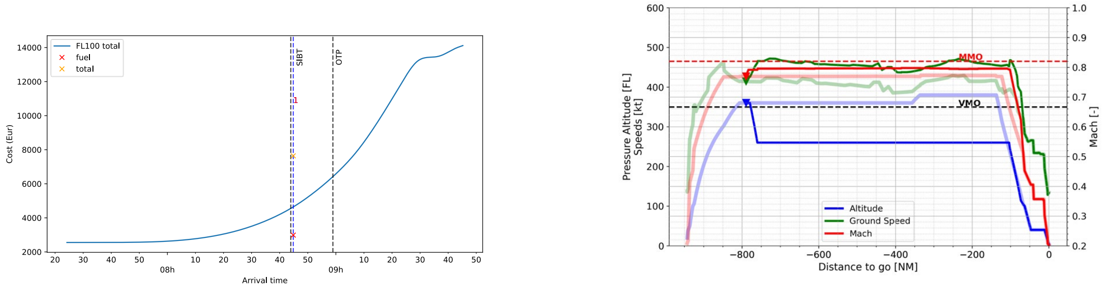
In the case of the first experiment (ID 403), Pilot3 provides a trajectory similar to the OFP (although the cruise Mach is slightly higher and we have a step-climb to FL390 in Pilot3 trajectory) as the strong headwind will exert some adverse effect on flight on-time performance causing a potential delay of the flight, which will be eventually balanced by the earlier arrival at the TOC.



a) Total expected costs as a function of arrival time at FL100

b) vertical profile with speed profile of Pilot3 optimised trajectory

Figure 80 Results of Pilot3 optimisation for Experiment 403 (**heavy head wind** update and **early departure of 20'**)



a) Total expected costs as a function of arrival time at FL100

b) vertical profile with speed profile of Pilot3 optimised trajectory

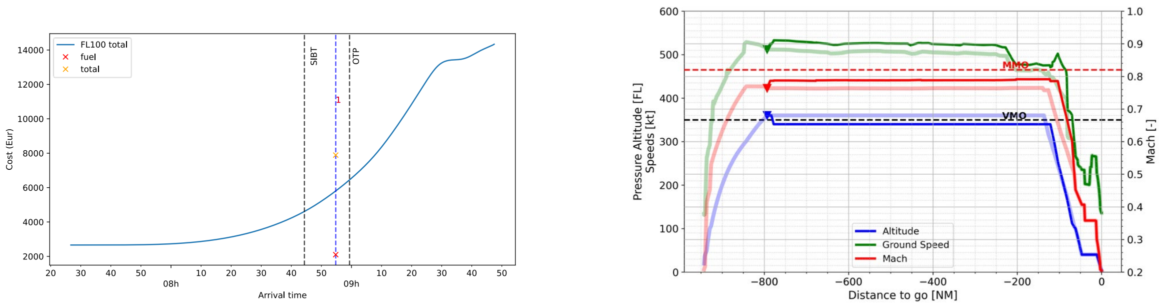
Figure 81 Results of Pilot3 optimisation for Experiment 404 (**heavy head wind** update and **on-time departure**)

In the case of the second experiment (ID 404), in order to compensate the potential delay occurred as a result of unexpected strong headwind, Pilot3 will suggest the pilot to gain additional time by speeding up at a lower altitude (descending from FL360 to FL240). It is worth observing that the resulting trajectory shows a ground speed profile similar to the profile found in the OFP. Hence, the extra headwind has been compensated by a higher true airspeed (and of course at the expense of burning more fuel).

As observed from Figure 80 (a) and Figure 81 (a), for both experiments, Pilot3 is able to provide the trajectories that meet OTP.

4.4.3.2.3 Results for CS60 - Heavy tail wind weather forecast update

In the case of experiments ID 412 and 413, due to the uncertainty (i.e., reactionary delay modelled by machine learning), there is still some cost greater than zero if the flight arrives at the SIBT. For this reason, even if tailwind benefits the flight and even if the arrival time is well before the SIBT (as observed from Figure 82 (a) and Figure 83 (a)) the Pilot3 solution still suggest to reduce trip time. This is achieved by descending to FL340 and slightly increasing the cruise Mach number. Yet, this speed increase is of lower magnitude than for the headwind experiments. However, as observed from Figure 82 and Figure 83, the trajectory profile of both experiments (i.e., ID 412 and 413) follow the same pattern as they speed up to the MMO -minus a safety/operational buffer.

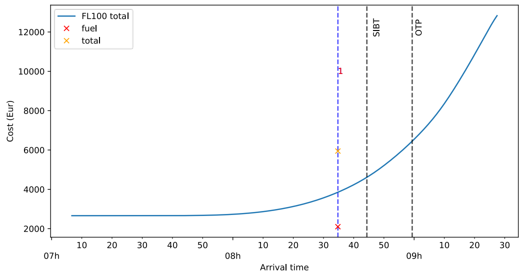


a) Total expected costs as a function of arrival time at FL100

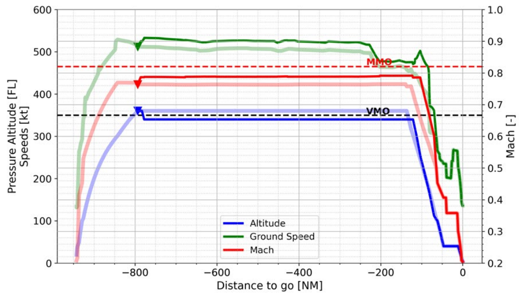
b) vertical profile with speed profile of Pilot3 optimised trajectory

Figure 82 Results of Pilot3 optimisation for Experiment 412 (**heavy tail wind** update and **late departure of 20'**)

EUROPEAN UNION



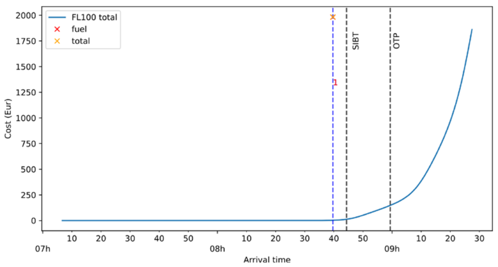
a) Total expected costs as a function of arrival time at FL100



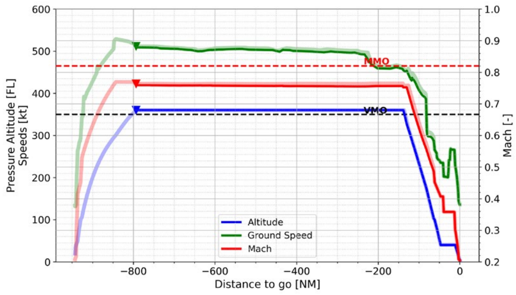
b) vertical profile with speed profile of Pilot3 optimised trajectory

Figure 83 Results of Pilot3 optimisation for Experiment 413 (**heavy tail wind** update and **on-time departure**)

However, the experiment ID 415 allows to observe the behaviour of Pilot3 when the reactionary delay is modelled by heuristics. The experiment ID 415 essentially reproduces the same conditions as in the case of ID 413, but estimating, in the PIE, the reactionary delay costs with heuristics instead of machine learning (see Figure 78 (b)). Due to the flat cost function, Pilot3 will slow down as maximum as possible in order to save fuel (corresponding to CI-0) (see Figure 84), still being able to arrive before the SIBT. Note that OFP was generated by using the CI equal to 10kg/min and this is the reason why there is not so much differences in the vertical profiles of Pilot3 and OFP trajectories.



a) Total expected costs as a function of arrival time at FL100



b) vertical profile with speed profile of Pilot3 optimised trajectory

Figure 84 Results of Pilot3 optimisation for Experiment 415 (tail winds, on-time departure and PIE with heuristics for reactionary delay)

4.4.3.3 Scenario 600: New York (KJFK) to Frankfurt (EDDF)

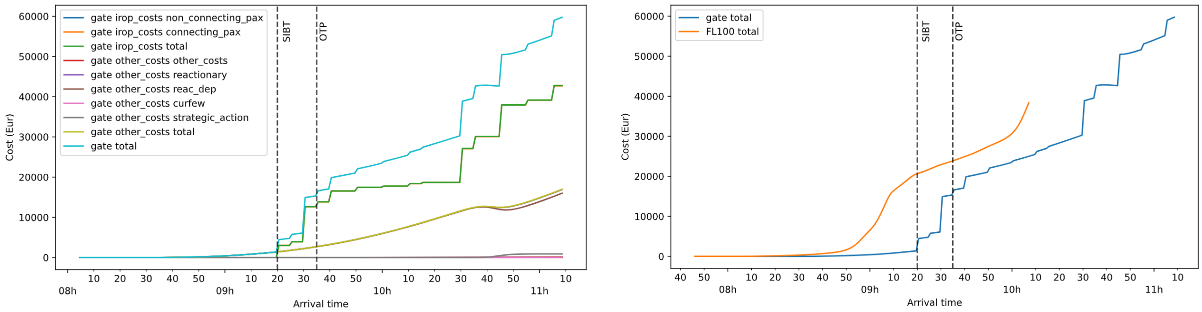
Scenario 600 will be used to assess the benefits of Pilot3 in three particular operational contexts, deemed as very important by the experts from the Advisory Board:

- departure delay of 30 minutes;
- turbulence ahead and no departure delay;
- turbulence ahead and departure delay of 30, 60 and 90 minutes

As a representative of long-haul flights, this scenario is particularly interested not only from the perspective of the length which enables more manoeuvrability along the trajectory, but also from the perspective of the high total cost that may accrue as a result of a great number of passengers that may potentially miss their connections due to delay.

4.4.3.3.1 Cost function

Figure 85 provides the cost function estimated for New York - London route. As observed in subfigure (a), the expected costs at gate is characterised by several jumps mainly stemming from a fact that there are different thresholds when the costs due to Regulation 261 have been materialised as well as due to several groups of passengers that may miss their connections at different time intervals. The costs at FL100 obtained by considering the uncertainty at arrival is presented on subfigure (b).



a) Total expected costs as a function of arrival time at the gate

b) Total expected costs at FL100 considering uncertainty

Figure 85 Cost function modelled without and with uncertainty for New York - Frankfurt route

Table 44: Rotation planned for Madrid – Frankfurt flight

Origin	Destination	SOBT	SIBT	Rotation time available from previous flight
KFJK	EDDF	01h45	09h20	-
EDDF	VIDP	11h45	19h10	2h25

Being an intercontinental flight, there is only one further rotation for the flight on the day (as shown in Table 44) with a planned rotation time of 2h25. This means that the propagation of reactionary delay is relatively low even if the rotation is also relatively short (see 4.4.3.4.1 for the description of the KJFK–EGLL flight where higher turnaround times are allocated). However, as the aircraft is a large aircraft with many passengers, this potential reactionary delay can be very costly. Therefore, as observed in Figure 85 even only after two hours of inbound delay some probability of performing a pre-tactical action to reduce that propagation of delay might be considered by the airline.

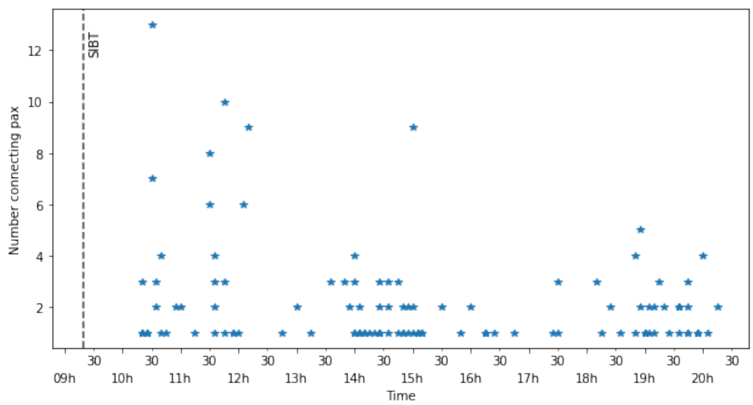


Figure 86 Passenger groups connecting at EDDF into follow up flights on the KJFK — EDDF flight

In any case, as it is considered that passengers are entitled to compensation due to Regulation 261 if the regulation delay thresholds are passed, these passengers' costs are the main driver of the expected

EUROPEAN UNION

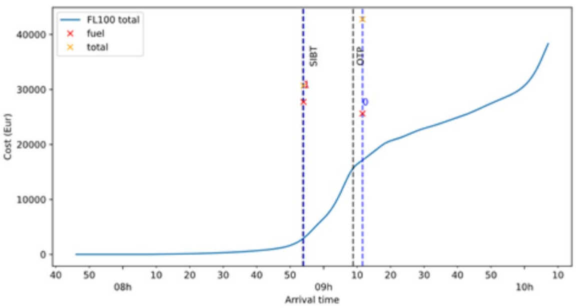
cost of delay. Note that it is not necessary to wait until 180 minutes of delay threshold of inbound delay for the flight to see already significant potential passenger related costs.

As presented in Figure 86, the first groups of passengers with further connecting flights have an outbound SOBT for their connecting flights around 10h20 (as a reminder the SIBT of the KJFK–EDDF flight is 9h20). This means that after considering the minimum connecting time required, if the flight is delayed even by a few minutes, there is a probability that some passengers might miss their connection, and depending on how late their subsequent alternative flight is, they will be entitled to some compensation due to Regulation 261. For this reason, we can observe in the cost function (Figure 85) how even arriving to the inbound gate before the on-time performance threshold some passenger related costs are expected.

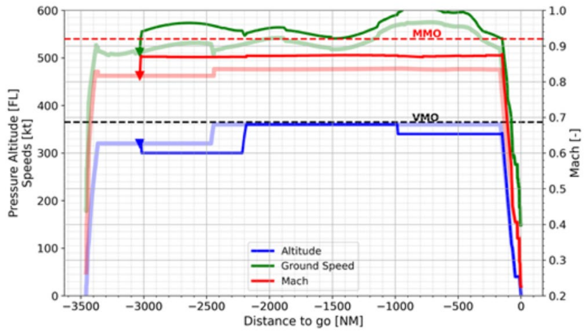
4.4.3.3.2 Results for CS10 - Departure delay, no turbulence

The description of the experiment ID405 for the scenario 600 is given in Table 39 above. As seen, the experiment assumes no changes in weather condition with respect to the OFP. The reason for triggering Pilot3 is that the aircraft has a departure delay of 30 minutes. The aircraft crew queries Pilot3.

Figure 87 shows the results for this experiment. Subfigure (a) shows the comparison between the OFP and the Pilot3 total expected costs as function of the arrival time at FL100, while the comparison of the vertical and speed profiles of these two trajectories is given in sub figure (b). Table 45, in turn, summarises the different KPIs obtained with both trajectories.



a) Total expected costs as a function of arrival time at FL100



b) vertical profile with speed profile of Pilot3 optimised trajectory

Figure 87 Results of Pilot3 optimisation for Experiment 405 (departure delay of 30', no turbulence)

Table 45: Different KPIs for Pilot3 optimised trajectory plan and OFP (from triggering point to FL100)

	Total fuel	Total trip time	Fuel costs	IROPs cost	Other costs
Pilot3 optimised trajectory	53,979 kg	315.85 m	27,729 EUR	1,514 EUR	1444 EUR
Keep flying OFP	49,765 kg	333.5 m	25,622 EUR	14,122 EUR	3,023 EUR
Difference between Pilot3 and OFP	+ 4,214 kg	-18 minutes	+ 2,107 EUR	- 12,608 EUR	- 1,579 EUR

Despite the initial delay of 30 minutes, the Pilot3 optimised trajectory plan is able to meet the OTP and thus, allowing a considerable savings of around EUR 12,600 in comparison to keep flying the OFP trajectory. This amount of savings was mainly generated by savings in IROPs cost as non-meeting the OPT will cause a large number of passengers to miss their connections. As in the case of Athens - London Heathrow route, in order to compensate a large amount of total cost that may accrue as a result of delay, Pilot3 optimised trajectory plan will burn around 4.2 tons of fuel more than the OFP.

As observed in Figure 87 (b), this is achieved by an increase of the cruise Mach number (and therefore the ground speed). This optimisation leads to slight changes in the vertical profile: after Pilot3 is triggered, it is suggested to descent to FL300; then, the step-climb that was planned to FL360 is delayed by approximately 300 NM; and finally, a step-descent to FL340 is performed at the end of the cruise. Recalling Figure 11 in Section 2, higher tailwind is experienced at this lower altitude when the aircraft is at approximately 1000 NM before the destination.

4.4.3.3.3 Results for CS80 - Turbulence ahead

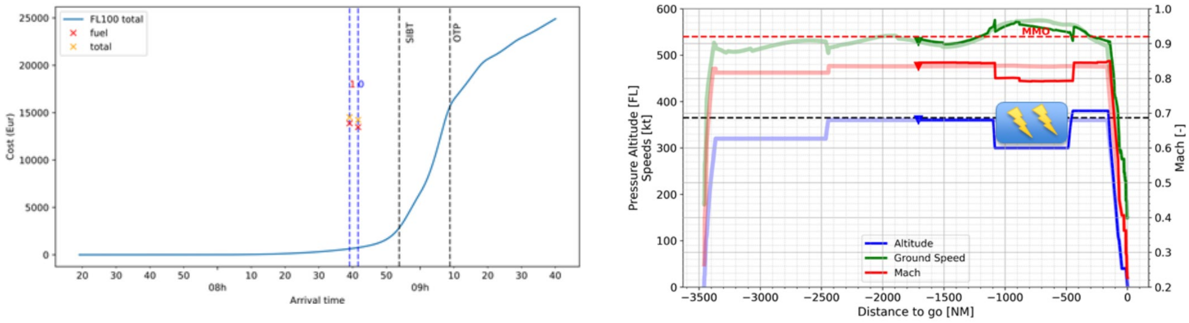
Turbulence along the route is a very common phenomenon that may occur during the execution of long-haul flight and thus, requires a special consideration by the pilot. The turbulence area typically needs to be avoided by the pilot in order to mitigate the discomfort of the passengers. However, it may have an impact on the total costs, as actions such as changing the flight level affect duration of the flight and fuel consumption.

In order to reproduce the behaviour of Pilot3 in case of a turbulence ahead, a csv has been used as an input file restricting an airspace volume. The airspace affected by turbulence has been defined from FL300 to FL400 and with the geographical span as depicted in Figure 88.

Figure 88 Restricted airspace affected by turbulence (i.e., airspace to be avoided by the Pilot3 trajectory)

Figure 89 (b) depicts the speed and vertical profile of Pilot3 optimised trajectory plan in the case of unexpected turbulence ahead and total expected costs as a function of arrival time at FL100. As observed from Figure 89 (a), with only negligible increase in total costs with respect to the "keep flying OFP" strategy (mainly induced by the increase in total fuel cost as a result of flying lower altitudes), the Pilot3 optimised trajectory plan was able to successfully avoid the turbulence area and arrive at destination before the SIBT.

EUROPEAN UNION



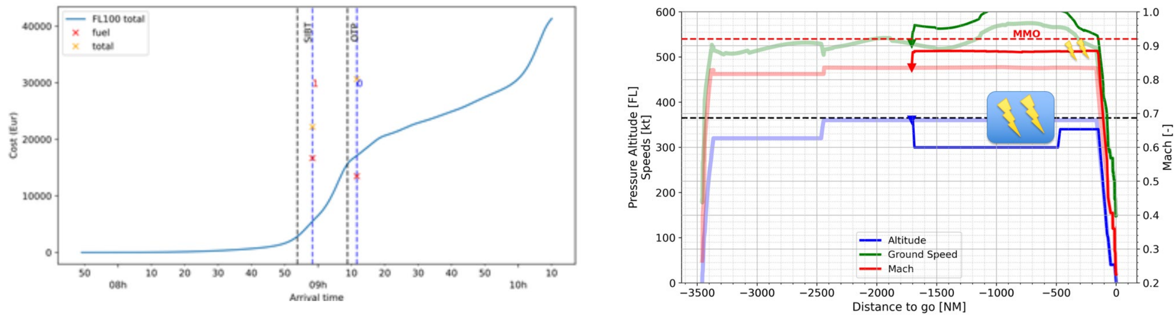
a) Total expected costs as a function of arrival time at FL100

b) vertical profile with speed profile of Pilot3 optimised trajectory

Figure 89 Results of Pilot3 optimisation for Experiment 406 (**turbulence** ahead, **on-time departure**).

The obtained results inspired us to additionally explore the benefits of Pilot3 in the case of turbulence, but imposing a significantly higher amount of delay. The objective of the set of experiments is to explore the behaviour of Pilot3 optimised trajectory plans with respect to OFP when the pilot faces a severe turbulence ahead in addition to certain amount of departure delay.

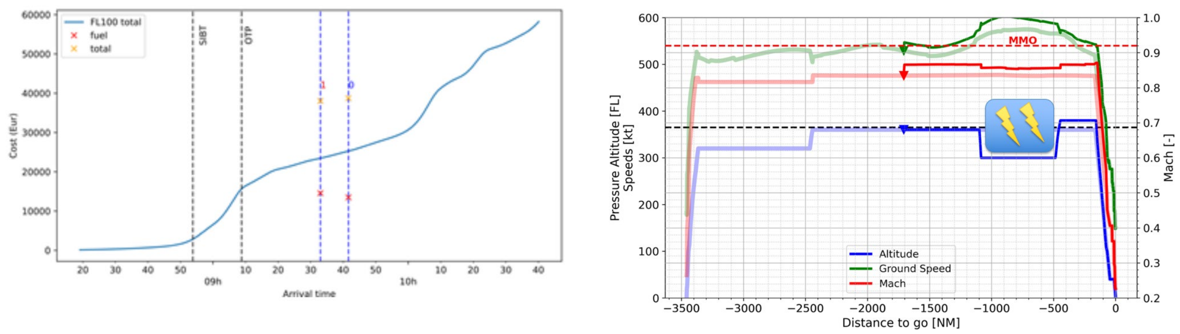
Table 46 summarises the benefits of Pilot3 trajectory plan with respect to "keep flying OFP" across different metrics.



a) Total expected costs as a function of arrival time at FL100

b) vertical profile with speed profile of Pilot3 optimised trajectory

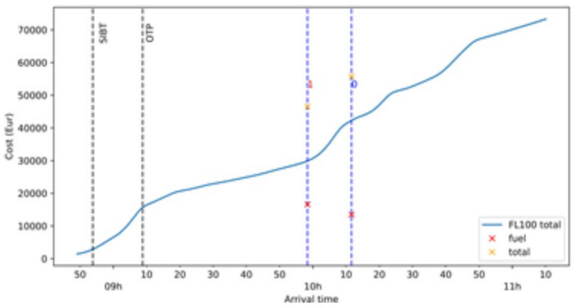
Figure 90 Results of Pilot3 optimization for Experiment 407 (**turbulence** ahead and **30' of departure delay**)



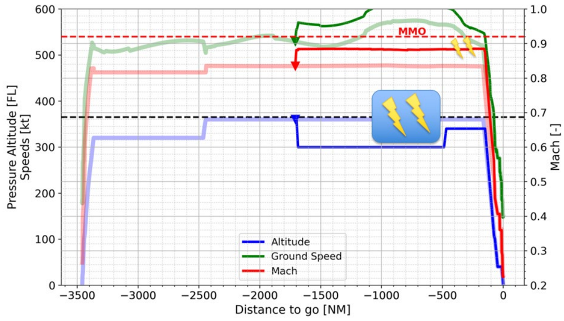
a) Total expected costs as a function of arrival time at FL100

b) vertical profile with speed profile of Pilot3 optimised trajectory

Figure 91 Results of Pilot3 optimization for Experiment 408 (**turbulence** ahead and **60' of departure delay**)



a) Total expected costs as a function of arrival time at FL100



b) vertical profile with speed profile of Pilot3 optimised trajectory

Figure 92 Results of Pilot3 optimization for Experiment 409 (**turbulence** ahead and **90' of departure delay**)

Table 46: Difference between Pilot3 optimised trajectory plan and OFP ((from triggering point to FL100)) across different metrics

	Departure delay of 0 minutes	Departure delay of 30 minutes	Departure delay of 60 minutes	Departure delay of 90 minutes
Difference in fuel consumption	+ 799 kg	+ 6,280 kg	+ 2,162 kg	6,289 kg
Difference in trip time	-3 minutes	-13 minutes	-8 minutes	- 13 minutes
Expected savings	-273 EUR	8,418 EUR	714 EUR	9,193 EUR

In the case when the departure is performed on-time and the turbulence is expected along the flight, the Pilot3 will not suggest to speed up well in advance (by flying at lower speed) the turbulence area, but rather right before it. This is the reason why Pilot3 will generate slightly higher costs than the OFP.

As seen, the benefit of Pilot3 would be particularly pronounced in the case of departure delay of 30 minutes, by offering the trajectory characterised by substantial savings obtained by reduction in IROPs and Other costs. In order to avoid the turbulence area at cruise phase, the Pilot3 trajectory will fly at lower altitudes for longer time than in the case of delay of 60 minutes (see Figure 91 and Figure 91). This action will allow to speed up and absorb some portion of initially assigned delay. With respect to the OFP, the Pilot3 optimised trajectory will arrive on-time at destination enabling the connecting passengers to be transferred to their subsequent flights. Note that absorbing some delay will require additional fuel burn of around 6 tonnes.

However, the cost benefit of Pilot3 will be considerably reduced in the case of delay of 60 minutes. Despite the fact that both OFP and Pilot3 trajectories will not be able to meet the OTP, the Pilot3 optimised trajectory plan will be still able to provide some savings in terms of total cost and arrive at destination 8 minutes before OFP. Moreover, despite the delay is bigger (60 minutes now with respect to 30 minutes before), the Pilot3 does not speed up "that much" because the extra fuel cost does not compensate the savings in delay cost (the cost function if flatter as observed in Figure 91(a)).

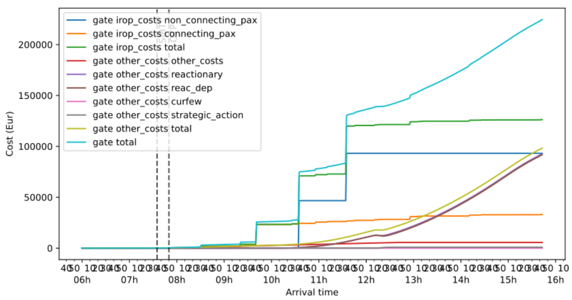
In the case of very large delay of 90 minutes, the output of Pilot3 in terms of speed and vertical profile will be very similar as in the case of delay of 30 minutes (see Figure 92). This implies that once the turbulence area is avoided, the Pilot3 will continue flying at lower altitude in order to gain additional time. By doing this, the Pilot3 will be able to avoid the next jump in the cost function which would significantly increase the IROPs cost.

4.4.3.4 Scenario 800: New York (KJFK) to London Heathrow (EGLL)

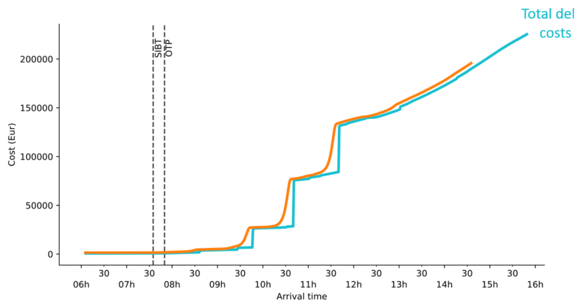
Scenario 800 aimed to demonstrate the effect of uncertainty modelled in the cost function on the final results of Pilot3 and their potential benefit with respect to OFP. For this purpose, two experiments were run - the first experiment encountering the uncertainty modelled in the cost function, whereas the second experiment does not model the uncertainty in the cost function. As already explained, the uncertainty modelled in the cost function mainly stem from the uncertainties inbuilt in OAE and PIE estimators.

4.4.3.4.1 Cost function

The cost function for Scenario 800 (KJFK - EGLL) shown in Figure 93 follows a similar pattern as in case of KJFK to EGLL: as it is assumed that passengers are entitled to compensation due to Regulation 261 these cost dominate the total cost of delay experienced by the airline (see Figure 93). For this particular flight, the following rotation is planned 5h40 minutes after the SIBT (see Table 47). This means that reactionary delay costs do not represent a significant value unless significant amount of delay is experienced (i.e., arrival with 3 hours or more of delay).



a) Total expected costs as a function of arrival time at the gate



b) Total expected costs at FL100 considering uncertainty

Figure 93 Cost function modelled for New York – London Heathrow route

Table 47 Rotations planned for KJFK-EGLL flight

Origin	Destination	SOBT	SIBT	Rotation time available from previous flight
KFJK	EGLL	00h40	07h35	-
EDDF	KSFO	13h15	00h10	5h40

EUROPEAN UNION

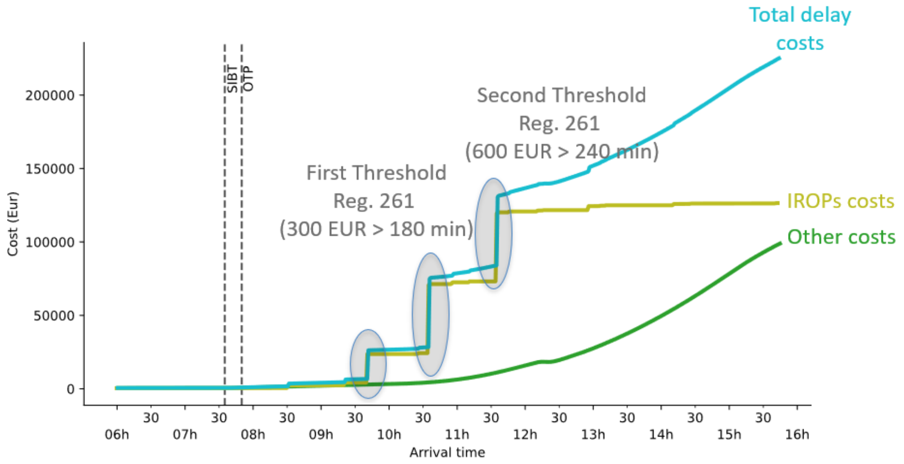


Figure 94 Cost function modelled for New York - London Heathrow route with Regulation 261 thresholds

Figure 94 represents the expected cost of delay at the gate but only focussing on the aggregated total, IROPs and other. As observed, there are three big discontinuities (around 9h41, 10h35 and 11h35). The 10h35 and 11h35 cost increments correspond to Regulation 261 thresholds of 180 minutes and 240 minutes of arrival delay, which as indicated in the Figure represent an increment on the compensation paid per passenger (300 EUR for delays > 180 minutes and 600 EUR for delays > 240 minutes). An analysis of the passenger groups with their connections will be required to understand the cost increment at 9h41.

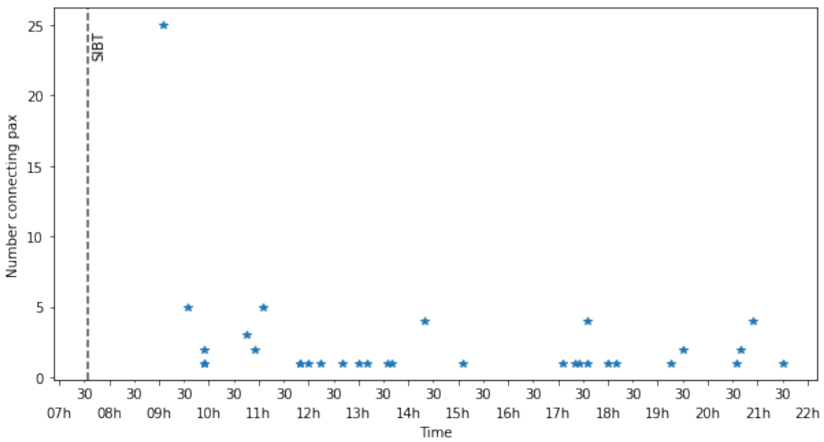


Figure 95 Passenger groups connecting at EGLL into follow up flights on the KJFK — EGLL flight.

Table 48 Rotations planned for KJFK – EGLL

Alternative	SOBT	SIBT	Delay with respect to planned trip	Latest arrival to EGLL to make connection	Expected IROP cost generated
1	9h05	10h40	0	7h41	0
2	11h05	12h35	115	9h41	0
3	13h30	15h00	260	12h01	15000 EUR (600x25)

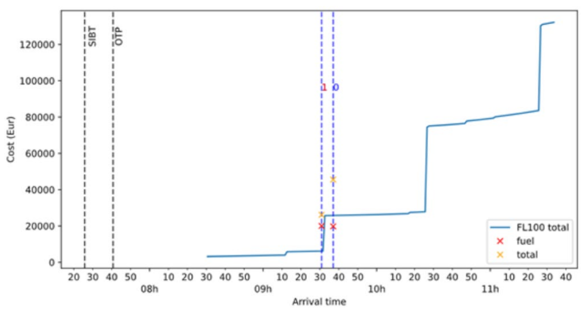
As shown in Figure 95 there is a group of 25 passenger with a follow up connection with a subsequent flight with an SOBT at 9h05. These passengers are connecting to a flight to EGPD (Aberdeen).

Considering that the minimum connecting time for international connections at EGLL is estimated to be 84 minutes, this means that if the flight arrives after 7h41 the passengers will miss their connection. This threshold of delay to miss the connection is earlier than 9h41. However, Table 48 presents the different possible subsequent alternative flights for these passengers. As observed, after 7h41 passengers will have to be reallocated to their second alternative which arrives to EGPD with a total delay with respect to plan of 115 minutes. However, after 9h41 this option will also be missed, and the subsequent available flight won't arrive to Aberdeen up to 15h00 producing an arrival delay of 260 minutes with entitle each passenger to 600 EUR of compensation. This is an example on how the delay experience by the flight can be different that the delay experienced by passengers and therefore have different implications for IROP costs, such as having to compensate for Regulation 261 even if the flight arrives earlier than the regulation delay threshold to the hub: 126 minutes of delay at arrival if arriving at 9h41 can represent 260 minutes for the passengers; on the contrary, if connections are not missed (e.g. passenger groups with connections at 17h) even a late arrival might not have a significant passenger related cost.

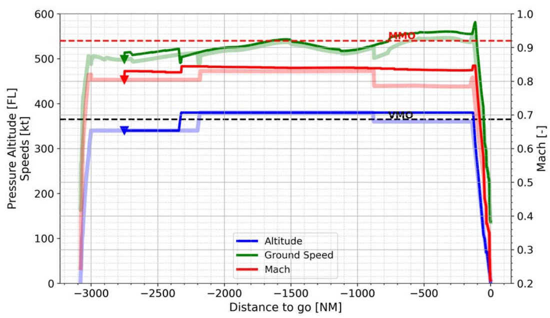
4.4.3.4.2 Results for CS10-Departure delay

Similar as in the case of SCN 600, we imposed a large amount of delay to enable the exploration of the challenging area of the cost function for New York - London route (Figure 96 and Figure 97).

As observed in Table 49, when uncertainty is not considered nominal arrival time of the flight is right is right after the passenger connections are lost. Thus, the Pilot3 is suggesting to save 6 minutes in order to be right on the left side of the jump in the cost function (i.e., no connections lost). In other words, Pilot3 is doing the "minimum” to deterministically save the connections resulting in a large amount of savings. In similar vein, when uncertainty is considered in the modelling of cost function, there is a high probability that a group of passengers will miss their connections. Consequently, Pilot3 will suggest to speed-up to lower the total expected cost (at the expense of higher fuel consumption). Regarding the vertical trajectory profile (see Figure 97 (b)), it can be observed that Mach is significantly increased with respect to the OFP Mach. Additionally, the altitude profile is similar, like before the step-descent at the end of the cruise is not done.

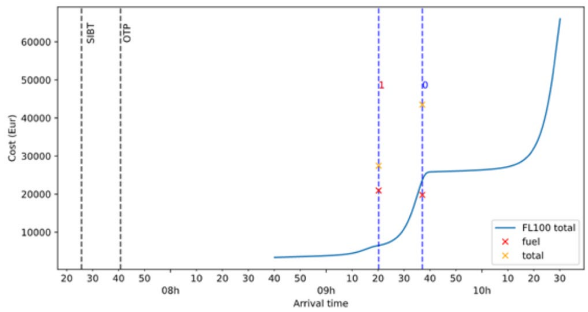


a) Total expected costs as a function of arrival time at FL100

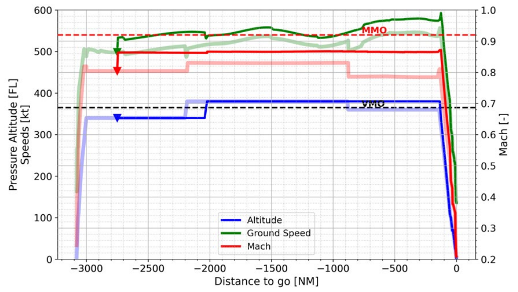


b) Vertical profile with speed profile of Pilot3 optimised trajectory

Figure 96 Results of Pilot3 optimisation for Experiment 410 (153’ of deaparture delay, cost function with no uncertainty)



a) Total expected costs as a function of arrival time at FL100



b) Vertical profile with speed profile of Pilot3 optimised trajectory

Figure 97 Results of Pilot3 optimisation for Experiment 411 (153' of deaparture delay, cost function with uncertainty)

Table 49: Difference between Pilot3 optimised trajectory plan and OFP ((from triggering point to FL100)) across different metrics

	Uncertainty not modelled in the cost function	Uncertainty modelled in the cost function
Difference in fuel consumption	+ 452 kg	+ 2, 314 kg
Difference in trip time	-6 minutes	-17 minutes
Expected savings	19,396 EUR	16,047 EUR

4.4.4 Summary of Research Questions and Hypothesis

After revision of the initially defined RQs and HPs aimed for validation of IVA4, with the previously discussed results for different scenarios, we are able to successfully validate the RQ-IV-070 (see Table 50).

Table 50: : Summary of research questions (RQ) and hypotheses addressed in IVA4

RQ ID	Rationale	Research question	Hypothesis	Success criteria	Status
P3-RQ-IV-070	Validate that the optimised planned trajectory performs equivalent or better than baselines.	For a given triggering event, will the optimised planned 4D trajectory(ies) generated by Pilot3 perform better than the integrated trajectories of the baselines (i.e., operational flight plan, basic pilot behaviour) with the updated information?	For triggering events which could not be foreseen at dispatch level, the pilot will be able to select the most appropriate trajectory from the set of 4D trajectories generated by Pilot3 which execution will provide either some savings in total costs and/or meeting OTP, than the different considered baselines (i.e., following the operational flight plan or basic pilot reaction).	The optimised trajectory plan generated by Pilot3 will contribute **to same or lower total cost** compared to the baselines **with equivalent reach of OTP** (both prioritising and not prioritising reaching OTP).	**Validated**

4.5 IVA7 – Validation of the HMI prototype

The scope of this validation activity was to assess overall accessibility and appropriateness of the HMI tool. The validation activity was conducted among the experts within consortium in order to gather their feedback. In this validation action, however, we collected additional feedback from individuals of the Pilot3 consortium institutions that have not been directly involved in the development and verification of the HMI prototype (besides individuals from the Topic Manager). The HMI prototype was developed by the consortium member, Innaxis. The internal validation of the HMI was performed iteratively during the course of the project with the active participation of all consortium members.

During the internal validation phase, the feedback received from the industrial partner PACE was of particular importance as this company already had a large experience in designing PACElab Flight Profile Optimiser (FPO), a tool of similar functionalities as Pilot3. The discussion between the two partners was conducted in a continues manner during October and November, 2021 by the means of online meetings. Once a first round of HMI design modification took place based on the comments provided by PACE, all other partners actively participated until the final consensus on the design of the HMI was achieved.

4.5.1 Approach

In order to embody the HMI design following the Agile approach, and given the fact that the tool had been continuously refined and modified with the active participation of all consortium members, the initial approach in which we supposed to distribute the questionnaires among the consortium members was abandoned for bilateral meetings and continuous feedback through the shared platform, inGrid.

4.5.2 Results

Table 51 compiles the workshop's outcome lead between PACE and Innaxis, providing the consortium members' feedback on specific aspects of HMI. The second round of feedback, with the final HMI designs resulting to the internal validation, is presented in Appendix B.

Table 51: Internal Validation Workshop's outcome

Slide	Questions	Discussion
1	1. What is the main goal? 1.a. Presenting multiple alternatives? 2.b. Presenting statistical operational values? 1.c. Should the app be continuously open? Should only the best alternative be presented? 2. Selection of flight needed 3. What are the actions for the pilot? 3.a. List FL/speed changes to follow proposed trajectory 4. What should be the main view for pilots?	**1.a.** The main goal is to optimise the trajectory and present the alternatives to the Pilot. As indicated in D1.1 (Pilot3 Consortium, 2020a) and D5.1 (Pilot3 Consortium, 2020c). The pilot will trigger Pilot3 (it could be done automatically if new relevant information is done but the idea is not to have it working in the background continuously). Then, alternatives will be computed based on the optimisation. These alternatives optimise the total cost but consider the sub-cost components (fuel, pax related costs (IROPs) and other (including reactionary and curfew). The alternatives are shown to the pilot who can disregard them (reject them), explore them (for example get more information on the sub-cost components, pax missing connections, OTP, parameters of the trajectory (e.g. flight level changes, speed profile, etc.)), add constraints and re-run if needed. The new trajectories would then be added to the previous and compared again, etc. **1.b.** The minimum results to present are OTP and total cost. Then for total cost the disaggregation by three sub-components. It could be possible go to 'deeper' on the disaggregation (e.g. differentiating between reactionary and curfew cost, cost per type of pax (connecting and not-connecting), fuel divided in flying (up to FL100, sequencing and merging, holding fuel, taxi-in fuel), etc.), also to explore operational parameters probability of curfew, holding, taxi-in time, etc.

EUROPEAN UNION

		1.c. Not only the best alternative is presented as this is a multi-objective problem. A set of trajectories are considered and the VIKOR algorithm rank them automatically. Therefore, the pilot will see the alternatives (including 'do-nothing') already 'ranked'. **2.** Given the maturity level aimed for the HMI, a real trajectory is not needed but rather a clear mock-up of the tool. **3.** Explore the solutions, reject trajectories, add/modify constraints, request re-optimisation. **3.a.** This would be part of exploring the solution (exploring the trajectory) **4.** Current trajectory with indication on OTP and total cost. This should be discussed with them on the External Validation.
2	1. Are the constraints linked to a certain alternative or are the valid for all alternatives? 2. Costs vs planned costs: a reference is missing	**1.** They could be alternative related. Different alternatives could have different constraints. **2.** Planned costs are 'alternative 0', i.e., do-nothing. Keep flying 'as planned' is one alternative for which the expected costs are also estimated.
3	1. What is the constraint? Is the selected FL the only available or not available?	**1.** Not clear in the current version of the HMI if the constraint is to use that FL or to avoid it, agreed. The constraints should be of the type 'do not use these FLs in this region'.
4	1. What is the intention of the estimators? Who should do what with this information? 2. Should the estimation to be considered be selected? For calculation an alternative? 3. Type not needed 4. Are the estimators and confidence intervals valid for all alternatives?	**1.** This should be part of exploring the alternatives. In this case we are providing information on what to expect in terms of operational uncertainty (holdings, sequencing and merging, taxi-in time, etc.). So that the pilot can understand better what to expect and why the optimiser is deciding (or not) to recover delay, for example. **2.** In this case, the uncertainties should be the same for all alternatives as they are linked with operational parameters (e.g. holdings at arrival) and these do not depend on the alternative. We can translate the times into fuel and cost for example, but again these should be the same for all solutions in theory. **3.** Agreed. **4.** We don't have confident interval but the distribution of possible values. So, we can provide expected (average) but also percentiles or even the whole distribution. Also, note that for now we have: • Holding time (and fuel) • Taxi-in time (and fuel) • Sequencing and merging time (and distance and fuel) • Distance in TMA
5	1. Which alternative considers which IROP? 2. What is the link between IROP/estimation and alternative? 3. Why four alternatives? Why not one only? 4. Which alternative brings which benefit. 5. Enable selection of Alternatives in the "Alternatives" view	**1.** Not sure what this means, each alternative might have a different expected arrival time and costs (including IROP costs). **2.** Each alternative has an expected cost, i.e., an expected IROPs cost (including expected cost per connecting/non-connecting pax, etc.). **3.** The alternatives will be already ranked by the system. At least we'll have two alternatives: 'optimised' and 'do-nothing', we can also have more than one (depending on the optimiser) and if the pilot, for example, adds constraints and re-optimise, that new optimised with constraint trajectory will be added. Note that VIKOR ranking algorithm might consider that only one (or x) alternative(s) are better overall and only present those. **5.** Per alternative we should present (easily visible) at least OTP and total cost.
6	1. Selection of flight needed	**1.** Pilot3 is for the 'current' flight on-board, so there are no more than one flight to select as it should be the current. Not sure what it's presented in this screen (or even if we need this), the current time? the SIBT?

4.5.3 Summary of Research Questions and Hypothesis

After revision of the initially defined RQs and HPs aimed for validation of IVA7, with the previously discussed feedback obtained by the experts within the consortium, we successfully validated the RQ-IV-120, 130 and 140 (see Table 52). As already discussed, the validation of the HMI was performed iteratively along the project, hereby abandoning the initial methodology based on the questionnaires. Consequently, the initial success criteria aimed for the validation of specific RQs was adjusted to encounter a new approach based on the informal consensus achieved among the partners.

Table 52: Summary of research questions (RQ) and hypotheses addressed in IVA7

RQ ID	Rationale	Research question	Hypothesis	Success criteria	Status
P3-RQ-IV-120	Validate the simplicity but completeness of the information presented to crew.	Is the information displayed to the pilot clear and easy to understand?	The information presented by the HMI will be simple and, as much as possible, predictable in its presentation, which means that appropriate balance will be found in terms of the amount of information so that the pilot can easily conceive (process) it.	• The partners reach an agreement that Pilot3 provides clear information to the pilot	**Validated**
P3-RQ-IV-130	Validate the facility of the HMI to convey the information computed by Pilot3.	Is the information given to the pilot informative enough and helps to take a more informed decision for a given triggering event?	Human-Machine Interface (HMI) will ensure that the pilot can easily understand the information on high level objectives (e.g. OTP and total costs), but also the information on different PIs and their trade-offs as well as the information on the confidence level provided for each trajectory displayed.	• The partners reach an agreement that Pilot3 aids the pilot to take a more informed decision	**Validated**
P3-RQ-IV-140	Validate the interface to receive input from the crew.	Is the mechanism which allows interaction with the tool acceptable (appropriate) enough from operational point of view?	Human-Machine Interface (HMI) will ensure that the pilot can easily interact with the tool in taking the actions such as rejecting/selecting solutions or based on the information provided, adding new constraints and requesting a re-evaluation of the alternatives in a concise and straightforward manner.	• The partners reach an agreement that the mechanism for the interaction is acceptable enough	**Validated**

4.6 EVA1 – Demonstrations of the HMI prototype and overall capabilities

The main objective of this action is to validate the prototype of the Human Machine Interface (HMI), interacting with the external experts, and to obtain an initial feedback regarding the overall capabilities of the Pilot3 prototype. As discussed in Section 4.5 (IVA7) the design of the HMI prototype and their main functionalities gradually evolved, with a great number of iterations between different partners within consortium. The aim of these iterations was to improve the design of the HMI prototype in order to better reflect the outputs generated by Pilot3 bringing the prototype closer to the operational environment.

4.6.1 Approach

The external validation campaign on the HMI prototype has been initiated during the Final Advisory Board meeting held on 13th January. As a starting action, the main goal is to put all the external experts in the context by introducing them with several important aspects of the tool, such as:

- the general concept of the Pilot3 tool (i.e., "*How is the tool working?*"),
- its specific features (i.e., *"What kind of information does the tool show to the pilot?"*), and
- mechanism implemented to interact with the pilot (i.e., "*How does it interact with the pilot?*").

It is worth mentioning that HMI prototype was presented in the form of mock-ups rather than an interactive dashboard including a number of screenshots for different HMI functionalities. The general feedback obtained during the meeting was that HMI prototype contains all important aspects relevant for the operations and may ease the decision-making process of the pilot. However, the detailed feedback was obtained by the survey sent to the Advisory Board members after the meeting held on 13th January 2022. By the time of closing this deliverable, the consortium received the feedback from two members of the Advisory Board, namely:

- the representative of the large full-service carrier operated at ECAC area
- the representative of an interdependent aviation consultancy company

The results of the survey will be summarised and analysed in the Section below.

4.6.2 Survey results

The respondents were asked to provide their feedback on two specific sets of questions within the survey. The first one referred to the evolution of the Pilot3 HMI prototype in terms of potential outputs that may be displayed to the pilot, while the second part of the survey aimed to obtain the feedback regarding the overall **capabilities of the HMI mock-up** presented during the Advisory Board meeting.

4.6.2.1 Potential information to be presented to the pilot

As already discussed, Pilot3 can generate a large set of outputs and our goal was to **identify the most relevant indicators for the pilot**. Some of these indicators might be available 'directly' to the pilot, while others might be provided 'upon request', or not provided as deemed unnecessary. The respondents were asked to rate the relevance of each of the output to be presented in HMI prototype. The outputs are divided into seven different categories. The results of the survey are provided in Appendix B. Table 53 summarises the main finding of the survey across seven different categories of indicators that may be presented to the pilot.

Table 53: The feedback obtained from the experts by the survey

Categories of indicators	Feedback obtained
Cost related indicators	Both respondents shared an opinion that the pilot should be presented with the information on total fuel cost and IROPs costs as they are deemed important in the decision-making process. Moreover, the respondents claimed that the pilot should not be overwhelmed with the information indicating sub-components and other cost and reactionary cost. Finally, the large discrepancy can be observed for "total cost" indicator, which appeared to be extremely important from the representative of the FSC, and slightly relevant from the point of view of representative from the consultancy firm. It is important to emphasise that the representative of FSC underlined the importance of integration of Pilot3 with the OPS system with clear responsibilities of decision.

EUROPEAN UNION

	In this vein, the sub-components of reactionary cost and other cost would be more relevant for the staff in the airline's operations control centre (AOCC), rather than for the pilot.
Time related indicators	The respondents agreed on the importance of displaying the information on OTP performance of the flight, as one of the upmost relevance for the pilot. In similar vein, the respondents also agreed on the importance of presenting "time at gate" which essentially complements the information on OTP. The large discrepancy among the respondents is observed for "time from FL100 to gate" indicator, which is considered as relevant to be shown by the independent expert and not relevant from FSC's representative point of view. Concerning the time related indicators, the representative from the FSC claimed that "*what is relevant for the pilot is the latest time at which the aircraft must dock in the gate, to avoid any reactionary delay through crew, passengers, aircraft rotation, …etc.*", while all others presented in the table above are *"nice to know but only upon the request of the pilots"*. The overload of information presented to the pilot may cases a safety issue on the long run.
Passenger missed connections indicators	The respondents provided their scores following similar logic as in the case of time-related indicators category. Whereas this indicator is seen as relevant from the perspective of independent aviation expert, the representative of FSC argued that this indicator should not be a first level information, and only available to the pilot upon request (potentially displayed at the pop-up window).
Operational ATM Uncertainties	Both respondents found that three different operational ATM indicators reflecting the uncertainties are, in general, relevant to be displayed to the pilot. however, the representative of FSC considered that this kind of information should be integrated in the OFP, rather than being presented in the HMI.
Other operational parameters	As observed, there is a large gap in the feedback concerning the indicators which indicate the arrival STAR, runway or gate. The representative of FCS found all these three indicator completely irrelevant to be presented to the pilot as they are already included in the OFP and updated if any changes occur. On contrary, the representative of the consultancy firm considered them as extremely relevant.
Other parameters	The experts generally agreed that the information on the "probability of breaching curfew by any subsequent rotation" would be relevant to be presented to the pilot. However, it is worth emphasizing that coordination with OCC must take place as both parties must have a clear definition of their roles and responsibilities. AOCC is typically responsible for the management of the entire fleet and planned flights, whereas pilots are responsible for ensuring a safe flight, if possible, economically and environmentally friendly and taking into account the airline strategic needs.
The full distribution for some indicators	While the representative of the aviation consultancy company found all the indicators listed in the table above as relevant to be presented to the pilot, the representative of the FSC argued that the pilot should not be overloaded with this kind of information. In addition, the AOCCs typically do not have sufficient time resources to train the pilots to understand such graphics nor to check it in a very tight schedule.

Based on the comprehensive feedback obtained from the experts, one would need to be very cautioned about the type of information and its level of details to be presented to the pilot at HMI. As a general remark, the pilot needs to be aware of the consequences which his/her action may have on the network, but ensuring the safe flight needs to remain of his/her utmost concern. It entails that a large burden of possible information may distract him/her from conducting the prime job which is the flying a flight. In this light, the analysis of the survey converges to the conclusion that among all possible indicators that could be generated by Pilot3, the information on the total cost (including fuel cost, IROPs costs and other costs) and OTP should have a significant relevance in decision-making and they are plausible to be displayed. This finding reinforces the feedback received from the experts during the First Advisory Board meeting on relevance of potential trade-offs between OTP and the total cost in the case of unexpected event/disruptions. In addition, some indicators initially intended for pilots may be of a high importance to AOCC as they have an insight into entire fleet and flights and may better understand the consequences of specific actions. In this regard, airlines need to establish the adequate level of coordination between the pilot and AOCC with strictly defined responsibilities and roles between the two parties.

4.6.2.2 Overall capabilities of the HMI mock-up

The respondents were asked to provide their feedback on two main functionalities of HMI mock-up: **Easiness of understanding of the information** and **Interaction with the system**. The survey also contains two final sections specifically designed **for pilots**: **General acceptability** and **Pilot's overall acceptance of the tool**. Each questionnaire contains several statements which were assessed on a 6-point Likert scale (mark with an X) from "Strongly disagree " to "Strongly agree". A summary on the main HMI capabilities with corresponding screenshots can be found in Appendix B.

Table 54: The feedback obtained from the experts by the survey

Questionnaire	Feedback obtained
1-Easiness of understanding the information	The representative of FSC strongly agreed that HMI will ensure that the pilot can easily understand the information on high level objectives (e.g. OTP and total costs), but also the information on different PIs and their trade-offs. Although slightly moderate in assessment in comparison to the FSC representative, the expert from the consultancy company generally agree that the HMI is capable to provide the information which can be easily understood.
2-Interaction with the system	Both experts share the same view that HMI will ensure that the pilot can easily interact with the tool in taking the actions such as rejecting/selecting solutions or based on the information provided, adding new constraints and requesting a re-evaluation of the alternatives in a concise and direct manner.
3- General acceptability – quantity of information provided to pilot (pilot only)	Although the representative of FSC provided a very positive feedback on the general acceptability of HMI prototype, the expert from the aviation consultancy company raised some concerns about the selection of the colours and the size of the font indicating that it needs to be a subject of further refinement.
4- Pilot's overall acceptance of the tool – quantity of information provided to pilot (pilot only)	The respondents generally agreed that the tool will substantially support the pilot to make the final decision on trajectory flown but still keeping him/her actively in the loop.

In addition to the scores given for each of the statements, we also received two additional comments from the representative of aviation consultancy company who is also a pilot:

- *"Size of font has to be revised, captains as myself tender to have 'problems' with small letters and numbers"*
- *Airlines focus basically on 2 factors: money and safety. I think that incorporating safety issues would help us. That is: weather in real time downloaded to the program, we would have a better picture that the weather radar (limited distance), and also others factors: topography.*

In addition, the representative of FSC provided us with the following comment:

- *"Knowing some airlines, size and colour fonts are always discussed extensively with the unions, so you should not target a common font satisfying all users"*

4.6.3 Summary of Research Questions and Hypothesis

The results of the survey together with the feedback obtained during the Final Advisory Board meeting clearly demonstrate that HMI prototype was very-well accepted and in accordance to the operational needs.

Having all these results, we are able to validate some of the RQs and HPs defined in D5.1.

Table 55: : Summary of research questions (RQ) and hypotheses addressed in EVA1

RQ ID	Research Question	Hypothesis	Success criteria	Status
P3-RQ-EV-040	Is the information given to the pilot informative enough and helps to take a more informed decision for a given triggering event?	Human-Machine Interface (HMI) will ensure that the pilot can easily understand the information on high level objectives (e.g. OTP and total costs), but also the information on different PIs and their trade-offs as well as the information on the confidence level provided for each trajectory displayed.	• The majority of the respondents should "agree" that Pilot3 is highly desirable decision support tool • None of the respondents should indicate "strongly disagree" and "disagree" option	**Validated**
P3-RQ-EV-050	Is the mechanism which allows the pilot to interact with the tool acceptable from the operational point of view?	Human-Machine Interface (HMI) will ensure that the pilot can easily interact with the tool in taking the actions such as rejecting/selecting solutions or based on the information provided, adding new constraints and requesting a re-evaluation of the alternatives in a concise and direct manner.	• The majority of the respondents should "agree" that Pilot3 is highly desirable decision support tool • None of the respondents should indicate "strongly disagree" and "disagree" option	**Validated**
P3-RQ-EV-030	Is the information given to the pilot simple (or concise) enough to allow their prompt reaction?	The information presented by the HMI will be simple and, as much as possible, predictable in its presentation, which means that appropriate balance will be found in terms of the amount of information so that the pilot can easily conceive (process) it.	• The majority of the respondents should "agree" that Pilot3 is highly desirable decision support tool • None of the respondents should indicate "strongly disagree" and "disagree" option	**Validated**
P3-RQ-EV-020	Given the overall concept of HMI presented, would the pilot be satisfied to have such decision support tool on-board?	With its user-friendly HMI interface which displays the large amount of information on the trajectories generated and with its interactive capabilities which still keep the pilot actively in the loop, the tool will substantially support the pilot to make the final decision on trajectory flown. Thus, pilots will highly regard having Pilot3 on-board	• The majority of the respondents should "agree" that Pilot3 is highly desirable decision support tool • None of the respondents should indicate "strongly disagree" and "disagree" option	**Partially** **Validated** (*one of the respondents disagreed)

4.7 EVA2 – Results obtained with stand-alone simulations at trajectory level

The purpose of this validation action was to show to the external experts some results obtained in IVA-4 -Pilot3 performance at generation of optimised trajectories plans of the internal validation plan in order to obtain feedback from stakeholders. However, during the development of Pilot3 software, additional Workshop with experts was organised in order to priorities the development of some Pilot3 functionalities that may be more relevant from operational point of view.

4.7.1 External Workshop

The External workshop was organised on 7th July, 2021 gathering experts from different airlines, aviation consultancy companies and EUROCONTROL.

4.7.1.1 Objectives of the meeting

The objective of the meeting was to present the experts the architecture of the Pilot3 given at that time and validate some of the assumptions imposed in the software development. In other words, at this stage, the purpose of the meeting was not to show the results of full Pilot3 scenario, but rather to explain the flow and processes within Pilot3 pipeline.

1. Present Pilot3 with architecture implemented.
2. Present current functionalities: cost function generation, uncertainty at arrival consideration, optimisation with CI.
    1. Cost function considering different cost components
    2. Uncertainties in TMA
    3. Optimisation divided in Trajectory Prediction (TP) and CI optimisation
3. Hypothesis on some modelling activities (e.g. how TP of arrival is performed).
4. Identify/Prioritise future evolution: what to improve.
5. Present the information to the pilot from the results of running Pilot3
6. Explore the HMI on different case studies
7. Gather feedback on what input should the pilot provide to the system. Note that this will link with future development of the optimiser too, e.g. how to add constraints to the optimisation, consideration of multi-criteria, etc.

The meeting was structured in seven main blocks, namely:

A- Introduction to Pilot3

B- The Architecture

1. Construction of cost function as a function of arrival time at gate.
2. Optimisation phases: optimisation up to where pilot 'cannot do anything', i.e., FL100 and TP for arrival.
3. Addition of uncertainties to cost function

C- Cost function

Discuss the components of the cost function. Some specific things to consider:

- Can airlines provide estimators of costs for each of the elementary costs elements we have identified? Should they be grouped?
- For each of the costs some questions, e.g.
    - Transfer costs, how to consider this?
    - Reactionary cost, is it worth it to model it explicitly? What about pax connections downstream, i.e., in subsequent flights? Importance of curfew.
    - Soft costs, should they be disaggregated?
    - Hard cost duty of care when connection takes too long?

EUROPEAN UNION

D- Uncertainties

How uncertainties and procedures which are not optimised are considered (taxi-in, holding, path stretching). Things to consider:

- What next? Focus on en-route uncertainty? focus on TMA uncertainty? focus on more info for the pilot but not currently needed in the model, e.g. runway at arrival, procedure in TMA?

E- Optimisation and Trajectory Prediction

Present current optimisation (based on CI), and TP approach. Things to consider:

- Focusing on the TP
    - Hypothesis, impact of too short distances, etc.
- Focusing on the optimiser
    - Use of CI only as control variable, limitations...
    - Shape of the optimisation, e.g. trade-offs on FL changes, should this be considered?
    - Fuel to be used by optimiser?
    - What next in terms of multi-objective optimisation? Focus on OTP, focus on sub-components of cost function?

F- Machine learning vs heuristic

Some things to consider:

- What is the added value of ML?
- Is it worth it?
- Try to get insight on which features to generate, e.g. which factors might affect each of the processes considered?

G- HMI

What to present to the pilot?

4.7.1.2 Results of the external workshop

The feedback form the External workshop is listed below. As observed from the comments, the experts generally emphasised that modelling the cost function is still one of the most complex tasks in airline operations. Different airlines have different strategies in managing their cost of delay which impose a particular challenge from a modelling perspective.

- The consideration of costs due to Reg. 261 differs from company to company. However, the costs due to Reg. 261 are increasing over the last years. The general remark is that modelling the cost function is very complex task and even dispatchers are not aware of its specification. Although Reg. 261 has not been explicitly modelled in the cost function, some airlines may compute the value of flight for each route taking into account the number of business passengers, connecting passenger, etc. Therefore, the characteristics of the given route will have impact on the application of different actions which may mitigate the effect of delay (ensuring the minimum connecting time for business-oriented routes). In addition, if the flight can save some additional minutes by speeding up in order to avoid the compensation due to Reg. 261, airlines would always opt for this strategy. However, it needs to be underlined that savings in time by speeding up the flight within Europe are limited up to 4 minutes, so very often, the shortcuts given by ATCo could benefit more in the reduction of delay.

- In addition, the experts acknowledged that impact of Reg.261 on cost function would differ not only on a day-by-day basis, but also by flight-to-flight basis. The development of machine learning models to predict the transfer time deemed as promising method to reduce the IROPs cost. In this regard, the machine learning models currently being developed in Dispatche3 project, may ensure to resolve some of these issues at pre-tactical phase of the flight. The particular challenge would be the selection of the adequate features to estimate the transfer time with a high accuracy. In addition, the airlines that operate under large airline group may allocate the passengers with critical connections to the flights of their alliance partners. In the case the airline is a member of airline alliances of joint-venture undertaking, the costs of the re-booking will highly depend on the type of arrangements between the partners.
- The airlines representative claimed that their costs are mainly time-based implying that some parts are credited by flight hours while some are included directly in the cost index. It further implies, that even in the case the aircraft is earlier than planned in OFP, it will not slow down as maintenance and crew costs may still arise. Moreover, it worth emphasising that some of the maintenance costs are cycle-driven (such as landing gear) and clearly depends on the number of take-offs.
- The consideration of curfew in Pilot3 can bring added value only if the pilot is informed about the particular action at the moment in which a proper decision may resolve the issue. However, only awareness of the possibility of hitting a curfew is not the real added value and may put additional pressure on the crew. There is a general consensus among the experts that problem of breaching the curfew should be addressed at dispatch level as they are responsible to make a decision on inserting another plane, or similar.
- The experts generally agreed that there is a great uncertainty associated with taxi-in time at large number of airports in Europe. There are a number of situations which may induce addition taxi-in time: blocked gate at arrival, long de-icing procedure, etc. However, the reduction in taxi-in time may be realised through proactive sequencing with Arrival Management tool (AMAN) or by changing the runway configurations.

4.7.2 Final Advisory Board meeting

The Final Advisory Board meeting was held on 13th January, 2022. Among ten external experts invited to attend this event, six of them eventually participated and provided their feedback on the results of Pilot3, namely:

- three representatives from three different FSCs
- a representative of aviation consultancy company

4.7.2.1 Objectives of the meeting

During the Advisory Board meeting, the experts were introduced with several important aspects of PIlot3:

- New capabilities that were developed in meantime from the Advisory Board meeting held in July, 2021;
- The results of Pilot3 optimised trajectory plan for several relevant scenarios;
- The capabilities of HMI prototype.

EUROPEAN UNION

Regarding the new capabilities developed for Pilot3, the experts had opportunity to have a deep insight into several main aspects of Pilot3, as follows;

- Modelling of **cost function** considering cost components at gate
    - Passenger related costs (connecting and non-connecting)
    - Other costs (including reactionary, curfew, etc.)
- Consideration of **operational uncertainties**
    - Uncertainties in TMA
    - Taxi-in procedures
- Estimators based on **heuristics and machine learning**
    - Possible en-route update of information
- Automatic **ranking of alternatives**

In order to keep this report as concise as possible, we will not show the results on different OAE and PIE estimators and their impact on modelling of cost function, as they are already discussed in details in Section 4.2 (IVA2).

It is worth mentioning that a new approach proposed by Pilot3 leveraging the flight optimisation on the complex cost function instead on the classical "cost index" approach received a very well feedback from the experts. However, the experts emphasised that the building the cost function is a complex task which requires comprehensive approach involving the synchronisation between different departments within an airline.

4.7.2.2 Feedback obtained

During the Advisory Board meeting, four different experiments were presented to the experts together with the corresponding metrics in order to conveniently assess the benefits of Pilot3. Although the number of experiments is small in size, they still very well reflect the potential capabilities of Pilot3 in different operational contexts. The objective of the first experiment (ID 401 from Table 39 in Section 4.4.2) is to demonstrate the benefit of Pilot3 when the aircraft reaches the cruise altitude (top of climb - TOC) earlier than planned. This case study was already identified as a relevant one during the consultation with the experts from the Advisory Board. The three remaining experiments (ID 404, 409 and 410) aimed to show the benefit of Pilot3 in the context of long-haul routes assuming different operational environment (i.e., passengers likely to miss their connections due to large departure delay, uncertainty modelled/not modelled in the cost function). For more details on each of four experiments indicated here, refer to Section 4.4 Validation (IVA4).

After presenting the results to the Advisory Board, they acknowledged that **results are, in general, meaningful and in line with current operational strategies/practice**. In this vein, Pilot3 has a considerable potential to support the pilot in making a proper decision during the flight execution. The experts pointed out the importance of reducing the CO_2 emission as a part "zero net carbon emission" paradigm shift in aviation leveraging on the "green" trajectories with lower fuel consumption. With its capability to be configured in different ways, Pilot3 is able to generate "environmentally friendly" trajectories and in this way, successfully embracing this ambitious goal.

Moreover, the experts emphasised the importance of the Trajectory Based Operations (TBO), as one of the main pillars of SESAR programme which aim to provide "high predictability and accuracy of the trajectory, which allows a seamless process from planning to execution and a seamless process from gate-to-gate". In this regard, TBO will require synchronization (and negotiation) of trajectories which may have direct implications to the uncertainty associated with TMA procedures. For instance, in the case of SCN 100 (Athens-London), if Requested Time of Arrival (RTA) was agreed in London TMA between the aircraft and the ATC, **some amount of TMA holding delay could be effectively transfer to linear holdings** (which is less costly and more fuel efficient). Note that the term "linear holding" is used to define the delay that can be absorbed during the cruise by flying at lower speeds.

Finally, the experts stressed that some of the results indicating the savings in fuel appear to be negligible in the case of long-haul flights. However, with a new version of Pilot3 software which development has been finalised after the Advisory Board meeting, the results for fuel savings have been substantially improved.

4.7.3 Summary of Research Questions and Hypothesis

After revision of the initially defined RQs and HPs aimed for validation of EVA2, with the previously discussed results for a set of four experiments, we were able to successfully validate the RQ-EV-070 and RQ-EV-080 (see Table 56). However, it is worth acknowledging that P3-RQ-EV-070 has been slightly modified with respect to the initial formulation, as Pilot3 was eventually able to generate a single trajectory as an optimisation output, rather than a set of alternatives. However, this complies to the Agile principle adopted in the project.

Table 56: : Summary of research questions (RQ) and hypotheses addressed in EVA2

RQ ID	Rationale	Research question	Hypothesis	Success criteria	Status
P3-RQ-EV-070	Validate that solutions provided are relevant for different experiments.	Is the solution provided by Pilot3 meaningful enough in the case of the given experiment presented?	Pilot3 will efficiently deal with a variety of issues imposed by different operational context that define the particular experiment by providing a meaningful solution.	• The **majority** of the respondents should "agree" that Pilot3 provides meaningful solution in the given operational context • None of the respondents should indicate "disagree" option	**Validated**
P3-RQ-EV-080	Validate overall acceptance of Pilot3 considering performance results for individual trajectories. Identify if improvements required.	Do experts find that Pilot3 worth it for an airline?	Given the benefit provided with respect to different experiments presented, Pilot3 will be worth acquiring by the airlines with different business models.	• The **majority** of the respondents should "agree" that Pilot3 is worth to be acquired by their companies.	**Validated**

EUROPEAN UNION

5 Conclusions and look ahead

This verification and validation report was conducted following the comprehensive plan specified in D5.1 (Pilot3 Consortium, 2020c). In order to efficiently manage different verification and validation activities emerged as a part of software development, the Agile principle adopted at the begging of the project proved to be an extremely useful approach. As already explained, during the course of the project, the trajectory optimiser, as a core of PIlot3, were gradually upgraded until we developed V1.3 of the Pilot3 release, which uses the Cost Index (CI) as a proxy for the optimisation, and then later V2.0, which implements a grid search of speed and altitude and embeds the expected total cost function within the optimiser. For each software releases (V1.3 and V2.0), a batch of the classical verification activities including software design technical reviews, code walk-through reviews, unit and interfaces testing, integration testing and functional testing were performed. In addition, system testing was conducted prior to both software releases in order to ensure that the requirement defined for Pilot3 were satisfied. The outcome of these activities demonstrate that the key of a successful verification and validation is to trace a good plan that foresees the resources and capacities of the project and traces all the preliminary work in an unambiguous approach. A good effort was well invested in the beginning of the project, and has been the baseline for both activities execution and reporting.

As defined in D5.1, internal validation activities aimed at addressing three objectives: to **validate the functionalities of the components** (IVA1, IVA2 and IVA3) of Pilot3, to **evaluate the operational benefits** (IVA4, IVA5 and IVA6) of the prototype and to **assess overall accessibility and appropriateness of the HMI** tool (IVA7). All validation activities aimed to validate the functionalities of the components were successfully accomplished during the validation campaign. The outcome of these activities is briefly summarised below:

- **IVA1,** which aimed to compare the results of the Pilot3 trajectory optimisation engine (i.e., Dynamo) and PACE FPO solution was carried out for two A320-231 flights in two different routes and employing the V1.3 release of Pilot3 software. The results provide a clear evidence that the optimisation engine, DYNAMO, as one of the main modules of Pilot3, was capable of providing trustworthy results. Moreover, the vertical and speed profile of the two trajectories did not differ substantially indicating the trajectories provided by Pilot3 are also meaningful from a real operational perspective.

- **IVA2** was performed independently of the Pilot3 optimisation framework focussing on improving the estimation of performance indicators which will be used to compute the objective functions, and the estimation of Operational ATM Estimator, which aimed at predicting operational uncertainties in the trajectory (e.g. arrival holdings). The evaluation of the model results is based on a set of standard metrics employed for different class of machine learning and heuristics models deployed for different indicators. In addition, the output of the predictive models was also assessed by the researcher experts from the consortium to ensure that the development is moving in the "right direction".

EUROPEAN UNION

- **The outcome of IVA3** revealed that the airline flight policy has an impact on the trajectories and their filtered and ranking. The performance assessment module incorporates the VIKOR algorithm and has been successfully validated by using a realistic example.

The second set of internal validation activities (IVA4, IVA5 and IVA6) aimed to quantify the operational benefits of Pilot3 in order to understand if it met the project's objectives involving the members of the consortium and the Topic Manager. The results of the associated activities are given below:

- **The outcome of IVA4** showed the added value of using the Pilot3 prototype in comparison to a baseline trajectory plan: executing the OFP regardless of the event that triggers Pilot3 (i.e., *do nothing*)). Following the Agile principle which allowed for the flexibility in the specification of different experiments along the project, the benefits of the Pilot3 optimised trajectory plan have been eventually validated on four different scenarios among the nine initially identified in D5.1. The results for some of the experiments provided an ample evidence of operation benefits of Pilot3 tool with respect to different time- and fuel-based metrics. In addition, Pilot3 tool demonstrated a great capability to cope with the change in meteorological conditions as one of the main causes of large disruption in the European network. In this way, the research questions addressing the benefit of Pilot3 with respect to the baseline could be successfully validated by adopting corresponding hypothesis.

Finally, the validation activities which aimed to assess the benefit of HMI prototype (IVA7) were performed in a continuous manner across the lifespan of the project involving all consortium partners. The results of the validation activities are given as follows:

- **The results of IVA7** indicated that the consortium members converged to the final consensus on the HMI design claiming that the current design cover most of the initially defined requirements. Note that the consortium partner PACE provided valuable feedback on the specification of different HMI functionalities as this company already owned a large experience in designing the tool of similar characteristics already exploited by European airlines.

As observed, among the seven initially identified internal validation actions, IVA5 and IVA6 failed to be executed by the end of the project. The main reason for the exemption of IVA5 stems from the fact that it requires further development of a stand-alone trajectory simulator that can "execute" the different trajectory plans (the Pilot3 solution and the baselines) and take into account realised uncertainty. Similarly, the execution of IVA6 would seek for the development of a fast-time simulator able to capture the impact of Pilot3 with system-wide metrics and integration of the Pilot3 prototype into this tool. Consequently, these two validation activities blocked the performance of some external activities, namely a part of EVA2 and EVA3 as, given by the definition, they were tightly related to the outcomes of IVA4 and IVA5, respectively.

However, throughout the whole duration of the project, the consortium members maintained a close interaction with the experts from the Advisory Board and the Topic Manager which facilitated the performance of two external validation actions (EVA1 and EVA3). The main outcome of these two activities is provided below:

- **EVA1** aimed to demonstrate the HMI prototype and the overall capabilities of Pilot3. In addition, the experts were asked to provide their feedback on the relevance of potential sets of the output generated by Pilot3. The HMI was well accepted by the experts from the Advisory Board confirming that HMI meets their expectations in terms of the interaction mechanism as well as its specific features. The experts emphasise the fact that the pilot should be presented

EUROPEAN UNION

with the essential information (e.g. OTP and total costs), but not to be overwhelmed with the information indicating sub-components and other cost and reactionary cost.

- **The main objective of EVA2** was to present the results obtained in IVA4 -Pilot3 performance at generation of optimised trajectories plans during the Final Advisory Board meeting. The feedback obtained from the experts during the Final Advisory Board meeting indicates that results are, in general, meaningful and in line with current operational strategies. Due to its capability to be configured in different way, the external experts acknowledged that Pilot3 tool can be of particular importance in the context of "net zero carbon emission" commitment which aims to dramatically reduce environmental impact of aviation. In addition, as a part of EVA2, the External workshop were organised with the aim to present the experts the architecture of the Pilot3 given at the given stage of the project and validate some of the assumptions imposed in the software development.

As seen from Table 57, with all activities that were performed during the validation campaign, we succeeded to **successfully validate 8 research questions aimed for the internal validation** and **6 research questions defined for external validation**, out of 14 and 10 initially designed in each of the two groups respectively. It is worth emphasising that the remaining RQs that have not been validated mainly stem from IVA5 and IVA6 validation actions which require additional development of the tool. Nevertheless, the successfully validated RQs proves that the results obtained by the Pilot3 tool met the expectations defined at the beginning of the project and highlight future lines of research with the exploration of the Pareto front (P3-RQ-IV-40), automatically providing a set of equivalent alternatives (P3-RQ-IV-50), validation and further development of consideration of uncertainty (IVA5) and the analysis of cost-benefit of a tool such as Pilot3 which could be derived from the activities defined in IVA6 and EVA3.

Table 57: Validation results

Activity	ID	Research question (RQ)	Status
IVA1	**P3-RQ-IV-10**	Are trajectories computed by the trajectory generator of Pilot3 realistic enough?	**Validated**
IVA2	**P3-RQ-IV-20**	Will Pilot3 enhance the estimation of the (K)PIs relevant to the airline?	**Validated**
	P3-RQ-IV-30	Will Pilot3 enhance the estimation of operational uncertainty parameters?	**Validated**
IVA3	**P3-RQ-IV-40**	For a given triggering event, will Pilot3 generate a meaningful set of alternative 4D trajectories when trade-off between objectives is present?	**Partially validated**
	P3-RQ-IV-50	For a given triggering event, will Pilot3 generate a meaningful set of alternative equivalent 4D trajectories?	**Not validated**
	P3-RQ-IV-60	For a given triggering event, will Pilot3 show different 4D trajectories for different airline policies configured in the tool?	**Validated**
IVA4	**P3-RQ-IV-70**	For a given triggering event, will the optimised planned 4D trajectory(ies) generated by Pilot3 perform better than the integrated trajectories of the baselines (i.e., operational flight plan, basic pilot behaviour) with the updated information?	**Validated**
IVA5	**P3-RQ-IV-80**	For a given triggering event, will the realised (executed) 4D trajectory(ies) generated by Pilot3 perform better than the realised trajectory of the baselines (i.e., operational flight plan, basic pilot behaviour) followed regardless of the different triggering events that might arise in flight considering the instantiation of uncertainty in the simulation?	**Not validated**
	P3-RQ-IV-90	For a given triggering event, will the advanced estimation of PI and operational ATM estimators provide more reliable outcomes?	**Not validated**
IVA6	**P3-RQ-IV-100**	Will Pilot3 show a benefit at network-wide level at the end of a day of operations with respect to airlines operational KPIs (cost, % of flights reaching OTP)?	**Not validated**

	P3-RQ-IV-110	Will Pilot3 show a benefit at network-wide level at the end of a day of operations with respect passengers' indicators (passenger delay and missed connections)?	**Not validated**
IVA7	**P3-RQ-IV-120**	Is the information displayed to the pilot clear and easy to understand?	**Validated**
	P3-RQ-IV-130	Is the information given to the pilot informative enough and helps to take a more informed decision for a given triggering event?	**Validated**
	P3-RQ-IV-140	Is the mechanism which allows interaction with the tool acceptable (appropriate) enough from operational point of view?	**Validated**
EVA1	**P3-RQ-EV-010**	From a very general point of view and based on the visual representation and information displayed by HMI, do experts find Pilot3 as a tool which is worth (or useful) having onboard?	**Not validated**
	P3-RQ-EV-020	Given the overall concept of HMI presented, would the pilot be satisfied to have such decision support tool on-board?	**Partially validated (*)**
	P3-RQ-EV-030	Is the information given to the pilot simple (or concise) enough to allow their prompt reaction?	**Validated**
	P3-RQ-EV-040	Is the information given to the pilot informative enough and helps to take a more informed decision for a given triggering event?	**Validated**
	P3-RQ-EV-050	Is the mechanism which allows the pilot to interact with the tool acceptable from the operational point of view?	**Validated**
	P3-RQ-EV-060	Is the information presented to capture the uncertainty on the planned trajectory considered adequate by the crew?	**Not validated**
EVA2	**P3-RQ-EV-070**	Are the solutions provided by Pilot3 meaningful enough in the case of the given experiment presented?	**Validated**
	P3-RQ-EV-080	Do experts find that Pilot3 worth it for an airline?	**Validated**
EVA3	**P3-RQ-EV-090**	Are benefit obtained by Pilot3 at network level relevant to airlines and passengers?	**Not validated**
	P3-RQ-EV-100	Do experts find that Pilot3 will provide benefits to airlines and passengers?	**Not validated**

(*) one of the respondents disagreed

Some of the validation activities were based on the very advanced functionalities of Pilot3 (e.g. a stand-alone trajectory simulator able to consider realised uncertainty, fast-time simulator, etc.), which development would have exceeded the timeframe of the project. Therefore, it is evident that some of the initially planned activities defined in D5.1 were too ambitious to be realised during the project timeline. However, they will be reported in D6.1 Model evolution and uptake serving as guidelines for some future work with more exhaustive verification and validation activities (Pilot3 Consortium, 2022b). Finally, it is worth noting that the interactions with the experts was critical for the final success of the project. These interactions were ensured by organising a **workshop** and **dedicated validation activities** (e.g. online site visits) which supported the refining and selection of experiments, and the prioritisation of the development of functionalities, while gaining further information on the airlines policies, operational approach and possibly datasets.

6 References

Boeing, 2019. Airports with noise and emissions restrictions

Brownlee, J., 2019. Probability for machine learning: Discover how to harness uncertainty with Python. Machine Learning Mastery.

Cook A., Tanner G., Cristóbal S. and Zanin M., 2012, Passenger-oriented enhanced metrics, 2nd SESAR Innovation Days

Cook, A. and Tanner, G., 2015. European Airline Delay Cost Reference Values, Updated and Extended Values (Version 4.1), University of Westminster

Dalmau, R., Melgosa, M., Vilardaga, S., & Prats, X., 2018. A fast and flexible aircraft trajectory predictor and optimiser for ATM research applications. Proceedings of the 8th International Congress on Research in Air Transportation (ICRAT). Castelldefels, Catalonia: EUROCONTROL and FAA.

DFS, 2020. AIP Germany, AIRAC-2013 2020-12-03-. Available at: https://aip.dfs.de/basicIFR/2022FEB24/pages/822E59611990348FBC2C37EE5C560066.html (accessed 09/12/2020)

EUROCONTROL, 2015a. DDR2 reference manual for generic users

EUROCONTROL, 2015b. BADA overview - Base of Aircraft Data (BADA) EUROCONTROL's aircraft performance model

European Commission, 2004. Regulation (EC) No 261/2004 of the European Parliament and of the Council of 11 February 2004 establishing common rules on compensation and assistance to passengers in the event of denied boarding and of cancellation or long delay of flights, and repealing Regulation (EEC) No 295/91

FAA, 2016. Federal Aviation Administration. Advisory Circular 00-45, Aviation Weather Services; Technical Report; Federal Aviation Administration: Washington, DC, USA

Gurtner, G., Delgado, L. and Vaput, D. (2021). An agent-based model for air transportation to capture network effects in assessing delay management mechanisms. Transportation Research Part C: Emerging Technologies (113), 103358

NATS, 2021. AIP Publications, AIRAC-2111 2021-11-04. Available at: https://www.aurora.nats.co.uk/htmlAIP/Publications/2021-11-04-AIRAC/html/eAIP/EG-AD-2.EGLL-en-GB.html#EGLL-AD-2.2 (accessed 18/11/2021)

Official Journal of European Union, 2008. COMMISSION REGULATION (EC) No 859/2008 of 20 August 2008 amending Council Regulation (EEC) No 3922/91 as regards common technical requirements and administrative procedures applicable to commercial transportation by aeroplane. L 254/1

Pilot3 Consortium, 2020a (May) Technical resources and problem definition, Tech. Rep. Ed. 01.02, Deliverable D1.1.

Pilot3 Consortium, 2020b. (Sep) Trade-off report on multi-criteria decision-making techniques, Tech. Rep. Ed. 01.01, Deliverable D2.1.

Pilot3 Consortium, 2020c. (Sep) Verification and validation plan, Tech. Rep. Ed. 01.01, Deliverable D5.1.

Pilot3 Consortium, 2022a. (Feb) Crew Assistant Decision model description (final release). Tech. Rep. Ed. 01.00, Deliverable D4.3.

Pilot3 Consortium, 2022b. (Feb) System evolution and uptake. Tech. Rep. Ed. 01.00, Deliverable D6.1.

Schultz, M., Lorenz, S., Schmitz, R. and Delgado, L.. 2018. Weather Impact on Airport Performance Aerospace 5, no. 4: 109

Shapley, L.S., Kuhn, H. and Tucker, A., 1953. Contributions to the theory of games. Annals of Mathematics studies 28, 307–317.

EUROPEAN UNION

7 Acronyms

AG: Alternatives Generator

AHP: Analytic Hierarchy Process

AIP: Aeronautical Information Publication

AMAN: Arrival Manger (instead of current Extended Arrival Manager)

AMAN: Extended Arrival Manager

AOBT: Actual Off-Block Time

AOCC: Airline Operating Control Centre

ATC: Air Traffic Control

ATCo: Air Traffic Control Officer

ATFM: Air Traffic Flow Management

ATH: Athens Airport

ATM: Air Traffic Management

ATOT: Actual Take-Off Time

BADA 4: Base of Aircraft Data version 4

CAS: Calibrated Airspeed

CI: Cost Index

CS2: Clean Sky 2

CS-x: Case Study x

DCO: Direct Cost Operations

DDR2: EUROCONTROL's Demand Data Repository 2

DLH: Lufthansa

DR: Domain Requirement

Dx.x: Deliverable x.x

ECAC: European Civil Aviation Conference

EDDF: Frankfurt Airport

EFB: Electronic Flight Bag

EGBB: Birmingham Airport

EGCC: Manchester Airport

EGLL: London Heathrow

ELDT: Expected Landing Time

ETA: Estimated Time of Arrival

EV: External Validation

EVA: External Validation Action

FL: Flight Level

FPO: Flight Profile Optimiser from Pacelab

FR: Functional Requirement

FRA: Frankfurt Airport

FSC: Full-Service Carrier

GS: Ground Speed

HMI: Human Machine Interface

HP: Hypothesis

IAF: Initial Approach Fix

ILS: Instrument Landing System

INX: Short name of Pilot3 partner: Fundación Instituto de Investigación Innaxis

IROPs: Irregular Operations costs

IV: Internal Validation

IVA: Internal Validation Action

JFK: John F. Kennedy (New York) Airport

JTI: Joint Technology Initiative

JU: Joint Undertaking

KJFK: John F. Kennedy (New York) Airport

KPI: Key Performance Indicator

KSFO: San Francisco International Airport

LAM: Lambourne fix

LCC: Low-Cost Carrier

LEMD: Madrid Airport

LGAV: Athens Airport

LHR: London Heathrow Airport

LIRF: Rome-Fiumicino International Airport

M: Mach

MAD: Madrid Airport

ML: Machine Learning

MMO: Maximum Mach in Operation

NFR: Non-Functional Requirement

NM: Nautical Mile

OAE: Operational ATM Estimator

OCK: Ockham fix

OD: Origin Destination

OEM: Original Equipment Manufacturer

OFP: Operational Flight Plan

OPS: Operations

OTP: On-time Performance

PACE: Short name of Pilot3 partner: PACE Aerospace Engineering and Information Technology GmbH

PAX: Passenger

PI: Performance Indicator

PIE: Performance Indicators Estimator

RBT: Requested Business Trajectory

RQ: Research Question

RTA: Requested Time of Arrival

RWY: Runway

SCN-x: Scenario x

SESAR: Single European Sky ATM Research

SIBT: Schedule In-Block Time

SOBT: Schedule Off-Block Time

STAR: Standard Terminal Arrival Route

SW: Software

SYS: Systems

TBO: trajectory Based Operations

TMA: Terminal Manoeuvring Area

TOC: Top of Climb

TOD: Top of Descend

TP: Trajectory Prediction

TTA: Target Time of Arrival

UAD: Upper Air Data

UoW: Short name of Pilot3 coordinator: University of Westminster

UPC: Short name of Pilot3 partner: Universitat Politècnica de Catalunya

VIDP: Indira Gandhi International Airport (Delhi Airport)

VIKOR: Visekriterijumsko KOmpromisno Rangiranje, a Serbian term for “multi-criteria optimization and compromise solution”

VMO: Maximum CAS in Operation

Vx: Version X

WPx: Workpackage x

Appendix A Weather analysis for SCN 201 (LEMD - EDDF)

This appendix presents the results of statistical analysis performed on the weather data for scenario 201 (LEMD-EDDF).

There are two computed ways for performing the weather analysis along the route:

1. Using coordinated of the GCD (Great Circle Distance) between LEMD – EDDF
2. Using the coordinates of the waypoints of the route described in P3-SCN-201: Madrid (LEMD) - Frankfurt (EDDF)

The results here presented are obtained by using the waypoints coordinates.

The analysis was performed separately for the year 2019 and 2018.

A.1.1 2019 Analysis

For the sake of computational effort, the ERA5 files used for the weather analysis are selected for a given time and a given altitude. The hour was selected according to the time of the first flight of the day performing the route. However, the impact of the hour of the day along the day on the weather forecast was analysed and estimated to be negligible. The additional information on two product types used for the analysis is provided in Figure 98 and Figure 99.

Open request form
Request ID: 600d111e-072b-44b9-b8d3-a451e7c70249

Product type:	Reanalysis
Variable:	Temperature, U-component of wind, V-component of wind
Pressure level:	300 hPa, 350 hPa
Year:	2019
Month:	January, February, March, April, May, June, July, August, September, October, November, December
Day:	01, 02, 03, 04, 05, 06, 07, 08, 09, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31
Time:	05:00
Sub-region extraction:	North 55°, West -10°, South 35°, East 10°
Format:	GRIB

Figure 98 Description of Reanalysis product type (2019 dataset)

Open request form
Request ID: ae5176d7-b7de-4ef7-a69b-1af5599606eb

Product type:	Ensemble mean
Variable:	Temperature, U-component of wind, V-component of wind
Pressure level:	300 hPa, 350 hPa
Year:	2019
Month:	January, February, March, April, May, June, July, August, September, October, November, December
Day:	01, 02, 03, 04, 05, 06, 07, 08, 09, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31
Time:	06:00
Sub-region extraction:	North 55°, West -10°, South 35°, East 10°
Format:	GRIB

Figure 99 Description of Ensemble means product type (2019 dataset)

Having all the dataset in place, the analysis was performed in order to:

1. Extract the days with the extreme weather conditions,

2. The relation between temperature gradient and projected wind,
3. Extract uncertain conditions,
4. Extracting mild-extreme weather conditions.

The results for each of the four items are provided below.

Extreme weather conditions

By using the reanalysis dataset, it is identified that 21st October 2019 corresponds to the day with the strongest tail wind (see Figure 100), while 10th December 2019 represents the day with the strongest headwind (see Figure 101). The colours represent different days in the different graphics.

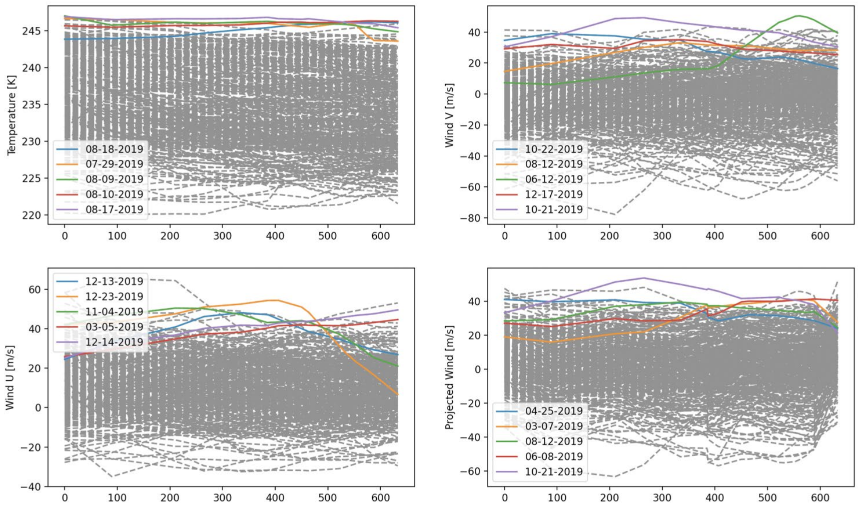


Figure 100 Strongest tail wind using reanalysis (21st October, 2019)

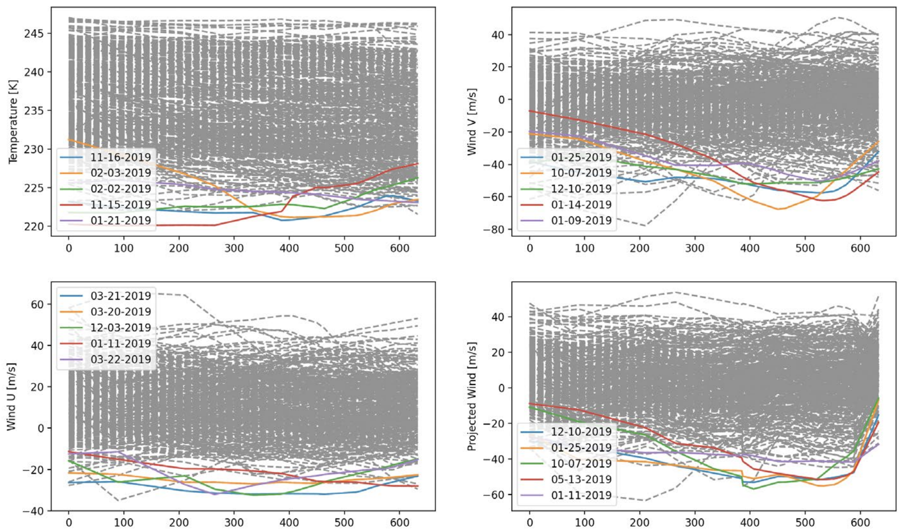


Figure 101 Strongest head wind using reanalysis (10th December, 2019)

The relation between temperature gradient and projected wind

The analysis aimed to determine whether the wind is related to the difference of temperature at the sea level and the temperature at the standard cruise level. The larger the difference between the two temperatures, the greater the head wind generated. The positive projected wind would be an indication for the tail wind, while negative projected wind would correspond to the head wind. The results of the analysis are shown in Figure 102and Figure 103.

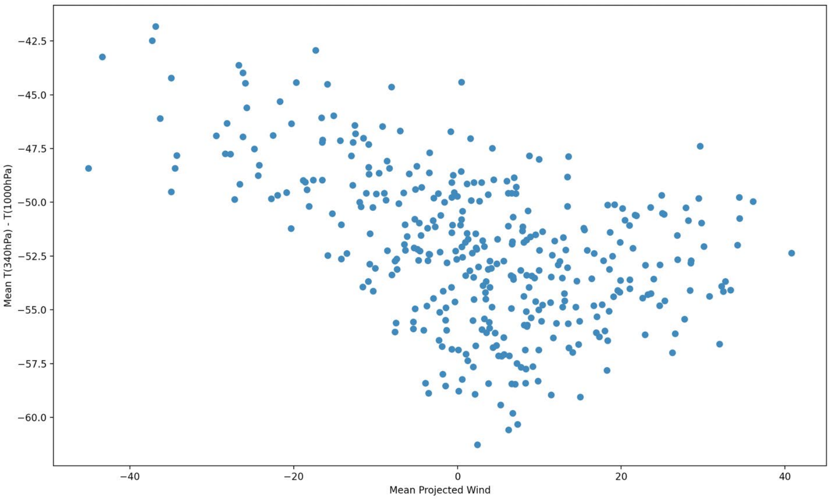


Figure 102 Relation between mean projected wind and mean flight altitude for LEMD to EDDF flights

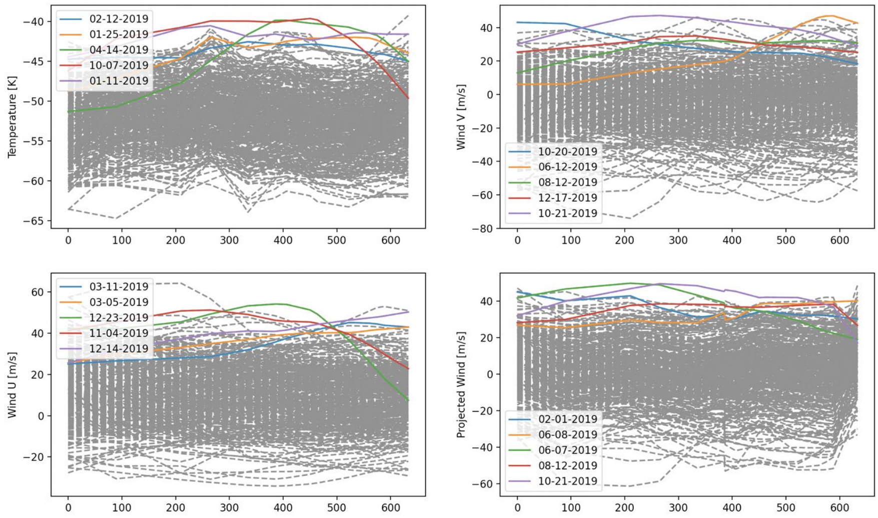


Figure 103 Temperature, wind components and projected wind as a function of distance for LEMD to EDDF flights

Extracting uncertain conditions

The results of the analysis (see Figure 104) indicated the days with greater uncertainties for the weather predictions based on the difference between the forecast at ensemble mean and the reanalysis data.

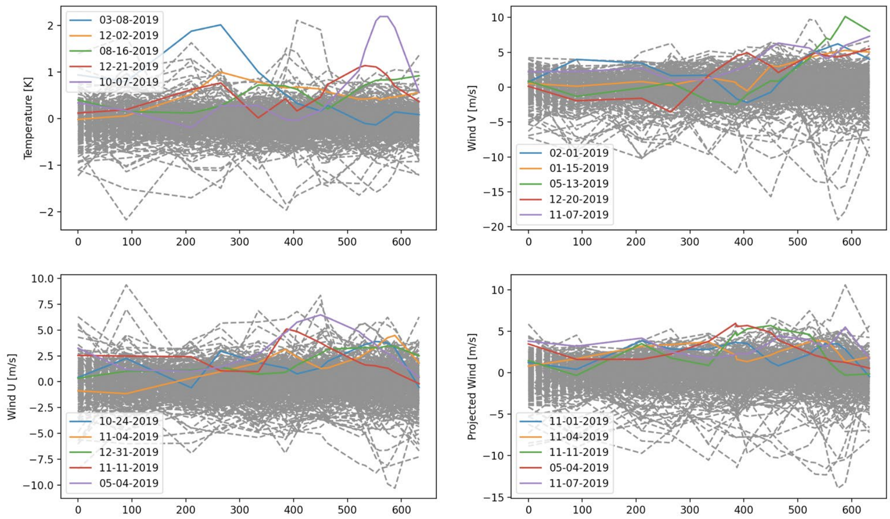


Figure 104 Days with uncertain conditions in 2019

Extracting mild-extreme weather conditions

Mild-extreme weather conditions were obtained by analysing the data from the second column of the histogram (see Figure 105). Figure 106 presents the days having a projected wind belonging to the second column of the histogram.

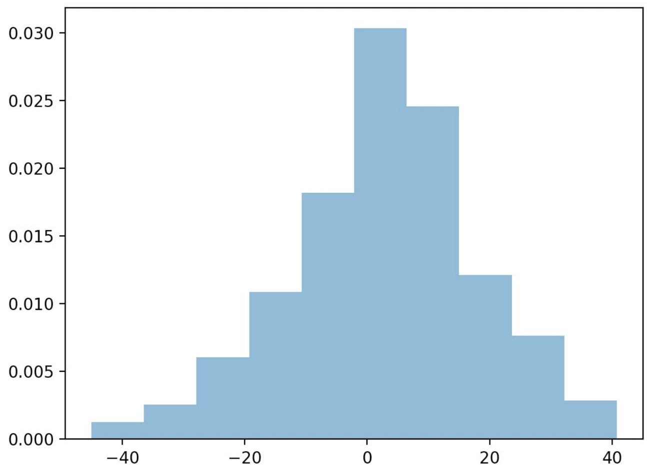
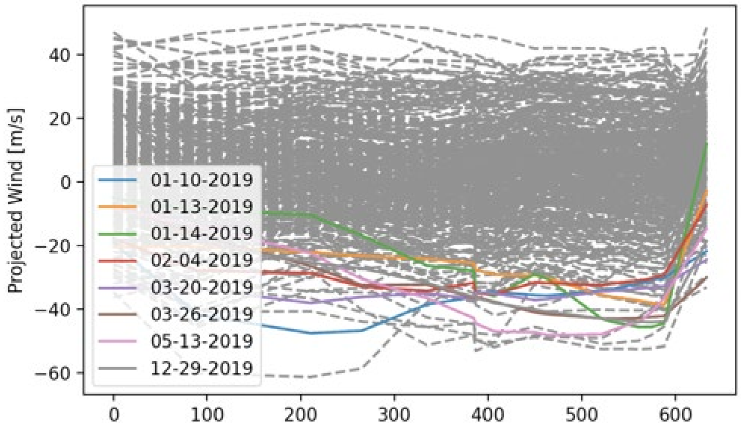


Figure 105 Wind histogram

EUROPEAN UNION



(b)

Figure 106 Days with the projected wind belonging to the second column of the histogram

A.1.2 2018 Analysis

Similar as in the case of 2019, the analysis performed for 2018 is based on the ERA5 files. The brief description of the dataset used is presented in Figure 107 and Figure 108.

Open request form
Request ID: 4f5a3d30-6b5e-45ca-868c-1ee625cecfda

Product type:	Reanalysis
Variable:	Temperature, U-component of wind, V-component of wind
Pressure level:	250 hPa
Year:	2018
Month:	January, February, March, April, May, June, July, August, September, October, November, December
Day:	01, 02, 03, 04, 05, 06, 07, 08, 09, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31
Time:	06:00
Sub-region extraction:	North 55°, West -10°, South 35°, East 10°
Format:	GRIB

Figure 107 Description of Reanalysis product type (2018 dataset)

Open request form
Request ID: bdc8e730-0038-4b15-90eb-dbdb5484274b

Product type:	Ensemble mean
Variable:	Temperature, U-component of wind, V-component of wind
Pressure level:	250 hPa
Year:	2018
Month:	January, February, March, April, May, June, July, August, September, October, November, December
Day:	01, 02, 03, 04, 05, 06, 07, 08, 09, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31
Time:	06:00
Sub-region extraction:	North 55°, West -10°, South 35°, East 10°
Format:	GRIB

Figure 108 Description of Ensemble means product type (2018 dataset)

The objective of this analysis was to identify the days with the following characteristics:

1. High projected wind, low uncertainty and low deviation from the ISA temperature,
2. Low projected wind, low uncertainty and low deviation from the ISA temperature,
3. Average projected wind, average low uncertainty and deviation from the ISA temperature,

4. Average wind and average temperature difference, large uncertainty on head,
5. Average wind and ISA temperature and large error tail (suggested day as maximum wind values).

The results for each of the five items are provided below.

Day with high projected wind, low uncertainty and low deviation from the ISA temperature

The results of the analysis indicated five different days (indicated by different colours in Figure 109) in 2019 characterised by high projected wind for an average temperature difference and weather forecast between 1 and -1. Due to its low fluctuation in the values, it was eventually suggested to select the day **12th November, 2018** as a representative day. Based on the historical meteorological information provided at Weather Underground, it can be concluded that fair conditions were prevailing that day at EDDF. The link is as follows:

- https://www.wunderground.com/history/daily/de/frankfurt/EDDF/date/2018-11-12

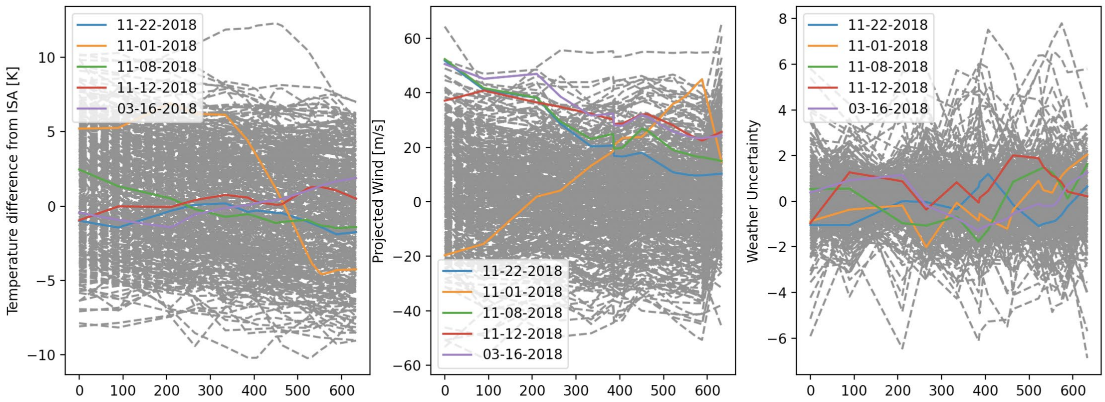


Figure 109 Days with high projected wind, low uncertainty and low deviation from the ISA temperature

Day with low projected wind, low uncertainty and low deviation from the ISA temperature

The results indicated three candidate days with the characteristics of low projected wind, low uncertainty and low deviation from the ISA temperature (see Figure 110). Those are:

- 15th January, 2018 which appeared to be mostly cloudy. The information on the historical weather for that day can be retrieved from the following link:

  https://www.wunderground.com/history/daily/de/frankfurt/EDDF/date/2018-1-15

- 2nd October, 2018 with fair weather conditions and light rain. The information on the historical weather for that day can be retrieved from the following link:

  https://www.wunderground.com/history/daily/de/frankfurt/EDDF/date/2018-10-2

- **17th April, 2018** which was the day with fair conditions. Eventually, this day was selected as the most representative one. The information on the historical weather for that day can be retrieved from the following link:

EUROPEAN UNION

https://www.wunderground.com/history/daily/de/frankfurt/EDDF/date/2018-4-17

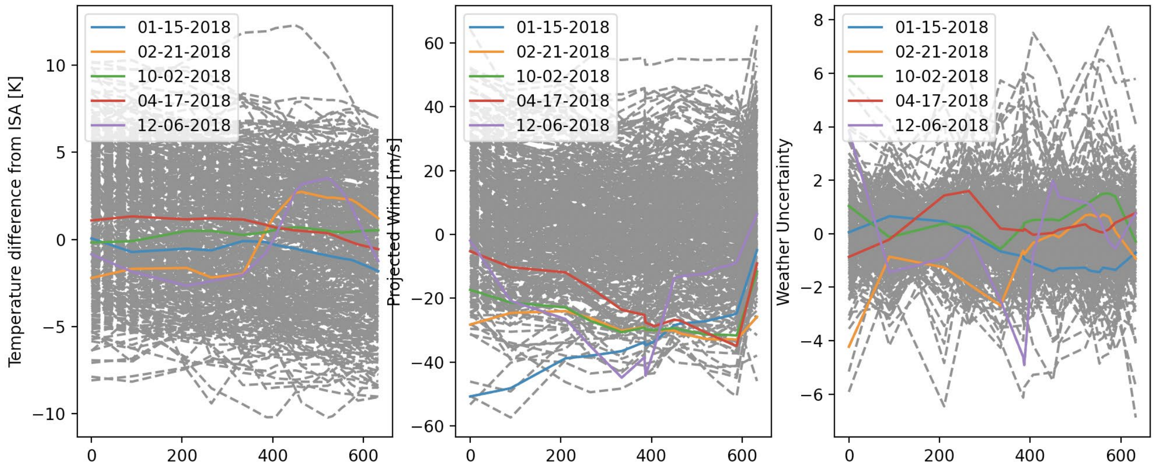


Figure 110 Days with low projected wind, low uncertainty and low deviation from the ISA temperature

Day with average projected wind, average low uncertainty and deviation from the ISA temperature

By analysing the data, **7th February 2018** was selected as a representative day (see Figure 111). The information on the historical weather for that day indicated fair weather conditions and can be retrieved from the following link:

- https://www.wunderground.com/history/daily/de/frankfurt/EDDF/date/2018-2-7

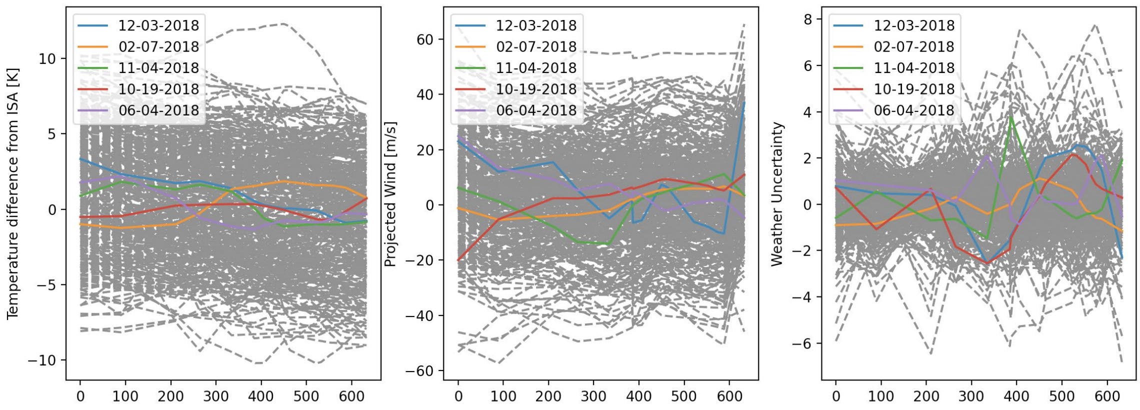


Figure 111 Days with average projected wind, average low uncertainty and deviation from the ISA temperature

Day with average wind and average temperature difference, large uncertainty on head

The results of analysis indicated three candidate days characterised by average wind and average temperature difference as well as large uncertainty on head (see Figure 112). Those are:

- 7th September, 2018 which appeared to be mostly cloudy. The information on the historical weather for that day can be retrieved from the following link:

https://www.wunderground.com/history/daily/de/frankfurt/EDDF/date/2018-9-7

- 8th June, 2018. The information on the historical weather for that day can be retrieved from the following link:

  https://www.wunderground.com/history/daily/de/frankfurt/EDDF/date/2018-6-8

- **12th April, 2018** which appeared to be mostly fair. Eventually, this day was selected as the most representative one:

  https://www.wunderground.com/history/daily/de/frankfurt/EDDF/date/2018-4-12

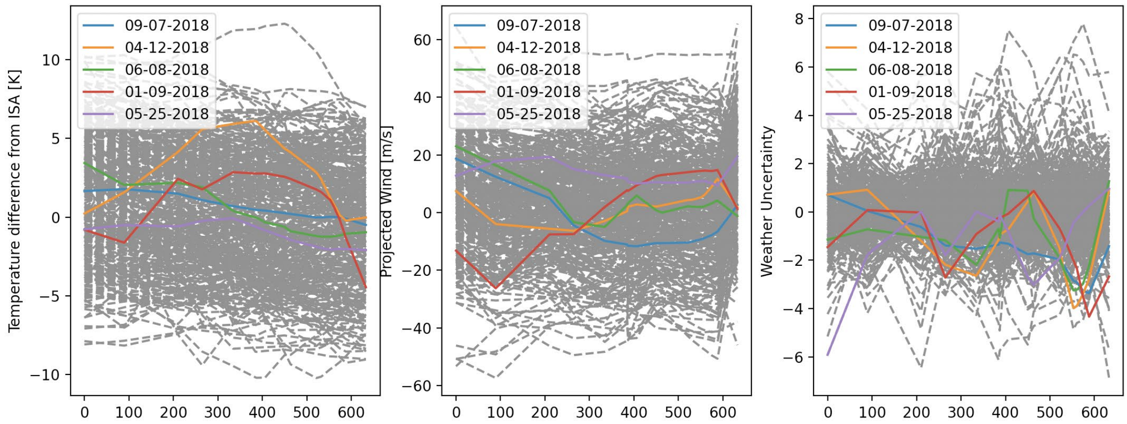


Figure 112 Days with average wind and average temperature difference, large uncertainty on head

Day with average wind and ISA temperature and large uncertainty on tail

By analysing a set of potential day candidates, it can be determined that **10th October, 2018** is most suitable candidate for the day with average wind and ISA temperature and large uncertainty on tail (see Figure 113). The information on the historical weather for that day can be retrieved from the following link: https://www.wunderground.com/history/daily/de/frankfurt/EDDF/date/2018-10-10 (Accessed March 2022).

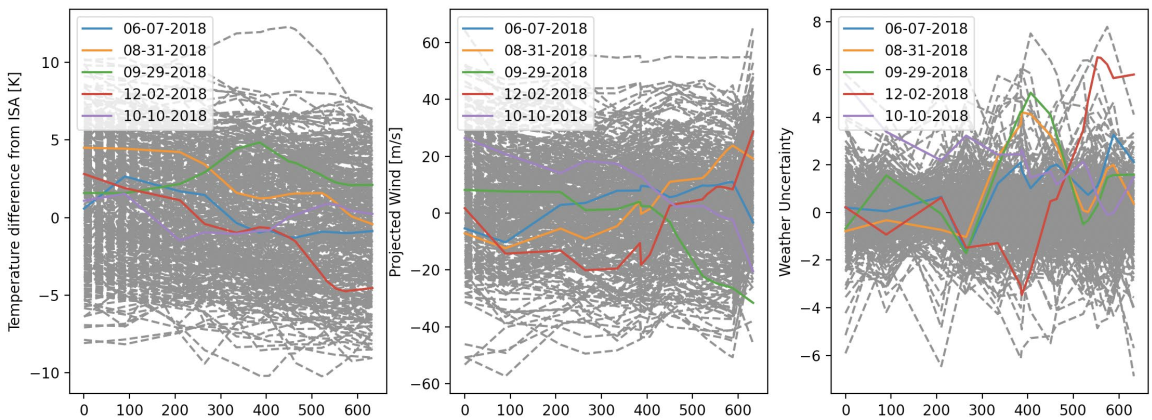


Figure 113 Days with average wind and ISA temperature and large uncertainty on tail

Finally, Table AI.1 below summarises the main results discussed above indicating the most representative day for each examined category.

Table AI.1 Summary and proposed days

Type	Requirements	Proposed Day	Conditions that day at EDDF	Bucket Nomeclature
tail_error	Avg wind and ISA temp and large error tail	**10-10-2018**	Fair	tail_high_error_forecast
tail	Tail wind and ISA temp and forecast error small	**11-12-2018**	Fair	heavy_tail
nominal	Avg wind and ISA temp and forecast error small	**02-07-2018**	Fair	nominal
head_error	Avg wind and ISA temp and large error head	**04-12-2018**	Mostly fair	head_high_error_forecast
head	Head wind and ISA temp and forecast error small	**04-17-2018**	Fair	heavy_head

Appendix B HMI validation and Advisory Board survey

This appendix presents the final version of the HMI designed in Pilot3 and the survey presented to the Advisory Board to gather feedback as part of IVA7 – Validation of the HMI prototype and EVA1 – Live or pseudo-live demonstration of the HMI prototype and overall capabilities..

B.1 Human machine interface design

This section presents the final design of a possible human machine interface for Pilot3 prototype as presented to the Advisory Board.

B.1.1 Main screen – Pilot and current flight information

Figure 114 Pilot profile selector view

Figure 114 shows the first screen of Pilot3 where the crew can select their own personalised profile. Figure 115 contains a depiction of the flight information and current trajectory view presented to crew informing of the main trajectory parameters of keep operating the current trajectory (OFP).

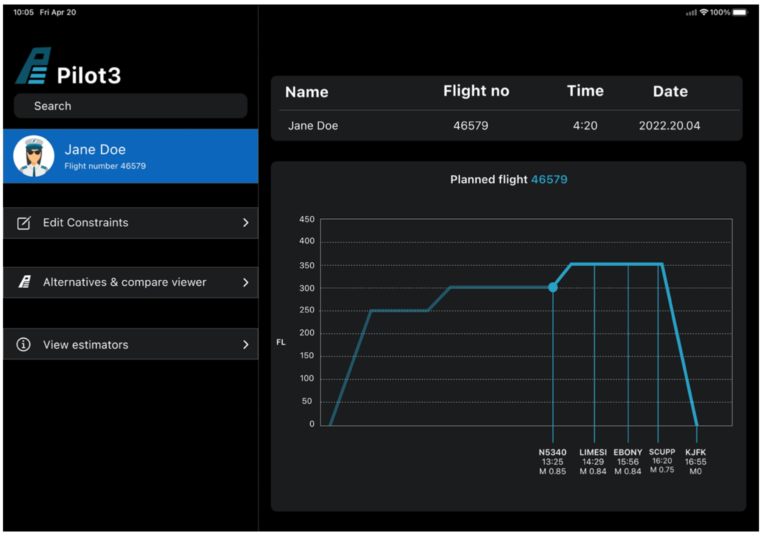


Figure 115 Flight information and trajectory view

B.1.2 Edit constraints functionalities

The crew can add constraints to Pilot3 to be considered during the optimisation. These constraints are divided into flight level, time (not implemented in current version of Pilot3), fuel and cost. Figure 116 shows how the crew can select which of these constraints to edit.

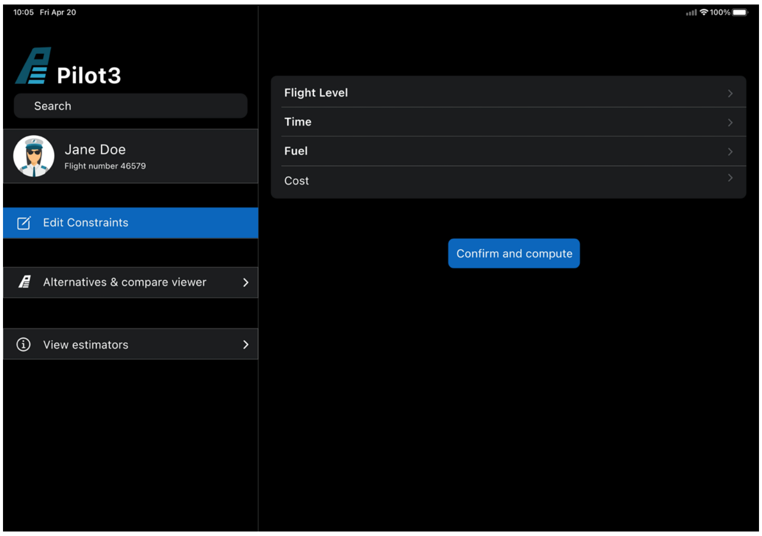
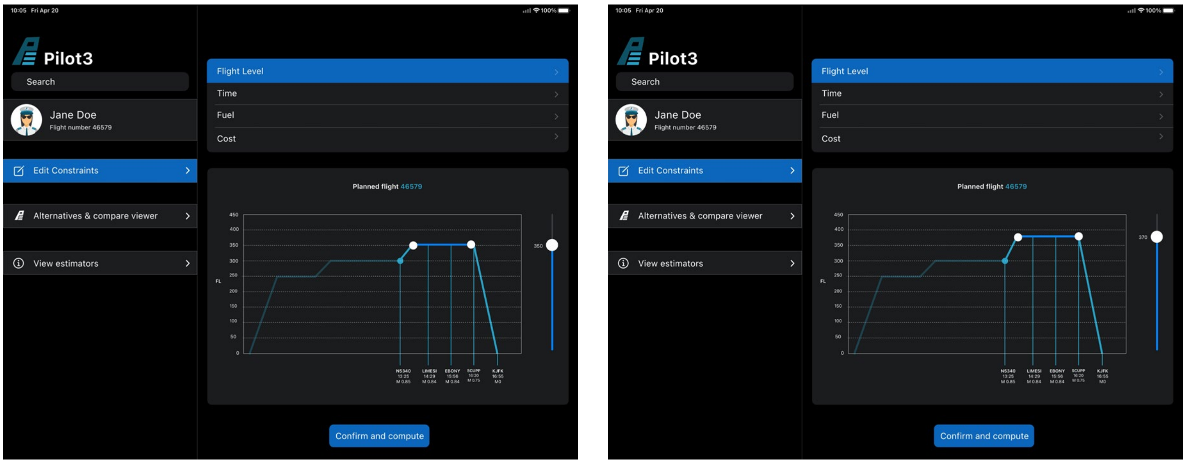


Figure 116 Edit constraints selector



a) Original FL350

b) Selected FL390

Figure 117 Maximum flight level constraint view

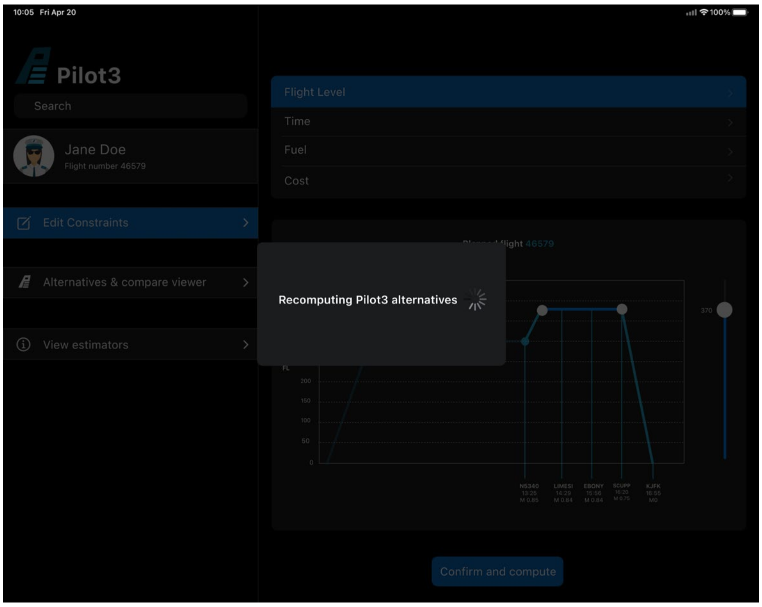


Figure 118 Confirm and compute constraint optimisation

Figure 117 shows an example of adding a flight level constraint (from FL350 to FL390). Figure 118 presents how the constraint is accepted and the optimisation triggered. Finally, Figure 119 presents the optimised trajectory.

EUROPEAN UNION

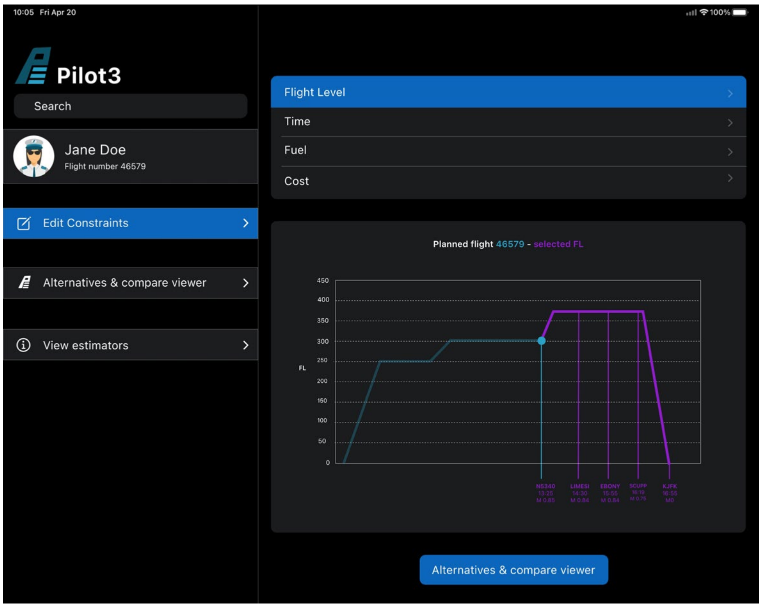


Figure 119 Modified flight plan with constraint

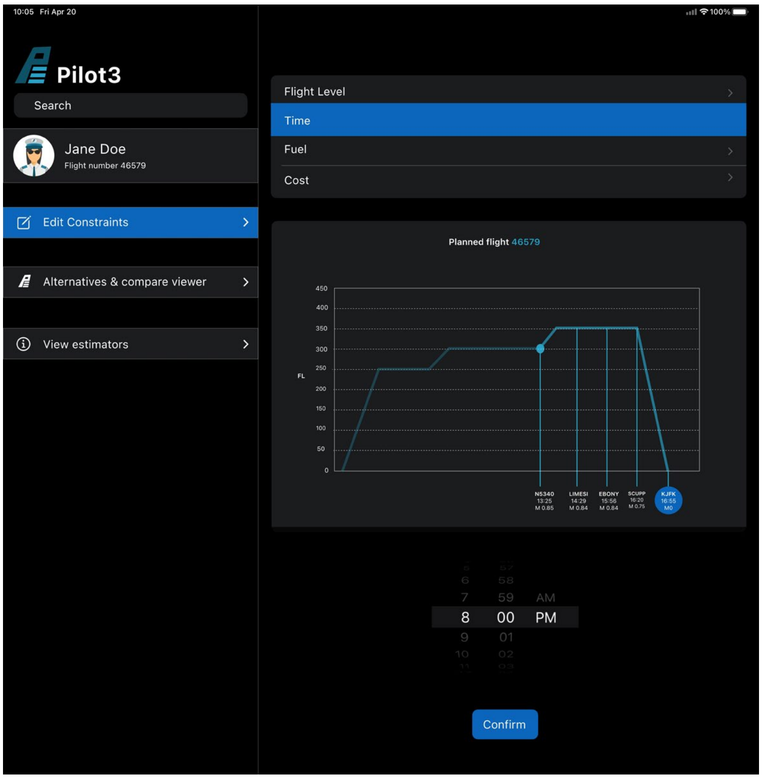


Figure 120 Edit time constraint

Figure 120 shows an example of a time constraint (not implemented in current version of Pilot3).

B.1.3 Alternatives and compare viewer functionalities

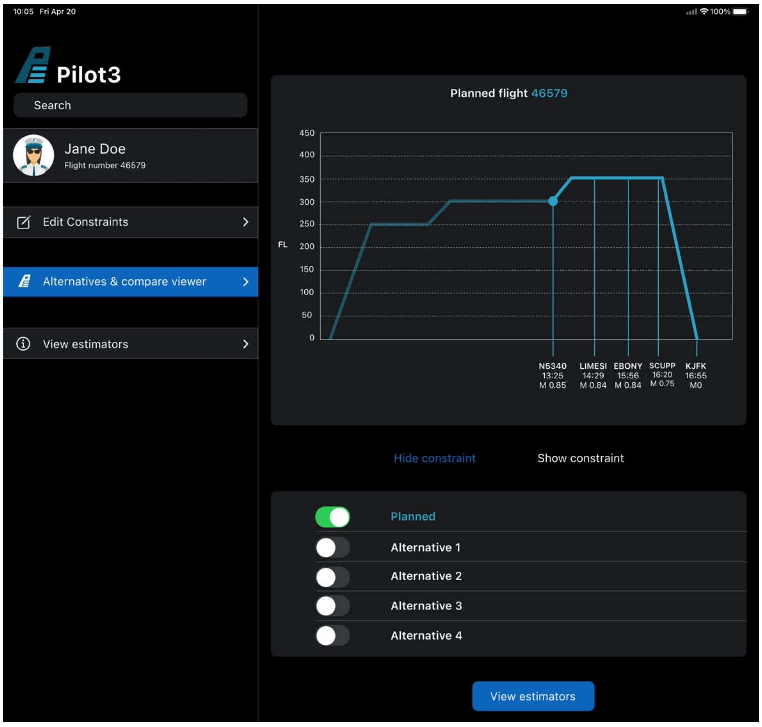


Figure 121 Comparison of alternatives selection tab

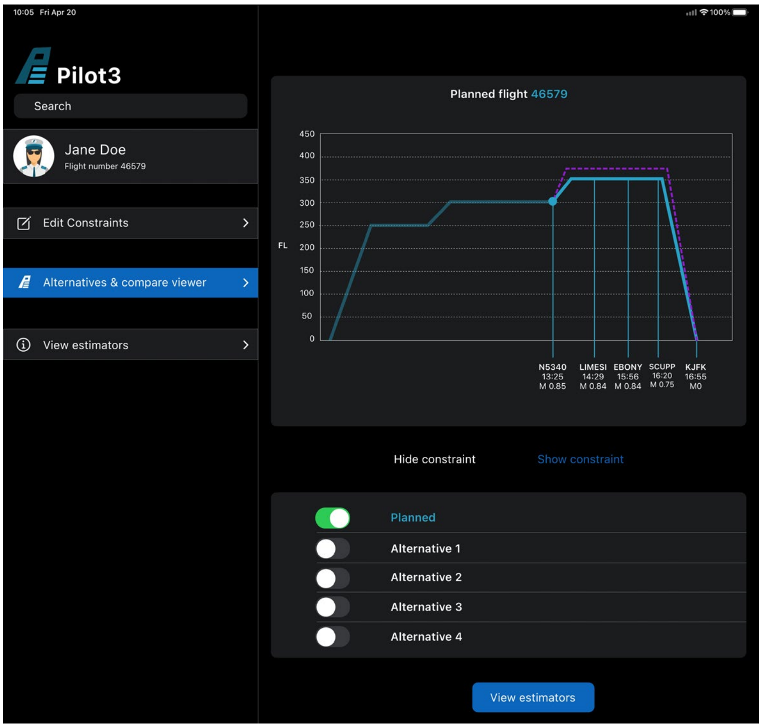


Figure 122 Hide or show previous constraint

EUROPEAN UNION

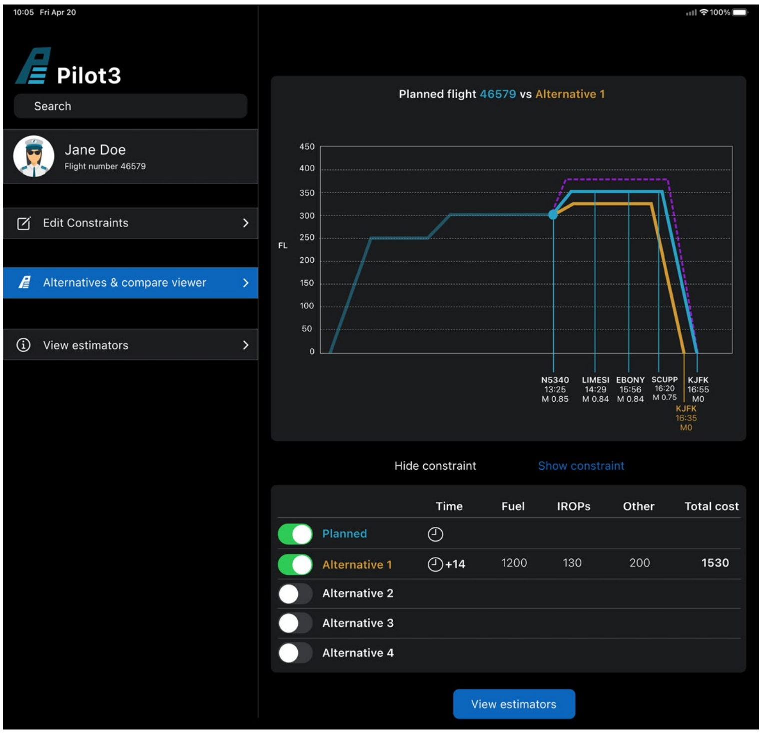


Figure 123 Comparison of three alternatives (original, OTP, constraint) view

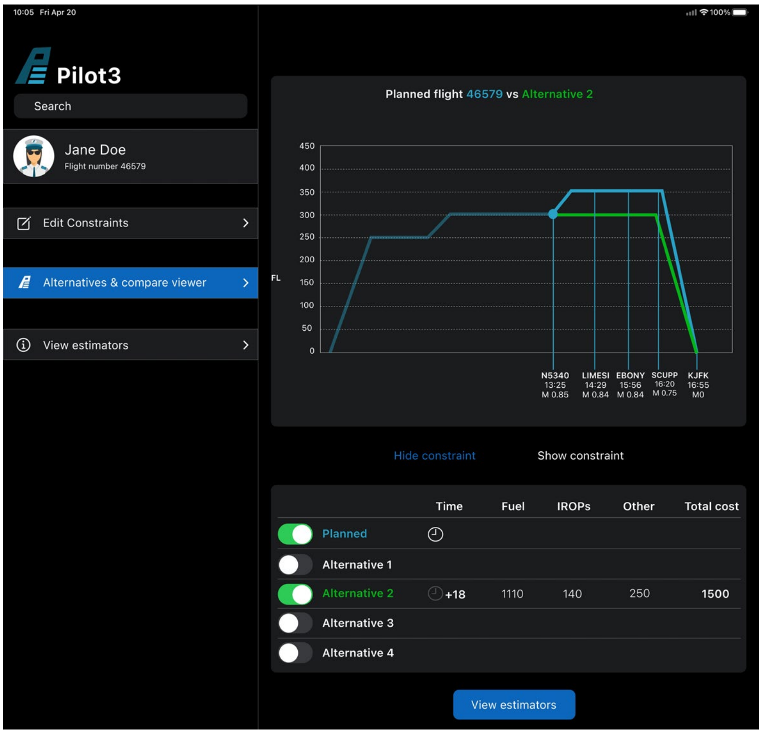


Figure 124 Two trajectories comparison (hidden constraint)

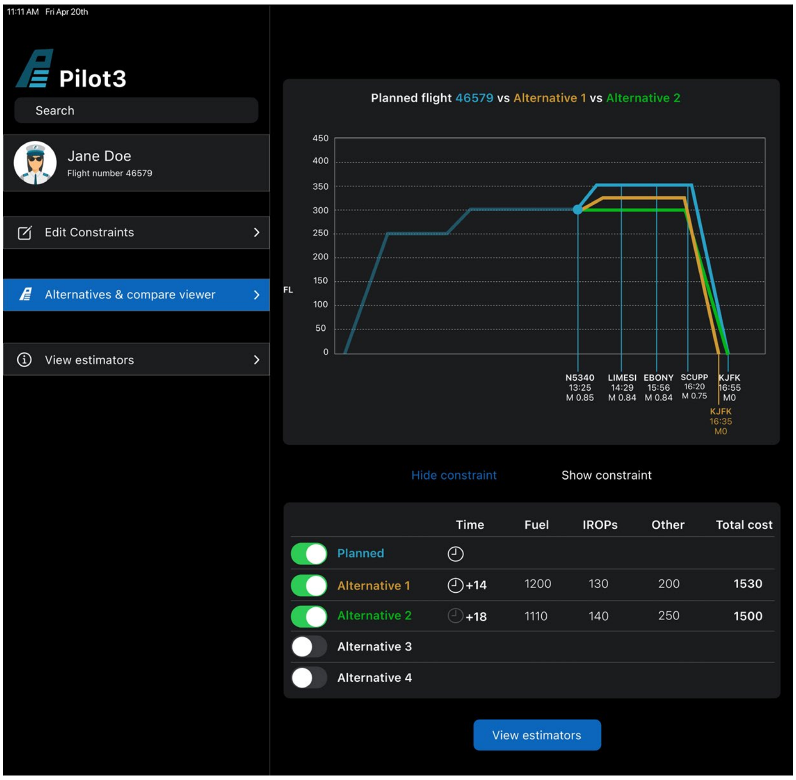


Figure 125 Multiple trajectory comparison view

Figure 121, Figure 122, Figure 123, Figure 124 and Figure 125 present how different alternative can be compared and explored by the crew.

B.1.4 Estimators view functionality

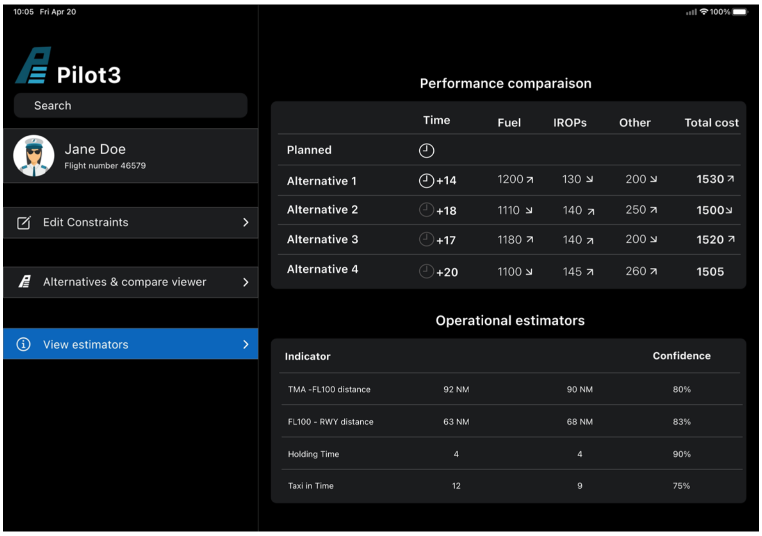


Figure 126 Alternatives results and uncertainties comparison

EUROPEAN UNION

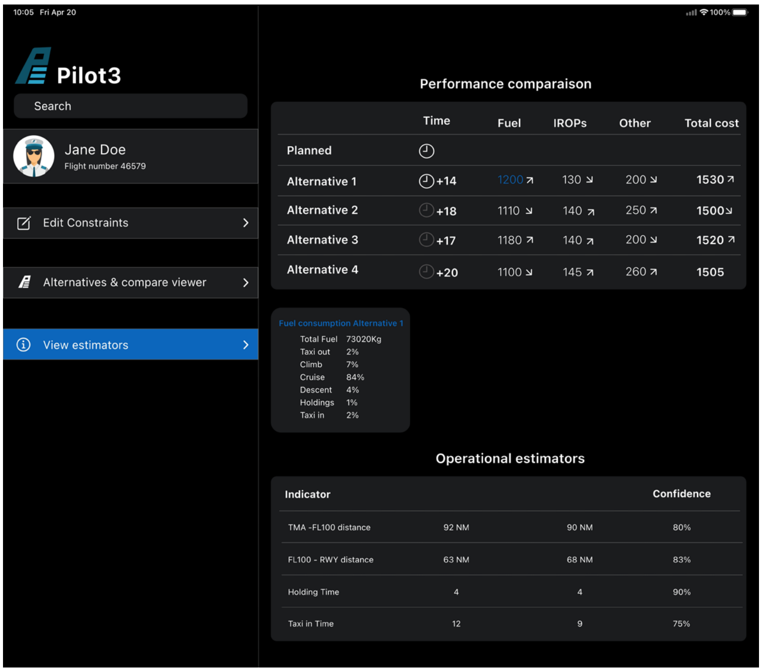


Figure 127 Exploration of indicators for a given alternative

As shown in Figure 126 and Figure 127 the crew can explore the information related to the expected performance for each alternative considering the key objectives identified: On-Time Performance and total cost divided in fuel, passenger related costs (IROP) and others. The crew can also see the uncertainties modelled in Pilot3 and their impact on the trajectories.

B.1.5 Generic application functionalities

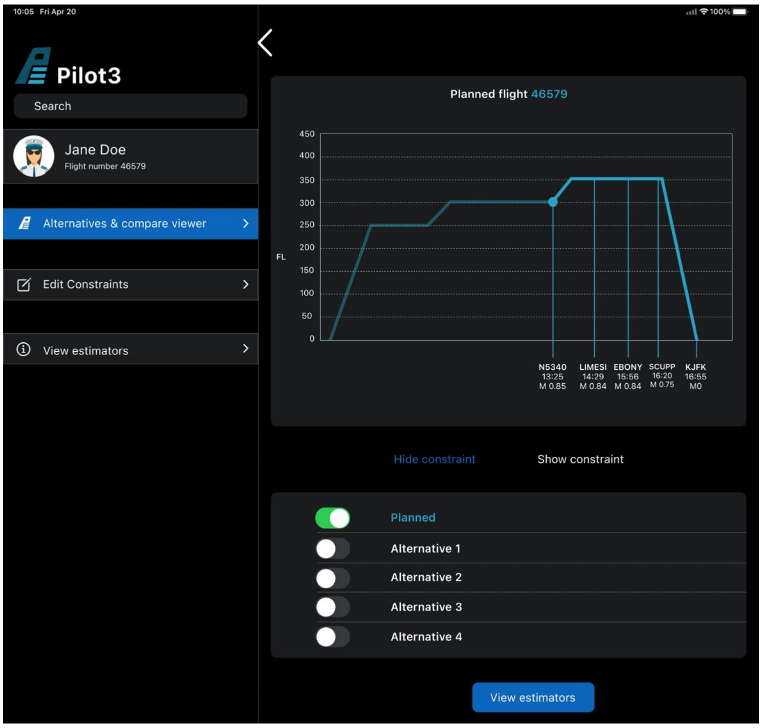


Figure 128 Explore current trajectory information

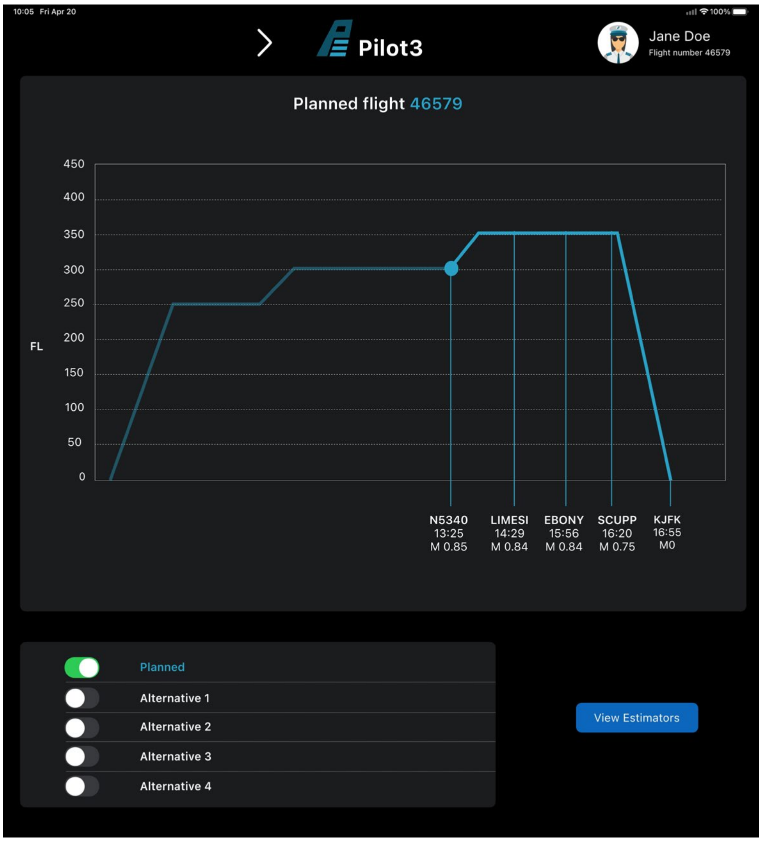


Figure 129 Detailed view of current trajectory waypoints

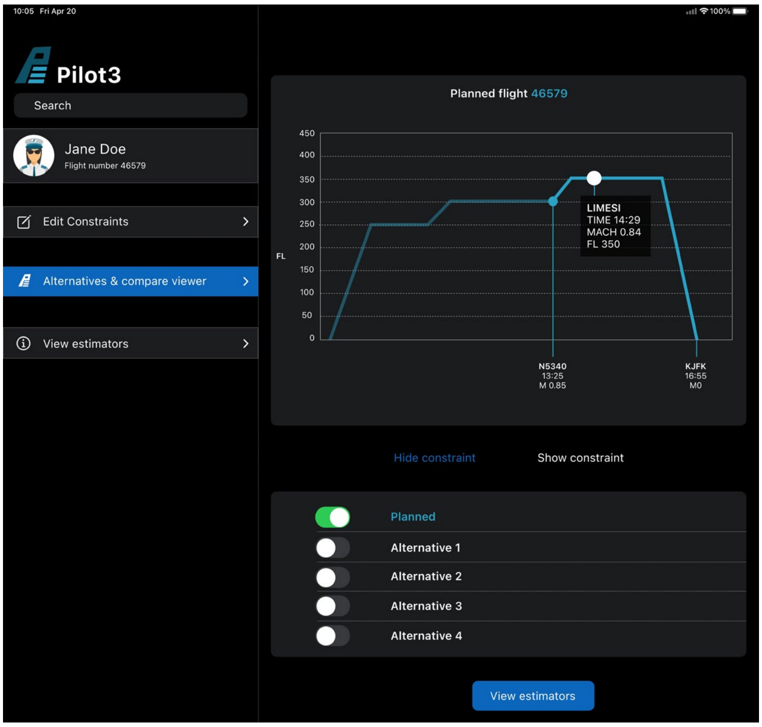


Figure 130 Information on a given point along the route

EUROPEAN UNION

Finally, Pilot3 provides a set of basic common functionalities such as exploring the current trajectory, e.g. waypoints to be passed (see Figure 128 and Figure 129) and information for a given point (see Figure 130).

B.2 Advisory Board survey

This section contains the results of the survey performed with the Advisory Board members in the context of EVA1. In particular, the survey is a follow up from the Advisory Board meeting held on the 13th of January 2022. The main goal of this survey is to **gather feedback on which information could be presented to the pilot** (to be used as potential future evolution of the prototype) and on the **current prototype proposal**.

As a reminder, Pilot3 aims at supporting tactically the crew providing information on which trajectory alternatives are possible. These alternatives are computing optimising the vertical profile considering the total expected cost for the flight. The system automatically ranks the alternatives considering their potential to meet on-time performance and the expected costs (differentiating in fuel, IROPs (passenger related) and other (including reactionary) costs).

This section contains the surveys with an indication of the responses obtained from the Advisory Board (number of x).

B.2.1 I PART – Potential information to be presented to the pilot

Pilot3 can generate a large set of outputs, we are interested in **identifying the most relevant indicators for the pilot**. Some of these indicators might be available 'directly' to the pilot, while others might be provided 'upon request', or not provided if deemed by you as unnecessary.

Could you please rate on a 5-point scale (mark with an X) the relevance of each of the output to be presented in HMI prototype indicated in Table below - from "Not relevant to be displayed at HMI" to "Extremely relevant to be displayed at HMI"?

1.Cost related indicators

	Not relevant at all	Slightly relevant	Relevant	Very relevant	Extremely relevant
Total costs (EUR)		X			X
Total fuel cost (EUR)				X	X
IROPs costs (EUR)				X	X
Other costs (EUR)			X		X
Sub-components of fuel (Up to FL100, holding, sequencing and merging fuel, taxi-in fuel)			X		
Sub-components of IROPs cost (non-connecting, connecting pax)		X	X		
Sub-components of other costs (reactionary, crew and maintenance)		X X			
Sub-components of Reactionary cost (due to propagation of delay, curfew, *strategic* action)		X X			
Cost of delay as a function of arrival time at gate (not for a particular trajectory/alternative but for the flight)	X X				

Cost of delay as a function of arrival time at FL100 (not for a particular trajectory/alternative but for the flight)	X	X

Other comments (FSC representative): *"Integration of Pilot3 must be with the OPS system and aligned with clear responsibilities of decision. This is still a main point, which is for me not so clear in the project."*

2. Time related indicators

	Not relevant at all	Slightly relevant	Relevant	Very relevant	Extremely relevant
Time at gate				X	X
Taxi in time		X		X	
OTP					X X
The time at FL100 (when the optimisation finished)	X		X		
Time from FL100 to gate	X				X

Other comments (FSC representative): *"What is relevant for the pilot is the latest time at which the aircraft must dock in the gate, to avoid any reactionary delay through crew, pax, AC rotation, …etc other ones are kind of "nice to know if requested by the pilots" but not mandatory (otherwise, overload of information to the pilot, which causes a safety issue on the long run)".*

3. Passenger missed connections indicators

	Not relevant at all	Slightly relevant	Relevant	Very relevant	Extremely relevant
Probability that the passengers with connections will make their connections per passenger group			X		X

Other comments (FSC representative): *"Relevant only if the pilot wants to know (so in a pop-up window maybe, but not on the standard screen. As said before, important is to not overload the pilot with information. He needs to know what is the clear plan which is the best for the global operations of its company. But knowing all the details is not helping him and could cause a safety problem".*

4. Operational ATM Uncertainties

	Not relevant at all	Slightly relevant	Relevant	Very relevant	Extremely relevant
Expected holding (min)			X		X
Expected taxi-in (min)			X X		
Expected distance of sequencing and merging (i.e., from FL100 to runway) (NM)			X		X

Other comments (FSC representative): *"Interesting but should be integrated in the OFP actually, not in an additional interface".*

5. Other operational parameters

	Not relevant at all	Slightly relevant	Relevant	Very relevant	Extremely relevant
Arrival STAR	X				X
Arrival runway	X				X
Arrival gate	X				X

Other comments (FSC representative): *"All these information are already incl. in the OFP and updated if any changes occurs".*

EUROPEAN UNION

6. Other parameters

	Not relevant at all	Slightly relevant	Relevant	Very relevant	Extremely relevant
En-route wind		X			X
Probability of breaching curfew by any subsequent rotation				X	X
Probability the AOCC will conduct a *strategic* action to cut the propagation of delay (e.g. a cancellation of a/c tail number swap) in any subsequent rotation			X		X
Expected reactionary delay in next rotation		X			X

Other comments (FSC representative): *"As said, coordination with OCC must take place and both parties must have a clear definition of roles and responsibilities. OCC is responsible for the management of the entire fleet and planned flights, pilots are responsible for ensuring a safe flight, if possible economically and environmentally friendly and taking into account the airline strategic needs (Last Time on Position for example).".*

7. For some indicators (e.g. taxi-in, arrival time at gate, meeting OTP) Pilot3 can generate not only the expected outcome but the full distribution (i.e., probability of obtaining different values) as uncertainties are modelled. Will the pilot be interested on these details? For example:

	Not relevant at all	Slightly relevant	Relevant	Very relevant	Extremely relevant
Full distribution of probability of meeting OTP	X				X
Expected probability of OTP (e.g. 70%)	X				X
Qualitative OTP (e.g. Yes/No)	X				X
Full distribution of arrival time at gate	X			X	
Expected arrival time at gate		X		X	

Other comments (FSC representative): *"As said, no overload of information, and no time for the pilot to be trained for understanding such graphics nor to check it in tight schedule (in which Pilot3 prototype would have an added-value".*

B.2.2 II PART – Feedback on HMI mock-up

This second part of the survey aims to obtain your feedback regarding the overall **capabilities of the HMI mock-up** presented to you during the Advisory Board meeting. You were introduced with several aspects of the tool, such as:

- the general concept of the Pilot3 (i.e., "*How is the tool working?*"),
- its specific features (i.e., *"What kind of information does the tool show to the pilot?"*), and
- mechanism implemented to interact with the pilot (i.e., "*How does it interact with the pilot?*").

This survey is divided in two parts: **Easiness of understanding of the information** and **Interaction with the system**.

There are two final sections **only for pilots**: **General acceptability** and **Pilot's overall acceptance of the tool**.

Each questionnaire contains several statements which will be assessed on a 6-point Likert scale (mark with an X) from "Strongly disagree " to "Strongly agree".

A summary on the main HMI capabilities with corresponding screenshots can be found in the Appendix.

1- Easiness of understanding of the information for the following aspects [1]

	Strongly Disagree	Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree
1. Information on a waypoint (box with performance conditions: waypoint name, time, Mach and flight level)				X		X
2. Information on the trajectories and their impact on the optimisation objectives (total cost and OTP)				X		X
3. Information on the trajectories and their impact on the different key performance indicators (cost of fuel, cost of IROPs, other cost)					X	X
4. Information on the trajectories and their impact on the different PIs (e.g. minutes of delay at arrival)				X		X
5. The trade-offs between OTP and total cost (i.e., the extra cost needed to achieve OTP)					X X	

[1] You can refer to the screenshots from “Function: Alternatives and compare viewer” and “Function: Main screen” in the Appendix.

2 - Interaction with the system (appropriate and easy to use) [2]

	Strongly Disagree	**Disagree**	**Slightly Disagree**	**Slightly Agree**	**Agree**	**Strongly Agree**
1. The mechanism which allows the pilot to set new trajectory constraints					X X	
2. The mechanism which allows to request a re-evaluation of the alternative trajectories					X	X
3. The comparison between alternatives					X	X

[2] You can refer to the screenshots from "Edit constraints" and "Function: Alternatives and compare viewer" in the Appendix.

<u>Pilots only</u>

3 - General acceptability – quantity of information provided to pilot [3]

	Strongly Disagree	**Disagree**	**Slightly Disagree**	**Slightly Agree**	**Agree**	**Strongly Agree**
1. The information provided to the pilot is simple and concise enough					X X	
2. The amount of information presented to the pilot is well balanced				X	X	

EUROPEAN UNION

3. The information provided to the pilot is predictable in its presentation				X	X	
4. The visual representation of the alternative trajectories presented is clear and well organised					X X	
5. The selection of the colours and the size of the font is appropriate and friendly		X			X	

[3] You can refer to the screenshots from "Edit constraints", "Function: Alternatives and compare viewer" and "View estimators" in the Appendix.

Other comments: *"Size of font has to be revised, captains as myself tender to have 'problems' with small letters and numbers"*

4- Pilot's overall acceptance of the tool – quantity of information provided to pilot [4]

	Strongly Disagree	Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree
1. The alternatives provided by Pilot3 will facilitate the pilot in their action to take the appropriate decisions				X	X	
2. With the alternatives provided, the pilot will have better awareness of their actions					X X	
3. The information on the probability of meeting OTP will aid the pilot to better assess the benefits of trajectories presented against each other.			X			X

[4] You can refer to the screenshots from "Edit constraints", "Function: Alternatives and compare viewer" and "View estimators" in the Appendix.

Other comments: *"Airlines focus basically on two factors: money and safety. I think that incorporating safety issues would help us. That is: weather in real time downloaded to the program, we would have a better picture that the weather radar (limited distance), and also others factors: topography"*

Appendix C Modelling uncertainty with machine learning

C.1 Literature review

This The characterisation of this individual uncertainty (or error) on each prediction is paramount in many fields. For example, when applying the outcome of these prediction model in dynamic and unstable systems, or when their outcome are combined with other models, as uncertainty can rapidly grow. This is the case for many applications in the field of Air Traffic Management (ATM) when integrating the prediction models into airlines and air traffic control support decision tools. There are different types of uncertainty and therefore different approaches to manage it. A common approach is the division between aleatory uncertainty, derived from the natural variability of the physical world, and epistemic uncertainty, originating from lack of knowledge or ability of modelling the physical world.

Different approaches can be found in the literature to estimate the uncertainty and reliability of the individual predictions, such as:

- Sensitivity analysis on the model to estimate the reliability of individual predictions observing the output response with respect to small change in the input data set (Bosnic and Kononenko, 2008).
- Delta method based on nonlinear regression but which is computationally highly demanding and assumes homogeneity on the error (Khosravi et al., 2011).
- Bayesian method allows to construct prediction intervals by considering the uncertainty due to both, the aleatory uncertainty of the data and the misspecification of the Neural Network (NN) parameters (related to epistemic uncertainty). However, this method, as the delta method, is computationally demanding and cumbersome for large NNs and datasets (Khosravi et al., 2011).
- Bootstrap methods which develops several NN models with subsets of the training set and combines them to obtain an indication of the range of possible values. However, frequently some of these models are biased, leading to inaccurate estimations and the total variance will be underestimated resulting in narrow performance indicators (Heskes, 1996).
- Local neighbourhood prediction interval using clustering of predictions based on error. There a prediction interval is computed per cluster using empirical distribution of errors associated with instances in the cluster which are combined with a regression model. This allows for non symmetrical upper and lower limit on the intervals. Fuzzy clustering techniques can also be used (Shrestha and Solomatine, 2006).
- Mean-variance estimation (MVE) method provides a NN with an indication of uncertainty and additional output to the expected value. The two outputs are computed using the same input features but are connected to two different hidden layers. This method assumes that the errors are normally distributed around the mean of the target allowing to construct the prediction interval easily from the parameters of the mean and variance of these distributions. However, the considered variance is only due to errors, not due to misspecification of model parameters. This can result in misleadingly narrow intervals (Nix and Weigend, 1994).
- Neural network to estimate interval described as multi-objective problem (intervals cover-age (maximisation) and width of these intervals (minimisation)) (Ak et al., 2013).

EUROPEAN UNION

- Gaussian processes (GPs) are widely applied within probability, statistics and machine learning (Rasmussen and Williams, 2006). Their nonlinear and nonparametric abilities make the GPs a powerful modelling tool in a wide range of regression and classification problems. Its Bayesian properties enable the quantification of the uncertainty present in its own predictions and in the data itself (Riis et al., 2021). This methodology has been used in ATM for example in the estimation of trajectory predictions (Graas et al., 2021) or as an approach to drive active learning for meta modelling on large simulations (Ri-is et al., 2021, 2022). One of its drawbacks is that an a priori distribution of error should be considered.

- Quantile regression estimates multiple quantiles simultaneously providing an indication of the distribution of the target variable for a given input. The number of studies on quantile function estimation has been increasing in recent years (Moon et al., 2021). Different techniques can be used such as: linear quantile regression, polynomial regression, Support vector regression (SVR), Quantile regression forest, or Quantile regression neural network (QRNN), which can be used to model data with heteroscedasticity as it is a non-linear approach (Meinshausen, 2006; Koenker and Hallock, 2001; Koenker, 2004; Taylor, 2000; Feng et al., 2010; Amalia et al., 2018; Muthusamy et al., 2016; Moon et al., 2021). Quantile regression forest is a generalisation of random forest not pruning trees and using outcome to estimate distribution with a weighted addition of results per tree based on number of elements in leaves (Meinshausen, 2006). Regression tree performances can be improved by using the bootstrap aggregation technique Vaysse and Lagacherie, 2017). This quantile regression forest provides information on full distribution and not only mean ena-bling the estimation of quantile distribution (Pevec and Kononenko, 2015). The cumu-lative distribution function estimated should be non-decreasing (e.g. 95% quantile should be greater than the 90% quantile). However, when multiple quantiles are estimated simultaneously this might not be respected leading to the crossing problem. This issue arises on non-linear models such as SVR or QRNN (Moon et al., 2021).

Most of these methods provide either an estimation of the variance of the error or an interval of reliability but are not able to describe the distribution of possible values. In Pilot3 is used the following methodology: a probabilistic classifier characterises the distribution of the error of a prediction relying on the estimation of this error on the training set, obtaining the discrete distribution of the possible expected values of the prediction (see Section 4.2.1.3).

C.2 References

Ak, R., Li, Y., Vitelli, V., Zio, E., Lopez Droguett, E., Magno Couto Jacinto, C., 2013. NSGA-II-trained neural network approach to the estimation of prediction intervals of scale deposition rate in oil & gas equipment. Expert Systems with Applications 40, 1205 - 1212.

Amalia, F., Suhartono, S., Rahayu, S., Suhermi, N., 2018. Quantile regression neural network for forecasting in ow and out ow in yogyakarta. Journal of Physics: Conference Series 1028, 012232.

Bosnic, Z., Kononenko, I., 2008. Estimation of individual prediction reliability using the lo-cal sensitivity analysis. Applied Intelligence 29, 187 - 203.

Feng, Y., Li, R., Sudjianto, A., Zhang, Y., 2010. Robust neural network with applications to credit portfolio data analysis. Statistics and its interface 3, 437 - 444.

Graas, R., Zun, J., Hoekstra, J., 2021. Quantifying accuracy and uncertainty in data-driven trajectory predictions with Gaussian process regression, in: Proceedings of the 11th SESAR Innovation Days, pp. 1- 8.

Heskes, T., 1996. Practical confidence and prediction intervals, in: Proceedings of the 9th International Conference on Neural Information Processing Systems, MIT Press, Cam-bridge, MA, USA. p. 176 - 182.
Khosravi, A., Nahavandi, S., Creighton, D., Atiya, A.F., 2011. Comprehensive review of neural network-based prediction intervals and new advances. IEEE Transactions on Neural Net-works 22, 1341- 1356.
Koenker, R., 2004. Quantile regression for longitudinal data. Journal of Multivariate Analysis 91, 74 - 89.
Koenker, R., Hallock, K.F., 2001. Quantile regression. Journal of economic perspectives 15, 143 - 156.
Meinshausen, N., 2006. Quantile regression forests. J. Mach. Learn. Res. 7, 983 - 999.
Moon, S.J., Jeon, J.J., Lee, J.S.H., Kim, Y., 2021. Learning multiple quantiles with neural networks. Journal of Computational and Graphical Statistics 0, 1 - 11.
Muthusamy, M., Godiksen, P.N., Madsen, H., 2016. Comparison of different configurations of quantile regression in estimating predictive hydrological uncertainty. Procedia Engineer-ing 154, 513 - 520.
Nix, D., Weigend, A., 1994. Estimating the mean and variance of the target probability distribution, in: Proceedings of 1994 IEEE International Conference on Neural Networks (ICNN'94), pp. 55 - 60 vol.1.
Pevec, D., Kononenko, I., 2015. Prediction intervals in supervised learning for model evaluation and discrimination. Applied Intelligence 42, 790 - 804.
Riis, C., Antunes, F., Gurtner, G., Camara Pereira, F., Delgado, L., Azevedo, C., 2021. Active learning metamodels for ATM simulation modeling, in: Proceedings of the 11th SESAR In-novation Days, pp. 1- 8.
Riis, C., Antunes, F., Gurtner, G., Camara Pereira, F., Delgado, L., Azevedo, C., 2022. Active learning for air traffic management simulation metamodeling, in: Proceedings of the 101st TRB annual meeting, pp. 1- 8.
Shrestha, D.L., Solomatine, D.P., 2006. Machine learning approaches for estimation of pre-diction interval for the model output. Neural Networks 19, 225 - 235.
Taylor, J.W., 2000. A quantile regression neural network approach to estimating the conditional density of multiperiod returns. Journal of Forecasting 19, 299 - 311.
Vaysse, K., Lagacherie, P., 2017. Using quantile regression forest to estimate uncertainty of digital soil mapping products. Geoderma 291, 55 - 64.

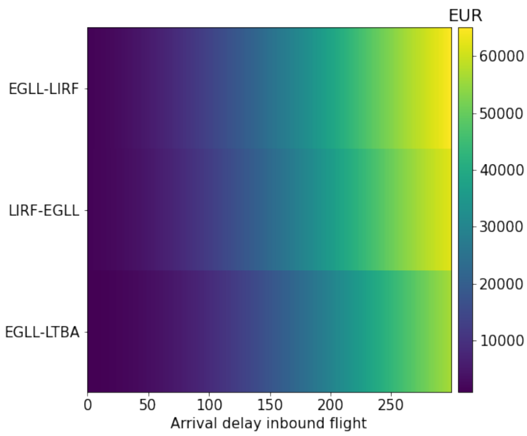
Appendix D Reactionary delay and cost estimation with strategic action

Section 4.2.2.2.2.4 Reactionary delay strategic action presented how the reactionary delay strategic action is computed and compared the impact of including this model on the reactionary cost. This Annex presents a step-by-step description on how the strategic action cost is computed.

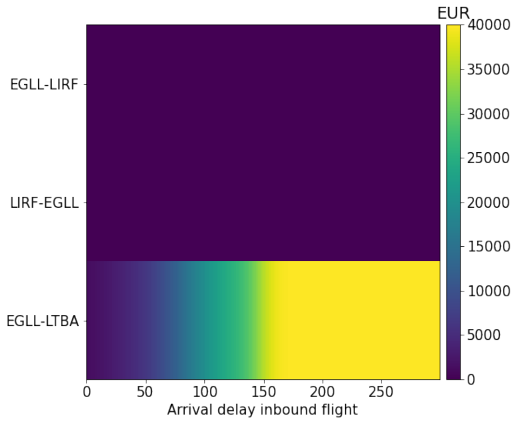
As described in Section 4.2.2.2.2.4 the strategic action model computes the probability of performing such action as combination of two factors:

- first a model which captures the **possibility of performing this action** which depends on the leg of the flight that the AOCC would like to apply the strategic action.
- the second model captures the willingness of the airline of doing this action: **probability of doing strategic action as a function of cost**.

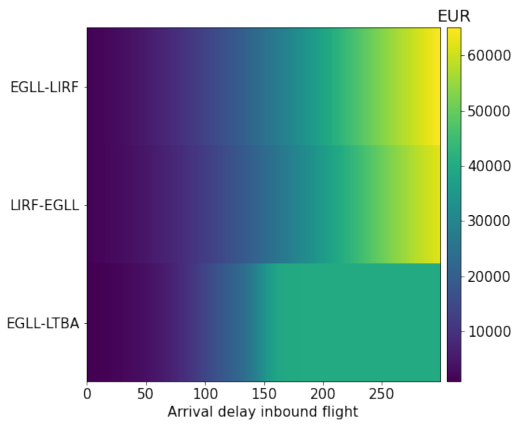
Combining both models the probability of the strategic action is computed. The same example as in Section 4.2.2.2.2.4 is used in this Annex as described in Table 30.



(a) departing delay cost



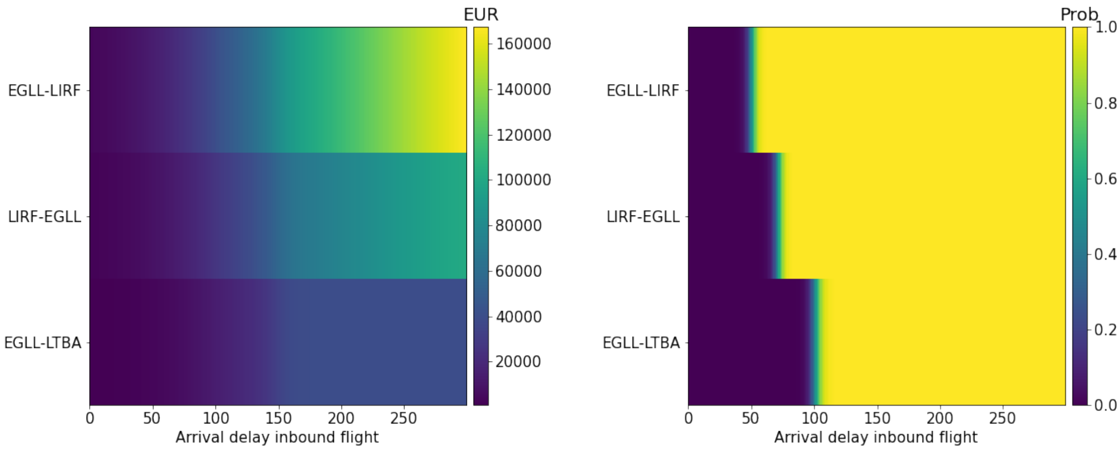
(b) curfew costs



(c)Departing and curfew costs

Figure 131 Raw expected cost per flight as a function of delay of current flight

Figure 131 presents the different expected *raw* cost of delay components for each subsequent leg as a function of the arrival delay of the first flight. These costs are composed of: departing delay cost (Figure 131 (a)) and curfew costs (Figure 131 (b)).



(a) Reactionary costs expected from that flight downstream

(b) Probability of strategic action based on cost

Figure 132 Probability of strategic action due to cost

Figure 132 (a) shows the expected cost from a given flight downstream (e.g. the first row represents the addition of costs for the second, third and fourth rotation) by adding the costs of Figure 131 (c). It therefore the figure indicates, if nothing is done at that point how much the propagation of delay will represent in total. This information will be used to compute the probability that the AOCC will do an action purely based on the expected cost (Figure 132 (b)).

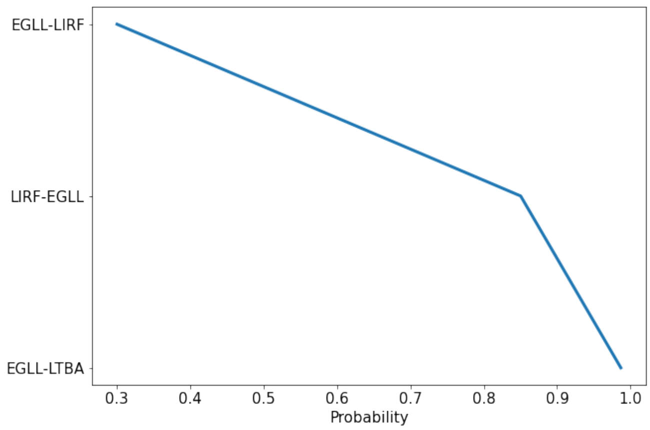
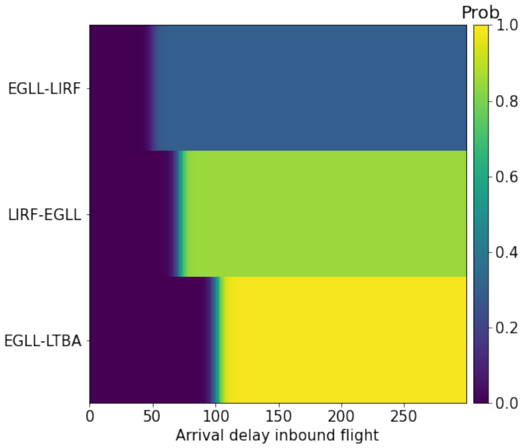


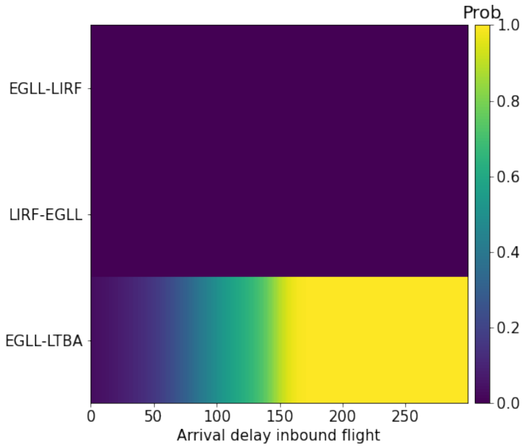
Figure 133 Possibility to do a strategic action as a function of rotation

Figure 133 describes the possibility to do a strategic action by the AOCC as a function of the rotation number (as indicated in Section 4.2.2.2.2.4).

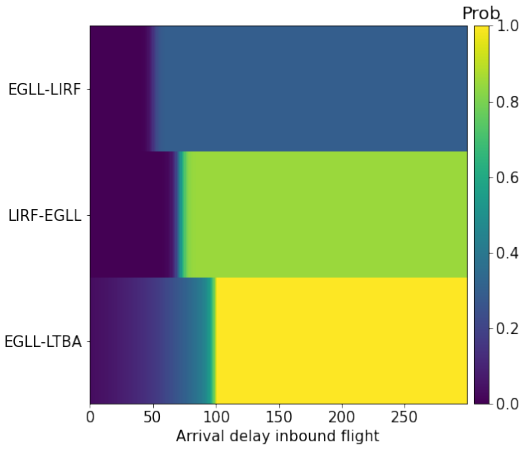
EUROPEAN UNION



(a) Probability of strategic action considering cost and possibility



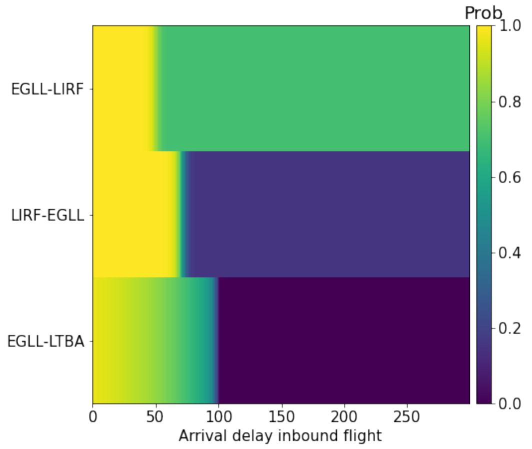
(b) Probability of breaching a curfew



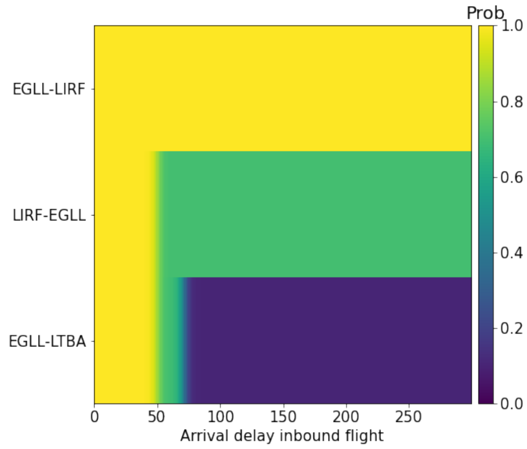
(c) Probability of strategic action or curfew

Figure 134 Probability of performing an action which will stop the propagation of delay

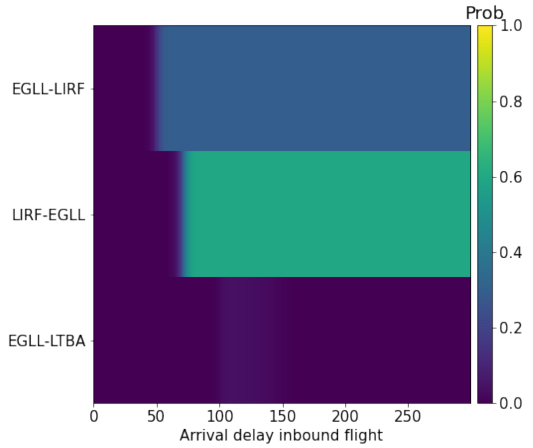
Now the probability of doing a strategic action is computed combining the probability of doing the action based on its costs (Figure 132 (b)) and the possibility of doing it (Figure 133) as shown in Figure 134 (a). Combining this probability with the probability of breaching the curfew (Figure 134 (b)), the probability of having an action which would stop the propagation of delay is computed (either strategic action or curfew) (Figure 134 (c)).



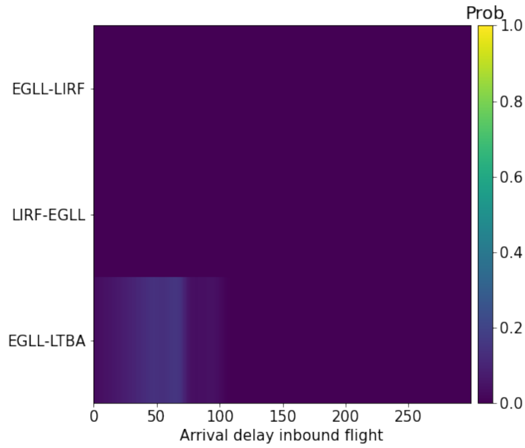
(a) Probability of doing departing (no strategic action done nor curfew)



(b) Probability of experiencing leg cost (no previous strategic action or curfew)



(c) Probability of strategic action and not curfew

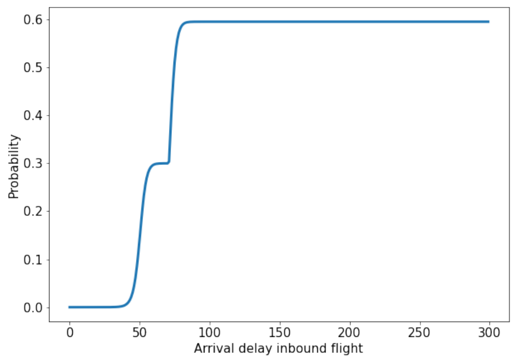


(d) Probability of curfew and no strategic action

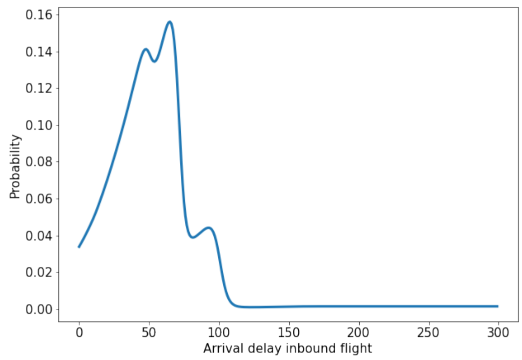
Figure 135 Probability of experiencing leg cost, strategic action and curfew considering previous actions

The complementary of Figure 134 (c) is the probability of experiencing the departing cost (as shown in Figure 135 (c)). Then considering that if a strategic action or a curfew is experienced upstream (Figure 135 (a)), the downstream leg won't be executed (Figure 135 (b)). Thus, the probability of experiencing the departing cost is computed (Figure 135 (d)).

EUROPEAN UNION



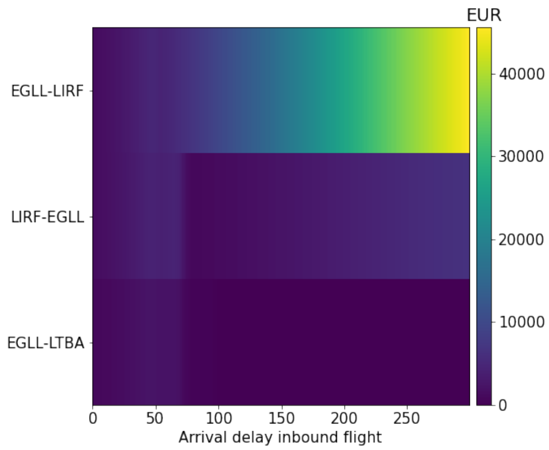
(a) Probability strategic action



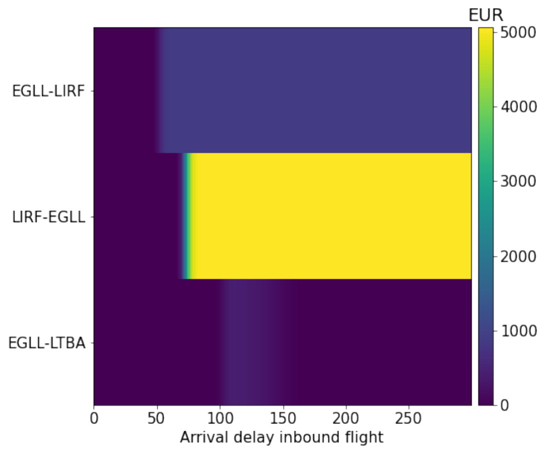
(b) Probability of curfew

Figure 136 Probability strategic action and curfew by any downstream leg

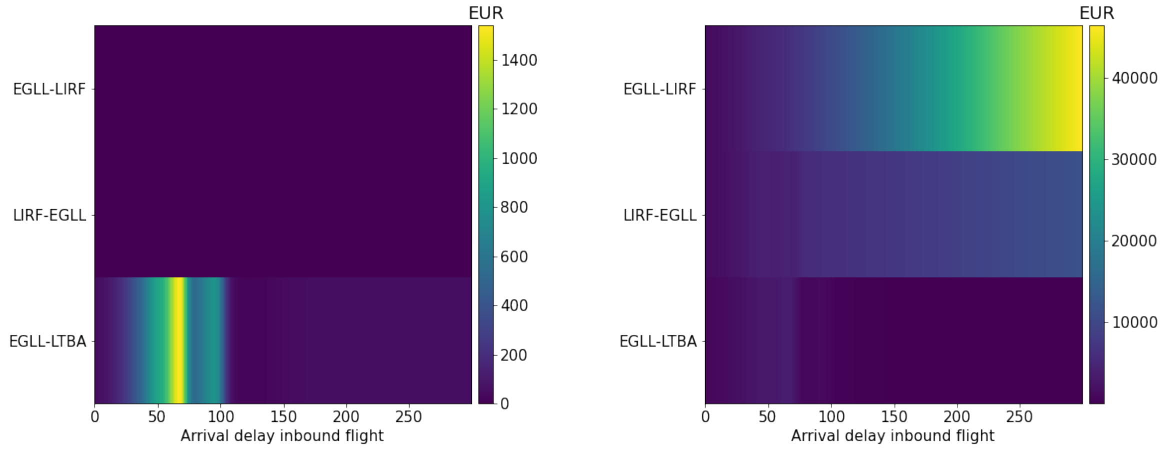
Considering the probabilities that some actions might be performed in previous legs, the probabilities of strategic action and curfew can be computed as presented in Figure 135; and from these, considering the maximum probability as a function of arrival delay of the inbound flight, the probability that downstream a strategic action or a curfew materialises can be computed as presented in Figure 136. Note how, even if the primary delay increases the probability of breaching the curfew does not increase, as the probability of preforming a strategic action in a previous leg would prevent the curfew from materialising.



(a) Expected departing delay cost



(b) Expected strategic action cost



(c) Expected curfew cost

(d) Expected total reactionary cost

Figure 137 Expected reactionary cost components

Figure 137 presents the expected costs due to the different possibilities: departing delay, strategic action or curfew.

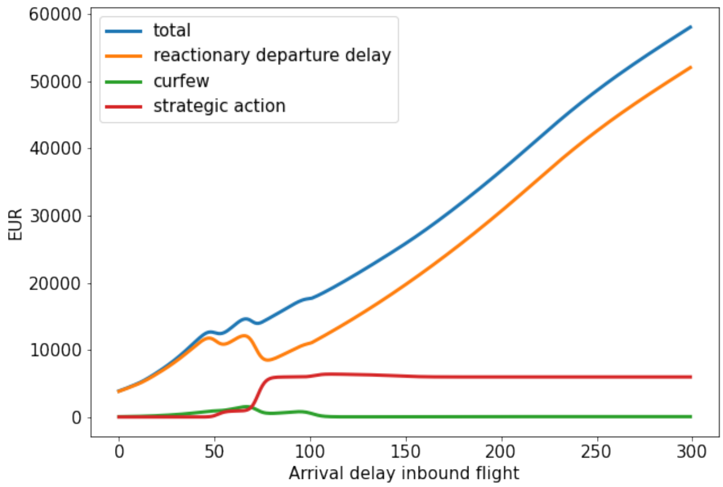


Figure 138 Total expected reactionary costs

Finally adding all the expected cost of all legs the total expected reactionary delay cost and components can be computed as presented in Figure 138.

EUROPEAN UNION

-END OF DOCUMENT-