D5.1 Verification and Validation plan

Deliverable 5.1

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:	**10 March 2021**
Edition:	**01.02**

Authoring & Approval

Authors of the document

Name/Beneficiary	Position/Title	Date
Jovana Kuljanin / Universitat Politècnica de Catalunya	Project member	10 September 2020
Xavier Prats / Universitat Politècnica de Catalunya	Project member	10 September 2020
Luis Delgado / University of Westminster	Project member	10 September 2020

Reviewers internal to the project

Name/Beneficiary	Position/Title	Date
Georgios Athanasiou / Universitat Politècnica de Catalunya	Project member	10 September 2020
Luis Delgado / University of Westminster	Project member	10 September 2020
Gérald Gurtner / University of Westminster	Project member	10 September 2020

Approved for submission to the CSJU

Beneficiary	Position	Date
University of Westminster	Project coordinator	10 March 2020

Document History

Edition	Date	Status	Author	Justification
01.00	7 August 2020	Release	Pilot3 Consortium	New document for review by Topic Manager
01.01	10 September 2020	Release	Pilot3 Consortium	Changes considering comments by Topic Manager; for review by CSJU
01.02	10 March 2021	Release	Pilot3 Consortium	Public version

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

This deliverable provides the methodological framework which will enable the execution of the verification and validation activities. The actions defined within framework plan will support the incremental development of the prototype based on the principle of Agile paradigm. The verification defines all activities that will ensure the thorough test of different prototype versions, while validation will assess the functioning hypotheses addressing the operational benefits of the tool. The validation campaign will be done primarily through the interaction with the internal and external experts to capture their feedback.

The deliverable presents the five-level hierarchy approach on the definition of experiments (scenario and case studies) that ensures the flexibility and tractability in their selection through different versions of prototype development.

The deliverable also details the organisation and schedule of the internal and external meetings, workshops and dedicated activities along with the specification of the questionnaires, flow-type diagrams and other instruments which aims to facilitate the validation assessments.

EUROPEAN UNION

Table of Contents

EUROPEAN UNION

List of figures

List of tables

Executive summary

The verification and validation plan describes the project verification and validation activities, which aim to provide a continuous assessment of the tool throughout the development of the model based on the principle of Agile paradigm. Different prototype versions of Pilot3 tool are defined during the project development considering high-level functionalities. In terms of the activities performed to reflect the notion of Agile approach:

1. **The high-level requirements** will be **disaggregated in low-level requirements** which need to be implemented considering the current model version. This activity will contribute to the additional tasks which need to be defined for different modules (or sub-module) of the tool and data acquisition.

2. **Scenarios** and **case studies** will be **prioritised** generating **low-level requirements** that need to be considered during implementation. The task stemmed from the selection of scenarios and case studies will be performed for the different modules and data acquisition.

3. Pilot3 **consortium team** will **prioritise** and **develop** the tasks from the first two steps.

4. **Verification** activities will be performed in parallel to the development of the prototype. For verification purposes simplified test-cases will be defined. Then, once a **scenario and case study** are ready to be executed, the associated **validation** activities will be performed.

5. A **revision of the development** will be carried out once a **version** is **released**.

Considering the iterative development of the prototype in which high-level of objectives are disaggregated in low-level requirements and implemented for each prototype version, the main objective of the verification and validation is to ensure that:

- thorough test of each prototype version is performed to check if it meets the specified requirements involving tests at subsystem and at integration levels.

- different set of experiments (instantiations of scenarios and case studies) are modelled to assess the results against the set of **research question** and functioning **hypotheses** which have been defined in this deliverable, but which might be subject to further modification as in compliance with Agile approach and based on the results obtained.

A **five-level hierarchy** is proposed to define the **experiments** that will be conducted in PIlot3:

1. **Scenario** is high-level item linked to **specificities of the routes and operations** (aircraft mission) that are modelled. A scenario specifies aircraft mission 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 particularisation of the experiments will be conducted with interaction with the Advisory Board.

The core of the verification and validation approach of Pilot3 prototype is composed of three different and inter-related phases, which must be combined progressively in an iterative and incremental way following the incremental development of the prototype, namely:

1. **Verification and Integration**, which focus on ensuring that the code is developed without errors and that the different requirements are satisfied. This activity will include **static**, **dynamic** and **system** verification tests. These tests will be conducted in parallel to the implementation of the system.

   The **static tests** can be in their turn divided between:

   a. **Software design technical reviews**: aiming at ensuring that the design approach for the software satisfies the different requirements that have been identified for the product.

   b. **Code walk-through reviews**: which provide a first step into the verification of the developed code by providing peer-review among members of the development team.

   The **dynamic verification of the software** is and incremental process which is conducted to verify the working of the software, with:

   a. **Unit and interfaces testing**: which verifies individual specific components of the software.

   b. **Integration testing**: which aims at verifying that the different modules operate correctly jointly.

   c. **Functional testing**: focused on testing functionalities of the software.

      i. **Smoke testing**: to quickly verify that the prototype does not have fatal errors (e.g., runtime execution errors) which need to be addressed before any further inspection of the functionalities.

      ii. **Functional test-cases testing**: focused on testing different functionalities defined as test-cases for which a known input/output has been defined.

      iii. **Regression testing**: aiming at verifying that when new releases are produced, previously accepted functional test-cases are still verified and that side effects have not been introduced in the code.

   Finally, high-level requirements are verified for the different prototype release in a **system testing** activity.

2. 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

EUROPEAN UNION

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. The internal validation actions are grouped between:

a. **Actions aiming at validating the different components of the model**

   i. **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.

   ii. **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).

   iii. **IVA3 - Assessment of the optimisation framework** - the objective of these actions is to assess how Pilot3 can generate different alternative trajectories and trade-offs.

b. Actions aiming at assessment the benefit of Pilot3

   i. **IVA4 - Pilot3 performance at generation of optimised trajectories plans** - IVA4 aims at assessing 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).

   ii. **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.

   iii. **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. This validation action does not focus on validating the prototype but its expected benefits at network level. Therefore, it is out of scope of the project and will not be prioritised.

c. Actions aiming at the validation of the HMI

   i. **IVA7 - HMI** -these action aims to ensure that HMI prototype is well designed with respect to the information and mechanism available to the pilot.

IVA1, IVA2, IVA3 and IVA7 are conducted in parallel to the model development and verification as they validate the working functionalities of the model and provide input into the development. IVA4, IV5 and IVA6 focus on evaluating the benefits of the prototype and therefore require fully working versions of the code to run a validation campaign with different experiments.

3. **External validation** aims to validate the benefit of the tool based on the experts judgement and perceptions toward different aspects of the HMI operational accessibility and appropriateness of the tool. This phase will highly use the **questionnaires** based on the 6-point Likert scale to gather the experts' feedback and eventually validate the research questions and their corresponding hypotheses. 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:

    a. **EVA1 - Live or pseudo-live demonstration of the HMI prototype and overall capabilities**.

    b. **EVA2 - Presentation of results obtained with stand-alone simulations at trajectory level** - based on the results from the internal validation actions IVA4 and IVA5.

    c. **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.

The validation campaign assumes the extensive interactions with the experts from the Advisory Board as well as tight coordination between the members within the consortium. This will be realised through the organisation of the following events:

- **Internal meetings** which will be periodically carried out to ensure synchronisation of the development, identification of bottleneck, identification of new tasks that are required, re-prioritisation of tasks, etc.

- **At least two internal workshops** which will discuss the results obtained by different prototype versions and potential improvements in the implementation of the model.

- **Workshop with external experts** involving the airlines from the Advisory Board, the Topic Manager and relevant stakeholders conducted with results produced with the first release of Pilot3. This workshop will also provide input to the consortium on which scenarios to further develop and provide input on how to improve the model for its final release.

- **Dedicated validation activities** with external experts will be performed by the means of bilateral meetings and on-line questionnaires. For this purpose, tailored scenarios/case studies could be formulated or scenarios for which the experts are more familiar with from the ones already generated be further discussed in more detail.

In order to effectively manage and track the progress of the verification and validation campaign, the results of the different experiments and feedback obtained (including suggestions, recommendations, limitations) during the workshops will be stored in the dedicated page created in a collaborative tool (inGrid).

1 Introduction

Pilot3 will develop a **software engine model** for supporting crew decisions for civil aircraft. This software will provide the crew with a **set of options** along with information to aid the crew to select the most suitable one considering the **multi-criteria business objectives** of the airline, including the **impact on the network** of flights of the airline of those decisions.

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 or the adherence to a given flight schedule. Moreover, understanding the full value of these indicators can be challenging as it does not depend solely on the disrupted flight but on the whole network. For example, 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 will 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**.

The optimisation performed by Pilot3 will consider the trajectory from the point of triggering the system to the arrival gate. Two objectives will be optimised: **on-time performance** (OTP) and **cost**. Cost is composed by three KPIs, which are considered as sub-objectives: fuel cost, passengers related costs (IROPs) and other costs (including among others reactionary delay costs, curfew).

Pilot3 comprises five sub-systems, namely the:

- **Alternatives Generator (AG)**, which will compute the different alternatives to be considered by the pilot; fed by the two independent sub-systems:
    - **Performance Indicators Estimator (PIE)**, which provides the Alternatives Generator with information on how to estimate the impact of each solution for the different performance indicators (PIs) needed to estimate the optimisation objectives which are relevant to the airline;
    - **Operational ATM Estimator (OAE)**, which provides the Alternative Generator with information on how to estimate some operational aspects such as tactical route amendments, expected arrival procedure, holding time in terminal airspace, distance

flown (or flight time spent) in terminal airspace due to arrival sequencing and merging operations, or taxi-in time;

- **Performance Assessment Module**, which, considering the expected results for each alternative on the different KPIs, is able to filter and rank the alternatives considering airlines and pilots preferences, and to show them to the pilot. This is part of the multi-criteria optimisation process; and
- **Human Machine Interface (HMI)**, which will present these alternatives to the pilot and allow them to interact with the system.

For more information on the different modules of Pilot3 the reader is referred to D1.1 - Technical Resources and Problem definition [10].

Prior to the flight, the Pilot3 engine will be configured (by the airline engineers/dispatchers); then tactically, when analysing and selecting alternatives, the pilot will perform a multi-criteria trade-off analysis. This will be driven by different factors: the airline's business objectives and flight policies, and operational aspects considered by the pilot. The process can be seen as an exploration of alternatives consisting on an optimisation framework that requires to generate trajectories considering objectives and constraints, filtering the solutions and adding/modifying them with operational constraints.

Pilot3 has two distinct **usage phases**: **configuration** and **execution**.

The **configuration phase** will be performed by the airline prior to the flight (configured by the airline engineers/dispatchers). This could be done strategically, or some parameters could be selected at dispatching level on a flight-by-flight basis. The objectives of this phase are to consider aspects related to **airline flight policies**: how the Performance Indicators and the Operational ATM Estimators should be computed (for example, using an heuristic or advanced machine learning models, with ground data or only considering data available directly airborne), and the prioritisation of sub-objectives for the cost function to be considered during the trajectory generation.

The **execution phase** (during the flight) is **triggered by the pilot** and consist of three different stages:

1. The generation phase where different trajectories are generated by solving a multi-objective optimisation problem. This is done considering:
    - the objectives function;
    - constraints: operational (e.g., airways) and *ad hoc* defined by the pilot (e.g., 'do not provide solutions which imply an altitude change');
    - environment data (e.g., weather, aircraft performance); and
    - information from the Performance Indicators Estimator and the Operational ATM Estimator on how to estimate these parameters.
2. The ranking of alternatives as a post-processing of the different available trajectories considering airlines preferences.
3. The selection phase where the pilot considers tactical operational related aspects (e.g., requesting new evaluations by adding new constraints).

More information on these phases can be obtained in D2.1 - Trade-off report on multi-criteria decision making techniques [11].

High-level requirements for Pilot3 are identified in D1.1 - Technical Resources and Problem definition [10]. For each of the modules that compose Pilot3, further lower-level requirements are being captured as part of the activities of WP4 - Model development. The consortium needs to ensure that adequate verification and validation activities are performed through the development of the prototype. Two different perspective should be considered (see **Figure 1** below):

1. to ensure that the prototype complies with the requirements set for Pilot3, and
2. to assess the prototype against a set of research questions and functioning hypotheses.

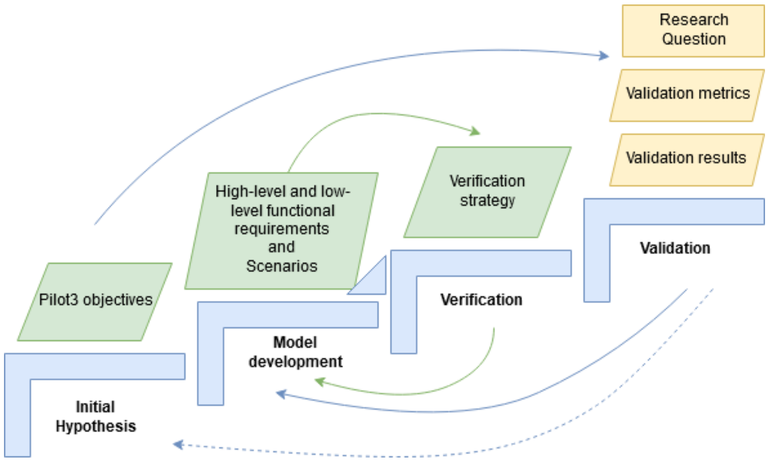


Figure 1. General concept of verification and validation

These two aspects are part of the verification and validation activities and need to be assessed after each development of the prototype. The first aspect, ensuring that the implementation covers the set requirements, relies on an adequate definition of versions which incrementally produce the software architecture and functionalities to Pilot3. The second point, assessing the functioning hypotheses of the tool and, more importantly, research questions, rely heavily on the implementation and assessment of relevant case studies. **Figure 1** shows how the initial hypotheses are underpinned by the objectives of Pilot3 tool and used to identify research questions to be assessed in the validation phase; the requirements and scenarios are used to drive the model development which is subject to verification activities. As depicted, this is not a unidirectional flow (waterfall) but some backward feedbacks are planned and expected (e.g., model development or scenario prioritisation affected by the results of verification and validation activities). In addition, such approach will enable the identification of potentially new hypotheses/research questions, or the modification of those initially defined, as results of the validation activities.

1.1 Development and testing methodology

Pilot3 will be developed following some of the principles of the Agile paradigm. Different target prototype versions are defined considering Pilot3 high-level functionalities (see Section 8), and scenarios and case studies to be modelled are defined in order to cover the different hypotheses and research questions that Pilot3 project will try to validate (see Section 6). Then, during the project development, we will:

1. **Disaggregate** the high-level **requirements** into lower-level requirements to be considered for implementation, considering the **current model version** that the consortium is developing. This activity will contribute to the definition of a backlog of tasks to be developed. These tasks will be defined for the different modules of Pilot3 (and sub-modules if required), and for the data acquisition and management needed to carry out the different validation activities.

2. **Prioritise** the **scenarios** and **case studies**, including variants or sub-scenarios or sub-case studies, to implement and analyse. This will also generate low-level requirements that need to be considered as part of the implementation. As with in the first step, the tasks derived from the selection and analysis of the scenarios and case studies will be performed for the different modules and data acquisition. Note that in both cases, tasks that affect more than one module (e.g., related with integration) could also arise. This instantiation of the scenarios will be decided considering feedback from ad-hoc interaction with the Advisory Board.

3. **Tasks** from the first two steps will be **prioritised and developed** by the different teams that compose the Pilot3 consortium. Track of the progress will be done using collaborative tools (e.g., inGrid and Trello) and meetings will be carried out as required to ensure the synchronisation of the development, identification of bottleneck and of new tasks that are required, re-prioritisation of tasks, etc.

4. **Verification** activities will be performed in parallel to the development of the prototype. For verification purposes simplified test-cases will be defined. Then, once a **scenario and case study** is ready to be executed, the associated **validation** activities will be performed.

5. A **revision of the development** will be carried out once a **sub-version** is **released**. This will include a re-prioritisation of functionalities to be implemented in the next version and scenarios and case studies to be modelled, in order to ensure that all relevant functioning hypotheses and research questions are validated. This may also trigger the identification of new hypotheses/research questions that will be further validated.

This approach ensures that the team is flexible to adjust the development to the functionalities, along with scenarios and case studies which are deemed more relevant. Thus, bottlenecks can be identified promptly for both the software development and the data acquisition and preparation (needed to prepare the scenarios).

1.2 Overview of the verification and validation activities

From software engineering point of view, verification and validation are independent procedures that are performed together to ensure that a product (i.e., system) meets requirements and specifications and that it fulfils its intended purpose. For instance, software engineering standards known as IEEE-STD-610 defines the verification as: "*a test of a system to prove that it meets all its specified requirements at a particular stage of its development.*", whereas the validation refers to "*an activity that ensures that an end product stakeholder's true needs and expectations are met*". As presented in **Table *1*** there are different formulations on the definition of the verification and validation procedures. However, the agreed view is the ultimate objective of these activities:

- Software verification aims at answering the question "*Are we building the product right?";* that is, does the software conform to its specifications?

- Software validation aims at answering the question, "*Are we building the right product?*"; that is, does the software do what the user really requires?

Table 1. Definitions for verification and validation

Source	Definition
Software Engineering standard - IEEE-STD-610 [7]	**Verification**: A test of a system to prove that it meets all its specified requirements at a particular stage of its development. [...] The verification of development refers to checking application that is still being developed to ensure that it adheres to these specifications. code reviews, walkthroughs, inspections but little, if any, actual testing.
	Validation: An activity that ensures that an end product stakeholder's true needs and expectations are met.
ToolsQA [18]	**Verification**: The process of evaluating software to determine whether the products of a given development phase satisfy the conditions imposed at the start of that phase. Verification is a static practice of verifying documents, design, code and program. It includes all the activities associated with producing high quality software: inspection, design analysis and specification analysis. It is a relatively objective process. Verification will help to determine whether the software is of high quality, but it will not ensure that the system is useful. Verification is concerned with whether the system is well-engineered and error-free.
	Validation: The process of evaluating software during or at the end of the development process to determine whether it satisfies specified requirements. Validation is the process of evaluating the final product to check whether the software meets the customer expectations and requirements. It is a dynamic mechanism of validating and testing the actual product.
Guru99 [6]	**Verification:** The verifying process includes checking documents, design, code, and program. It does **not** involve executing the code. Target is application and software architecture, specification, complete design, high level, and database design etc.
	Validation is a dynamic mechanism of Software testing and validates the actual product. It checks whether the software meets the requirements and expectations of a customer. Target is an actual product.
Wikipedia [19]	**Verification**: Software verification is a discipline of software engineering whose goal is to assure that software fully satisfies all the expected requirements. [...] When it is defined more strictly, verification is equivalent only to static testing and it is intended to be applied to artifacts. And, validation (of the whole software product) would be equivalent to dynamic testing and intended to be applied to the running software product (not its artifacts, except requirements). Notice that requirements validation can be performed statically and dynamically

Source	Definition
	Validation: Software validation checks that the software product satisfies or fits the intended use (high-level checking), i.e., the software meets the user requirements, not as specification artifacts or as needs of those who will operate the software only; but, as the needs of all the stakeholders (such as users, operators, administrators, managers, investors, etc.).

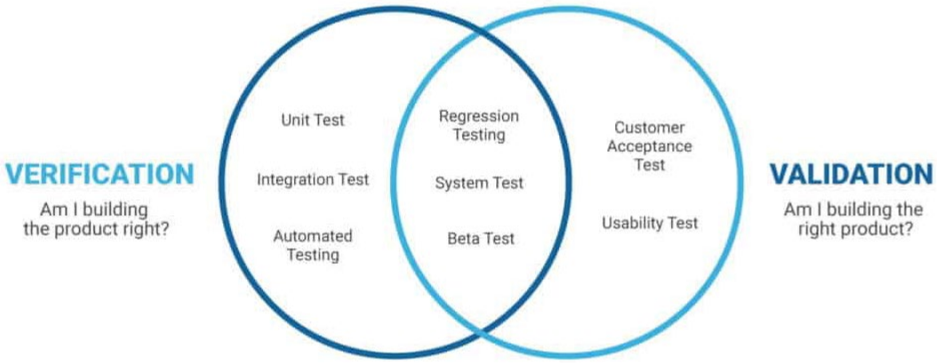


Figure 2. Verification and validation concept and accompanying tests (Source: [13])

Although the distinction between verification and validation are clear enough from their ultimate objective it very often remains quite vague when it comes to their practical applications. For instance, some authors consider that verification should remain as a static process which focus at reviewing the code and the system, while the dynamic execution of the software and the analysis of its output could be considered as part of the validation activities; while others incorporate some dynamic testing as part of the verification to ensure that the code is error free. Therefore, the border between verification and validation becomes fuzzy when the details which specific activities should be performed as part of verification or as part of validation are defined. This is particularly the case in the domain of testing code at a functioning and system level. This is very well illustrated in **Figure 2**. Eventually, which is of utmost concern during the verification and validation process is to perform all the activities associated with producing high quality software: inspection, design analysis and specification analysis in order to secure that final product is worth of using by final customer.

According to the definitions provided above, the software developed in Pilot3 will be first subject to a verification process, which will be carried as part of the developed following the Agile methodology explained above. On each iteration, when a new functionality is implemented and before the release of the new version, the verification of the code will be performed. We consider that dynamic tests should be performed as part of these verification activities (e.g., unit, integration and functional tests) but they should focus on identifying and solving errors on the code and verifying functionalities, rather than validating the outcome of the prototype with targeted scenarios.

The software validation campaign will aim at checking if the Pilot3 prototype meets the expectations of the project. For this purpose, a set of research questions (RQ) is formulated, each one with its corresponding hypothesis. First, an internal validation campaign is proposed, involving only interactions with the members of the consortium and the Topic Manager. Then an external validation campaign will follow, involving interactions with external experts and members of the Pilot3 Advisory Board. The validation activities can be categorised between activities which aim at ensuring that the software produces the right outcome, and activities which aim at evaluating (and quantifying) the quality and impact of the Pilot3 prototype.

1.3 Technological Readiness Level and context within Clean Sky 2

Clean Sky 2 (CS2) is a Joint Technology Initiative (JTI) that aims to develop and mature breakthrough ‘clean technologies’ for Air Transport. The CS2 Programme, will serve society’s needs, contributing to Europe’s strategic environmental priorities and simultaneously, promoting competitiveness and sustainable economic growth. It will enable cutting edge solutions for further gains in decreasing fuel burn (and CO2) and reducing NOx and noise emissions. It will contribute strongly to the renewed ACARE Strategic Research and Innovation Agenda. The CS2 Programme consists of four different elements [1]:

- three Innovative Aircraft Demonstrator Platforms (**IADPs**) for Large Passenger Aircraft (LPA), Regional Aircraft and Fast Rotorcraft, operating demonstrators at vehicle level;
- three Integrated Technology Demonstrators (**ITDs**), looking at Airframe, Engines and Systems, using demonstrators at system level;
- the Technology Evaluator (**TE**), assessing the environmental and societal impact of the technologies developed in the IADPs and ITDs; and
- two Transverse Activities (Eco-Design, Small Air Transport), integrating the knowledge of different ITDs and IADPs for specific applications.

Pilot3 fits within the activities of CS2 Systems ITD WP1.3 “FMS and functions” of the systems (SYS) ITD, and it addresses some of the high-level objectives and challenges for this ITD defined by the CS2 Joint Technical Programme [1], in particular the extension of FMS capabilities. SYS-ITD aims to further mature some of the incipient developments and demonstrators done in CS SGO in the first CS programme, raising them to TRL5 or TRL6, while accommodating the needs of the next generation of aircraft, such as those foreseen in CS2 IADPs (Innovative Aircraft Demonstrator Platforms), and considering the specificities of air transportation in different key performance areas (KPA) involving a diversity of stakeholders.

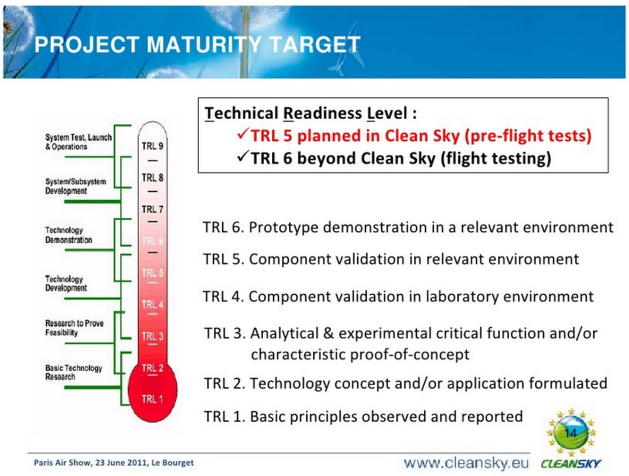


Figure 3. TRL levels and transition phases (Source: [2])

Pilot3 with operational functionalities aims at reaching TRL4 as it is expected to be tested in the relevant laboratory environment (see **Figure 3**). Namely, stand-alone prototyping implementation and test with integration of technology elements will be validated by conducting experiments with full-scale problems or data sets, achieving this way the necessary criteria for TRL4 specified in SESAR JU [16]. However, some of the advanced functionalities of Pilot3 (e.g., the use of machine learning to predict performance indicators or to estimate operational uncertainties, or the consideration of uncertainty during the optimisation process) will remain at a lower level as they encompass more exploitative research activities aiming at presenting a proof-of-concept. Pilot3 aims at reaching TRL3 for these advanced components.

1.4 Deliverable structure

This document is organised in **12 sections** and one Appendix:

- Section 1 introduces the context of Pilot3 decision support tool for crew support on trajectory management. Then it presents the development and testing methodology and the definition and approach considered for verification and validation. The targeted TRL is also presented.
- Section 2 lays out the verification and validation concept and approach that will be followed in Pilot3. Preliminary considerations and definitions are also described in this section.
- The different activities that will be performed for the verification of the code and system are presented in Section 3.
- Section 4 and 5 describe the actions that could be conducted to validate the system with internal and external validation activities respectively.
- As mentioned before, the validation activities aim at answering some research questions. These are described in Section 6.
- Section 7 presents the different scenarios and case studies that will be considered in the project.
- Once the approach for verification and validation has been described, and the different activities for verification and validation are identified, the materialisation of the verification and validation approach into a timely schedule is presented in Section 8.
- The document closes with some conclusions (in Section9) and next steps and look ahead (Section 10).
- References and acronyms are provided in Sections 11 and 12 respectively.
- Finally, an appendix (Appendix A) is provided with questionnaires to be used for the different validation activities.

2 Verification and validation concept and approach

This section describes the verification and validation approach planned for Pilot3. As explained in Section 1, verification and validation activities need to be tailored to reflect the iterative development of the prototype and ensure that:

1. Each prototype version developed meets all its specified requirements by performing tests at subsystem and at integration levels.
2. Prototype meets the expectation of the projects assessed against the set of the research questions and corresponding hypotheses.

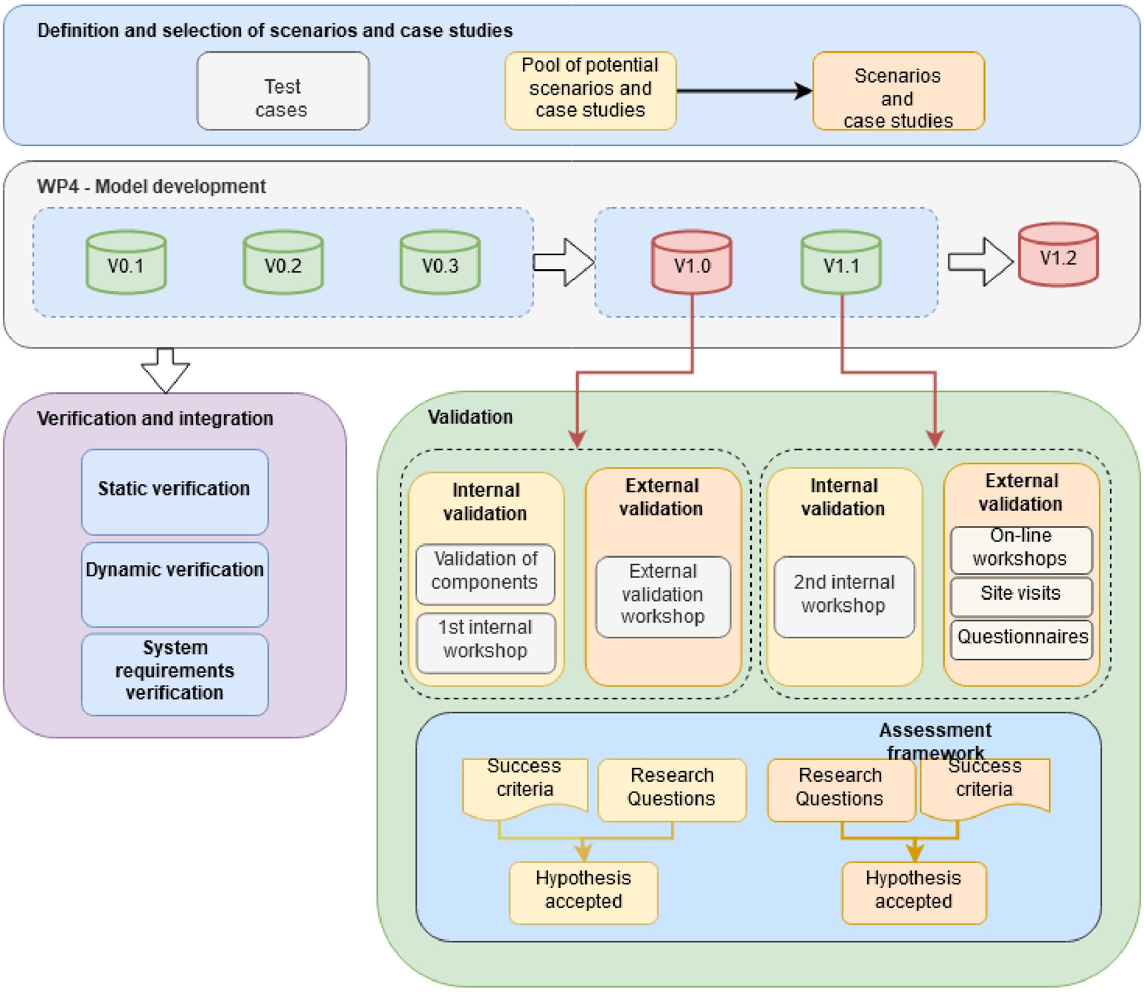


Figure 4. Verification and validation concept and approach

The overall concept of the verification and validation is illustrated in **Figure 4**. As observed, the core of the verification and validation approach of Pilot3 prototype is composed of three different and inter-related phases, which have to be combined progressively in an iterative and incremental way following the incremental development of the prototype and scenarios, namely:

- **Verification and integration**, which needs to ensure that the prototype is properly verified. These activities will be performed continuously along the development of the prototype. Note that different parts of the verification will be more relevant at different stages of this implementation. The objective is to verify that the software does not contain errors and that the requirements are considered in the implementation.
- **Internal validation**, which validates that the inner parts of Pilot3 provide adequate optimisations and quantifies the operational benefits of Pilot3 in order to understand if it meets the project's objectives involving the members of the consortium and the Topic Manager.
- **External validation**, which validates the benefit of the tool through external expert judgement with the support of the Advisory Board and relevant stakeholders and experts.

In order to achieve the goals envisaged by each of the three phases a set of supporting activities are required:

1. **Definition and selection of scenarios and case studies**. The scenarios and case studies are used as part of the verification, but they have a specific relevance for the internal and external validation, and they will support the prioritisation of some development activities. Simplified test-cases might be required for some of the verification activities (as described in Section 3), but specific scenarios and case studies will be considered for the validation of the prototypes. The specific scenarios and case studies that will be modelled in Pilot3 will evolve as part of a consultation process to filter, select and particularise them with requirements and feedback of the Advisory Board members. For this purpose, a **pool of potential scenarios and case studies** is defined in this deliverable following a **five-level hierarchy framework** (see Section 2.3.1 and 7). This will facilitate the Agile approach which assumes the iterative development of the tool.
2. **Definition of assessment framework containing the research questions, hypotheses and success criteria**. Both the internal and external validation are focused on the assessment of the results obtained by respective prototype versions involving internal and external experts. In order to facilitate and drive the assessment, a set of research questions and its corresponding functioning hypotheses are designed aiming to estimate the operational benefits of the Pilot3 tool. These hypotheses/research questions might be subject to modifications as the results obtained by the respective prototype version could generate the need for additional hypotheses/research questions. The experts will assess the results of the scenarios and case studies selected for the given prototype version based on **proposed metrics** (computed from the execution of the prototype) or specific **questionnaires** (e.g., using 6-point Likert scale) to capture the impression of the experts.
3. **Workshops and dedicated validation activities**. The workshops present the main collaborative instruments which will enable to gather the feedback from the great number of experts involved during the validation campaign. The consortium members will organise **at least two internal workshops** which will assess the results obtained by different prototype versions and different set of scenarios and case studies. The external validation campaign envisions the organisation of **one external workshop** which will gather the **Advisory Board members, Topic**

Manager and other experts and stakeholders. In addition to the external workshop, the extensive interaction with the external experts will be conducted by the means of dedicated validation activities (**on-line workshops, site visits and questionnaires**).

For the model development, a set of intermediate versions are planned as part of the pre-release of Pilot3, these versions will be primarily subject to different verification activities and the validation of some of its functionalities. Once the first release is produced (V1.0), the first set of validation activities to quantify the outcome of Pilot3 will be conducted. With the feedback obtained, modifications will be performed producing an intermediate version (V1.1), which will be further validated creating the final release of Pilot3 (V1.2). See Section 8 for more details on the versioning and schedule for the different verification and validation activities.

2.1 Verification and Validation phases approach

This section provides a brief description on the approach followed for the verification and validation phases. The timely definition of the different activities is detailed in Section 8.

2.1.1 Verification and integration

This phase aims at checking that the software prototype modules are working together (or have been developed) correctly according to the high-level requirements specified in WP1 and low-level requirements and operational workflow specified in WP4 by applying software engineering principles.

First, as part of the static verification of the software design a review of this will be conducted. Then as the model is developed, a set of static and dynamic verification activities will be conducted. When new code is developed, walk-through reviews will be conducted, unit testing and integration testing will ensure that the different components of Pilot3 are verified and working in an integrate form. Functional tests will be conducted to assess the different functionalities of the tool. These will be done considering quick smoke-tests and more detailed functional tests which will consider specificity designed test-cases. The number and exhaustiveness of the test-cases evaluated might vary during different parts of the model development, being as exhaustive as possible prior software releases, but allowing for more targeted, and simpler, test-cases for incremental software developments. Note that the objective of these functional tests is to support the identification of errors in the code, and verify some lower level functionalities of Pilot3, and not the validation of the output of the prototype. Prior any release, regression tests will be conducted to ensure that previously passed functional tests are still verified and that no side effects have been introduced in the code due to new developments.

Finally, the last step of the verification activities will consider the testing of system requirements focusing on verifying that the high-level requirements specified in D1.1., and that the low-level requirements from the different modules are satisfied.

Each institution in charge of developing a module within Pilot3 will be responsible to ensure that the verification activities are conducted as part of this development.

2.1.2 Internal validation

The internal validation has two objectives: to **validate the functionalities of the components** of Pilot3 and to **quantify the benefits** of Pilot3 in order to understand if the main goals of Pilot3 have been successfully achieved. This will be performed considering different **scenarios and case studies**.

The internal validation will purpose a set of quantifiable metrics which will enable the **objective quantification** of the benefits of Pilot3 by **expert judgement within the consortium** (including the Topic Manager). This activity will involve the organisation of **internal workshops** in order to gather the feedback from the experts who will perform detailed assessments of certain scenarios/case studies and interpretation of the results as part of Task 5.2. In the light of observed interim results, the continuous refinement of case studies, or specification of new ones, will emerge in order to explore as much as possible the benefits of the tool and capture its limitations. Finally, the sensitivity analysis will be performed by running batch studies which will eventually enable to assess the robustness of the tool with respect to the proposed metrics.

The different activities that will be considered as part of this internal validation include:

- the **validation of the different components** of the model:
    - the trajectory optimiser,
    - the machine learning predictors, and
    - the optimisation framework
- the **assessment of the benefits** of Pilot3 with respect to different baselines considering:
    - the planning of trajectories,
    - the realisation of those trajectories, and
    - potentially at network level with a full day of operations, note that this will be carried out only if resources are available and it is not a priority
- the **validation of the proposed interface**.

2.1.3 External validation

The internal validation provides a quantitative valuation of the tool. The external validation focuses on a rather qualitative outcome (i.e., acceptance/appropriateness of the tool). The aim is to discover the experts' judgement and perceptions toward different aspects of the HMI operational accessibility and appropriateness of the tool and thus, they are rather qualitative based.

The external validation will be performed through **expert judgement with Advisory Board members, Topic Manager and other experts and stakeholders** as part of Task 5.3. Note how, depending on the resources available by the Topic Manager, their input will be considered either as part of the internal validation, the external validation or both. The Topic Manager will be invited to participate in both the internal and the external validation activities.

A **workshop** will be organised to support the external validation of the first release of the Pilot3 prototype. In this workshop, a demonstration of the prototype will be performed (e.g., including live or pseudo-live demonstration of the HMI prototype and overall prototype capabilities), and the preliminary results from the scenarios and cases studies evaluated in the internal validation presented to gather the feedback from the external experts. This activity could be seen as "**customer general acceptance** testing" as the results of the scenarios and case studies shown to the experts to access the operational benefit of Pilot3.

With the feedback of this validation workshop and considering inputs from the external experts on further functionalities to refine and scenarios and case studies to implement, the final version of the prototype will be developed. Then, various form of **dedicated validation activities** (e.g., bilateral meetings (site visits), on-line questionnaires) will be performed with the airlines members who will assess the obtained results to finalise the external validation of the final release and to identify further potential improvement of the tool to be considered towards its future industrialisation.

2.2 Management and tracking of the verification and validation activities

This section thoroughly explains the methodology adopted to effectively manage the progress achieved in the verification and validation activities. The adequate execution of the verification and validation campaign requires substantial effort to synchronise and monitor the large number of activities defined above.

2.2.1 Management and tracking of verification activities

Due to its complexity and interdependencies with other tasks within WP5, as well as those defined in WP4, the verification must be efficiently managed to ensure the seamless progress of all activities performed. The management of all verification activities must be based on the concept of transparency which requires the certain level of information sharing among the members within consortium and providing relevant information for each of them. In addition, the execution of the activities in the verification and integration calls for a high level of coordination between the partners in order to fulfil the goals in time efficient and cost-effective manner.

In order to provide the transparent insight into the progress of the verification activities, all experts within consortium must have access to specifically designed files in the collaborative decision support tool (i.e., inGrid). For tracking the progress of system testing for each prototype version a set of **living documents** will be created:

- Each **document** will consist of an integrated Excel file in the separate page in inGrid which contains the information on the different tests to be performed (code walk-through, unit and interface testing, integrated testing and system testing). Bearing in mind the level of details required and very sophisticated information specified in the file, the following access rules apply to the documents:
  - All Pilot3 partners will have read privileges to all documents
  - Task 5.2 leader(s), Pilot3 coordinator and members involved on the development of different parts of Pilot3 will have write privileges for code walk-through reviews, unit and interfaces, integration and f testing.
  - Only Task 5.2 leader(s) and Pilot3 coordinator will have write privileges for the system testing document.

2.2.2 Management and tracking of internal validation activities

The principles of transparency and sharing information “just in time and place” will be applied during the execution of the internal validation campaign. In order to facilitate the execution of the activities within internal validation, the consortium will need to prepare:

- a definition of time schedule to run the different experiments.
- the design of the internal workshops and other internal validation mechanisms.
- the setup of a living document to track the progress of the internal validation campaign and to facilitate the refinement and definition of new case studies.

WP5 leader will setup the **living document** by creating a dedicated page in a collaborative tool (e.g., a dedicated inGrid page or shared spreadsheet), which will help to monitor the progress of the validation activities in a transparent manner.

As already explained in Section 1.1, different target prototype versions will be iteratively developed following the principle of the Agile methodology by modelling scenarios and case studies in order to validate the different research questions. In light of the Agile methodology proposed, a **five-level hierarchy** on the definition of scenarios is adopted (see Section 2.3.1). This hierarchical framework aims at providing certain level of flexibility and tractability in the selection of the scenarios during different activities in the validation campaign. There is no need to define the scenario(s) which will address the particular research questions as each research questions can be addressed by most of the scenario and case study defined in the pool. Pilot3 will aim at evaluating as many relevant scenarios and case studies as possible to increase the confidence on the findings for the different research questions. The refinement, definition and selection of **scenarios and case studies** will be monitored in the living document considering input from the Advisory Board.

In addition to tracking validation activities in the live document, the consortium members will periodically carry out **internal meetings** (at least monthly) in order to:

- ensure that all activities are well synchronised within the partners.
- share the information of common interests among different teams that compose the Pilot3 consortium.
- identify the potential bottleneck which can occur during data acquisition and preparation.
- identify the potential bottleneck which can occur during the development of particular model version.
- identify the potential new scenarios and case studies (and/or different parametrisations) and/or prioritise the existing ones.
- identify the new tasks required for different modules (and sub-modules) of Pilot3 and/or re-prioritise the existing ones.

Finally, the consortium members will organise **at least two internal workshops** which will discuss the results obtained by different prototype versions and different set of scenarios and case studies. Namely, the results of obtained prior to the first release of the prototype to the Topic Manager will be discussed at the first internal workshop, while the second workshop will focus on the results obtained once the model has been improved and new scenarios and case studies developed following the feedback from the external validation activities (i.e., external validation workshop).

2.2.3 Management and tracking of external validation activities

In order to fully derive the benefits from the interaction with the external experts, the consortium members will need to carefully perform the following tasks:

- Design of dedicated activities to **identify and select scenarios and case studies** to be modelled in Pilot3. From the pool of scenarios further feedback and information will be gathered using dedicated site visits (or teleconferences) with members of the Advisory Board. These scenarios will be then considered as part of the validation of the first release of Pilot3 prototype.
- Design of the **external validation workshop (**with the Advisory Board members, Topic Manager and other experts and stakeholders)
    - The main aim of the workshop is twofold - first, to **briefly introduce the capabilities of the tool** to the experts and second, to **validate the first release of the prototype**. A set of scenarios and case studies will be refined and selected as part of the internal validation activities but considering feedback obtained from the Advisory Board with dedicated **site visits**. These scenarios will be comprehensive enough to capture different aspects of airlines operations (e.g., type of airlines, operational environment, type of triggering events, such as unexpected changes in weather conditions, large network disruptions). The scenarios are defined with the objective of defining a realistic context in which Pilot3 would be used so that the experts can be introduced to the capabilities of the tool. The results from the internal validation will be used to gather from the experts information using **questionnaires** based **on the six-point Likert scale** and **flow-chart diagrams** which will be distributed during the workshops. These validation activities will also include the external validation of the HMI.
    - In addition to these validation activities, the workshop will support Pilot3 with three additional aspects:
        - Prioritise, select and refine further scenarios and case studies to be modelled with the final release version. This will allow the experts to refine the scenarios to better capture their particular operational experience. Note that this process will be also conducted on a more continuous manner by conducting dedicated site visits.
        - Prioritise and select functionalities that should be incorporated into the tool. The expert will provide their feedback on the potential improvements of the current version of the prototype, but also contributing to the identification of potential modifications required for the future industrialisation of the tool or of some of its modules.
        - Contribute to better define further dedicated validation activities. The experts will provide their feedback on the potential mechanism which will enable the efficient interaction and communication during the dedicated validation activities that will be performed with the new version of the model.
- Design of **dedicated validation activities** for final release prototype version and a time schedule for them
    - Once the feedback is obtained from the external validation workshop and new functionalities, scenarios and results modelled, a final round of dedicated validation activities will be performed with the aim of showing the final results obtained by Pilot3 to experts who will validate the results. This will be done with the result of the final internal validation activities and using targeted small surveys and/or bilateral site visits (or teleconferences). These interactions will also support the identification of further

improvements, which could be either integrated in the development of Pilot3 or gathered and specified in D6.1 - System evolution and uptake.

As with the internal validation, the management of external validation campaign must be based on the principle of transparency. For this purpose, the collaborative tool (e.g., inGrid) will be used:

- to store all feedback information obtained during the external validation campaign.
- to setup the **living document** by creating the dedicated page which will help to monitor the progress of the validation activities in a transparent manner to facilitate the refinement and definition of new case studies.
- The results of the different experiments and feedback obtained (including suggestions, recommendations, limitations) during the validation activities will be stored in the dedicated page created in inGrid.

2.3 Preliminary considerations and definitions used in this validation plan

This section describes some preliminary considerations and definitions required as part of the validation activities.

2.3.1 Scenarios and case studies hierarchy

A **five-level hierarchy** on the definition of **experiments** to be evaluated is proposed as follows:

1. **Scenario** is high-level item linked to specificities of the routes and operations (aircraft mission) that are modelled. A scenario specifies aircraft mission 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. The triggering event refers to any type of unexpected event that may potentially distort the execution of the filed flight plan.
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.

In this way, will be able to specify an **experiment which refers to the exact combination of the variables that define scenario, sub-scenario, case-study and the parametrisation**. This five-level hierarchy framework aims at considering all the different aspects that could be used to define a specific characterisation of the conditions to execute Pilot3.

An instantiation and execution of the experiment with a given prototype version will provide a simulation **run**. Note that different **runs** could be performed to the same experiment on the same version of the model to capture **different realisation** of uncertainty in order to obtain a sensitivity analysis of the results.

Finally, for the verification of functional test-cases and for the internal validation activities that aim at validating the different components of the model, ad-hoc **test-cases** might be required. These will be based, when possible, on fully functional experiments designed for the validation of the tool, but adjusting them as needed, and containing the minimum information required for the verification or validation purpose. For example, defining the minimum parameters required to test the optimisation of a trajectory for which the expected outcome is known.

Table *2* summarises the different definitions used.

Table 2. Definition of scenarios, case studies, experiments elements

Elements	Description
Scenario	Definition of aircraft mission parameters which define route: origin-destination, airline characteristics, baseline defined to use scenario
Sub-scenario	External factors defining particularities of operational environment: type of weather, ATFM environment, curfew, imposition of TTA.
Case-study	Triggering events for Pilot3: event that distort the execution of planned flight plan and requires the execution of Pilot3.
Sub-case study	Configuration of Pilot3 tool: Performance indicator estimator, ATM Operational estimator and Optimisation ranking.
Parametrisation	Parameters to specify scenario and case study to study sensitivity.
Experiment	Combination of variables to specify scenario, sub-scenario, case study, sub-case study and parametrisation.
Run	Execution of an instantiation of an experiment.
Test-cases	Ad-hoc definition of parameters needed to verify specific functionalities or to validate the components of Pilot3 model.

2.3.2 Trajectories definitions

In order to execute Pilot3, first an **operational flight plan (OFP)** will be required for each scenario. This flight plans will be generated considering:

1. **Schedules**: flight schedule information will be defined, including parameters such as origin, destination, Schedule Off-Block Time (SOBT), Schedule In-Block Time (SIBT), aircraft type. Even if these flight schedules could be synthesised (e.g., using historical data or a trajectory optimiser such as Dynamo to estimate the block times), Piltot3 aims at using historical schedules when possible, for this reason the use of DDR2 historical datasets is preferred. This information will be captured as part of the scenario definition.
2. **Flight dispatching parameters**: parameters such as initial Cost Index, take-off weight or route, will be required to integrate an OFP. These will be estimated (or gathered from historical data).
3. **Environment data:** the impact of the forecast weather will be considered when generating the operational flight plan.

EUROPEAN UNION

4. With 1, 2 and 3, **a first baseline trajectory plan** will be computed (generating the **Operational Flight Plan - OFP**).
    1. UPC's Dynamo could be used to compute this OFP. Dynamo has some limitations such as the consideration of RAD restrictions, however, due to the limited number of scenarios (and hence routes), manual constraints could be added to ensure that these limitations are respected.
    2. Alternatively, the OFP could come from an external flight planning tool.
    3. Finally, historical OFP could be considered. The Advisory Board could be approached to see if this can be obtained.
5. The trajectory needs to be simulated to the point prior the execution of Pilot3. For this, a **trajectory simulator** (developed by UPC) will be used to simulate the portion of the trajectory **from take-off to the point Pilot3 is triggered.**
    1. This simulation is needed to compute the initial state of the aircraft (i.e., position, altitude, time, fuel on board, etc), which is needed as input by Pilot3.
    2. It could be the same simulator that will be used to simulate some trajectory executions (see below).
    3. Therefore, the definition of the experiment shall also include information to run this simulation. For example: actual take-off time, actual weather encountered, deviations from OFP trajectory until the reach of the triggering point.
6. With the initial aircraft state computed, and all the additional information that defines the experiment, Pilot3 will be triggered. As a result, an **optimised trajectory plan** (**Pilot3 trajectory plan**) will be obtained.

Finally, it is worth noticing that information on **passengers** will also be required to optimise the trajectories as passengers (and their itineraries) have an impact on the expected costs of delay (e.g., due to missed connections). These data will be gathered from the datasets developed in previous research projects (ER3 - Domino [3]), or when possible from data from members of the Advisory Board.

In order to assess the benefit and impact of Pilot3, different **baseline trajectory plans** could be considered from the OFP to the assumption of some typical pilot's reactions based on their experience and in the absence of Pilot3. The comparison of planned trajectories should however be done with caution as misinterpretations could arise. For example:

- In a case when Pilot3 is triggered in cruise with a new weather forecast update that shows much more headwind than the weather forecast used during the dispatch (i.e., used to compute the OFP). The new Pilot3 trajectory plan will likely involve more fuel consumption than the fuel computed in the OFP. But this does not mean that the Pilot3 outcome is worse than the OFP. In fact, it is quite likely that after executing the OFP with the new weather forecast, the resulting trajectory will end up using more fuel than the execution of the new optimised plan.

- Pilot3 is launched just before take-off with exactly the same weather forecast used in the OFP and assuming an on-time departure. A different trajectory (saving fuel) is obtained by Pilot3 since RAD constraints are not considered in the optimisation (leading for example, to the use of an optimal Flight Level which was restricted by the RAD when the OFP was generated). This directly does not mean that the trajectory generated at dispatching (i.e., the OFP) is not adequate, as ATC might not allow the pilot to fly the Pilot3 optimised trajectory plan.

For this reason, as explained in Section 4.4 a comparison with an **integration** of the different baselines with the available information is preferred. Note that if the information available is uncertain (e.g., a weather ensemble) then stochastic integrations might be required. This is the **planned uncertainty**, i.e., the uncertainty known at the moment of triggering Pilot3 (e.g., weather ensembles, congestion at TMA leading to holdings). This integrated trajectory will be an **updated trajectory results** for the baselines (e.g., updating OFP) and the **optimised trajectory** and **optimised trajectory results** due to applying the optimised trajectory plan.

Finally, when Pilot3 is triggered during the flight, even if the trajectory is integrated, the result is the expected outcome of the **planned trajectories**. Theses planned trajectories will be then operated by the crew by using its information in the flight management and guidance system. Due to external disturbances (e.g., actual weather conditions and conditions at the network), modelling errors (e.g., actual aircraft performance) and the flight guidance logic itself (i.e., how the aircraft is actually steered to follow the plan), the actual flown trajectories might differ from the planned trajectories. Thus, the actual performance achieved (e.g., arrival time, fuel consumed) will not be exactly the same as in the plans. In order to assess these effects, the optimised trajectories will need to be flown, or simulated. The **trajectory simulation** will instantiate these uncertainties with the **realised uncertainties** (e.g., actual weather, modified uncertainty distributions of arrival holding time with updated TMA congestion information) to compute the **realisation** of the planned trajectories (for the optimised and for the baselines). This simulation is a stochastic simulation which captures the uncertainty on the prediction and that therefore, different runs will be required to compute statistical results. Note that the process is equivalent to the one used for the integration of the planned trajectories. However, a differentiated name, trajectory simulation rather than trajectory integration, has been used to facilitate the distinction between planned uncertainty and realised uncertainty.

Table *3* summarises the different elements on trajectories definition.

Table 3. Definition of trajectories elements

Elements	Description
Operational flight plan (OFP)	The original operational flight plan which is considered for a given flight
Baseline trajectory plan	Baseline trajectories plans used as reference for comparison of the outcome of the optimised trajectory by Pilot3. This can be either the OFP or assuming some typical pilot's reactions based on their experience and in the absence of Pilot3.
Optimised trajectory plan - Pilot3 trajectory plan	Optimised trajectory (intends) outcome of Pilot3.
Trajectory integration	Integration of the trajectory from the point where Pilot3 is triggered until its arrival to the gate considering that the information available at the moment (i.e., planned uncertainty).

Elements	Description
Trajectory simulation	Simulation of the trajectory from the point where Pilot3 is triggered until its arrival to the gate but instantiating the uncertainty with the realised uncertainty. I.e., a realisation of the trajectory
Planned trajectory	Planned trajectory (intends) either baseline or optimised with Pilot3.
Integrated trajectory	Outcome of integrating a trajectory plan. For the optimised plan it will create the optimised trajectory, for a baseline it will update the expected result of the baseline plan.
Optimised trajectory	Integrated optimised trajectory plan.
Updated trajectory	Integrated (with updated information, e.g., weather) baseline trajectory plan.
Realised trajectory	Outcome of simulating a trajectory plan.
Planned uncertainty	Uncertainty modelled/considered at the triggering of Pilot3.
Realised uncertainty	Uncertainty realised once the trajectory is simulated. It might be different from the planned uncertainty as the environment conditions (e.g., network, weather) might evolved.

3 Verification and integration

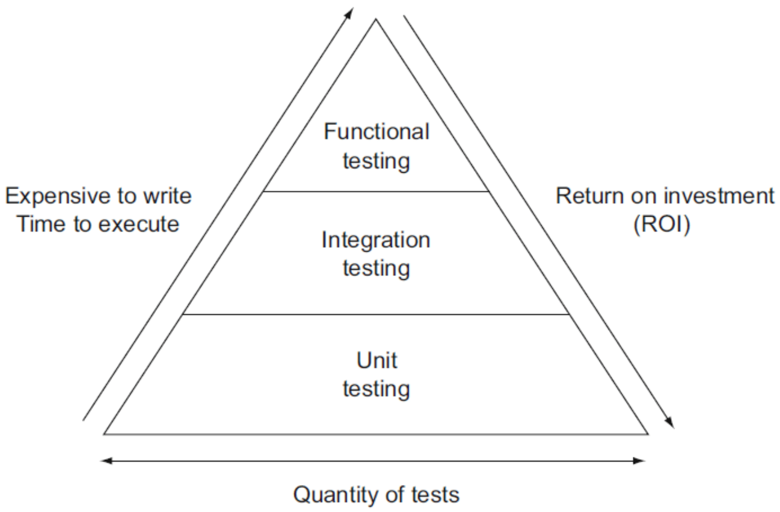


Figure 5. Software verification testing (Source: [17])

Different types of testing are typically conducted in a software verification process as depicted in **Figure 5**. These different tests have also different characteristics in terms of quantity of tests, time required to prepare and execute the tests and impact on the overall verification of the software. In Pilot3, we differentiate between **static**, **dynamic** and **system** verification tests. These tests will be conducted in parallel to the implementation of the system.

The **static tests** can be in their turn divided between:

1. **Software design technical reviews**: aiming at ensuring that the design approach for the software satisfies the different requirements that have been identified for the product.
2. **Code walk-through reviews**: which provide a first step into the verification of the developed code by providing peer-review among members of the development team.

The **dynamic verification of the software** is conducted to verify the working of the software. As described in the introduction, some views are that this should be considered as part of the validation and that the verification should remain static. However, we consider that these low-level dynamic tests are required to verify the code developed and basic functionalities. Note that the full system execution and analysis of functionalities will be performed as part of the internal validation activities. Therefore, we define the dynamic verification of the software as an incrementally process on complexity and integrated verifications:

1. **Unit and interfaces testing**: which verifies individual specific components of the software.

2. **Integration testing**: which aims at verifying that the different modules operate correctly jointly.
3. **Functional testing**: focused on testing functionalities of the software.
    1. **Smoke testing**: to quickly verify that the prototype does not have fatal errors (e.g., runtime execution errors) which need to be addressed before any further inspection of the functionalities.
    2. **Functional test-cases testing**: focused on testing different functionalities defined as test-cases for which a known input/output has been defined.
    3. **Regression testing**: aiming at verifying that when new releases are produced, previously accepted functional test-cases are still verified and that side effects have not been introduced in the code.

Finally, high-level requirements are verified for the different prototype release in a **system testing** activity.

Figure 6 present some example of the different verification activities that will be performed in Pilot3.

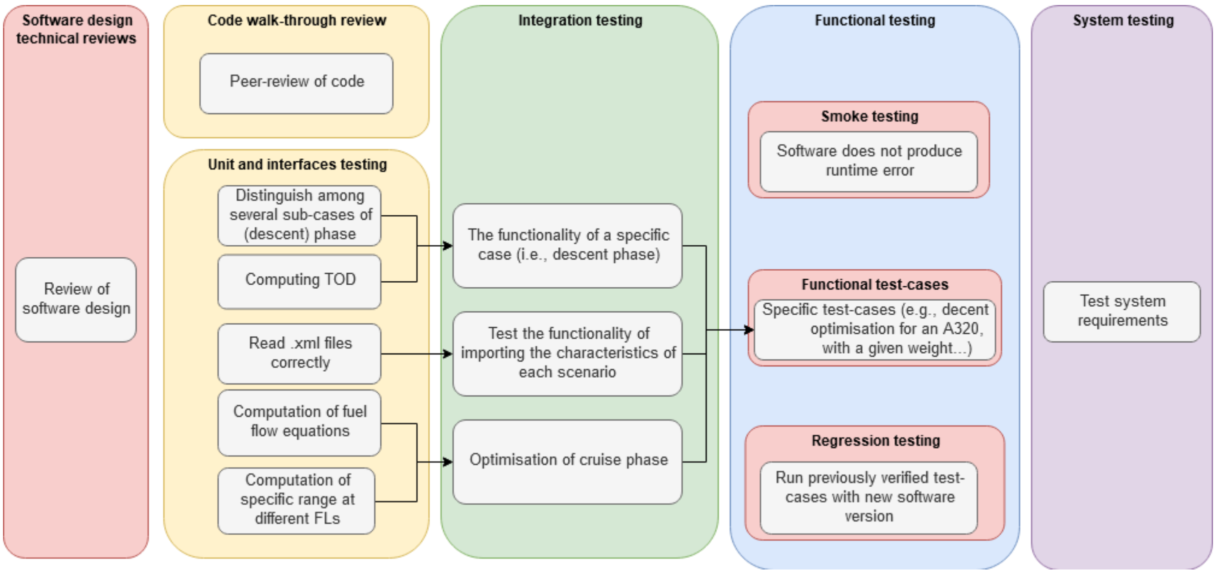


Figure 6. Example of verification activities performed in Pilot3

3.1 Software design technical review

The purpose of software design technical review (SDTR) is to examine the suitability of the software design for its intended use and identify discrepancies from specifications and requirements. More formally, SDTR is a software quality control activity performed by software engineers based on the software design documentation. Although there are various designs for SDTR, the common features for most of them assume the following tasks [15]:

- detecting errors in software design at any stage of the life cycle, and
- obtaining secondary benefit through a two-stage process, in which:

- software engineers first independently inspect the software design for errors, and then
- combine their efforts in a group meeting in which participants adopt roles with the final goal of producing a report in which all the errors agreed upon by the group are identified.

Following these general recommendations in SDTR design on the one hand, and considering particularities of the project on the other, three major objectives are envisioned by the first verification activity in Pilot3:

1. Expert researchers from Pilot3 consortium will analyse the software design document and compare it against all requirements.
2. These experts need to verify that all requirements defined are properly covered by the software design of the given version.
3. Task 5.2 leader will thoroughly track the progress and status of the verification of these requirements. A collaborative spreadsheet will be used and, based on the information provided, all Pilot3 partners will have insight into the progress made in the verification of requirements, whereas only Task 5.2 leader and the Pilot3 coordinator will have write privileges.

3.2 Code walk-through reviews

Code walk-through is a form of "peer-review" in which a software developer leads the review process and the other team members ask questions and spot possible errors against development standards and other issues. The main purpose of walk-through is to enable learning about the content of the document under review to help team members gain an understanding of the content of the document and to find defects.

During the walk-through reviews process, the expert researchers from Pilot3 will review the software product, particularly certain parts of the code, to gain feedback about its technical quality. Coding guidance will be developed and followed as part of the development of the prototype. Their usage will be verified as part of these walk-through reviews.

3.3 Unit and interfaces testing

As depicted in **Figure 5**, the unit testing is a basic layer in the verification process which is generally efficient in terms of the execution time and easiness to be performed. During this first round of testing, the given version of the software is assessed focusing on specific units or components of the software to determine whether each one is fully functional.

Individual and independent testing of parts of the Pilot3 prototype functions and/or interfaces will be performed to ensure that each individual part functions as designed (i.e., to confirm that the code is doing things right). This activity will be split into three sub-tasks with the objectives to:

- perform the verification of each model version developed as well as the data used for this purpose
- verify the functionality added in each new version of the model

- verify the interfaces in order to ensure the consistency in naming conventions, exchange formats, etc.

In order to accomplish the above-mentioned goals, the developer will acquire the methodology which contains three steps specified below:

1. For each version, the developer will have to define the inputs and expected outputs of the model.
2. The expected outputs will be validated by using the external tools (models) that have been already validated. If this is not possible, expert judgment will be applied.
3. The verification step, which will ensure that the actual output matches the expected output.

3.4 Integration testing

As an upper second layer in the verification of the working software, integrating testing aims to test the combinations of individual software modules. In other words, integration testing checks whether different modules are working fine when combined as a group. In contrast to unit testing that considers checking a single component of the system, integration testing aims at checking integrated modules in the system.

The aim of this activity is two-fold:

- to verify integrated models and data,
- to ensure and verify that different modules that compose Pilot3 are correctly integrated.

The methodology applies here will follow the similar approach as in the case of unit testing, although integration testing aims at checking different combinations of the individual modules.

3.5 Functional testing

Functional testing refers to evaluating test-cases which provide a definition of input/output for the full functionalities of the model. This will be performed in three stages: smoke tests, functional test-cases and regression testing. Note that the objective is to support the verification of the code to identify and fix errors rather than to assess the output of the full model and Pilot3 framework.

3.5.1 Smoke testing

The objective is not to perform exhaustive testing but to use high-level test-cases that are executed to preliminarily identify simple failures, severe enough to reject a prospective release.

3.5.2 Functional test-cases testing

When new functionalities are implemented, they need to be verified. This will be done by verifying pre-defined test-cases:

- Defining a subset of test-cases that covers the most important functionalities of the Pilot3.
- Define expected outputs by using the external tools, models or data (already validated). If this is not possible, expert judgment will be applied.

EUROPEAN UNION

- Verify whether the actual outputs match the expected outputs.

In order to efficiently test complete system with full workflow working, the following activities have to be performed:

- several **test-cases** that address all system level requirements will be specified in order to verify that no errors are incorporated for all the software functionalities,
- the selection of these test-cases shall be controlled to test the most important ones, as some previous knowledge of the expected results is already available,
- the verification that the actual outputs match the expected outputs.

Test-cases will be defined ad-hoc to address one or several system level requirements. They will be based on one or several scenarios and case studies that are foreseen for the validation campaign (see Section 7), but with the sufficient amount of data and/or detail to address the functionality requirement to be tested. Note that this activity focuses on the verification on the code and not on the potential of the tool, therefore, the number of scenarios verified will be reduced with respect to the assessment that will be performed as part of the validation activities. Also, note that the validation of the outcome of the different components of the Pilot3 framework will be assessed as part of the internal validation (see IVA1, IVA2 and IVA3 in Section 4).

3.5.3 Regression testing

When a new functionality is added to Pilot3, a set of functional test-cases will be required to verify that this functionality is verified. However, we need to ensure that the implementation of this new functionality has not inadvertently incorporated side effects on previously verified functionalities. Therefore, the regression testing will re-evaluate previously passed functional tests to ensure that they are still verified when new functionalities are added prior to new releases.

3.6 System testing

System testing considers **the complete, integrated system as a whole**. Thus, the given test has to be performed on some of the matured versions of the prototype. System Testing is important as it verifies that the application meets the technical, functional, and business requirements that were set at the onset of the project. The main goals at this level is to evaluate if the system has complied with all the high-level requirements specified in D1.1. (i.e., SYS requirements) and lower-level requirements specified for each module composing Pilot3 (as captured in WP4) to see if it meets quality standards.

4 Internal 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 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).
  - **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.

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 with PACE FPO

4.1.1 Purpose

The purpose of the fist validation action is to check that the trajectory generation engine and data of Pilot3 is able to create trajectories which are realistic and representative. It should ensure that the accuracy of the generated solutions is similar to those generated by state-of-the-art tools.

4.1.2 Tools and methodology

To validate the trajectory generated by Pilot3, Pacelab Flight Profile Optimiser (FPO), developed by PACE, will serve as a benchmark tool. Pilot3 and FPO differ in terms of the aircraft performance model used in their respective trajectory generation engines (i.e., Pilot3 uses BADA4.x while FPO uses Original equipment manufacturer - OEM - data). This may have an impact on the results obtained as different aircraft performance models may result in potential discrepancy between trajectories generated by these two tools. Nevertheless, the main goal of this validation action is to ensure that trajectories provided by Pilot3 are of the similar magnitude as those generated by FPO. In other words, to draw the evidence that the solutions provided by Pilot3 are reliable and of trustworthy qualities.

In order to ensure the consistency of the comparison between the Pilot3 trajectory plan and the PACE FPO trajectory plan, some preconditions must be met prior to running both trajectory generation engines. Both shall use:

1. the same weather information,
2. the same operational flight plan,
3. the same objective function in the trajectory optimisation process,
4. the same operational constraints in the trajectory optimisation process.

Different experiments (test-cases) will be carried out to explore different trajectories with different trip fuel or time priorities. For instance, trajectories that minimise fuel (maximum range cruise operations) will be compared. Likewise, trajectories minimising time will be also compared and some *trade-off* trajectories (i.e., at given cost indices) will be also generated. Alternatively (or additionally) trajectories that minimise fuel, but with a given arrival time fixed as input constraint, could be considered. All this will be achieved by properly defining the objective function and constraints in the optimisation problem.

Besides full trajectories, performances with specific conditions (altitude, speed, weight, wind) might be computed in order to estimate the discrepancies on fuel consumption and time due to aircraft model discrepancies. This might be used to adjust the target of discrepancy for the whole optimisation.

4.1.3 Main metrics for validation

To validate the results obtained and to efficiently compare the two trajectory plans, a set of measurements is proposed encompassing the following quantitative metrics:

- The relative difference in total fuel consumption (in %) computed from the executed PACE FPO trajectory plan and executed Pilot3 trajectory plan.
- The relative difference in total flight time (in %) computed from the executed PACE FPO trajectory plan and executed Pilot3 trajectory plan.
- Difference in the number of speed changes between the executed PACE FPO trajectory plan and executed Pilot3 trajectory plan.
- 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.

Besides these quantitative measurements, some expert judgement will be also a valuable input in examining and comparing the trajectories, especially by comparing both vertical trajectory profiles (i.e., note that only vertical profile is optimised in both tools).

4.2 IVA2 – Validation of indicators and estimators prediction

4.2.1 Purpose

Pilot3 counts with two modules which focus on improving the estimation of parameters: Performance Indicators Estimator, which focus on estimating performance indicators which will be used to compute the objective functions, and Operational ATM Estimator, which aims at predicting operational uncertainties in the trajectory (e.g., arrival holdings). The purpose of these validation actions is to assess the performance of the advanced estimators independently of the Pilot3 optimisation framework.

4.2.2 Tools and methodology

Different type of predictors will be used for the different indicators and estimators: based on heuristics or on the use of machine learning tools, and using only information available in the air or including information from the ground (AOC, ATM).

In both cases (performance indicators and operational estimators) a set of data will be used to create the heuristics and the machine learning predictors. These datasets will be divided between training and validation sets. In the case of the performance indicators, as real datasets are not available, synthetic data generated with Mercury (mobility model developed in previous projects [3]) will be used. For the ATM operational estimators, datasets from historical flight operations (e.g., ADS-B data) will be gathered and processed.

EUROPEAN UNION

Different situations will be considered as part of the validation as the predictors might be tailored for specific operation characteristics (e.g., reactionary delay for a given airline at a given airport, or expected holding time for a given airport at a given time of the day).

4.2.3 Main metrics for validation

To validate the result the predictions of the estimators will be compared with the dataset maintained for validation. Finally, qualitative expert judgement will be also used. The metrics that will be computed are:

- The relative error in the prediction for the different performance indicators (in %) between the predicted value and the validation dataset.
- The relative error in the prediction for the different Operational ATM Estimations (in %) between the predicted value and the validation dataset.

4.3 IVA3 – Assessment of the optimisation framework

4.3.1 Purpose

This validation action focuses on the assessment of the optimisation framework. The objective is twofold: validate that Pilot3 is able to find different trajectories in the Pareto front, and that the airline flight policy has an impact on the trajectories and their filtered and ranking.

4.3.2 Tools and methodology

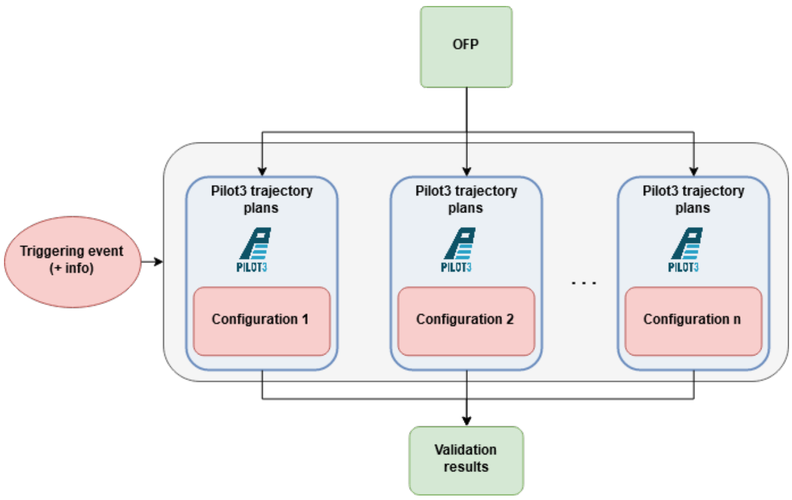


Figure 7. IVA3 – evaluation diagram

The consortium will create a set of experiments (test-cases) to test the capabilities of the model to generate Pareto solutions. Note that in some cases, the test-cases (experiments) (i.e., exact definition of the scenarios and case studies with all parametrisations) will be designed in such a way it is known beforehand that a trade-off between Total Cost and OTP exist, and/or that potential trade-off between costs components KPIs (fuel, IROPS and other) exits.

Finally, different test-cases with different sub-case studies, which reflect different configurations of Pilot3 (and hence different flight policies, e.g., different priorities between cost KPIs), will be defined and executed as presented in **Figure 7**.

4.3.3 Main metrics for validation

- Number of trajectories in the Pareto generated by Pilot3 considering the optimisation objectives (OTP and cost), and cost KPIs (fuel, IROPs and other).
- Filtering and ranking of trajectories generated by Pilot3.

4.4 IVA4 – Pilot3 performance at optimised trajectories plan

4.4.1 Purpose

One of the key elements to capture the benefits of the Pilot3 optimised trajectory plans is to validate them against several baseline trajectory plans. The objective of this validation action is to compare the trajectory generated by Pilot3 at the moment of triggering the system with respect to a set of alternatives considering the information available at that moment, i.e., it will provide the comparison of the planned trajectories. This comparison represents a high-level validation, as it captures the characteristics of a single flight at the moment of triggering Pilot3.

As explained in Section 2.3.2, 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 in order to provide more meaningful comparisons. Note that this integration will be done with the most updated information, but it does not necessarily imply that this will be deterministic as some uncertainty might be present (e.g., using weather ensembles). This can be considered as a **planned uncertainty**. Therefore, the result of the integration might require the results of a stochastic simulation.

4.4.2 Tools and methodology

Several **baseline trajectories plans** for reference are proposed:

- 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.

Example of this second set of baselines could be to accelerate when facing a situation involving delay; changing a certain number of Flight Levels when facing turbulence (based on Pilot's experience or subjective judgement), etc. Thus, this second set of baseline trajectories plans will be scenario and/or case-study dependent.

EUROPEAN UNION

For the Pilot3 plan, as well as for each baseline trajectory plan, **a stand-alone trajectory integration** will be run **without** considering any **noise** on the trajectory execution due to **uncertainties**. This does not imply not considering uncertainty, but not considering the presence of other sources of uncertainty (e.g., performance variations), or that the uncertainty distributions used are not the realised ones (e.g., actual weather different from the ones available in a weather ensemble). Therefore, for a given event that triggered Pilot3 (e.g., update weather forecast, updated estimation for holding in TMA, tromboning delay estimated, etc.,), the integration (simulation) of the trajectory will be run considering the distributions as the actual ones (i.e., the weather forecast is considered as "actual" weather, the holding/tromboning delay estimation is perfect, etc.).

The outcome of the trajectory integration will be an updated trajectory per trajectory plan. Note that stochastic simulations might be required. Thus, in this validation step the updated planned trajectories corresponding of the reference baseline plans will be compared against the optimised planned Pilot3 trajectory plan. This process is summarised in **Figure 8**.

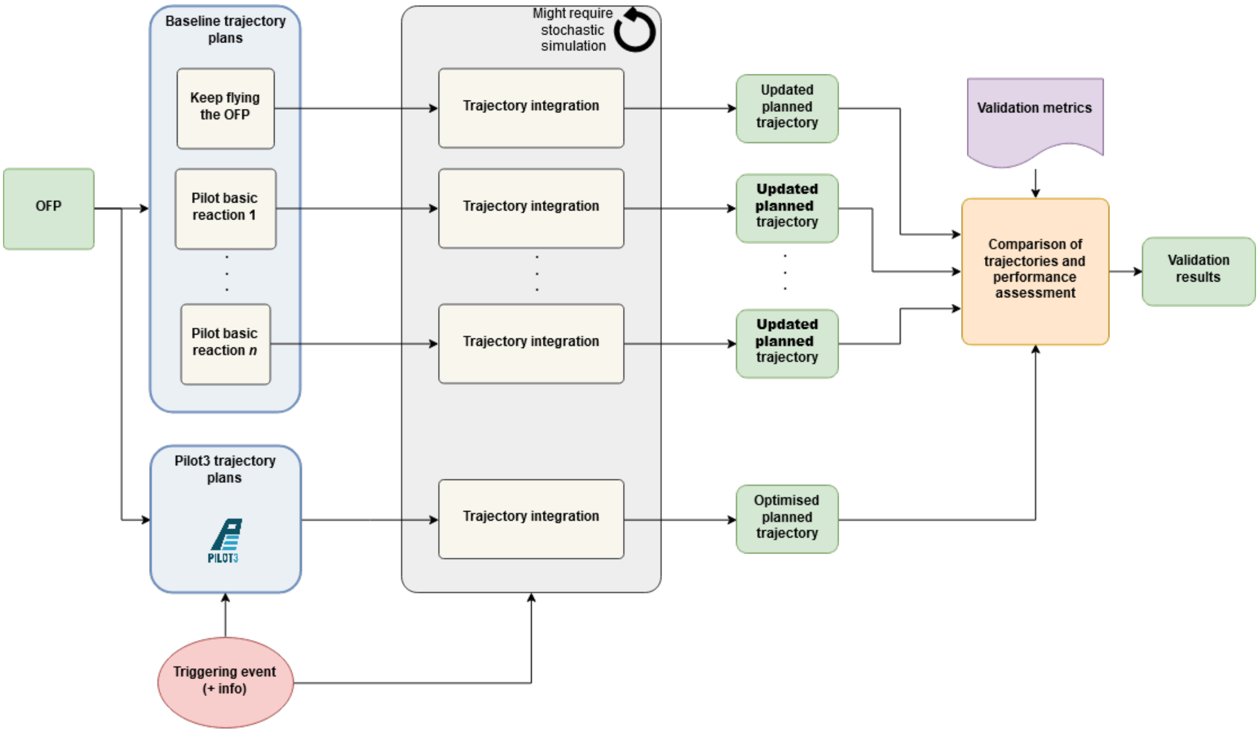


Figure 8. IVA4 – evaluation diagram

4.4.3 Main metrics for validation

The metrics defined below will be used in this validation step to compare the integrated trajectories:

- The relative difference in the total cost (in %) computed from the integrated baseline trajectory plans and integrated Pilot3 trajectory plan.
- The relative difference in other (K)PIs (in %) computed from the integrated baseline trajectory plans and integrated Pilot3 trajectory plan.

- Indication whether OTP is achieved in the case of the integrated baseline trajectory plan and integrated Pilot3 trajectory plan.

4.5 IVA5 – Pilot3 performance at trajectory realisation

4.5.1 Purpose

The purpose of this validation action is to assess the effectiveness of the planned trajectory when realised (i.e., when operated by the crew until arrival at the destination gate), considering the materialisation of different uncertainties, with respect to different baseline alternatives.

4.5.2 Tools and methodology

In contrast to the previous validation step, this validation action will perform **several simulations** considering different realisations of uncertain parameters which might be based/improved with respect to the distributions used at planning (e.g., actual weather which might differ from the forecast, TMA or taxi-in delay different from estimated values as the network situation evolves, etc.). Note that in this way, IVA5 addresses the uncertainty which lays on the estimators that have been used in IVA4, i.e., the materialisation of uncertainty which might differ from the models used at the triggering of Pilot3 point. In this validation action, the uncertainty is therefore **realised uncertainty** and might differ from the planned one. In this way, it is expected to achieve meaningful statistical results, addressing as well, the robustness of the Pilot3 trajectory plans.

For this purpose, the same **baseline trajectories plans** considered in the previous validation step are proposed:

- 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.

For the Pilot3 plan, as well as for each baseline trajectory plan, **a stand-alone trajectory simulator** will be run considering trajectory execution uncertainties. The outcome of each simulation will be a realisation of the executed trajectory. Like in previous validation step, realised trajectories corresponding of the realised baseline trajectories plans will be compared against the realised Pilot3 trajectory plan and some statistics will be drawn. Note that in this case, for each planned trajectory, several runs of the realisation of the trajectory will be required to provide a set of statistics. Finally, different sub-case studies, i.e., configurations of Pilot3, could be assessed to evaluate the realisation of different indicators predictors and operational ATM estimators. This process is summarised in **Figure 9**.

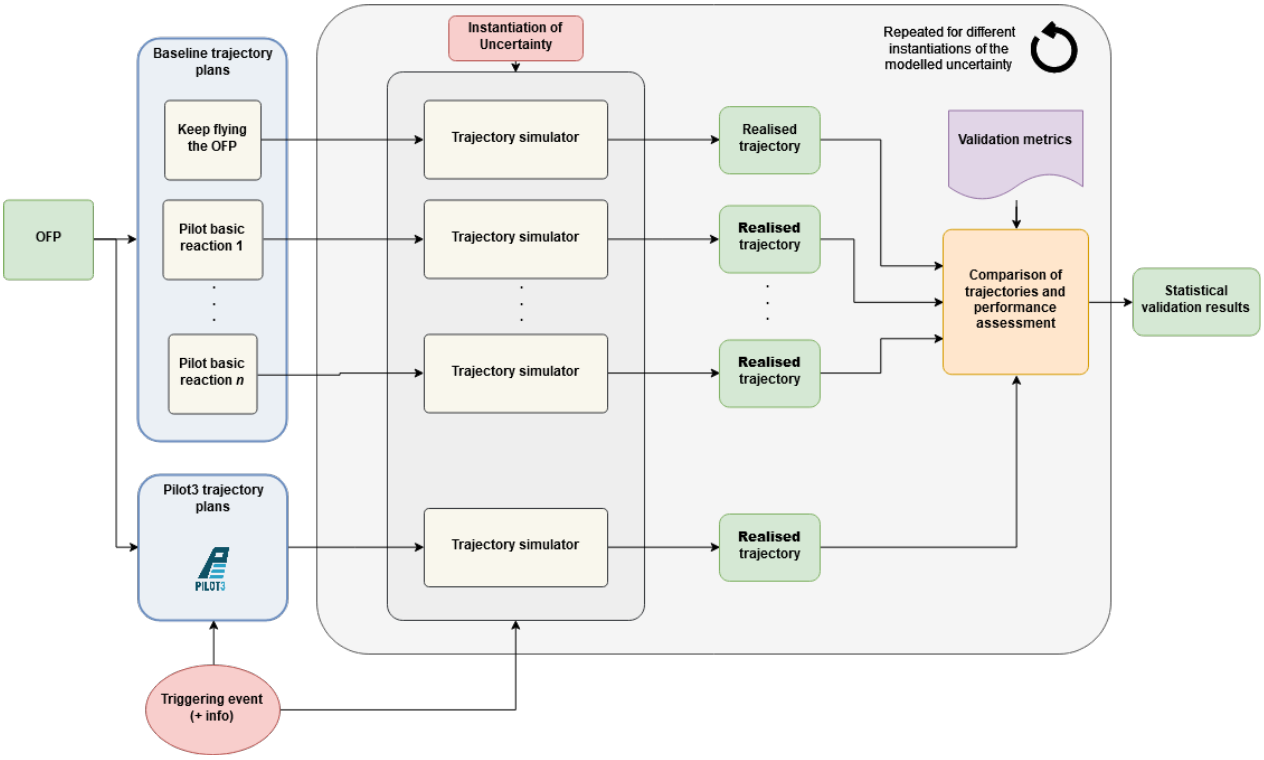


Figure 9. IVA5 – evaluation diagram

4.5.3 Main metrics for validation

The metrics defined below will be used in this validation step to compare the executed trajectories:

- The statistical characterisation of the relative error in the total cost (in %) computed from the realised baseline trajectory plans and realised Pilot3 trajectory plan in the case of uncertainty.
- The statistical characterisation of the relative error in other (K)PIs (in %) computed from the realised baseline trajectory plans and executed Pilot3 trajectory plan in the case of uncertainty.
- Percentage of time when OTP is achieved in the case of the realised baseline trajectory plans and realised Pilot3 trajectory plan in the case of uncertainty.

4.6 IVA6 – Pilot3 performance full day of operations

4.6.1 Purpose

Evaluate the impact of operating a fleet of aircraft equipped with Pilot3 through a day of operations with respect to a baseline situation. This means to consider air transportation network interactions and dynamics (such as quantifying reactionary delay at the end of a given operational day) and/or to consider fleet aspects when assessing benefits (such as computing passenger miss-connections considering the status of connecting and feeder flights). This validation action focuses on the evaluation of the impact of a tool such as Pilot3 on airlines operations, and it does not target the validation of the trajectory optimiser prototype itself. Therefore, it is considered as an optional validation action.

EUROPEAN UNION

4.6.2 Tools and methodology

This validation requires fast time simulations to capture the impact of a tool such as Pilot3 with system-wide metrics. The fast time simulator **Mercury**, a pre-tactical/tactical mobility agent-based simulator developed by UoW and Innaxis which considers flights and passengers itineraries, is considered for this. In the current implementation of Mercury (re-designed and re-implemented as part of ER3 - Domino project [3]), Mercury is an Agent Based Model which captures the behaviour of key stakeholders: AOC, flights, E-AMAN, DMAN, Network Manager, among others. It can track flights and individual passenger itineraries across Europe for a whole day of operations, and provides metrics at airline, flight and passenger level. It allows the evaluation of mechanisms to assess key metrics such as delay and cost. Mercury already counts with different mechanisms to tactically manage flight disruptions, in particular 4DTA (4D Trajectory Adjustments) which provides a combination between waiting for passengers and dynamic cost indexing. Three levels of implementation for the 4DTA mechanism are available (for more detail on this mechanism see [3, 8]):

- Level 0 - Baseline: following current practices where wait for passengers is seldom performed, and flights are tactically speeded up following basic rules defined by the AOC.
- Level 1 - where the decision of waiting for passengers and increasing the speed are conducted independently but based on expected costs.
- Level 2 - where wait for passengers is coupled with dynamic cost indexing, which is reassessed at TOC and which allows for slowing down flights to save fuel if that is the alternative with the highest expected cost reduction.

Mercury focuses on the representation of the processes among the different stakeholders and therefore is able to capture complex behaviour including the decision-making processes for flight and AOCs. However, it has some limitations regarding the detail of the modelling of some aspects of the flight. In particular, the trajectory execution considers BADA 4.x performances but includes some probabilistic distributions for aspect such as weather or location of TOD. One possibility could be to expand Mercury to incorporate the detailed representation of the trajectory through the flight and then incorporate the explicit model of Pilot3. This could, however, require a significant effort to be allocated to a validation action which is not critical for Pilot3. Therefore, a second possibility consisting on modelling the impact of Pilot3 could be envisioned. In this case, the results from previous validation actions (IVA4, IVA5) will be used to generate a model of how Pilot3 would affect the trajectories in case of disruption. Then, this would be translated into a new Level of implementation of the 4DTA mechanism in Mercury. This approach would not be able to indicate specific detailed characteristics of Pilot3, but might be sufficient to provide some insight of the potential impact of Pilot3 on cost and delay metrics at network level for both airlines and passengers.

Figure 10 represents the approach that could be followed to produce this validation.

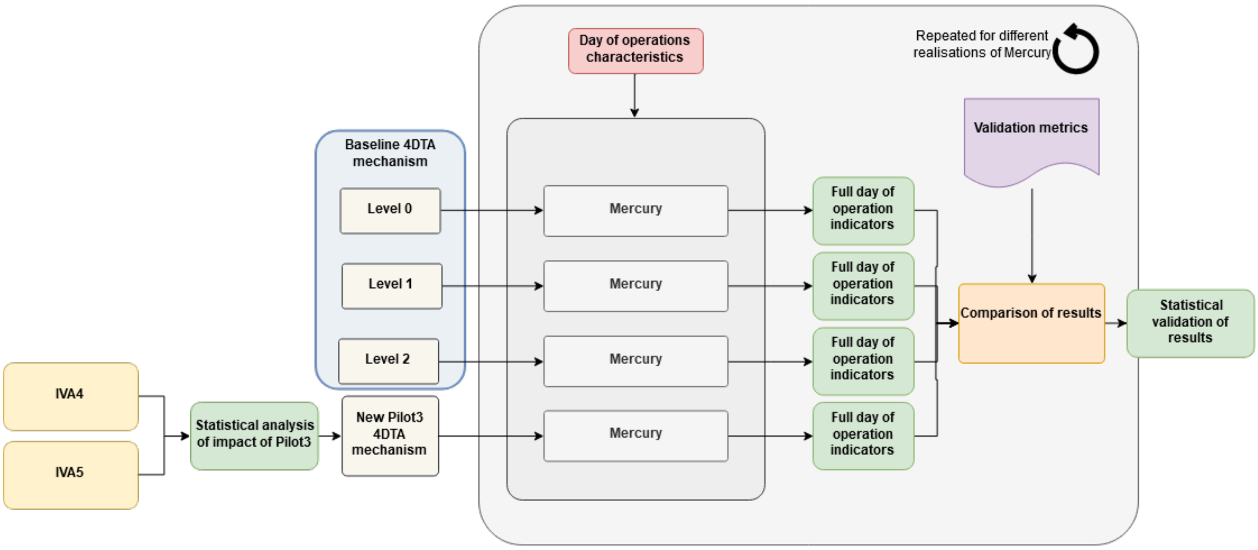


Figure 10. IVA6 – evaluation diagram

4.6.3 Main metrics for validation

In order to assess the benefits of Pilot3 on a network level (with respect to whether Pilot3 is installed in a single airline or on all aircraft in the network), the following metrics are used computed at the end of a given operational day between network of flights operated with Pilot3 and network of flights without Pilot3:

- Average relative error in cost (in %)
- Average relative error in OTP reached (in %)
- Average relative error in fuel cost (in %)
- Average relative error in passengers disruption (IROPs) cost (in %)
- Average relative error in other costs (in %)
- Average relative error in the reactionary delay (in %)
- Average relative error in the number of missed connections (in %)

4.7 IVA7 – Validation of the HMI prototype

4.7.1 Purpose

As already discussed in deliverable D2.1 - Trade-off report on multi-criteria decision making techniques [11], the interaction with the pilot via the Human-Machine Interface (HMI) is considered as the final step of the Performance Assessment Module. The purpose of the HMI is to present information on the trajectories and their impact on the different performance indicators to the pilot, who will be able to interact, rejecting solutions or, based on the information provided, adding new constraints and requesting a re-evaluation of the alternatives. Thus, the information presented should be simple and, as much as possible, predictable in its presentation.

The aim of this validation step is to gather information on experts' general judgements about overall accessibility and appropriateness of the tool based on the illustrative example.

4.7.2 Tools and methodology

In order to accomplish this task, the HMI will be particularised to one (or several) scenario/case study, which will be shown to some expert researchers from Pilot3 consortium and from the TM in order to assess the following HMI features:

- General acceptability of the tool concerning the quality of information displayed to the pilot.
- Easiness of understanding the information on the solutions obtained.
- Appropriateness of mechanism which allows the interaction between the pilot and the tool.

It should be noted that during the verification process of the Pilot3 tool (see Section 2), the HMI prototype will also be subject to verification, meaning that feedback from Pilot3 consortium members will be collected at a regular basis. In this validation action, however, we will seek for 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).

Each expert will be asked to specify their level of the agreement to the statement addressing specific operational related aspects of HMI prototype (design and functionalities). For this purpose, three different questionnaires using a six-point Likert scale will be used. Appendix A contains these questionnaires, that address the general acceptability of the tool, the easiness of understanding of the information and the interaction with the system.

4.7.3 Main metrics for validation

A six-point Likert scale will the following items: “Strongly disagree”, “Disagree”, “Slightly Disagree”, “Slightly Agree”, “Agree” and “Strongly agree”.

5 External validation

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:

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 and IVA5 will be used. The objective is to validate the relevance of the findings.
3. **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 describes with more detail the different external validation actions.

5.1 EVA1 – Demonstrations of the HMI prototype and overall capabilities

5.1.1 Purpose

The aim of this action is to validate the prototype of the Human Machine Interface (HMI), interacting with the external experts (with live or pseudo-live demonstrations), and to obtain an initial feedback regarding the overall capabilities of the Pilot3 prototype. As the initial action in the external validation campaign, 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?*").

5.1.2 Tools and methodology

In order to demonstrate the HMI prototype and the overall capabilities of Pilot3, some results obtained in the internal validation action IVA4 - Pilot3 performance at planning of trajectories (see Section 4.4) will be used and presented to the external experts using different instances of the HMI mock-up (as internally validated in IVA7 (see Section 4.7).

As explained in Section 4.4 of the internal validation, the optimised planned trajectories of Pilot3 will be compared against several baseline reference plans involving the **operational flight plan (OFP)** and some **new plan(s)** assuming some typical pilot's reactions based on their experience and in the absence of Pilot3. The comparison between the optimised trajectory planned by Pilot3 and the reference trajectories will enable the experts to easily identified the differences in the speed/altitude profile between the two executed trajectories (i.e., Pilot3 trajectory plan vs. baseline reference plan) and observe how the benefits obtained by Pilot3 are translated into operational context (as it would have been presented to the pilot). By using the outcome of these validation actions to populate the HMI, the experts will have insight into the plans generated by Pilot3. At least two distinct scenarios will be presented:

1. focusing on a short/medium-haul flight.
2. considering a long-haul flight.

With this approach (using the result from experiments validated in IVA4 to generate the instances of the interface), two objectives are achieved: to introduce the experts with the overall Pilot3 concept and at the same time, to demonstrate the functionality and design of the HMI prototype.

As already introduced in D2.1 – Trade-off report on multi-criteria decision making techniques [11] , the amount of information presented to the pilot should be considered in the context of easiness of usage while providing the required output. In other words, more information may increase the confidence of the pilot in the solutions obtained, but it may also lead to less percentage of the information used, and thus the quality of the solution may be worse. The balance among the amount of information that will be finally presented to pilot is of utmost importance as it will directly affect the acceptance of the tool.

The HMI will illustrate for each scenario presented:

- A set of alternative trajectories and their impact on objectives (cost and OTP) and the different KPIs/PIs to allow the pilot to compare the alternatives
- Visual arrangements of the alternatives presented
- Mechanisms available to the pilot, which will allow them to interact with the tool assuming the following set of actions:
    - selecting the preferred solution,
    - rejecting solutions,
    - imposing new constraints on the trajectory based on the information provided,
    - requesting a re-evaluation of the alternatives.

A flow-chart diagram will be used to evaluate the acceptability of the Pilot3 tool and 6-point Likert scale survey used as gathered in Appendix A.

5.1.3 Main metrics for validation

- Outcome of flow-chart diagram to validate the HMI.
- A six-point Likert scale will the following items: “Strongly disagree”, “Disagree”, “Slightly Disagree”, “Slightly Agree”, “Agree” and “Strongly agree”.to validate the HMI.

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

5.2.1 Purpose

The purpose of this validation action is to show to the external experts some results obtained in IVA-4 - Pilot3 performance at generation of optimised trajectories plans (see Section 4.4) and IVA-5 - Pilot3 performance at trajectory realisation (see Section 4.5) of the internal validation plan in order to obtain feedback from stakeholders.

5.2.2 Tools and methodology

As explained in Section 4.4 and 4.5, a **stand-alone trajectory integrator and simulator** will be run for the Pilot3 plan, as well as for each baseline plans, with major differences of considering uncertainties in IVA-5, producing the realisation of the different plans, while IVA-4 considers the information available at the moment of triggering Pilot3 and hence integrates the trajectory presenting the expected outcome if the information is correct (i.e., not uncertainty). Based on the metrics defined in Section 4.4 and 4.5 of the internal validation plan, the external experts will be able to:

- assess the benefits of the tool with the information available when making the decision (i.e., when the trajectory is generated),
- assess the benefits of the tool when operated and being subject to the materialisation of different sources of uncertainties that may arise during the flight execution (i.e., assessing the realisation of the trajectory),
- provide the feedback of the overall benefits of the solutions with respect to the metrics proposed, and
- provide the feedback on the appropriateness of trajectory planned with respect to operational perspective.

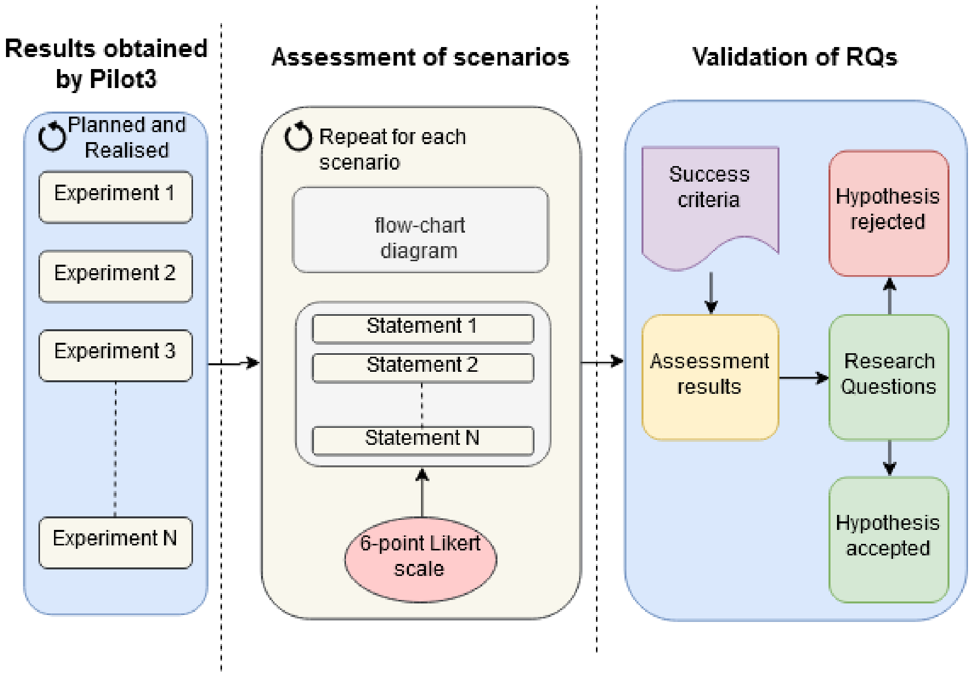


Figure 11. Methodology for validation of stand-alone simulations

In order to accomplish the given objectives, the following steps will be carried out (summarised in **Figure 11**):

- The results for each of the experiments defined in IVA-4 will be shown to the experts. In other words, for each of the experiments defined, the experts will be able to assess the benefits of Pilot3 against the baseline reference plans based on the set of quantitative indicators.
- The results for each of the experiments defined in IVA-5 will be shown to the experts. In order to assess the benefits of Pilot3 when the trajectories are *flown* against the baseline reference plans when executed based on the set of quantitative indicators.

In order to facilitate the assessment of the results presented, the experts will **be asked to provide their feedback on a set of statements** addressing their personal attitudes towards operational benefits of the solutions obtained. For this purpose, a six-point Likert scale is employed. In addition, a flow-chart acceptance diagram is also designed to assess the experts' opinion on the benefit of Pilot3 tool in terms of the solutions obtained.

After obtaining the feedback for each scenario proposed, the results will be gathered and the tool in general will be validated based on defined success criteria.

5.2.3 Main metrics for validation

- Average outcome of a six-point Likert scale will the following items: “Strongly disagree”, “Disagree”, “Slightly Disagree”, “Slightly Agree”, “Agree” and “Strongly agree”.
- Outcome of flow-chart diagram

5.3 EVA3 – Presentation of results obtained with network-wide simulations

5.3.1 Purpose

The purpose of this validation action is to introduce the external experts with the results obtained in IVA6 - Pilot3 performance full day of operations (see Section 4.6) of the internal validation plan in order to obtain their feedback. As described in Section 4.6 this validation action goes beyond the validation of Pilot3 and considers the benefit of Pilot3 at network level. Therefore, as with IVA6, this will not be prioritised.

5.3.2 Tools and methodology

As explained in Section 4.6, **the fast time simulator Mercury** will be executed considering the benefits provided by Pilot3 with respect to different baselines implementations of 4DTA mechanism. This will allow us to quantify the benefits on Pilot3 at network level. Based on the statistical analysis obtained, the external experts will be able to assess the benefits of the tool at the system level.

In order to facilitate the assessment of the results obtained in IVA-6, the experts will be asked to provide their feedback on the statement addressing operational related aspects of solutions generated. For this purpose, a six-point Likert scale is employed.

5.3.3 Main metrics for validation

- Average outcome of a six-point Likert scale will the following items: “Strongly disagree”, “Disagree”, “Slightly Disagree”, “Slightly Agree”, “Agree” and “Strongly agree”.
- Outcome of flow-chart diagram.

EUROPEAN UNION

6 Research Questions

A set of research questions (RQs) and their corresponding hypotheses (HPs) are designed to address the benefits of Pilot3. The RQs aim at being quantifiable rather than qualitative, as we try to estimate the real operational benefits of the tool. Objective and quantifiable success criteria will be defined for each RQ in order to validate or refute the corresponding hypotheses. As previously indicated, the research questions can be addressed by most of the scenarios and case studies considered in Pilot3 (see Section 7). Therefore, we will try to validate each question with as many experiments as possible (i.e., using different scenarios, case studies, etc.). Modifications to the hypotheses/research questions or the inclusion of new ones might be required in in the light of the obtained feedback. This section summarises the different research questions for the internal and the external validation actions.

- **Table *4*** summarises the research questions for the internal validation activities. The internal validation activities will aim at quantifying some of the results of the different planned experiments.
- **Table *5*** summarises the research questions for the external validation activities. Note that for the external validation activities some of them will be validated as part of the planned external workshop while others will require ad-hoc interaction with members of the Advisory Board. Finally, some of the research questions relate to the different experiments that are validated as part of the internal validation activities. The goal is to answer research questions which aim at obtaining the impression and feedback from stakeholders and experts on the quantified results of the experiments performed.

Table 4. Research questions and hypotheses for internal validation

ID	Rationale	Research question (RQ)	Hypotheses (HP)	Success criteria	Methodology
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.	**IVA1** - Graphical and numerical comparison between Pilot3 trajectory plans and PACE FPO trajectory plans. Computation of trajectories with different operational conditions.
P3-RQ-IV-020	Validate that advanced estimations on performance indicators are more accurate than simpler approaches.	Will Pilot3 enhance the estimation of the (K)PIs relevant to the airline?	Using advanced estimation techniques and larger data sources, Pilot3 generates more accurate estimation of different (K)PIs	• The error on the prediction of performance indicators (e.g., passenger missed connections) using advanced techniques and more data sources will be **lower than** with deterministic estimations.	**IVA2** - Computation of error on prediction of variables for deterministic, heuristic, machine learning, air and ground performance estimators.

EUROPEAN UNION

ID	Rationale	Research question (RQ)	Hypotheses (HP)	Success criteria	Methodology
P3-RQ-IV-030	Validate that advanced estimations on operational uncertainty are more accurate than simpler approaches.	Will Pilot3 enhance the estimation of operational uncertainty parameters?	Using advanced estimation techniques and larger data sources, Pilot3 generates more accurate estimation of different operational ATM estimators.	• The error on the prediction of the Operational ATM Estimators (e.g., expected holding time) using advanced techniques and more data sources will be **lower** than with deterministic estimations.	**IVA2** - Computation of error on prediction of variables for deterministic, heuristic, machine learning, air and ground performance estimators.
P3-RQ-IV-40	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.	**IVA3 -** Visual inspection based on the results arising from the different experiments.

ID	Rationale	Research question (RQ)	Hypotheses (HP)	Success criteria	Methodology
P3-RQ-IV-50	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.	**IVA3 -** Visual inspection based on the results arising from the different experiments.
P3-RQ-IV-60	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	**IVA3 -** Visual inspection based on the results arising from the different experiments.

EUROPEAN UNION

ID	Rationale	Research question (RQ)	Hypotheses (HP)	Success criteria	Methodology
P3-RQ-IV-70	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).	**IVA4 -** Comparison of the results obtained by the optimised trajectory plan of the Pilot3 versus the integration of the baselines with the most updated information.

ID	Rationale	Research question (RQ)	Hypotheses (HP)	Success criteria	Methodology
P3-RQ-IV-80	Validate that the optimised planned trajectories perform better than baselines when the trajectories are realised. I.e., the optimisation validated in P3-RQ-IV-70 when executed is still maintained.	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?	On average, it is expected to obtain the same results as for P3-RQ-IV-070.	• 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).	**IVA5-** Comparison of the results obtained by simulation of the Pilot3 trajectory plan versus the simulation of the operational flight plan considering uncertainty

ID	Rationale	Research question (RQ)	Hypotheses (HP)	Success criteria	Methodology
P3-RQ-IV-90	Validate that having advanced estimators of PI and ATM operational parameters provide less variance on the executed trajectory. I.e., the improved estimation of performance indicators and operational ATM parameters validated in P3-RQ-IV-20 and P3-RQ-IV-30 provide a benefit when the trajectories are executed.	For a given triggering event, will the advanced estimation of PI and operational ATM estimators provide more reliable outcomes?	The more advanced prediction of indicators and operational uncertainty will lead to lower variance between planned and realised trajectory as uncertainties will be better modelled at the triggering point leading to lower differences between planned uncertainties and realised uncertainties.	• Pilot3 will provide, on average, a **lower variance between predicted and realised** predictions for objectives (cost and OTP) and cost components (KPIs) when advanced estimation of indicators and operational ATM estimators are used.	**IVA5 -** Comparison between planned and realised predictions considering different sub-case studies.

ID	Rationale	Research question (RQ)	Hypotheses (HP)	Success criteria	Methodology
P3-RQ-IV-100	Validate the benefit of Pilot3 when deployed at network level for airlines.	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)?	The benefits of Pilot3 when operated at network level will provide better values for airlines KPIs with respect to baselines mechanisms to tactically manage disruptions.	On **average**, **airlines** operational KPIs of **cost** and percentage of **flights reaching OTP** will improve with respect to baselines.	**IVA6 -** Simulation in the fast time simulator Mercury
P3-RQ-IV-110	Validate the benefit of Pilot3 when deployed at network level for passengers.	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)?	The benefits of Pilot3 when operated at network level will provide better values for passengers KPIs with respect to baselines mechanisms to tactically manage disruptions.	• On **average**, **passengers'** indicators (**delay** and percentage of **missed connections**) will improve with respect to baselines.	**IVA6 -** Simulation in the fast time simulator Mercury

EUROPEAN UNION

ID	Rationale	Research question (RQ)	Hypotheses (HP)	Success criteria	Methodology
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 **majority** of the respondents should "agree" that Pilot3 provides clear information to the pilot • None of the respondents should indicate "strongly disagree" and "disagree" option	**IVA7** - 6-point Likert scale for the "General acceptability" questionnaire (see Appendix A).
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?	The 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 aids the pilot to take a more informed decision • None of the respondents should indicate "strongly disagree" and "disagree" option	**IVA7 -** 6-point Likert scale for the "Easiness of understanding of the information" questionnaire (see Appendix A).

EUROPEAN UNION

ID	Rationale	Research question (RQ)	Hypotheses (HP)	Success criteria	Methodology
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?	The 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 **majority** of the respondents should "agree" that the mechanism for the interaction is acceptable enough. • None of the respondents should indicate "strongly disagree" and "disagree" option	**IVA7 -** 6-point Likert scale for the "Interaction with the system" questionnaire (see Appendix A).

* A recent study [9] shows that average relative error of elapsed time and fuel consumption measured in the integrated optimisation results (i.e., the determination of the optimum speed over the whole cruise phase) obtained by BADA4 performance model accounts for 1.1% and 0.81% respectively with respect to the results obtained by aircraft manufacturer software which use the most accurate source of aircraft performance data. In this study, the aircraft performance of 3 different narrow-body and 4 different wide-body aircraft were assessed for two different optimisation criteria: Maximum Range Cruise (MRC) and Long Range Cruise (LRC) conditions.

All aircraft types were from the same manufacturer (Boeing), although the exact aircraft types and models are not specified due to confidentiality issues. It was found that the **maximum relative error of the fuel consumption** and **elapsed total time** for **narrow-body aircraft** accounts for **1.40%** (corresponding to LRC conditions) and **4.72%** (corresponding to MRC conditions) respectively, whereas in the case of **wide-body aircraft** the respective errors for **fuel consumption** and **elapsed time** estimation account for **2.17%** and **2.85%** (both for LRC conditions). Having in mind that the typical values of CI used by airlines are comprised between MRC (i.e., CI=0) and LRC conditions [14], the given errors can provide a solid foundation for setting the success criteria when comparing the performance of trajectory plans generated by Pilot3 and PACE FPO that are due to discrepancies in the aircraft performance models. However, in the absence of the relevant results for CIs other than those corresponding to MRC and LRC conditions (which are typically selected in the case of operations in normal conditions) and with the lack of information on the actual aircraft types and conditions on which these discrepancies have been evaluated in the literature, it is anticipated that discrepancy in results of the respective indicators should be **at least equal** or **even higher** in the case of other CIs selected for the operations in a disruptive environment. Pilot3 will also evaluate the trajectories at different operational conditions, as presented in Section 4.1 in order to adjust these targets if required.

Table 5. Research questions and hypotheses for external validation

ID	Rationale	Research question (RQ)	Hypotheses (HP)	Success criteria	Methodology
P3-RQ-EV-010	Validate the overall acceptance of Pilot3 by experts from interface.	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?	Given its user-friendly interface as well as a broad amount of well-structured information provided to the pilot, Pilot3 is deemed as a very desirable decision support tool for commercial use by the airlines with different business models.	• The final score provided by the individual experts should range between 8 and 10	**EVA1** - Flow-chart diagram for global acceptance. (see Appendix A)
P3-RQ-EV-020	Validate the overall acceptance of Pilot3 by crew from interface.	Given the overall concept of HMI presented, would the pilot be satisfied to have such decision support tool onboard?	With its user friendly HMI interface which displays the large number 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 onboard.	• 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	**EVA1** - 6-point Likert scale for the "Pilot's overall acceptance of the tool" questionnaire. (see Appendix A)

ID	Rationale	Research question (RQ)	Hypotheses (HP)	Success criteria	Methodology
P3-RQ-EV-030	Validate the simplicity but completeness of the information presented to crew.	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 provides clear information to the pilot • None of the respondents should indicate a "strongly disagree" and "disagree" option	**EVA1 -** 6-point Likert scale for the "General acceptability" questionnaire. (see Appendix A)
P3-RQ-EV-040	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 **majority** of the respondents should "agree" that Pilot3 aids the pilot to take a more informed decision • None of the respondents should indicate "strongly disagree" and "disagree" option	**EVA1 -** 6-point Likert scale for the "Easiness of understanding the information" questionnaire. (see Appendix A)

EUROPEAN UNION

ID	Rationale	Research question (RQ)	Hypotheses (HP)	Success criteria	Methodology
P3-RQ-EV-050	Validate the interface to receive input from the crew.	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 variability for each alternative will aid the pilot to make final decision on trajectory selected • None of the respondents should indicate "strongly disagree" and "disagree" option	**EVA1 -** 6-point Likert scale for the "Interaction with the system" questionnaire. (see Appendix A)
P3-RQ-EV-060	Validate the information provided in terms of uncertainties.	Is the information presented to capture the uncertainty on the planned trajectory considered adequate by the crew?	The information on the confidence on the outcome of each trajectory planned and displayed in the HMI will aid the pilot to better assess the benefits of trajectories presented against each other.	• The **majority** of the respondents should "agree" that variability for each alternative will aid the pilot to make final decision on trajectory selected • None of the respondents should indicate "strongly disagree" and "disagree" option	**EVA1 -** 6-point Likert scale for the "Pilot's overall acceptance of the tool" questionnaire. (assess third statement only) (see Appendix A)

ID	Rationale	Research question (RQ)	Hypotheses (HP)	Success criteria	Methodology
P3-RQ-EV-070	Validate that alternatives provided are relevant for different experiments.	Are the solutions 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 the set of meaningful solutions.	• The **majority** of the respondents should "agree" that Pilot3 provides meaningful solutions in the given operational context • None of the respondents should indicate "disagree" option	**EVA2 –** 6-point Likert scale for the "Goodness of solutions in different operational context" questionnaire, based on the outcome of experiments validated in IVA4 and IVA5. (see Appendix A)
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 final score provided by the individual experts should range between 8 and 10	**EVA2 -** Flow-chart diagram, based on the outcome of experiments validated in IVA4, IVA5. (see Appendix A)
P3-RQ-EV-090	Validate overall results from Pilot3 at network level.	Are benefit obtained by Pilot3 at network level relevant to airlines and passengers?	Pilot3 will provide benefits that are relevant for both passengers and airlines when considered at network level.	• The **majority** of the respondents should "agree" that the benefits obtained by Pilot3 at network level are relevant for airlines and passengers. • None of the respondents should indicate "disagree" option	**EVA3 -** 6-point Likert scale for the "Goodness of solutions with respect to airline and passenger key metrics" questionnaire, based on the outcome of experiments validated in IVA6. (see Appendix A)

ID	Rationale	Research question (RQ)	Hypotheses (HP)	Success criteria	Methodology
P3-RQ-EV-100	Validate overall acceptance of Pilot3 considering performance network level results.	Do experts find that Pilot3 will provide benefits to airlines and passengers?	Given the benefit provided with respect to the network validations Pilot3 will be accepted by airlines.	• The final score provided by the individual experts should range between 8 and 10	**EVA3 -** Flow-chart diagram for "overall acceptance of Pilot3 considering network-wide level", based on the outcome of experiments validated in IVA6. (see Appendix A)

7 Scenarios and case studies

As explained previously, the validation of Pilot3 will be based on the simulation of specific flights in given operational conditions. For this purpose, different aspects need to be considered such as, elements related to flight characteristics (e.g., type of aircraft, route length), operational aspects (e.g., airline type, characteristics of arrival TMA), environmental considerations (e.g., weather, ATFM conditions), event which triggers the use of Pilot3 (e.g., late arrival at TOC with respect to planned), or the configuration of Pilot3 tool.

In order to structure all these different considerations, a five-level hierarchy has been defined, recalling the definitions from Section 2.3.1:

1. **Scenario** is high-level item linked to **specificities of the routes and operations** (aircraft mission) that are modelled. A scenario specifies aircraft mission 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**.

A unique identifier will be produced per experiment so that it is possible to refer to it during the different development and validation activities. This identifies will be composed of sub-identifiers for the different components of the experiment in the following manner:

1. **Scenario:** A unique scenario will be identified by an id in the form of P3-SCN-xxx. Section 7.2.1 presents a set of potential scenarios already defined in this document.
2. **Sub-scenario**: A unique identifier will be created for the different sub-scenarios in the form of SBSCN-xx. In this case, this deliverable does not provide an exhaustive list of potential sub-scenarios but only the components that define a sub-scenario (see Section 7.2.2). Therefore, their naming will be created as the sub-scenarios are defined for their use and reused as appropriate, i.e., when the same combination of parameters is considered in a different experiment.
3. **Case study**: A unique identifies in the form of CS-xx will be used. Section 7.2.3 presents the potential case studies considered in Pilot3.

EUROPEAN UNION

4. **Sub-case study**: The identifier of the sub-case studies will be of the form of SBCS-xx. As with the sub-scenarios this deliverable does not present an exhaustive list of all the potential sub-case studies but the parameters considered to define them (see Section 7.2.4).
5. **Parametrisation**: The parametrisation of a given scenario, sub-scenario, case study, sub-case study will be identified by a numerical id (xxx).

Following this convention an experiment will be defined by an id in the form of P3-SCN-xxx-SBSCN-xx-CS-xx-SBCS-xx-xxx, e.g., P3-SCN-100-SBSCN-01-CS-10-SBCS-08-001.

Finally, note that as mentioned in Section 3.5, 4.1 and 4.3, for the verification activities and the internal validation activities which focus on the validation of the components of the prototype (IVA1, IVA2 and IVA3), **test-cases** will be defined ad-hoc. These will be based, when possible, on the experiments previously described but adjusting them as needed, and containing the minimum information required for the verification or validation purpose.

7.1 Methodology to define experiments

As explained in Section 2, the definition and selection of the scenarios and case studies play an important role during the validation campaign; and as described Section 1.1, an interrelation between the specification of particular scenarios and case studies and particular functionalities exists, as each scenario will generate low-level requirements that need to be considered as part of the implementation (e.g., in which arrival airport the machine learning to predict the holding should be trained). Moreover, when a given experiment is defined, a set of data requirements will arise. For this reason, a consultative approach is suggested. The consortium will define the characteristics of the different components of the experiment (scenario, sub-scenario, case studies, sub-case study and parametrisation) which are relevant but the specific characteristics of these will be consulted with ad-hoc interactions with the Advisory Board members. This will facilitate the selection of experiments which are more relevant but also for which insight and data could be acquired from the Advisory Board.

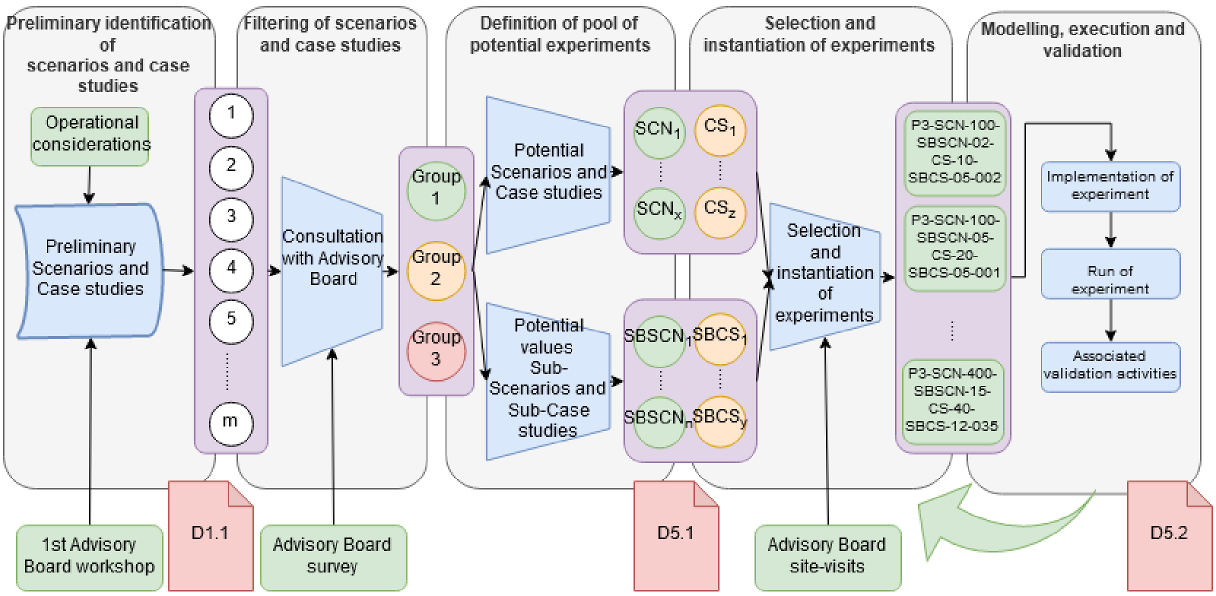


Figure 12. Definition, selection, instantiation and evaluation of experiments

Figure 12 presents the approach followed to create the pool of potential experiments to be modelled and evaluated in Pilot3. First, the consortium carried out a **preliminary identification of scenarios and case studies** by identifying a set of potential operational aspects to consider when defining experiments to evaluate in Pilot3. These included elements such as airline characteristics, type of flight (short/medium-haul, long-haul), event triggering the need of using Pilot3. During the first Advisory Board meeting (held in London on the 7th February 2020), feedback was gained from airlines and experts on which aspects are more relevant. These included, for instance, the consideration of aspects such as unexpected changes of weather conditions (e.g., wind ahead), unexpected events or/and major disruptions, a weather forecast update ("recalculate the trajectory"). The Advisory Board also indicated the need of considering sensitivity studies of the Pilot3 solutions to capture the robustness of the solutions and the fact that TOC is a major flight milestone, and thus a good point for querying Pilot3. Information was also gathered on which operational environments (within ECAC and intercontinental) might be suitable for a tool such as Pilot3. With all this information, a list of potential scenarios/case studies/triggering elements was created and reported in D1.1 - Technical Resources and Problem definition [10]. These included:

- two types of flights: short/medium-haul (within Europe) and long-haul (with oceanic segment).
- for the short/medium-haul example a total of 12 different case studies were described considering different triggering events.
- for the long-haul flight, a total of 13 different case studies were identified.

Then, a survey was conducted to members of the Advisory Board. This survey was part of a **filtering of scenarios and case studies** to identify which case studies were found more relevant. Ad-hoc site visits were also conducted. The details of these interactions with the Advisory Board are reported in D3.1 - Airlines data collection report [12]. The outcome of these actions was a prioritisation on three groups the different case studies.

In this deliverable, all the information gathered have been used to **define the pool of potential experiments** that can be modelled and executed in Pilot3. This includes, the definition of potential scenarios and case studies (as presented in Section 7.2.1 and 0) and the identification of values to be used to model the different sub-scenarios and sub-case studies (as reported in Section 7.2.2 and 0). Producing an exhaustive and strict definition of all the potential experiments is avoided to enable the flexibility of further select and instantiate experiments with feedback from the Advisory Board.

The next step on the modelling of experiments is the **selection and instantiation of experiments**. Following the Agile approach described in this document (see Section 1.1), interactions within the consortium (and the Topic Manager) and with the Advisory Board (e.g., ad-hoc site visits) will be used to select which experiments to model and prioritise. Consideration such as relevance, differentiation with respect to previous experiments, possibility of obtaining data from Advisory Board members and re-usability of previous experiments to maximise validation activities will be considered. Note that in this phase, some changes might be produced on the actual instantiation of the experiments (e.g., modifying the reference flight of the scenario (see Section 7.2.1)). The number of different experiments executed will increase as the project progresses and the prototype matures. This will be reflected with the feedback obtained from the first release of the prototype and the activities conducted until the final release (see Section 8).

Finally, the experiments will be implemented, run and feed the different validation activities. The outcome of these validation activities will be reported in D5.2 - Verification and validation report.

7.2 Potential values for definition of experiments

7.2.1 Potential scenarios

The first hierarchy level specifies the most basic variables which help to particularise the specific flight. The following variables have been considered to provide a range of scenarios which cover different operational characteristics in which Pilot3 can be used:

1. **Operating airline and O/D pair** - specifies the airline operating the route.
2. **Type of route operated** - different scenarios are defined to cover short/medium-haul route operated in the intra-ECAC airspace and long-haul routes operated in the oceanic airspace (e.g., North Atlantic, South Atlantic).
3. **En-route uncertainty** - captures different level of uncertainty which can be faced during en-route phase.
4. **Destination characteristics** - captures different levels of complexity (e.g., multi-airport TMA, dense TMA, etc.,) and procedures (e.g., tromboning) that exist at different TMA airspace.
5. **Airline type** - defines the airline type (i.e., low cost airlines, full-service airline)
6. **Destination type** - distinguish between hub and non-hub airports.
7. **Time of the day** - distinguishes between different periods of the day in which the flight takes place.

Other aspects that will be also considered during the selection process include variability on **taxi-in times**, airlines **buffers**, and data availability that can be obtained from external sources (e.g., Advisory Board).

It is worth mentioning that creating a new scenario requires a significant amount of effort, as data acquisition, preparation and in some cases model training of some of the advance indicators and operational estimators will be required. For this reason, we attempt to confine to a reduced number of scenarios, but providing a wide range of operational environments. Considering scenarios which are operated (or similar to operated routes) by members of the Advisory Board is also considered of relevance as more in-depth feedback might be acquired from the results during the preparation of the experiments and the validation activities.

Finally, it is worth noticing that in order to properly model the impact of Pilot3, not only the route and the flight characteristics are needed, but the whole network for the airline (including follow up rotations of the aircraft and passengers itineraries) might be required. For this reason, when possible, and unless more specific data can be obtained, Pilot3 will base the scenarios on historical flights as recorded in the DDR2 database with the passengers itineraries as created in the ER3 project Domino [3]. This will also facilitate estimating other operational parameters required to generate the OFP such as initial Cost Index or take-off weight.

Table *6* presents nine different scenarios which will be further particularised through sub-scenarios defined at next level.

Table 6. The description of scenarios identified for the 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	Multi-Airport TMA	FSC	Hub	Morning Afternoon
P3-SCN-200	VLG: ENGM (OSL) - LEBL (BCN)	Intra-ECAC	Normal	Tromboning	LCC	No Hub	Afternoon Evening
P3-SCN-300	NAX: ENGM (OSL) - LEAL (ALC)	Intra-ECAC	Normal	Conventional arrival - secondary airport	LCC	No Hub	Morning
P3-SCN-400	VLG: LEBL (BCN) - LEAM (LEI)	Intra-ECAC	Normal	Conventional arrival - military airspace interference	LCC	No Hub	
P3-SCN-500	ICE: BIKF (KEF) - EFHK (HEL)	Intra-ECAC	High	Open STAR	LCC	Hub	Afternoon
P3-SCN-600	DLH: KJFK (JFK) - EDDF (FRA) or EDDM (MUC)	North - Atlantic	High	Tromboning	FSC	Hub	Morning
P3-SCN-700	SWR: KJFK (JFK) - LSZH (ZRH)	North - Atlantic	High	TTA	FSC	Hub	Morning
P3-SCN-800	BAW: KJFK (JFK) - EGLL (LHR)	North - Atlantic	High	Multi-Airport TMA	FSC	Hub	Morning Afternoon Evening
P3-SCN-900	BAW: SAEZ (EZE) - EGLL (LHR)	South-Atlantic	High	Dense TMA	FSC	Hub	

*As mentioned, the actual flight used as a reference might vary once the scenarios are instantiated as part of an experiment. However, each scenario id will univocally relate to a specific flight.

7.2.2 Potential sub-scenarios

The sub-scenarios aim to further particularise the given scenario in terms of operational environment and weather. Thus, the sub-scenario specifies the four additional variables:

- **ATFM conditions** - captures the difference between operations in normal conditions and those when the network is very disrupted. The network disruption will be based on the amount of ATFM regulations, and the delay assigned by these regulations, on historical operational days. This congestion has an impact on the expected cost of different amounts of arrival delay, as in disrupted operational environments higher propagation of delay could be expected.
- **Weather** - differentiates between the three types of weather conditions that may occur during the flight execution.
- **Curfew** - distinguishes whether the flight is constrained by the curfew at destination airport or at the end of operational day.
- **Entitlement to compensation (Reg. 261)** - distinguishes from scenarios where if the required delay threshold is met passengers are entitled to claim compensation due to Regulation 261 [5] from scenarios where the airline is except to pay these compensations.
- **Target Time of Arrival** (TTA) - distinguishes whether the airport applies TTA or not**.**

Table *7* presents the different values that can be used per variable to define the sub-scenarios. As previously described, in this deliverable, the possibilities are presented but not exhaustively instantiated. Note that with the number of potential values that can be selected per variable, there are a total of 72 different sub-scenarios. Once the sub-scenarios are instantiated, as required, a name will be given to them (e.g., SBNC-01 - ATFM - Normal; Weather - Normal; Curfew - No; TTA - Yes), this naming will then be maintained for any other scenario which is particularised with this specific values for the sub-scenario parameters.

Table 7. The description of sub-scenario identified for the validation campaign

AFTM	Weather	Curfew	Entitlement to compensation (Reg. 261)	TTA
• Normal • Network very disrupted	• Normal • Strong beneficial wind • Strong decremental wind	• No • Yes, at arrival • Yes, at the end of the day	• Yes • No	• Yes • No

7.2.3 Potential case studies

The case study refers to different events that may trigger Pilot3. **Table *8*** identifies the nine major events with a brief description for each of them and the potential parametrisation that can be performed for each one of them in order to create different experiments. Note that some parameters are common to all case studies, namely: departure delay and cost of fuel.

These potential case studies have been derived from the consultation activity (surveys) with the Advisory Board.

Table 8. 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	–
CS-20	Wrong en-route estimates with respect to last planned trajectory	• Magnitude of time deviation • Location en-route
CS-30	Significant route shortcut in cruise (i.e., conditional route, MIL area inactive, etc.)	• Magnitude of distance deviation • Location en-route
CS-40	Delay at destination TMA updated in cruise	• Delay at TMA
CS-50	TTA at destination TMA updated	• Variation with respect estimated time of arrival
CS-60	Updated weather forecast	• Weather characteristics
CS-70	Convective weather ahead	• Weather characteristics
CS-80	Turbulence in current FL	• Weather characteristics
CS-90	Oceanic clearance changed	• Magnitude of deviation • Location

7.2.4 Potential sub-case studies

Sub-case study particularises the configuration of Pilot3 and in particular 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. For example, indicating if heuristic or an advanced model should be used to estimate a given parameter with air or ground information, in case of equivalent impact on different indicators, which ones should be prioritised.

For the naming convention a similar approach will be used as for the sub-scenarios, i.e., once a sub-case study is instantiated a code is assigned to it and reused if the same configuration is applied in a different experiment. However, note that in this case the number of potential sub-case studies is larger than the combination of the high-level variables described below (which would produce a total of 48 combinations), as for example, it could be possible to indicate individually which PI should be computed with each of the different possibilities (heuristic or machine learning, with air or including ground information).

Table 9. 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 with air information • Heuristic with ground information • Machine Learning with air information • Machine Learning with ground information	• Heuristic with air information • Heuristic with ground information • Machine Learning with air information • Machine Learning with ground information	• Cost of fuel • Cost of IROPs • Other cost

8 Schedule of the verification, integration and validation

The section provides the schedule of the activities that will be performed during verification and validation campaigns. The schedule of all activities is highly related to the development and functional readiness of different prototype versions of the tool. In other words, particular prototype versions will enable the validation of specific internal and external activities at specific points in time. In order to effectively follow the schedule, the brief introduction on the different type versions and the description on their main functionality is provided in Section8.1. It is worth mentioning that some of activities in the verification, internal and external validation are performed in parallel and have a substantial impact on each other.

8.1 Prototype software versioning

Table 10 presents a summary of the different versions that are planned for the Pilot3 prototype. Note that due to the Agile development approach this might be adjusted during the development phase and it should be considered only as an indication of the main functionalities that will be added/considered for each version. Versions might be combined and the functionalities that are considered for inclusion adjusted based on the internal prioritisation of functionalities and experiments, along with feedback from interaction with the Advisory Board and Topic Manager. As described in the Development methodology in Section 1.1, the tasks that are implemented will also be subject to the selection of case studies that will be prioritised.

Table 10. Prototype versions

Version	Date aim at	Description	Characteristics – Main functionalities

V0.1	M11 - SEP20	Software architecture integrated between the different modules.
		
		Interface between modules and simple/dummy behaviour.
			• Performance Indicators Estimator (PIE)
			  ○ Low level indicators estimators defined as simple heuristics
			• ATM Operational Estimator (AOE)
			  ○ Low level operational estimators defined as simple heuristics
			• Alternative Generator (AG)
			  ○ Cost function build from PIE outcome
			  ○ Trajectory optimisation architecture without uncertainty
			• Performance Assessment (PA)
			  ○ Comparison solutions from AG
			• Data
			  ○ Data for 1st experiment identified and gathered
			  ○ Input for model prepared for 1st experiment.

V0.2	M13 - NOV20	Improved heuristics for Performance Indicators Estimator and ATM Operational Estimator.	• Performance Indicators Estimator (PIE)
			  ○ Better estimated heuristics
			• ATM Operational Estimator (AOE)
			  ○ Better estimated heuristics

V0.3	M14 - DEC20	First advanced PIE and AOE estimators.
		
		Trajectory generator validated
			• Performance Indicators Estimator (PIE)
			  ○ Use of machine learning techniques for prediction of some indicators
			• ATM Operational Estimator (AOE)
			  ○ Use of machine learning techniques for prediction of some indicators
			• Alternative Generator (AG)
			  ○ Validated with respect to commercial trajectory generator
			  ○ Incorporation of some uncertainty on the optimisation
			• Data
			  ○ Data required for machine learning gathered.

Version	Date aim at	Description	Characteristics – Main functionalities

V1.0	M16 - FEB21	Full first prioritised case studies simulated.
		
		Alternative Generator with trade-off capabilities.
		
		Performance Assessment with VIKOR.
		
		HMI prototype.
			• Alternative Generator (AG)
			  ○ Lexicographic optimisation
			• Performance Assessment (PA)
			  ○ Implementation of VIKOR for comparison of solutions.
			• Data
			  ○ Data required for prioritised case studies.

V1.1	M21 - JUL21	Changes to incorporate feedback from External Validation (external workshop). Prepare and execute new case studies.	
V1.2	M24 - OCT21	If changes are required based on final verification and validation activities.	

8.2 Development, verification and validation schedule

The diagram below (**Figure 13**) presents a planning on the development, verification and validation schedule. Once again, note that this will be reviewed as the project is developed as different implementation cycles are planned within each version considering the prioritisation of tasks from the backlog to be implemented.

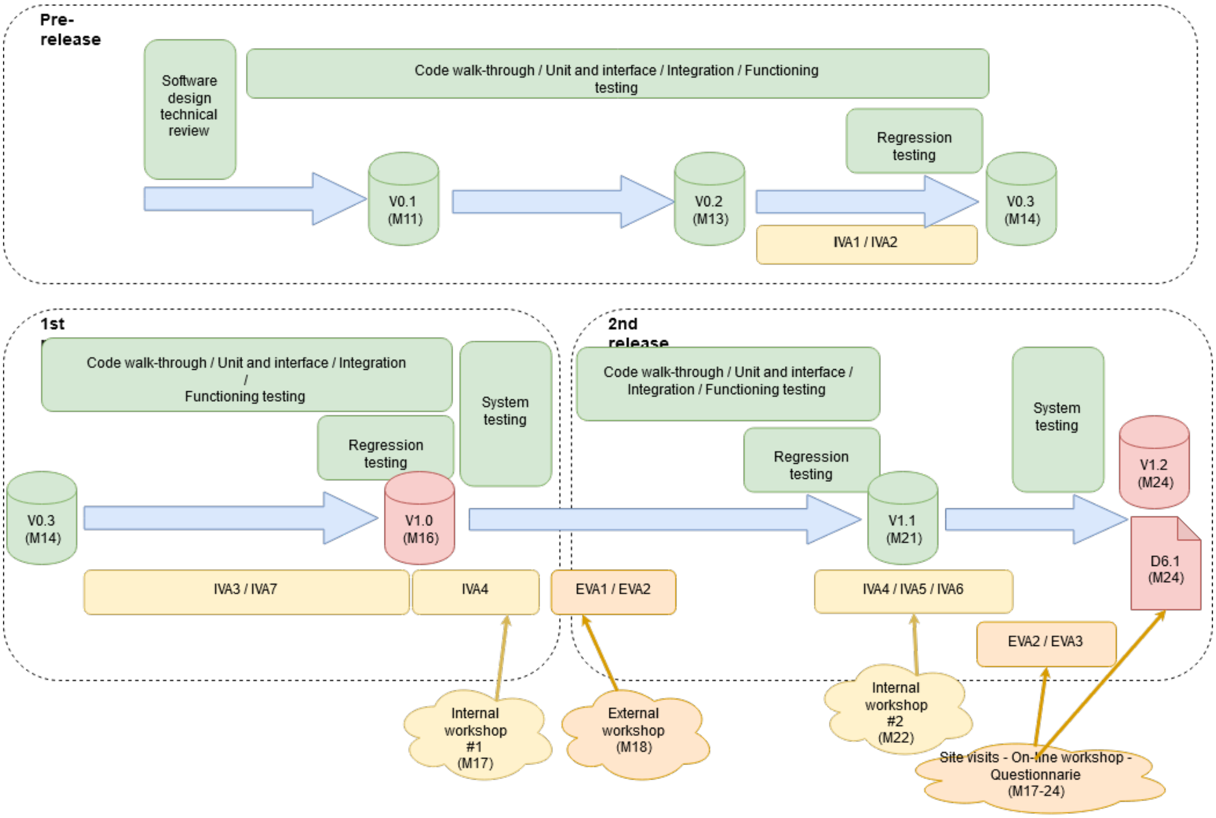


Figure 13. Diagram on software development, verification and validation activities planned

As presented, the verification activities will be performed in parallel to the development of the prototype. Some verification actions might require a given level of maturity of the prototype (e.g., the execution of functional test-cases). Prior to the completion of a version which provides new functionalities, regression tests will be conducted to ensure that previous verified functionalities have not inadvertently being affected. System testing will be conducted before the release of the prototype to the Topic Manager.

Two different type of internal validation activities are appreciated: validation activities aimed at validating the different components of the model (IVA1, IVA2 and IVA3) and the HMI (IVA7), and validation activities aimed at estimating the benefits of Pilot3 (IVA4, IVA5 and IVA6). In the first case, these internal validation activities will be performed in parallel with the model development and verification, as they influence these activities. The internal validation actions which aim at estimating the benefit of the fully working prototype, however, are planned once the versions are stabled and executed as part of internal validation campaigns (with attached internal workshops).

External validation actions are planned after the internal validations are conducted (as the outcome of these is required). Note that in some cases, particularly for the final release, an overlap might exist as new experiments are internally validated and then feedback seek as part of external validation. The first external validation campaign will be carried out in a dedicated workshop. The second external validation actions will be conducted in ad-hoc interactions as the model and the results are produced with particular focus at the end of the development phase.

Finally, continuous interaction with the Advisory Board will be maintained to ensure that their feedback is incorporated in the prioritisation, selection and concretisation of experiments and functionalities. Feedback on which functionalities to modify or incorporate between the first and the final release will be gathered, and further actions that should be conducted after the project competition to facilitate the industrialisation of the prototype (in its total or partial form) will be gathered in a dedicated deliverable (D6.1 - System evolution and uptake), which will be released with the final prototype version.

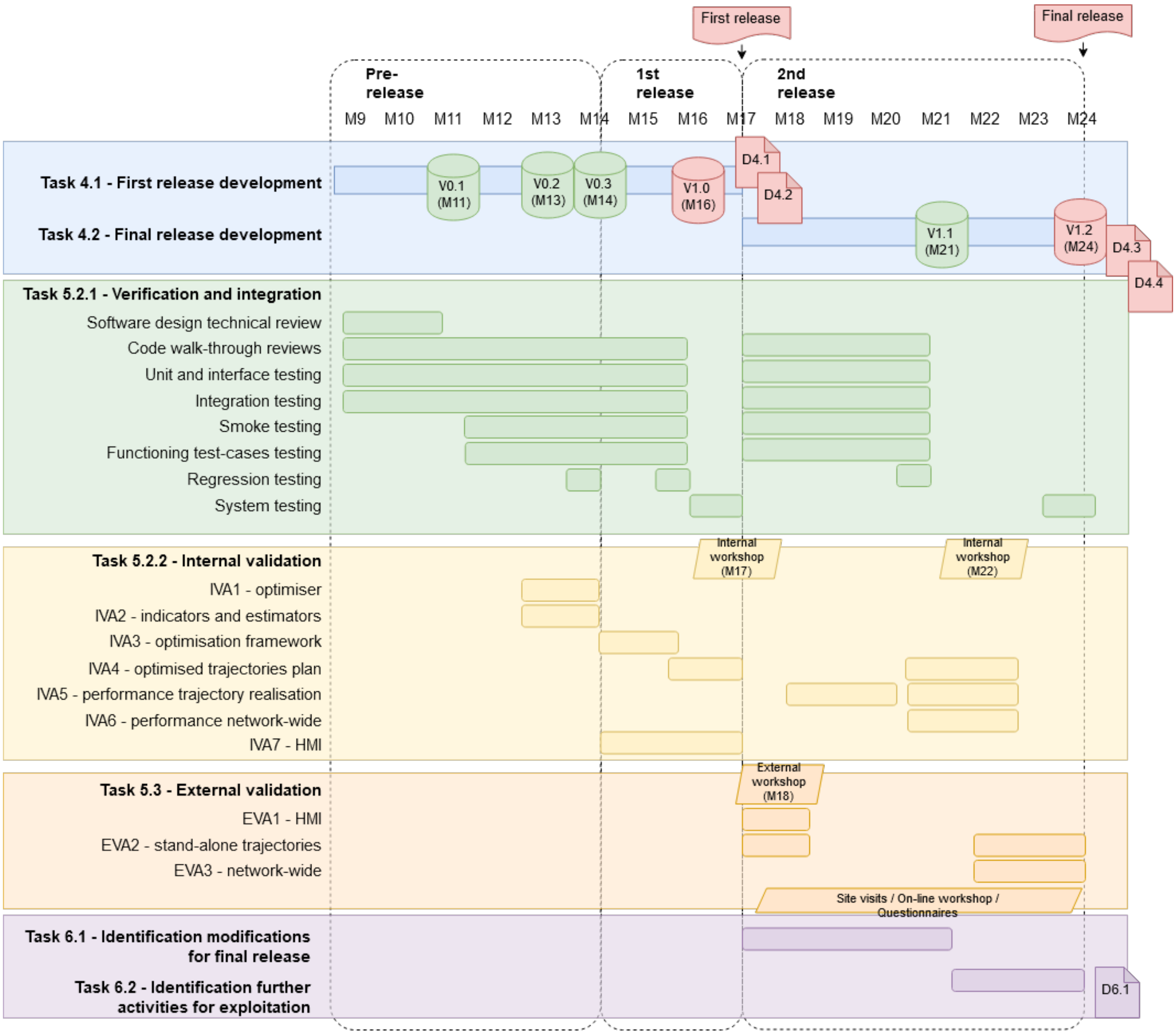


Figure 14. Gantt diagram with activities for development, verification and validation

Figure 14 presents a more detailed Gantt diagram on the different tasks that need to be performed. The schedule of the verification and validation plan are highly related to the schedule defined for different prototype version development specified in WP4 and can be monitored by two separate tasks:

- Task 5.2 - Internal verification and validation, which is composed of two sub-tasks:

EUROPEAN UNION

  - Task 5.2.1 - Verification and Integration (**M9-M24**), which aims at ensuring the continuous **verification of software design and testing**. This sub-tasks needs to be performed in an iterative manner to reflect the development of prototype versions based on the Agile principle evolving from very basic prototype version (i.e., V0.1) towards mature fully functional prototype (i.e., V1.2).
  - Task 5.2.2 - Internal validation (**M13 - M22)**, which aims at ensuring the validation of each of the seven different actions (IVA1 - IVA7) defined in Section 4, involving the interaction with the experts within consortium and the Topic Manager.
- Task 5.3 - External validation (**M17 - M24**) will involve the interaction with Advisory Board members, Topic Manager and other experts and stakeholders to ensure the validation of each of the three different actions (EVA1 - EVA3) defined in Section 5.

Task 5.2.1 will involve several verification tests which will be performed in an iterative manner:

- **Software design technical review (M9-M11)**, which presents the initial test which will ensure that design approach for the software satisfies the different requirements. This test will be performed at the onset of the verification and integration process, as it verifies that the architecture designed for the prototype is adequate.
- **Code walk through review** provides the first step into the verification of the developed code by providing peer-review between members of the development team and will be conducted for each prototype version developed. However, the code walk through review can be performed at any point to provide the insight into the developed code, if needed.
- **Unit and interface testing**, as a basic layer, will be performed for each functionality to ensure individual and independent testing of parts of the Pilot3 prototype functions and/or interfaces. As it can be observed from Gantt diagram, the testing will be performed in a continuous manner along the prototype development.
- **Integration testing** will ensure that different modules are working fine when combined as a group. Similar to unit and interface testing, integration testing will be performed for each prototype version along the prototype development.
- **Smoke testing** will require certain level of prototype maturity, as it verifies that no critical/simple failures, severe enough such as runtime errors are produced. Therefore, this type of tests will be conducted after some functionalities are implemented (i.e., after V0.1).
- **Functional test-cases testing** will need to ensure that new functionalities implemented in the particular prototype version do not generate errors in the code, as well as to ensure that the actual outputs match the expected outputs. As in the case of smoke testing, functional test-cases requires to have at least low-level prototype version released (at least, V0.1). In addition, the consortium members will need to prepare several **test-cases** to test the different functionalities of the prototype. Note that the focus here is on fixing errors in the code rather than validating the prototype.
- **Regression testing** will be performed on the functional mature prototype (i.e., V0.3, V1.0 and V1.1) to ensure that the implementation of the new functionality into each respective

prototype version has not generated the side effects on previously verified functionalities. These tests will consist on re-verifying previously performed functional test-cases.

- Finally, **system testing**, as the highest in the hierarchy, will be performed considering the complete, integrated system as a whole to ensure that the system requirements have been considered. Thus, this test will be performed prior to the software releases to the Topic Manager (i.e., with V1.0 (first release) and V1.2. (final release)).

Task 5.2.2 will be carried out internally among the experts within consortium and in the tight coordination with the Topic Manager. As the validation of actions defined in Section 4 requires different levels of functionality to be implemented in different prototype versions, it can be assumed that depending on the readiness of the respective prototype version, the validation of results will be performed gradually, namely:

- The first two internal validation actions that refer to the validation of the optimiser and indicators and estimators predictions (i.e., **IVA1** and **IVA2**) will be addressed in parallel to the prototype development. Note that the primer aim of the verification testing is to provide the test of platform to fix errors and to enable the incremental development of the prototype versions. The validation of the models will be conducted as part of these validation activities. These two initial internal validation actions are anticipated to be carried out between **M13** and **M14**, aiming at having a validated trajectory optimiser by M14 (December 2020).

- In the same vein, the internal validation action than tackles the assessment of the optimisation framework (i.e., **IVA3**) will stem from prototype development after implementing Alternative Generator with trade-off capabilities as a new functionality in the system. The validation of this action is envisioned to be performed between **M14** and **M16**. This approach will produce a fully working optimisation framework by M16 (February 2021).

- The internal action that refer to validation of the HMI prototype (i.e., **IVA7**) will be performed between **M14** and **M17**. The consortium partner Innaxis will be involved in the development of the interface and some progress have already been done in the HMI development. Feedback gathered from these internal validation actions will contribute to the improvement and development of the suggested design of the HMI.

- The validation of the benefits of Pilot3 optimised trajectories plans against several baseline plans defined in **IVA4** will require the fully functional prototype and stable, frozen, version of the code (i.e., V1.0), which will be used to evaluate the different experiments. This activity will be performed between **M16** and **M17**, and finally, the results will be presented to **the first internal workshop** that will take place at **M17** and discussed internally with the consortium members (as illustrated in **Figure 14**). All this information will be used to produce the first release of the prototype planned for M17 (March 2021).

- The actions that refer to performance trajectory realisation and performance of the tool at network-wide level (**IVA5** and **IVA6**) will require the fully functional prototype version and frozen version of the prototype which incorporates the feedback of the first external validation actions (i.e., V1.1) in order to assess their benefits. Note that, however, some validations of the prototype planned for IV5 could also be conducted with the prototype with functionalities as developed in V1.0. Additionally, the re-assessment of the optimised trajectory plans defined in **IVA4** will be performed with the new scenarios which are more relevant. The results of these three activities (IVA4, IVA5 and IVA6) will be presented at **the second internal workshop** which will take place at **M22.**

Wrapping up the information provided above, internal validation campaign will rely on the feedback from the internal experts gathered by the means of:

1. **Internal meetings** which will be held periodically to ensure the synchronisation of the tasks distributed within the consortium partners and discuss the results obtained by different prototype versions. For instances, the results of IVA1, IVA2, IVA3 and IVA7 activities will be presented at the internal meeting as they will be available earlier than the results of other IVA activities **according to the schedule**.

2. **First internal workshop** will be organised around **mid month 17** of the project. In this way, the team within consortium will have substantial time to prepare and run the experiments, and to present the obtained results of IVA4 to the rest of the team and the Topic Manager.

3. **Second internal workshop** will take place at **month 22** of the project. During this workshop, the experts within consortium together with the Topic Manager will be shown the results of activities IVA4, IVA5 and, if performed, IVA6. The consortium team will thoroughly review the feedback obtained from the workshops with the external experts and decide which improvements can be still implemented in the "V1.2" prototype version, and which items will be left for the future development (and specified in D6.1 - System evolution and uptake).

Task 5.3 will involve the interaction with Advisory Board members, Topic Manager and other experts and stakeholders. The activities in Task 5.3 will be mainly interrelated to the activities performed in the internal validation campaign. The results obtained during the internal validation will serve as an input to the activities performed in the external validation, and vice versa, the results obtained during external validation campaign will shape some of the activities foreseen by the internal validation and model development of the following version of the prototype. During the external validation campaign, the interaction with the external experts will be carried out through:

1. The **external workshop** will be organised at the end of **month 18** of the project (April 2021), by presenting the results obtained with first release (i.e., V1.0). Note that depending on the COVID-19 outbreak, this workshop might be adjusted to be held on-line in one or several events. The first external workshop will gather the Advisory Board members, Topic Manager and other experts and stakeholders who will assess the benefits of the Pilot3 tool, its potential limitations and possible place for improvements. To accomplish this goal, during the workshop the experts will be introduced with the tool developed by providing them with the insight into:

    - Live or pseudo-live demonstration of the functionalities of the tool and validation of the HMI prototype (including comparison between "plans").

    - Presentation of results obtained with stand-alone simulations at trajectory level with the set of selected scenarios and case studies **based on the results of IVA4**. The majority of these experiments will be the same as those used for the first internal workshop, although some additional scenarios may emerge.

In this way, the external experts will be able to address two external validation actions - **EVA1** and **EVA2** and validate the results by using the specially designed questionnaires.

2. In addition to the external workshop, **dedicated validation activities with experts** will be running in parallel with all activities explained above within the time horizon starting at **month 17** and finishing at **month 23**. The aim of the dedicated validation activities is to use a variety of mechanism (e.g., questionnaires, on-line workshops, site-visits, etc.), so that, if possible, specific feedback on experiments which are relevant for the different airlines members of the Advisory Board so that the results presented can be closer to their real operational context. Once the specific knowledge on the experiments relevant for different airlines member are obtained, the dedicated validation activities will be performed with V1.1 of the prototype consisting on a new round of **EVA2** actions (with results from IVA4 and IVA5) and, if results from IVA6 are available, validation of the network-wide benefits of Pilot3 (**EVA3**). These interactions will also support the identification of further improvements which might be integrated in V1.2 prototype version, or gathered and specified in D6.1 - System evolution and uptake.

Finally, it is worth noticing that even if the external validation actions EVA1, EVA2 and EVA3 revolve around the answering of specific research questions with defined formularies, as indicated previously, this is not the only type of validation and interaction that Pilot3 will incorporate from experts. A more continuous interaction with the Advisory Board will be carried out to ensure that their feedback is promptly incorporated into either the model functionalities or the experiments considered. Note that the Advisory Board might also provide insight on their operations, needs and, in some cases, even datasets that can be used for Pilot3. During the validation campaign (e.g., external validation workshop) general feedback beyond the structured formularies defined in this deliverable will be also gathered.

9 Conclusions

The verification and validation plan is a complex document structured around the concept of Agile principle that will be used during the development of the tool prototype. The **Agile** methodology relies on the idea that different prototype versions are incrementally developed by adding new functionalities into each version, until we reach the fully functional prototype. However, such development triggers a backlog of tasks that need to be developed and performed in a synchronised manner in order to ensure the adequate implementation of the tool. These tasks mainly stem from the process of disaggregation of high-level requirements into lower-level requirements imposed by the implementation of the different functionalities of Pilot3, and on the prioritisation and selection of experiments to be modelled.

As acknowledged by a great number of experts who deal with the verification and validation activities across different fields of applications, even if both activities are distinct on their purpose, some overlap exists. This is particularly the case for functional testing. In Pilot3 we consider that the testing of functionalities on simple test-cases is part of the **verification** activities, while dedicated **validation actions** are defined to validate the **prototype components** (trajectory optimiser, machine learning predictors and optimisation framework). Classical **verification** activities (both **static** and **dynamic**) will be 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, will be conducted prior to both software releases.

The remaining validation actions focus on 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). Specific **research questions** have been defined and will be answered during these validation campaigns. The external validation will focus on the operational benefits of Pilot3 by using some of the outcomes of the internal validation actions, which will quantify and **assess the benefits** of Pilot3. Overall, these validation actions target the functionalities of Pilot3 (considering the optimised trajectories and the expected performance once they are operated under uncertainty) and the **interface** designed for the tool. Further validation actions could be conducted to evaluate the benefit of a tool such as Pilot3 when deployed at network level. This would provide a network view for metrics such as cost and delay metrics for both airlines and passengers. However, these activities do not tackle any specific detailed characteristics of Pilot3, but rather provide an insight on their potential benefit at network level, therefore, even if designed in this deliverable, they will only be performed if the time-frame of the project allows it.

The verification and validation plan underpins its concept and approach on **interaction** with airlines from the Advisory Board and more broadly relevant stakeholders. These interactions are ensured by organising a **workshop** and **dedicated validation activities** (e.g., site visits) which will support 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.

The pool of experiments defined in this deliverable using a five-level hierarchy framework allows the required flexibility to further particularise and select the scenarios to be modelled as part of this Agile approach.

Overall, the proposed verification and validation approach ensures that the team can prioritise the development to the functionalities and scenarios and case studies which are deemed more relevant as part of the different consultation activities. In addition, bottlenecks can be identified promptly for both the software development and the data acquisition and preparation. Note that two internal workshops and interaction between the members of the consortium are also planned to support this promptly detection of potential issues.

10 Next steps and look ahead

The deliverable has presented the comprehensive framework which thoroughly describes the actions for the verification and validation of the prototype. As presented in this plan, the project will follow an incremental development for both its functionalities and experiments (scenarios and case studies) to be modelled. The development in the project has already started as part of WP4 activities. The first activities consist on disaggregating some of the requirements into the different sub-modules of Pilot3 and the definition of the software architecture. The verification activities will start closely monitoring the progress in WP4, and verification actions will start as soon as adequate in a continuous manner, starting with a software design technical review once the software architecture is finalised.

The particularisation and selection of scenarios continue with ad-hoc interactions (site visits) with members of the Advisory Board. Data acquisition and preparation have already started and will be carried out as required to be able to execute the different experiments designed and prioritised. This data acquisition includes the dataset required for the training and validation of the machine learning models in the Operational ATM Estimator and the Performance Indicator Estimator, and substantial work has been already done in the selection of datasets required to model the different case studies (e.g., days with high or low uncertainty in weather). Simplified test-cases will also be produced for the verification of the lower level functions of Pilot3 engine.

The current development of the model focuses on the production of the pre-release functionalities but once the first components of the Pilot3 prototype are implemented, internal validation activities to validate the trajectory optimiser and the machine learning algorithms to predict the indicators and the operational estimator will be conducted. The consortium aims at having the main functionalities of the trajectory generator implemented and validated by the end of 2020 (M13). The full optimisation framework and a prototype of the HMI is aimed at M16, so that the first results from the internal validation of the optimisation of trajectories can be ready for the first internal workshop (planned at M17). After this internal workshop, the first release of the software to the Topic Manager is planned (D4.1 - Crew Assistant Decision model description (first release) and D4.2 - Crew Assistant Decision model software package (first release)). The outcome of the internal validation will be used during the external validation workshop scheduled for April 2021 (M18).

Between the external validation workshop and the end of the project, a continuous interaction with the Advisory Board is planned in order to obtain support on the selection of experiments and functionalities, and on the validation of the further developed model. Due to the COVID-19 outbreak, all these interactions might be arranged through online meetings. A second internal workshop is planned for M22 so that an internal validation can be conducted prior a final round of external validation with the Advisory Board and the final release of the prototype at M24 (D4.3 - Crew Assistant Decision model description (final release) and D4.4 - Crew Assistant Decision model software package (final release)). Further modifications and improvements to the system required in order to facilitate its industrialisation will be compiled in D6.1 - System evolution and uptake (due in M24).

EUROPEAN UNION

All the verification and validation activities conducted through the project will be adequately recorded and reported in D5.2 - Verification and validation report.

Finally, it is worth mentioning that the members of the consortium are also involved in the Innovative Action Dispatcher3 [4]. Dispatcher3 will look at the use of machine learning techniques to provide advice to dispatchers. The members of Dispatcher3 include the members of the consortium of Pilot3 with the addition of Vueling airlines and Skeyes. This will provide access to data and resources from airlines and ANSPs. As with Pilot3, Dispatcher3 has an Advisory Board which includes all the members of Pilot3 with the addition of dispatching experts. Therefore, the actions planned in Dispatcher3 (workshops, interactions) will provide, when adequate, opportunities to further foster the relationship with relevant stakeholders and gain a full view of the pre-tactical and tactical operations which can be used to support some of the prioritisation of scenarios and functionalities of Pilot3.

11 References

1. Clean Sky 2, 2015 (Mar), Joint technical programme. Technical report. Version 5.
2. Clean Sky, 2011 (Jun), HIPE AE 440 - Diesel pistol engine design and prototyping, Astro Engine TEOS. Presentation at Paris Air Show (23rd June 2011), Le Bourget.
3. Domino Consortium, 2020 (Feb), D1.2 – Final project results report, Tech. Report. Ed. 01.01.00
4. European Commission, 2020 (Apr), Grant Agreement number: 886461 - Dispatcher3 - H2020-CS2-CFP10-2019-01, April 2020
5. European Parliament and the Council, 2013. 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.
6. Guru99. Difference Between Verification and Validation with Example. Available at: https://www.guru99.com/verification-v-s-validation-in-a-software-testing.html (accessed 23/07/2020).
7. IEEE,1990, IEEE 610.12-1990 - IEEE Standard Glossary of Software Engineering Terminology
8. Mazzarisi, P., Zaoli, S., Lillo, F., Delgado, L., Gurtner, G., Cook, A., Valput D., 2019 (Dec), Network-wide assessment of 4D Trajectory Adjustments using an Agent-based model, 9th SESAR Innovation Days, 2019, Athens.
9. Mouillet, V., Nuic, A., Casado, E., Leones, J. L., 2018, Evaluation of the Applicability of a Modern Aircraft Performance Model to Trajectory Optimization. 37th Digital Avionics Systems Conference (DASC). IEEE/AIAA.
10. Pilot3 Consortium, 2020 (Mar), Technical Resources and Problem definition, Tech. Report. Deliverable D1.1. Ed. 01.00.
11. Pilot3 Consortium, 2020 (May), Trade-off report on multi-criteria decision making techniques, Tech. Report. Deliverable D2.1. Ed. 01.00.
12. Pilot3 Consortium, 2020 (Jul), Airlines data collection report, Tech. Report. Deliverable D3.1. Ed. 01.00.
13. Plutora, 2019, Verification vs Validation: Do You know the Difference? Available at: https://www.plutora.com/blog/verification-vs-validation (accessed 31/03/2020).
14. Roberson, W., Root, R., Adams, D., 2007, Fuel conservation strategies: cruise flight. Boeing Commercial Airplanes AERO Quarterly 04/07/2007.
15. Sauer, C., Jeffery, D. R., Land, L., Yetton, P., 2000, The effectiveness of software development technical reviews: a behaviorally motivated program of research. IEEE Transactions on Software Engineering, 26(1), 1–14.
16. SESAR Joint Undertaking, 2015. Introduction to the SESAR 2020 Programme Execution. Available at: https://ec.europa.eu/research/participants/data/ref/h2020/other/guides_for_applicants/jtis/h2020-pr-exec-intro-ta-ir-vld-sesar-ju_en.pdf (accessed 30/07/2020).

EUROPEAN UNION

17. Software Testing Help, 2020, The Differences Between Unit Testing, Integration Testing and Functional Testing. Available at: https://www.softwaretestinghelp.com/the-difference-between-unit-integration-and-functional-testing (accessed 23/07/2020)
18. ToolsQA. Difference between Verification and Validation. Available at: https://www.toolsqa.com/software-testing/difference-between-verification-and-validation/ (accessed 23/07/2020).
19. Wikipedia. Software verification and validation. Available at: https://en.wikipedia.org/wiki/Software_verification_and_validation (accessed 23/07/2020).

12 Acronyms

4DTA: 4D Trajectory Adjustments

ACARE: Advisory Council for Aviation Research and Innovation in Europe

ADS-B: Automatic Dependent Surveillance - Broadcast

AG: Alternatives Generator

AOC: Airline Operating Centre

AOE: ATM Operational Estimator
ATC: Air Traffic Control

ATFM: Air Traffic Flow Management

ATM: Air Traffic Management

BADA 4.x: Base of Aircraft Data version 4.x

CI: Cost Index

CS2: Clean Sky 2

CS-x: Case Study x

DDR2: Data Demand Repository Version 2

DMAN: Departure Manager

Dx.x: Deliverable x.x

E-AMAN: Extended Arrival Manager

ECAC: European Civil Aviation Conference

EFB: Electronic Flight Bag

ER3: Exploratory Research 3

EVA: External Validation Action

FMS: Flight Management System

FPO: Flight Profile Optimiser from Pacelab

FSC: Full-Service Carrier

HMI: Human Machine Interface

IADP: Innovative Aircraft Demonstrator Platforms

IEEE: Institute of Electrical and Electronics Engineers

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

IROPs: Irregular Operations costs

ITD: Integrated Technology Demonstrators

IVA: Internal Validation Action

JTI: Joint Technology Initiative

JU: Joint Undertaking

KPA: Key Performance Area

KPI: Key Performance Indicator

LCC: Low Cost Carrier

LPA: Large Passenger Aircraft

LRC: Long Range Cruise

ML: Machine Learning

MRC: Maximum Range Cruise

OAE: Operational ATM Estimator

OEM: Original Equipment Manufacturer

OFP: Operational Flight Plan

OTP: On-time Performance

PA: Performance Assessment

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

PI: Performance Indicator

PIE: Performance Indicators Estimator

RAD: Route Availability Document

RQ: Research Question

SCN-x: Scenario x

SDTR: Software Design Technical Review

SESAR: Single European Sky ATM Research

SGO: Systems for Green Operations

SIBT: Schedule In-Block Time

SOBT: Schedule Off-Block Time

SYS: Systems

SYS-ITD: Systems Integrated Technology Demonstrator

TE: Technology Evaluator

TMA: Terminal Manoeuvring Area

TOC: Top of Climb

TOD: Top of Descend

TRL: Technology Readiness Level

TTA: Target Time of Arrival

UoW: Short name of Pilot3 coordinator: University of Westminster

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

WPx: Workpackage x

Appendix A Questionnaires for validation

This appendix presents the questionnaires and flow-chart diagrams that have been prepared for the different validation actions.

A.1 Internal validation

A.1.1 Internal validation of the HMI prototype (IVA7)

The next three questionnaires below are defined to facilitate the assessment of the HMI prototype by the internal experts (**IVA7**). **Table *11*** provides the information on the match between specific research questions and questionnaire used.

Table 11. Methods to address different RQs defined in IVA7

RQs ID	Questionnaires/Flow chart diagram
P3-RQ-IV-120	“General acceptability” questionnaire
P3-RQ-IV-130	“Easiness of understanding of the information” questionnaire
P3-RQ-IV-140	“Interaction with the system” questionnaire

EUROPEAN UNION

1- General acceptability

Please indicate, by ticking the bullets, whether you agree or disagree with the statements given below when considering general acceptability of the tool with the respect to quantity of information provided to the pilot.

General acceptability	Strongly Disagree	Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree
1. The information provided to the pilot is simple and concise enough	O	O	O	O	O	O
2. The amount of information presented to the pilot is well balanced	O	O	O	O	O	O
3. The information provided to the pilot is predictable in its presentation	O	O	O	O	O	O
4. The visual representation of the alternative trajectories presented is clear and well organised	O	O	O	O	O	O

Please indicate any additional comments relevant for the above set of statements

____

2- Easiness of understanding of the information

Please indicate, by ticking the bullets, whether you agree or disagree with the statements given below when considering the easiness of understanding of the information provided

Easiness of understanding of the information	Strongly Disagree	Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree
1. Information on the trajectories and their impact on the high-level optimisation objectives (total cost and OTP) is easy to understand	O	O	O	O	O	O
2. Information on the trajectories and their impact on the different key performance indicators (e.g., cost of fuel, cost of IROPs, other cost) is easy to understand	O	O	O	O	O	O
3. Information on the trajectories and their impact on the different PIs (e.g., minutes of delay at arrival) is easy to understand	O	O	O	O	O	O
4. The trade-offs between OTP and total cost (i.e., the extra cost needed to achieve OTP) are clear and easy to understand	O	O	O	O	O	O
5. The information on the confidence level provided for each trajectory is clear and easy to use	O	O	O	O	O	O

Please indicate any additional comments relevant for the above set of statements

____

3- Interaction with the system

Please indicate, by ticking the bullets, whether you agree or disagree with the statements given below when considering acceptability of the mechanism for interaction between the pilot and the tool

Interaction with the system	Strongly Disagree	Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree
1. The mechanism which allows the selection of the solution is appropriate and easy to use	O	O	O	O	O	O
2. The mechanism which allows the rejection of the solution(s) is appropriate and easy to use	O	O	O	O	O	O
3. The mechanism which allows the pilot to set new trajectory constraints is appropriate and easy to use	O	O	O	O	O	O
4. The mechanism which allows to request a re-evaluation of the alternative trajectories is appropriate and easy to use	O	O	O	O	O	O
5. The comparison between alternative trajectories is easy to use	O	O	O	O	O	O

Please indicate any additional comments relevant for the above set of statements

____

A.2 External validation

A.2.1 External validation of the HMI prototype (EVA1)

Considering the capabilities of the HMI prototype and HMI demonstration of the specific scenario defined above, the experts will be asked to provide their feedback by:

- expressing the overall acceptance of the HMI prototype by using the flow chart diagram (e.g., flow-chart diagram)
- specifying their level of the agreement to the statements addressing specific operational related aspects of HMI prototype (design and functionalities). For this purpose, a six-point Likert scale is employed. The external experts will be asked to assess four different sets of statements, each of which reflects different aspects of HMI prototype, as follows:
    1. The first set of statements refers to the general acceptability of the tool with the respect to quantity of information provided to the pilot.
    2. The second set of statements refers to the easiness of understanding the information provided
    3. The third set of statements refers to acceptability of the mechanism which allows the interaction between HMI and the pilot
    4. The fourth set of statements refers to general applicability of the tool from pilots' point of view (designed for the pilots only)

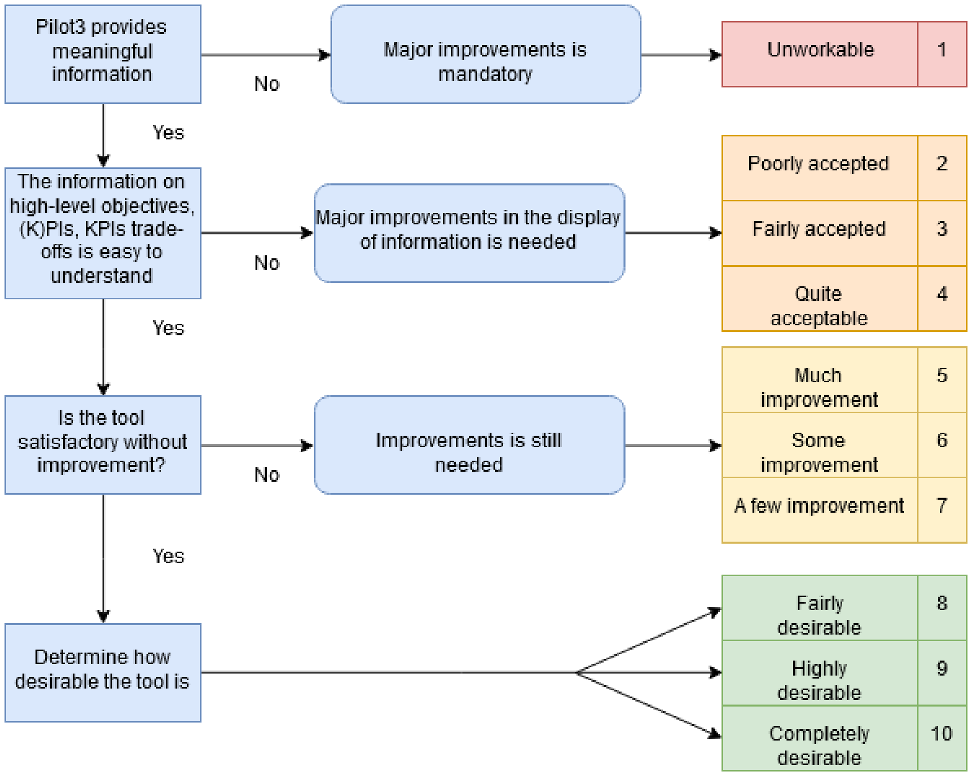
In order to facilitate to track the link between different RQs and mechanism for their assessment, **Table *12*** below explains how different RQs are assessed during EVA1.

Table 12. Methods to address different RQs defined in EVA1

RQs ID	Questionnaires/Flow chart diagram
P3-RQ-EV-010	Flow-chart diagram for global acceptance
P3-RQ-EV-020	“Pilot's overall acceptance of the tool” questionnaire
P3-RQ-EV-030	“General acceptability” questionnaire
P3-RQ-EV-040	“Easiness of understanding of the information” questionnaire
P3-RQ-EV-050	“Interaction with the system” questionnaire
P3-RQ-EV-060	“Pilot's overall acceptance of the tool” questionnaire (3rd statement only)

Flow-chart diagram for global acceptance

Considering the solutions presented for the given scenario, express your overall acceptance of the Pilot3 HMI prototype by going through the scheme given below, indicating the final score by circling the appropriate numeric value (on the provided 1 – 10 scale).



Please provide your main reasons for this score and/or detail which improvements would be needed:

____

1- General acceptability

Please indicate, by ticking the bullets, whether you agree or disagree with the statements given below when considering general acceptability of the tool with the respect to quantity of information provided to the pilot.

General acceptability	Strongly Disagree	Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree
1. The information provided to the pilot is simple and concise enough	O	O	O	O	O	O
2.The amount of information presented to the pilot is well balanced	O	O	O	O	O	O
3. The information provided to the pilot is predictable in its presentation	O	O	O	O	O	O
4. The visual representation of the alternative trajectories presented is clear and well organised	O	O	O	O	O	O

Please indicate any additional comments relevant for the above set of statements

____

EUROPEAN UNION

2- Easiness of understanding of the information

Please indicate, by ticking the bullets, whether you agree or disagree with the statements given below when considering the easiness of understanding of the information provided

Easiness of understanding of the information	Strongly Disagree	Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree
1. Information on the trajectories and their impact on the high-level optimisation objectives (total cost and OTP) is easy to understand	O	O	O	O	O	O
2. Information on the trajectories and their impact on the different key performance indicators (e.g., cost of fuel, cost of IROPs, other cost) is easy to understand	O	O	O	O	O	O
3. Information on the trajectories and their impact on the different PIs (e.g., minutes of delay at arrival) is easy to understand	O	O	O	O	O	O
4. The trade-offs between OTP and total cost (i.e., the extra cost needed to achieve OTP) are clear and easy to understand	O	O	O	O	O	O
5. The information on the confidence level provided for each trajectory is clear and easy to use	O	O	O	O	O	O

Please indicate any additional comments relevant for the above set of statements

____

EUROPEAN UNION

3 – INTERACTION WITH THE SYSTEM

Please indicate, by ticking the bullets, whether you agree or disagree with the statements given below when considering acceptability of the mechanism for interaction between the pilot and the tool

Interaction with the system	Strongly Disagree	Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree
1. The mechanism which allows the selection of the solution is appropriate and easy to use	O	O	O	O	O	O
2. The mechanism which allows the rejection of the solution(s) is appropriate and easy to use	O	O	O	O	O	O
3. The mechanism which allows the pilot to set new trajectory constraints is appropriate and easy to use	O	O	O	O	O	O
4. The mechanism which allows to request a re-evaluation of the alternative trajectories is appropriate and easy to use	O	O	O	O	O	O

Please indicate any additional comments relevant for the above set of statements

____

4- Pilot's overall acceptance of the tool

The fourth set of statements is particularly **designed for the pilots only**:

Please indicate, by ticking the bullets, whether you agree or disagree with the statements given below when considering the statements designed for the pilots about his/her general acceptance of the tool

Pilot's overall acceptance of the tool	Strongly Disagree	Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree
1. The alternatives provided by PIlot3 will facilitate the pilot in his/her action to take the appropriate decisions	O	O	O	O	O	O
2. With the alternatives provided, the pilot will have better awareness of his/her actions than in the case he/she needs to take the decision by him/herself	O	O	O	O	O	O
3. The information on the confidence level provided for each trajectory will aid the pilot to better assess the benefits of trajectories presented against each other.	O	O	O	O	O	O

Please indicate any additional comments relevant for the above set of statements

____

A.2.2 External validation of stand-alone simulations (EVA2)

Table *13* below explains how different RQs are assessed during **EVA2**.

Table 13. Methods to address different RQs defined in EVA2

RQs ID	Questionnaires/Flow chart diagram
P3-RQ-EV-070	“Goodness of solutions in different operational context” questionnaire
P3-RQ-EV-080	“Overall acceptance of Pilot3” flow-char diagram

1 - Goodness of solutions in different operational context

Please indicate, by ticking the bullets, whether you agree or disagree with the statements given below when considering the entire results obtained for the given scenario:

Goodness of solutions in different operational context	Strongly Disagree	Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree
1. In the light of the obtained results and the operational context of the given solutions, do you believe that Pilot3 is worth acquiring by your company?	O	O	O	O	O	O
2. In the light of the obtained results and operational benefits of the given solutions, do you believe that Pilot3 will contribute to better translate the strategic goals of your company policy to the tactical execution of the flight?	O	O	O	O	O	O
3. In the light of the obtained results and operational effort of the given solutions (i.e., actions required to carry out in order to perform the trajectory), the benefits provided by selecting the solution generated by Pilot3 will be more worthy than flying the trajectory based on the pilot's typical reaction?	O	O	O	O	O	O
4. In the light of the obtained results and operational effort of the given solutions (i.e., actions required to carry out in order to perform the trajectory), the benefits provided by selecting the solution generated by Pilot3 will be more worthy than flying the OFP plan?	O	O	O	O		O

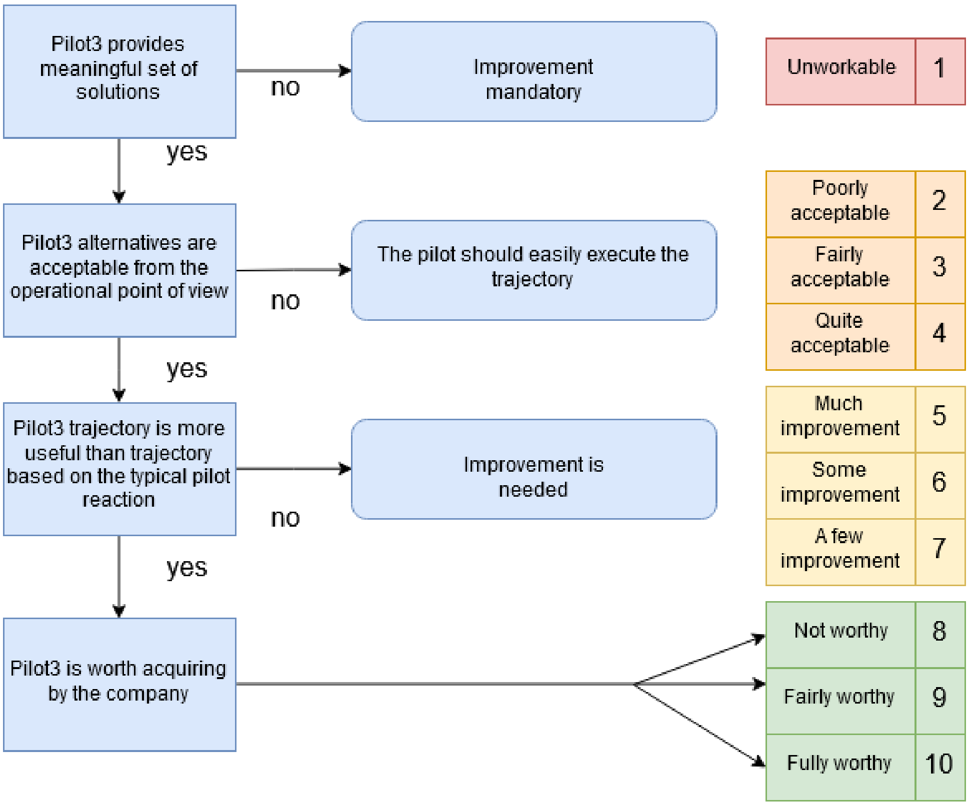
Please indicate any additional comments relevant for the above set of statements

____

EUROPEAN UNION

FLOW-CHART DIAGRAM FOR OVERALL ACCEPTANCE OF PILOT3 CONSIDERING PERFORMANCE RESULTS FOR INDIVIDUAL TRAJECTORIES

Considering the solutions presented for the given scenario, express your overall acceptance of Pilot3 considering performance results for individual trajectories by going through the scheme given below, indicating the final score by circling the appropriate numeric value (on the provided 1 – 10 scale).



Please provide your main reasons for this score and/or detail which improvements would be needed:

____

EUROPEAN UNION

A.2.3 External validation of network simulations (EVA3)

Table *14* below explains how different RQs are assessed during EVA3.

Table 14. Methods to address different RQs defined in EVA3

RQs ID	Questionnaires/Flow chart diagram
P3-RQ-EV-090	"Goodness of solutions with respect to airline and passenger key metrics" questionnaire
P3-RQ-EV-100	"Overall acceptance of Pilot3 considering network-wide level" flow-char diagram

1 - Goodness of solutions with respect to airline and passenger key metrics

Please indicate, by ticking the bullets, whether you agree or disagree with the statements given below when considering the entire results obtained for the given scenario:

Goodness of solutions with respect to airline and passenger key metrics	Strongly Disagree	Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree
1. With Pilot3 deployed at network level, the airline cost is reduced with respect to baseline implementations of 4DTA mechanism	O	O	O	O	O	O
2. With Pilot3 deployed at network level, the airline delay is reduced with respect to baseline implementations of 4DTA mechanism	O	O	O	O	O	O
3. With Pilot3 deployed at network level, the number of missed connections is reduced with respect to baseline implementations of 4DTA mechanism	O	O	O	O	O	O

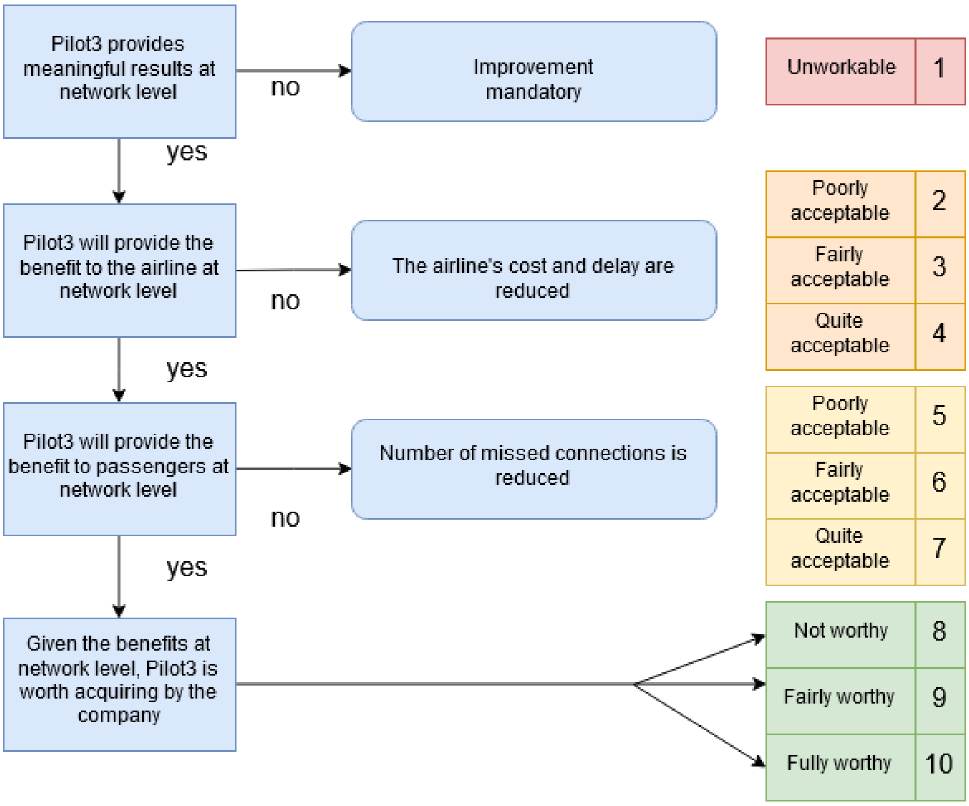
Please indicate any additional comments relevant for the above set of statements

____

EUROPEAN UNION

FLOW-CHART DIAGRAM FOR OVERALL ACCEPTANCE OF PILOT3 CONSIDERING NETWORK-WIDE LEVEL

Considering the solutions presented for the network simulations, express your overall relevance of Pilot3 to achieve benefits for airlines and passengers by going through the scheme given below, indicating the final score by circling the appropriate numeric value (on the provided 1 – 10 scale).



Please provide your main reasons for this score and/or detail which improvements would be needed:

____

-END OF DOCUMENT-