EXPLORATORY RESEARCH

sesar JOINT UNDERTAKING

Final Project Results Report

Deliverable ID:	**D1.3**
Dissemination Level:	**PU**
Project Acronym:	**SIMBAD**
Grant:	**894241**
Call:	**H2020-SESAR-2019-2**
Topic:	**SESAR-ER4-03-2019**
Consortium Coordinator:	**Nommon**
Edition date:	**09 February 2023**
Edition:	**01.00.00**
Template Edition:	**02.00.05**

EUROPEAN PARTNERSHIP

Co-funded by the European Union

Authoring & Approval

Authors of the document

Name / Beneficiary	Position / Title	Date
Raquel Sánchez (Nommon)	Deputy Project Coordinator	02/02/2023
George Vouros (UPRC)	WP3 leader	21/11/2022
Rubén Rodríguez (CRIDA)	WP4 leader	21/11/2022
Jordi Pons (UPC)	WP6 leader	21/11/2022

Reviewers internal to the project

Name / Beneficiary	Position / Title	Date
David Mocholí (Nommon)	Project Coordinator	02/02/2023

Reviewers external to the project

Name / Beneficiary	Position / Title	Date

Approved for submission to the S3JU By - Representatives of all beneficiaries involved in the project

Name / Beneficiary	Position / Title	Date
David Mocholí (Nommon)	Project Coordinator	02/02/2023
Rubén Rodríguez (CRIDA)	WP4 Leader	02/02/2023
George Vouros (UPRC)	WP3 Leader	02/02/2023
Jordi Pons (UPC)	WP6 Leader	02/02/2023
Gennady Andrienko (Fraunhofer)	General Assembly	02/02/2023

Rejected By - Representatives of beneficiaries involved in the project

Name and/or Beneficiary	Position / Title	Date

Document History

Edition	Date	Status	Name / Beneficiary	Justification
00.00.01	21/11/2022	Draft	Raquel Sánchez, George Vouros, Rubén Rodríguez, Jordi Pons	Initial draft for internal review
00.01.00	24/11/2022	Submitted to the SJU for approval	Raquel Sánchez	Integration of internal review comments for SJU submission

00.02.00	16/12/2022	Submitted to the SJU for approval	Raquel Sánchez	Corrected version for SJU delivery
00.03.00	02/02/2023	Submitted to the SJU for approval	Raquel Sánchez	Corrected version for SJU delivery
01.00.00	09/02/2023	Final	Raquel Sánchez	Approved by the SJU

SIMBAD

COMBINING SIMULATION MODELS AND BIG DATA ANALYTICS FOR ATM PERFORMANCE ANALYSIS

This document is part of a project that has received funding from the SESAR3 Joint Undertaking under grant agreement No 894241 under European Union's Horizon 2020 research and innovation programme.

Abstract

This document is the Final Project Report of the SIMBAD project. The document provides an overview of the project, including its objectives and the work performed, in particular: (i) the developed models for the estimation of two hidden variables, the cost index and the landing weight, and for the trajectory modelling; (ii) the methodology for the identification of traffic patterns and the selection of representative traffic samples at different geographical scales for each particular problem under study; (iii) the definition of performance metamodels for the DYNAMO and R-NEST simulation tools; and (iv) the validation of the three developments through two case studies. Then, the link between the work done and the objectives of the SESAR program is discussed, providing an assessment of the overall maturity of the project. Finally, the project conclusions and lessons learnt are presented to the reader, together with a set of proposals for the future extension of the research carried out by SIMBAD.

Table of Contents

List of Tables

List of Figures

1 Executive Summary

The development of performance modelling methodologies able translate new ATM concepts and technologies into their impact on high-level, system wide KPIs has been a long-time objective of the ATM research community. Bottom-up, microsimulation models are often the only feasible approach to address this problem in a reliable manner. However, the practical application of large-scale simulation models to strategic ATM performance assessment is often hindered by their computational complexity. The goal of SIMBAD is to develop and evaluate a set of machine learning approaches aimed at providing state of-the-art ATM microsimulation models with the level of reliability, tractability and interpretability required to effectively support performance evaluation at ECAC level.

The specific objectives of the project are the following:

1. Explore the use of machine learning techniques for the estimation of hidden variables from historical air traffic data, with particular focus on airspace users' preferences and behaviour, in order to enable a more robust calibration of air traffic microsimulation models.

2. Develop new machine learning algorithms for the classification of traffic patterns that enable the selection of a sufficiently representative set of simulation scenarios allowing a comprehensive assessment of new ATM concepts and solutions.

3. Investigate the use of active learning metamodelling to facilitate a more efficient exploration of the input output space of complex simulation models through the development of more parsimonious performance metamodels, i.e., analytical input/output functions that approximate the results of a more complex function defined by the microsimulation models.

4. Demonstrate and evaluate the newly developed methods and tools through a set of case studies in which the proposed techniques will be integrated with existing, state-of-the-art ATM simulation tools and used to analyse a variety of ATM performance problems.

The work carried out by SIMBAD has translated into one solution, the SIMBAD Performance Modelling Framework, which enables the comprehensive exploration and understanding of the performance impact of ATM concepts/solutions under different traffic conditions.

The framework comprises: 1) a model for the estimation of hidden variables related to airspace user behaviour that are necessary inputs for the ATM microsimulation models (e.g., cost index and payload mass); 2) a methodology to identify representative traffic patterns at different scales (ECAC, ANSP, ACC and airport) for each particular problem under study; and 3) a metamodelling framework that enables the approximation of the results of a microsimulation model to facilitate a more efficient exploration of its input-output space.

The potential of this solution has been evaluated through two case studies, the SESAR solutions Free-Routing and Demand and Capacity Balancing. The evaluation of the developed techniques has delivered a number of valuable results:

- The hidden variable estimation models and trajectory prediction models showed a very good performance when compared to the brute force approach (i.e., DYNAMO simulations), indicating the usefulness of the machine learning methods for hidden variables estimation and trajectory prediction. In particular, both hidden variables (cost index and payload mass) were

estimated with high accuracy. Using the predicted trajectories and the estimated hidden variables, some trajectory KPIs were estimated. This approach allowed a quite accurate estimation of three trajectory KPIs: fuel consumption, flown distance, and gate-to-gate time.

- The clustering-based methodology for the identification of traffic patterns and selection of traffic samples at different geographical scales allowed the identification of representative traffic demand patterns (such as the days with air traffic controller staff strike in France, or the winter and summer weekends and the international holidays) for both case studies. This methodology provided high quality and trustful results, outperforming in some cases the ATM experts' judgement and days' selection, as it revealed the subtle traffic patterns which can be easily omitted by the experts due to their small shares in the total traffic sample. In addition, the set of representative days provided a reliable approach for calculating KPIs, such as the fuel consumption.

- The three metamodels developed (one for DYNAMO and two for R-NEST) showed that they can imitate the behaviour of the microsimulation tools. In particular, the performance of the DYNAMO metamodel is quite close to that of the simulation tool, while the R-NEST metamodels reach a good performance for one output, but for the other one, the predictive error is quite high. The three metamodels outperform the ATM simulation tools in terms of execution time, computational resources and expert knowledge. These metamodels enable the exploration of the two SESAR solutions more deeply, amplifying the exploration of the simulation input and output behaviour space. The results of the metamodels indicate the performance gain of employing a metamodeling strategy compared to a traditional approach based on the exhausting simulations.

2 Project Overview

2.1 Operational/Technical Context

One of the cornerstones of the ongoing ATM modernisation programmes, including SESAR, is performance orientation. A performance-driven approach entails the need for: (i) performance measurement, including the definition of appropriate indicators able to translate performance objectives into measurable quantities, as well as data and methodologies allowing the calculation of such indicators; (ii) performance evaluation, which requires the expertise to assess the impact of management actions (new operational concepts, new technologies, etc.) on performance indicators; and (iii) decision support, which involves the use of methods and tools for trade-off analysis and decision-making. Among these three elements, performance evaluation is arguably the most challenging: ATM performance results from the complex interaction of interdependent policies and regulations, stakeholders, technologies and market conditions, leading to the emergence of trade-offs not only between KPAs, but also between stakeholders, as well as between short-term and long-term objectives.

The development of performance modelling methodologies able to grasp the interdependencies between different Key Performance Areas (KPAs) and translate new ATM concepts and technologies into their impact on high-level, system-wide Key Performance Indicators (KPIs) has been a long-time objective of the ATM research community. Generally speaking, the modelling approaches to this problem can be classified into two main categories: macroscopic and microscopic. Macroscopic models represent the behaviour of a system by formulating the relationships between aggregated variables without explicitly modelling the individual system components; they are usually built from the top down and involve a high-level of abstraction. Macroscopic models are the most commonly used for strategic decision-making because of their simplicity, speed and interpretability, but they suffer from important drawbacks: first, their formulation is, to a large extent, based on expert judgement and supported by little empirical evidence; second, their top-down, aggregated nature limits their ability to detect emergent behaviours, which play a fundamental role in a heavily networked system of systems like ATM. As opposed to the previous approach, microscopic models adopt an explicit representation of the actions and interactions of the individual elements that compose a system with the aim to observe the performance that emerges at the macroscopic level. These models have shown their ability to capture a rich variety of behaviours in a very realistic manner. However, the practical application of complex simulation models to strategic ATM performance assessment and decision-making is hindered by several factors: (i) hidden variables: even if they could, in principle, be measured, certain aspects of the ATM system may not be observable for practical reasons. This is the case, for example, of business sensitive data related to the behaviour of airspace users (AUs) that are of paramount importance for the construction of microsimulation models, such as aircraft take-off weight (TOW), selected cost index (CI), etc.; (ii) computational complexity: the process to obtain each input-output pair (i.e., the set of KPI values that correspond to a particular scenario) is extremely costly, due to the need to run complex simulations that can take minutes or even hours, which limits the number of scenarios that can be explored; (iii) interpretability issues: the high number of variables involved hampers the analysis, interpretation and communication of the modelling results, which in turn constitutes a barrier for their use in evidence-based decision-making.

The SIMBAD project was conceived to explore how recent advances in machine learning can help overcome these problems, by developing data-driven simulation techniques that facilitate hidden

variable estimation and increase the tractability and interpretability of large-scale, complex, fast-time ATM simulation models.

2.2 Project Scope and Objectives

The goal of SIMBAD is to develop and evaluate a set of machine learning approaches aimed at providing state-of-the-art ATM microsimulation models with the level of reliability, tractability and interpretability required to effectively support performance evaluation at network level. The project is focused on three fundamental problems: (i) how to estimate hidden variables that are not directly observable; (ii) how to select a sufficiently representative set of traffic scenarios; (iii) how to use complex microsimulation models to build performance metamodels that are conceptually simpler and less computationally costly, in order to allow a more efficient exploration of the simulation space and an easier interpretation of the modelling outcomes.

The specific objectives of the project are the following:

1. (O1) - Explore the use of machine learning techniques for the modelling of trajectories and estimation of hidden variables from historical air traffic data. Particular attention is paid to the estimation of variables related to AUs' preferences and behaviour (e.g., airline cost functions), which are one of the major unknowns when assessing the actual performance benefits delivered by a certain ATM concept or solution.

2. (O2) - Develop new machine learning algorithms for traffic pattern classification. Given the complexity of large-scale, microscopic air traffic simulations, running simulation models for each and every day of the year is usually prohibitive. On the other hand, limiting simulations to one or few particular days may not be representative of the impact of a certain operational improvement under other scenarios. SIMBAD investigates how different clustering and machine learning classification techniques can be used to identify a representative set of demand patterns that allows a comprehensive impact assessment of new SESAR solutions at ECAC level.

3. (O3) - Investigate the use of active learning metamodelling to enable a more efficient exploration of the input-output space of complex ATM simulation models. Given the computational cost of realistic air traffic simulations, a goal when exploring the simulation space should be picking only the most informative instances. The questions are: if we have a limited number of data points that we can obtain, where in the input space shall we place them? How can we approximate the rest of the points? In recent years, the concept of active learning has been proposed to address these questions. SIMBAD explores how active learning can be used to translate a complex simulation model into a performance metamodel, i.e., an analytical input/output function that approximates the results of a more complex function defined by the simulation model itself, improving computational tractability and interpretability of results.

4. (O4) - Demonstrate and evaluate the newly developed techniques in order to assess their maturity, derive recommendations on how to apply them to ATM performance assessment, and propose a roadmap for the transition of the project results to the next stages of the R&D cycle. To this end, two case studies are developed in which the proposed techniques are integrated with existing, state-of-the-art ATM simulation tools (namely the R-NEST and APACHE tools developed by EUROCONTROL and UPC, respectively) and used to analyse a variety of ATM performance problems.

2.3 Work Performed

The SIMBAD work packages and their interdependencies are shown in the figure below.

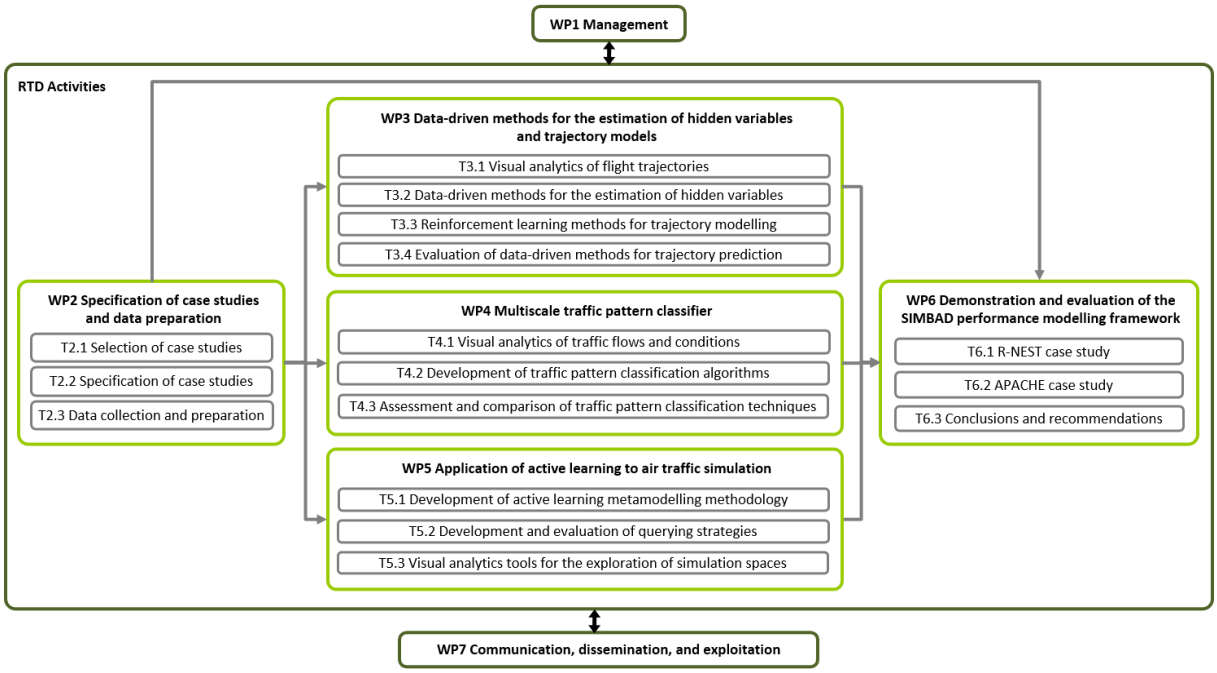

Figure 1 Work flow and interdependencies between work packages

Table 1 presents the relation between the project's objectives and the different work packages.

Table 1 Relation between objectives and work packages

Related objective	Associated tasks	Associated deliverables
1) Explore the use of machine learning techniques for the modelling of trajectories and estimation of hidden variables from historical air traffic data	WP3	D3.1 Data-Driven Methods for Trajectory Modelling
2) Develop new machine learning algorithms for traffic pattern classification	WP4	D4.1 Methodologies and Algorithms for the Selection of Representative Traffic Samples
3) Investigate the use of active learning metamodelling to enable a more efficient exploration of the input-output space of complex ATM simulation models	WP5	D5.1 Active Learning Metamodelling
4) Demonstrate and evaluate the newly developed techniques in order to assess their maturity, derive recommendations on how to apply them to ATM performance assessment, and propose a roadmap for the transition of the project results to the next stages of the R&D cycle	WP2, WP6	D2.1 Specification of Case Studies D6.1 Evaluation of the SIMBAD performance modelling framework and implementation guidelines

The detailed description of the work performed in each work package is presented below.

2.3.1 WP1 Management

The purpose of WP1 Management was to manage and coordinate the project so as to ensure the achievement of the project goals within agreed time, cost, and quality. WP1 was led by Nommon, as Project Coordinator, with the support of the rest of SIMBAD partners.

The work done in WP1 includes:

- Overall project management and coordination: Nommon, as Project Coordinator, was in charge of the planning and monitoring of the project, tracking of deliverables and daily management (coordination, distribution of documents, meeting set-up, etc.).
- Set up of a Project Information System, a wiki-like tool that provides on-line, real-time visibility on all areas of programme management. The Project Information System was set up before T0+1 and is continuously updated throughout the project.
- Preparation of D1.1 Project Management Plan: D1.1 was prepared by Nommon and submitted in February 2021. The document was later on updated in accordance with the comments received from the SJU and was approved in July 2021.
- Preparation of D1.1.010 Progress Report 1, which was submitted in September 2021 covering the project period from 1st January 2021 until 30th June 2021. The document was approved by the SJU in September 2021.
- Elaboration and submission of D1.2 Data Management Plan (DMP). The DMP describes the data management life cycle for the data collected, processed and generated by the project. It also includes a data inventory, the proposed approach for handling research data during and after the end of the project (including security requirements), the methodologies and standards that will be applied, and the specification of the data that will be made open access. Initially submitted in July 2021, the document was updated according to the comments received from the SJU and approved in November 2021. D1.2 is a living document that will be periodically updated to ensure its validity and alignment with the project objectives.
- Preparation of D1.1.020 Progress Report 2, which was submitted in July 2022 covering the project period from 1st July 2021 until 30th June 2022. The document was approved by the SJU in September 2022.
- Selection of the External Experts Advisory Board (EEAB): the role of the EEAB is to provide independent advice and peer review of the project results. A tentative list of members was prepared by the consortium and then contrasted with SJU members. The EEAB is composed by 15 experts: 10 end users, which include representatives from airlines, air navigation service providers (ANSPs), SESAR PJ19.04 project and EUROCONTROL, and 5 academic experts, including at least one expert in each of the core research areas involved in WP3, WP4, and WP5, namely reinforcement learning, traffic patterns, metamodelling, and visual analytics.

2.3.2 WP2 Specification of case studies and data preparation

The goal of WP2 was the selection and specification of the two case studies, one based on EUROCONTROL's R-NEST tool and a second one based on UPC's DYNAMO simulator, which will be used

to develop and evaluate the new methodologies and techniques investigated by SIMBAD. The specific objectives of this WP were to: (i) define which of the SIMBAD developments will be tested in each case study, ensuring that each of the three main contributions of the project (data-driven methods for trajectory modelling and hidden variable estimation, traffic pattern classifier, active learning metamodelling) are tested in at least one of the two case studies; (ii) select the ATM solutions and/or operational concepts that will be simulated; (iii) specify the associated operational environments; (iv) specify the KPAs and KPIs that will be evaluated; and (v) identify, collect and prepare the input data required for the implementation of the case studies.

WP2 was led by CRIDA.

For the selection and definition of the SIMBAD case studies, a list of possible solutions to be used for the evaluation of the methods and tools developed in the project was elaborated collaboratively with the participation of all the members of the consortium. Several criteria, including compatibility with SIMBAD developments, availability of a mature performance assessment and potential gap between the expected and actual benefits, were established by the consortium to select the more appropriate SESAR solutions. The selected case studies are Free-Routing (FR) and Demand and Capacity Balancing (DCB). This proposal was contrasted and validated with a panel of experts at the 1st SIMBAD Stakeholder Workshop.

Regarding the specification of the geographical regions for the FR case study, four European regions where a first approach to FRA is already deployed were identified: FRAIT (Free Route Airspace Italy), FRAL (Free Route Airspace Lisbon FIR), FRASAI (Free Route Airspace Santiago & Asturias), and MUAC (Free Route Airspace Maastricht). The FRAIT scenario was considered especially interesting by the experts that participated in the 1st SIMBAD Stakeholder Workshop, due to the fact that it has several levels of implementation.

While for the specification of the geographical regions for the DCB case study three European regions were identified: French airspace - Bordeaux (LFBB), MUAC airspace - MUAC (EDYY), and Swiss airspace – Geneva (LSAG). The Bordeaux scenario was especially interesting for the experts in the 1st SIMBAD workshop because an algorithm for sector configuration optimisation is currently used in operations. Additionally, it was decided that the project will focus on the Dynamic Airspace Configuration (DAC) levels 1 and 2.

All this information is collected in the deliverable D2.1 Specification of Case Studies, which describes the criteria followed for the selection of the case studies and the selected case studies. This document was elaborated by CRIDA with the support of the rest of the partners.

Finally, concerning the last objective of the WP, i.e., the identification, collection and preparation of the input data, several working sessions took place with the participation of all the project partners to identify and collect the data sources required for the achievement of the project objectives. The collected datasets have been analysed and classified to assess the quality of the information provided. A factsheet summarising the different features of each dataset has been produced. The factsheets can be found in Appendix A of D1.2.

To collect all the data, a Data Repository was set up with the aim of providing the consortium members with secure, efficient and reliable access to the project datasets. The repository was implemented by Nommon using state-of-the-art open-source technologies. The selected storage technology is Nextcloud, a widely used open-source platform for data storage and transfer that is freely available under GNU AGPLv3 license, enabling community add-ons and support.

Deliverable D2.2 SIMBAD Data Repository describes the specifications and the implementation of the Data Repository. This document was elaborated by Nommon with the support of the rest of the partners.

2.3.3 WP3 Data-driven methods for the estimation of hidden variables and trajectory models

The goal of WP3 was to explore new approaches based on artificial intelligence and machine learning for the estimation of hidden variables and trajectory models from historical trajectory data. The specific objectives were to: (i) design and implement solutions for estimating relevant hidden parameters and variables of mechanistic trajectory prediction models; (ii) design and implement solutions for data-driven modelling of trajectories; and (iii) compare the results of predicting flights KPIs using data-driven and mechanistic approaches.

WP3 was led by UPRC, and the work performed includes the following.

In first place, the overall WP3 methodology was specified. The methodology was defined in several working sessions with the participation of all the project partners. As part of the methodology, the hidden variables related to AUs to be estimated were identified. These variables are the payload mass (PL) and the cost index (CI).

Several working sessions and consultations among the WP3 members took place to discuss about data sources, and the structure of data for the application of visual analytic techniques.

In order to generate the datasets needed for the machine learning (ML) algorithms, flight plans for different values of PL and CI within specific ranges were created using DYNAMO. These flight plans were enriched with parameters regarding aircraft state, weather conditions, positional information and performance data. Then, the data management, pre-processing, and association of data sources tasks needed to train and test the ML algorithms were performed. These datasets were used for both the estimation of hidden variables and multi-modal trajectory modelling. Additionally, visual analytics techniques were applied for a visual investigation of data properties.

Next, the ML algorithms for the estimation of hidden variables were designed, implemented, trained, and tested. Different machine learning algorithms, including deep neural networks, support vector regression, kernel ridge and lasso, were considered. The accuracy of these methods to estimate hidden variables was evaluated in different settings with enriched trajectories provided by mechanistic models and with enriched trajectories provided by the pre-processing and data integration methods implemented.

Regarding the data-driven trajectory modelling, two imitation learning methods, InfoGail and TripleGail, were designed, implemented, trained, and tested. Data sources, including weather data, ALLFT+ trajectories (flight plans) and flight plans generated with DYNAMO were used to train and test these reinforcement learning methods.

Finally, comparisons for the predictions of KPIs using mechanistic and data-driven models were performed to assess (a) how the estimations of hidden variables provided by the ML models, for given flight plans, support the mechanistic prediction of trajectory KPIs using standard prediction error metrics; and (b) how trajectory predictions and estimated hidden variables for these trajectories support the mechanistic prediction of trajectory KPIs, using standard prediction error metrics, compared to the KPIs estimated for known trajectories. For that, the following KPIs are considered:

fuel consumption, flown distance, and gate-to-gate time. The mechanistic KPI predictions were made using DYNAMO.

All the work developed in WP3 is collected in the deliverable D3.1 Data-Driven Methods for Trajectory Modelling. This document was elaborated by UPRC with the support of the rest of the WP members.

2.3.4 WP4 Multiscale traffic pattern classifier

The goal of WP4 was to investigate the application of different clustering and machine learning classification techniques for the identification of a set of traffic demand patterns that is sufficiently informative for the study of a particular ATM performance problem. The specific objectives of WP4 were to: (i) identify the most relevant traffic features to be considered for the classification of air traffic patterns; (ii) investigate the application of different clustering and classification techniques for traffic pattern identification at different spatio-temporal scales; and (iii) evaluate and compare the performance of the investigated techniques. WP4 was led by CRIDA.

Next, the work done in this WP is described.

In first place, a review of the state-of-the-art in traffic pattern identification was performed. For that, a literature review was conducted for the identification of the most used clustering and classification techniques together with the variables used to derive the traffic patterns.

Next, a methodology for the identification of traffic patterns and the identification of representative traffic samples at different geographical scales was defined. This methodology was validated through the two case studies addressed by SIMBAD (FR and DCB). For that, it was applied to both cases independently in the following way.

First, the most relevant traffic features and KPIs to be considered for each one of the case studies were selected and defined. Table 2 shows the variables identified for each case. This identification was contrasted with a panel of experts at the 1st SIMBAD Stakeholder Workshop. Next, the geographical scales to be studied for traffic patterns identification were specified. These include ECAC, ANSP, ACC and airport level.

Table 2 Traffic features selection for the SIMBAD case studies

Performance Indicator	DCB	FR
Average flown distance per flight		
Gate-to-gate flight time		
Actual average fuel burnt per flight		
Direct air navigation services (ANSs) Gate-to-gate cost per flight / Route Charges		
Average of difference between flown trajectories and flight plans (Predictability)		
Number of instrumental flight rules (IFR) movements		
Non-nominal regulation causes		
Average departure delay per flight (On – Time Performance)		

Performance Indicator	DCB	FR
ATFM regulations (MET reason)		
Average minutes of en-route ATFM delay per flight attributable to air navigation services		
Convective phenomena		
Estimated/Actual Time of Arrival		
Weather variables		
Sectorization		

After these identifications, the selected variables were computed for each geographical scale. For each variable selected, different statistics were computed to obtain more robust patterns, including sum, average, standard deviation, variance, maximum, minimum, mode, percentiles 25, 75, 90 and 95, median, skewness, and kurtosis. Validations were performed to ensure that the datasets were computed correctly.

A visual exploration of those features was performed to gain insights of the data through visual analytics and clustering techniques. This analysis allows the identification of abnormal days, special days of weeks, and types of airports. This information was very useful to derive the final traffic patterns.

Feature engineering processes were implemented for the selection of traffic features prior to the clustering analysis. For the DCB use case, a correlation analysis to remove correlated features and an agglomerative feature algorithm were applied. For the FR use case, a correlation analysis to remove correlated features and a principal component analysis (PCA) were applied.

Then, the clustering algorithms were implemented for the identification of traffic patterns for the different case studies. Two clustering algorithms were applied: kmeans and hierachical clustering. The technical evaluation of the clustering results was made using three state-of-the-art assessment metrics, the silhouette score, the Davies Bouldin score, and the Calinski Harabasz score, together with the elbow method.

Once the clusters were computed, the results were interpreted using the distribution within each cluster of the variables computed and two auxiliary temporal variables indicating whether a day is a weekend day or not, and whether it belongs to the IATA winter season or not.

Finally, a set of representative traffic samples was selected by taking two days from each cluster. The days considered were the day with the highest silhouette value of the cluster, i.e., the most representative day for the cluster, and the day with the lowest silhouette value of the cluster, representing the highest potential deviation from the cluster centroid.

All the work developed in WP4 is collected in the deliverable D4.1 Methodologies and Algorithms for the Selection of Representative Traffic Samples. This document was elaborated by CRIDA with the support of the rest of the WP members.

2.3.5 WP5 Application of active learning to air traffic simulation

The goal of WP5 was to explore the application of active learning metamodelling techniques to air traffic simulation in order to enable a more efficient exploration of the simulation space. The specific objectives of WP5 were to: (i) explore and develop different metamodelling interpolation methodologies; (ii) explore and develop different active learning querying strategies; and (iii) develop visual analytics tools to support the analysis of the metamodel results. WP5 was led and executed by Nommon.

Regarding the work performed in this WP, in first place, a literature review of active learning methodologies and querying strategies was carried out. Also, a revision and selection of active learning Python libraries was done. Several libraries devoted to active learning were analysed and compared, including modAL, acton, alp, libact and ALiPy. Finally, modAL was chosen due to its easy installation, its integration with sklearn and its straightforward and intuitive use.

Next, three metamodels were defined, one for DYNAMO and two for R-NEST. The DYNAMO metamodel was validated through the FR SESAR's solution. The proposed metamodel estimates the fuel efficiency (measured as the FEFF1 KPI of the SESAR Performance Framework) of the set of flights crossing the Free Route Airspace Italy (FRAIT) region and its input variables are the fuel cost and the FRAIT activation. In order to implement this metamodel, two airspace configurations were selected, one before FRAIT implementation (specifically, the first day of the last AIRAC before the FRAIT implementation, i.e., 10th November 2016) and one after (specifically, the first day of the sixth AIRAC after the FRAIT implementation, i.e., 22th June 2017). The airspace configuration used is encoded with the FRAIT variable as follows. When the airspace configuration used for the DYNAMO simulations is the one before the FRAIT implementation, the variable is 0, and when it is the one after the FRAIT implementation, the FRAIT variable is 1.

This metamodel was implemented with the support of UPC, that provided an updated release of the tool and helped with the installation and integration processes.

The two R-NEST metamodels were tested with the DCB SESAR's solution. The first metamodel is a simpler one developed with the purpose of exploring the input-output space and validate the active learning metamodelling methodology. This metamodel estimates the average departure delay (measured as the PUN1 KPI of the SESAR Performance Framework) and the flights per air traffic controller (ATCo) hour in duty (measured as the CEF2 KPI of the SESAR Performance Framework) for the set of flights crossing the lower and east cluster of the Bordeaux ACC in a fixed day. The inputs of this metamodel are the minimum opening duration of the configuration and sector in the configuration of the opening scheme (OS). The metamodel was implemented for the 5th July 2019.

The second metamodel is an extension of the first one to include more days. For that, the input variables are extended with more variables. In particular, this metamodel includes the hourly entry counts of the day in the considered ACC (Bordeaux ACC) with the purpose of extending the metamodel to more days. This metamodel was implemented for the 7º AIRAC of 2019, in order to extend the temporal scope of the previous metamodel. This way, it estimates both KPIs for any combination of the OS parameters and any day of AIRAC.

In order to select the days to train this metamodel with, the methodology for the selection of representative traffic samples defined in WP4 was applied. This led to six representative days in the Bordeaux ACC, which were used to train the metamodel.

Then, an active learning framework was defined to train the three metamodels. A machine learning model, a querying strategy and a stopping criterion were selected. The machine learning model selected is a Gaussian Process and the querying strategy, to label the instance with higher predictive variance (i.e., the point for which the model predicts with more uncertainty). The stopping criterion was based on two conditions. First, at least the 10% of the input space has to be explored (i.e.,

labelled). Second, the predictive variance on a separated dataset should be reduced in at least 80% from the first iteration of the training cycle and this reduction should be maintained at least 5 consecutive iterations. This provides a measure of stability during the training process.

The metamodels, together with the active learning training cycle, were implemented in Python. All the elements of the active learning process were implemented and connected so that the iterative process was run automatically. In the case of DYNAMO, the tool was fully integrated within the active learning cycle, making the labelling process completely automatic. Unlike DYNAMO, R-NEST cannot be integrated into the training cycle. So, the labelling process has to be performed manually for each request of the active learning process.

During this time, several collaboration meetings with the NOSTROMO project took place. Given the great synergies between the metamodelling activities of both projects, it was finally decided to use the NOSTROMO metamodelling framework (through the NOSTROMO metamodelling API) to train the metamodels developed in SIMBAD. This work was done in collaboration with the NOSTROMO team.

The initial AL methodology defined by the SIMBAD project was used to generate the labelled data needed to train the metamodels with the NOSTROMO API.

Finally, the complete input-output space of the metamodels defined is computed for its visualisation and analysis. The work was performed by Fraunhofer. This visualisation allows the analysis of the relationships between the input and output variables of the metamodel.

All the work developed in WP5 is collected in the deliverable D5.1 Active Learning Metamodelling. This document was elaborated by Nommon with the support of the rest of the WP members.

2.3.6 WP6 Demonstration and evaluation of the SIMBAD performance modelling framework

The goal of WP6 was to evaluate the SIMBAD performance modelling framework through a set of case studies, demonstrating the added value of the methodologies developed within the project, highlighting future research areas and identifying the next steps for the exploitation of project results. The specific objectives of WP6 were to: (i) integrate the methodologies developed within SIMBAD with the R-NEST and DYNAMO simulation tools; (ii) conduct the case studies specified in WP2, evaluating the SIMBAD performance modelling framework; and (iii) synthetise the knowledge extracted from the results of the case studies into a set of guidelines to facilitate the transition of the project results to the subsequent stages of the R&D cycle. WP6 was led by UPC. The work done in this WP is described next.

WP6 deals with the evaluation and validation of the three previous work packages. A validation methodology has been proposed for each WP, aimed to evaluate its particular objectives. In each case, the evaluation exercise is adapted to the techniques used. Next, the evaluation exercises carried out for each WP are described:

- Data-driven methods for the estimation of hidden variables and trajectory models (WP3): this WP analysed the use of data driven models and machine learning techniques for the estimation of hidden variables and the prediction of trajectories. The proposed techniques enable the estimation of two hidden variables, CI and PL. Both variables are known only by the company and the dispatcher; however, these two variables have a relevant influence on the configuration of the flying trajectory. The evaluation activities for WP3 consist in comparing the performance of the methods developed for the estimation of hidden variables, as well as analysing the estimation of trajectory KPIs using the estimated hidden variables. For that, the

same three trajectory KPIs used for WP3 (see Section 2.3.3) were selected, namely, fuel consumption, flown distance, and gate-to-gate time.

- Multiscale traffic pattern classifier (WP4): this WP proposed a clustering-based methodology to identify traffic patterns and select representative days at different geographical scales. In order to evaluate the methodology, the results obtained for the two case studies were compared to a manual selection made by experts of 30 days of the year. For each case study, 30 days were selected using the following criteria:
    - FR: considering the significance of the efficiency variables in FR environment, the days have been selected based on the number of flights (top 30);
    - DCB: considering the significance of the ATFM delays variables in the DCB environment, the days have been selected based on the total ATFM delay (top 30).

  For the comparison, three different criteria were used:
    - Analysis of the distribution within the clusters of the 30 days selected by experts. The aim of this criterion is to see whether all the traffic patterns found are covered by the expert's selection.
    - Analysis of the total amount of traffic covered by the experts' selection for each cluster obtained. The aim of this criterion is to assess, based on specific traffic features related to the case study, the portion of traffic covered by the selected days based on their membership to different clusters and their respective shares. For each case study and geographical scale, the following traffic features were used:
        - FR: at ECAC and ANSP levels, the actual average fuel burnt per flight, the gate-to-gate flight time, the average flown distance per flight, and the average minutes of en-route ATFM delay per flight attributable to ANS were considered. At ACC level, the gate-to-gate flight time and the average minutes of en-route ATFM delay per flight attributable to ANS were considered.
        - DCB: at ECAC and ANSP levels, the number of flights, the average of difference between flown trajectories and flight plans, the average departure delay per flight, and the average minutes of en-route ATFM delay per flight attributable to ANS were considered. At ACC level, only the average of difference between flown trajectories and flight plans and the average minutes of en-route ATFM delay per flight attributable to ANS were considered. At airport level, the average departure delay per flight, and the average minutes of en-route ATFM delay per flight attributable to ANS were considered for the analysis.
    - Estimation of the annual value of some selected KPIs using the selected representative days and the experts' selection, and comparison of these values with the actual annual value of the KPI. This way, the representativity of the days selected using the methodology proposed is assessed.
- Application of active learning to air traffic simulation (WP5): this WP analysed the application of active learning techniques to develop metamodels for two ATM simulation tools: DYNAMO and R-NEST. In order to evaluate the metamodels developed, their performance is compared to that of the simulation tools based on four performance metrics:
    - Time: the execution time, measures in seconds, required to obtain one value of the corresponding output KPIs.
    - Computational resources: the required computational resources to obtain one value of the corresponding output KPIs, i.e., the memory, RAM, hard disk, processor speed and similar parameters related to the computer enabling the run.
    - Expert knowledge: the level of expertise and skills necessary to run the simulations/metamodels; three different levels are defined, namely basic, intermediate and advanced. "Basic" level means no specific knowledge is required,

"intermediate" means some knowledge about ATM and trajectory modelling is required, and finally "advanced" means that a clear understanding of the simulation is required.

- Accuracy: comparison between the metamodel predictions and the simulation results and computation of the error as the difference between them.

All the work developed in WP6 is collected in the deliverable D6.1 Evaluation of the SIMBAD performance modelling framework and implementation guidelines. This document was elaborated by UPC with the support of the rest of the WP members.

2.3.7 WP7 Communication, dissemination and exploitation

The goal of WP7 was to facilitate a fruitful and efficient exchange of information with different stakeholders, and prepare for the exploitation of the project results. The specific objectives were to: (i) establish efficient communication channels with external partners, such as industrial actors, policy makers and academic researchers in order to gather their inputs and feedback; (ii) disseminate the findings of the project to encourage exploitation of the results; (iii) to create the conditions for the transfer of the project results to the subsequent stages of the R&I cycle.

WP7 was led by Nommon, with the support of the rest of SIMBAD partners. The work done in WP7 is described in Table 3.

Table 3 Communication, dissemination, and exploitation actions

Action	Description	Date	Place
Creation of the SIMBAD project website (www.simbad-h2020.eu)	The website, implemented and maintained by Nommon, is described in deliverable D7.1 Project Website. This document was prepared by Nommon and submitted in March 2021. It was updated in accordance with the comments received from the SJU and was approved in April 2021	March 2021	online
1st SIMBAD Stakeholder Workshop	The workshop was dedicated to discussing the case studies defined in WP2. The workshop participants included different EEAB members from airlines, air navigation service providers (ANSPs), the SESAR PJ19.04 project, and EUROCONTROL. UPC, CRIDA, Nommon and UPCR participated in the preparation of the workshop	2nd July 2021	online
2nd SIMBAD Stakeholder Workshop	The workshop was dedicated to validate the machine learning methodology defined in WP3, WP4 and WP5. The workshop participants included several academic experts from all the core research areas involved in WP3, WP4, and WP5 (including reinforcement learning, traffic patterns, metamodelling, and visual analytics), as well as representatives from SESAR ER4 TAPAS project and from the SJU. All the partners of the	11th February 2022	online

	project participated in the preparation of the workshop		
3rd SIMBAD Stakeholder Workshop	The workshop will be a joint event with the NOSTROMO project and will be dedicated to disseminate the results of the project and consolidate the SIMBAD White Paper (D7.2)	January 2023	online
Social media	Publication of news about the project in social media such as Twitter and LinkedIn	Spread throughout the project	online
Joint events with other SESAR projects	• The project participated in a workshop with PJ19.04 project to find synergies between both projects and be aligned with their objectives • The project participated in a workshop with PJ19.04 and NOSTROMO project to present the project results and exploit synergies	Spread throughout the project	Online & Brussels
Communications and press releases	• Publication of SIMBAD results at the EC portal CORDIS • Elaboration of a project summary for inclusion in a SESAR 3 JU ER4 results publication distributed at the SESAR Innovation Days	Spread throughout the project	online
Bilateral meetings and presentations	The project was presented individually to Swiss, in a dedicated bilateral meeting	13th July 2021	online
SIDs2021	Presentation of a poster to the SESAR Innovation Days 2021, describing the SIMBAD objectives and approach	December 2021	online
Scientific publications	• Publication of a paper entitled "Visual Analytics for human-centered Machine Learning" in the journal IEEE Computer Graphics & Applications • Submission of the paper entitled "Modelling Flight Trajectories with Multi-modal Generative Adversarial Imitation Learning" to the Intl. Journal Expert Systems with Applications, Elsevier. This paper is currently under revision	-	online
SIDs 2022	Three papers were accepted and presented in the SIDs 2022: • Data-Driven Estimation of Flights' Hidden Parameters • Identification of Traffic Patterns and Selection of Representative Traffic Samples for the Assessment of ATM Performance Problems • Active Learning Metamodelling for R-NEST	December 2022	Budapest

White paper	The white paper "Combining Simulation Models and Big Data Analytics for ATM Performance Analysis: Lessons Learnt from the SIMBAD Project and Way Forward" (D7.2) provides a high-level view of the main results and conclusions of the project. This document is currently being elaborated by Nommon with the support of the rest of the partners	December 2022	online

2.3.8 WP8 Ethics requirements

The objective of WP8 was to ensure compliance with the 'ethics requirements' for the project. WP8 was led by Nommon and the work done includes:

- D8.1 H - POPD – Requirements No. 2. This deliverable describes the mechanisms and procedures put in place by SIMBAD to ensure full compliance with the applicable EU and national laws on data protection. The document includes the names of the Data Protection Officers (DPO) appointed by the partners, together with the DPO for the whole project. It describes SIMBAD's personal data protection policy, including the specific principles that must apply to the processing of any personal data and the technical and organisational measures implemented to safeguard the rights and freedoms of the data subjects and research participants. It also deals with the data anonymisation/pseudonymisation process.

2.4 Key Project Results

The work conducted within the different WPs contributes to the achievement of the project objectives through the provision of the following key results:

- **A methodology for the estimation of hidden variables through the combination of mechanistic trajectory models and data-driven machine learning techniques.** Using this methodology, two hidden variables related to AUs, PL and CI, were estimated for trajectories generated with DYNAMO connecting the Charles de Gaulle airport and the Istanbul Ataturk airport. Both hidden variables were estimated with high accuracy, in particular, the CI was estimated with an error under 4% and the PL with an error under 2%. These models allow the improvement of the trajectory simulations of microsimulation models, as they provide information about the AUs behaviours and preferences.

- **The development of trajectory prediction models**. Reinforcement learning techniques were used to predict trajectories in 4D and patterns of trajectory evolution (in space and time) chosen by AUs, given initial conditions and weather forecast. Trajectories connecting the Charles de Gaulle airport and the Istanbul Ataturk airport were predicted with high accuracy in all dimensions using the Triple-GAIL method, in particular, the trajectories are modelled with an accuracy of more than 99% both in distance and flight time. The prediction of these trajectories allows the reduction of the simulation time of ATM trajectory simulation models, such as DYNAMO, as they allow to disentangle and model the airspace users' trajectories. They also allow the modelling of the landing phase.

- **A clustering-based methodology for the identification of traffic patterns and selection of traffic samples at different geographical scales**. The methodology consists in 6 steps:

1. definition of relevant scenarios and traffic features,
2. data management,
3. application of clustering algorithms,
4. technical evaluation of the results,
5. interpretation of the results, and
6. selection of the traffic samples.

The methodology was validated using two different operational case studies (DCB and FR) for four different geographical scales, ECAC, ANSP, ACC, and airport. The methodology allowed the identification of representative traffic demand patterns, including different types of days, such as the days with air traffic controller staff strike in France, the winter and summer weekends, or some the international holidays. An analysis of the results allowed the characterisation of each pattern obtained in terms of the different variables used. Finally, for each traffic pattern found, two representative days are selected based on the silhouette score, the day with the highest silhouette score (i.e., the most representative day) and the day with the lowest silhouette score (representing the biggest deviation from the pattern). This methodology allows a more comprehensive and efficient performance assessment of ATM solutions and concepts, as it provides a method to systematically select a set of days representing specific traffic conditions and behaviours.

- **The definition of three metamodels for two state-of-the-art ATM simulation tools**. One metamodel for DYNAMO and two metamodels for R-NEST were designed and implemented. For each simulation tool, the metamodels are defined in terms of a different SESAR Solution, namely, FR for the DYNAMO metamodel and DCB for the R-NEST metamodels. They were developed using the active learning technique. The metamodels developed have a very good predictive performance. In particular, the DYNAMO metamodel reaches an accuracy of more than 98%, reducing the computational time from 3 hours to a few seconds. While the first R-NEST metamodel reaches an accuracy of 97% for the punctuality variable and of 99% for the cost-efficiency variable, reducing the computational time from 1.5 hours to a few seconds. These metamodels enable a more efficient exploration of the simulator's input-output space, allowing the identification of those regions with more information or interest for the assessment of an ATM solution, reducing the number of needed simulations.

- **The combination of the metamodelling methodology and the methodology for the selection of representative traffic samples**. Both methodologies were combined to enlarge the temporal scope of the R-NEST metamodel from one day to one AIRAC cycle. The combination of this methodology enables to train the metamodel using only six days of the AIRAC. For those six days, the metamodel reaches a predictive performance of almost 70% for the punctuality variable and of 98% for the cost efficiency variable. While the predictive performance for the rest of the days of the AIRAC is of 75% for the punctuality variable and of more than 96% for the cost efficiency variable. The combine methodology allows the generation of metamodels for large temporal periods, reducing the training cost.

- **An assessment of the results**. The developed methodologies and models have been used to evaluate a set of case studies and assess the performance of the proposed framework as well as the usefulness of the methodologies themselves. The evaluation of results was performed

in three different and independent activities, one per technical WP. The evaluation activities showed that:

1. the methodology proposed in WP3 to estimate trajectory KPIs using the trajectory predictions and the estimated hidden variables for these trajectories allows a quite accurate estimation of three trajectory KPIs, namely the fuel consumption, flown distance, and gate-to-gate time, compared to the KPIs computed for known trajectories. In particular, the three KPIs were computed with an accuracy of 99% using the trajectories simulated with the estimated values of the hidden variables. While, when using the predicted trajectories, the accuracy reached is above 96% for the three KPIs. This allows the improvement of trajectory simulation and KPIs estimation, enhancing the performance assessment of ATM solutions.

2. The clustering methodology proposed in WP4 provides high quality and trustful results, and, in some cases, it allows the identification of more traffic patterns than those provided by the ATM experts through their 30-days selection. Moreover, the representative days selected provide a better estimation of the annual value of the analysed KPIs than the experts' selection, i.e., a better estimation with less days is obtained. For instance, the annual fuel consumption of 2019 at ECAC level is estimated with an accuracy of 99.5% using the representative days selected (12 days), while using the experts' judgment (30 days), the accuracy is of 95.5%. This allows a more comprehensive and efficient assessment of ATM solutions.

3. The metamodels developed in WP5 outperform the ATM simulation tools in terms of execution time, computational resources and expert knowledge. In particular, the execution time is reduced from more than an hour to a few seconds in all the cases and the expert knowledge required to run the simulation tools is reduced significantly, as the metamodels only required to set the input values; hence, a much shallower knowledge is needed. In the DYNAMO metamodel, the computational resources are greatly reduced as well (from a cluster of 6 computers to one computer). In terms of accuracy, the performance of the DYNAMO metamodel is pretty close to that of the simulation tool, while the R-NEST metamodels reach a good performance for one output (cost-efficiency), but for the other (punctuality), the predictive error is quite high. Nevertheless, all the metamodels allow the identification of the more informative regions of the models for which simulations should be run, reducing the number of required simulations for a performance assessment.

- **A performance modelling framework for the impact assessment of ATM concepts/solutions**. The SIMBAD toolset provides a performance modelling framework able to address the three main drawbacks of ATM microsimulation models: the presence of hidden variables involving business sensitive data related to the behaviour of airspace users (AUs), the computational complexity, and the interpretability issues.

2.5 Project Deliverables

Table 4 Project Deliverables

Reference	Title and description	Delivery Date[1]	Dissemination Level[2]
D1.1	Project Management Plan *The Project Management Plan defines how the project is managed, executed, monitored and controlled, including project organisation, schedule, management procedures, and reporting mechanisms.*	06/07/2021	Confidential
D1.2	Data Management Plan *The Data Management plan describes the data management life cycle for the data to be collected, processed and generated by the SIMBAD project, in order to ensure that all the research data is findable, accessible, interoperable and reusable (FAIR) as well as that ethics and data security aspects are properly addressed.*	26/10/2022	Confidential
D1.3	Final Project Report *The Final Project Report covers all the research activities performed by the project, including the final publishable summary report, the plan for use and dissemination of the foreground, and a selfassessment of the TRL achieved at the end of the project. It is available in the project website.*	24/11/2022	Public
D2.1	Specification of Case Studies *This document specifies the case studies and the required input data. It is available in the project website.*	18/11/2021	Public
D2.2	SIMBAD Data Repository *This document presents the SIMBAD Data Repository, which provides secure, efficient and reliable access to the data collected and generated by the project.*	29/07/2021	Confidential
D3.1	Data-Driven Methods for Trajectory Modelling *This document describes the new data-driven methods for the estimation of hidden variables and trajectory models and comparing the performance of the different proposed approaches. It is available in the project website.*	01/09/2022	Public
D4.1	Methodologies and Algorithms for the Selection of Representative Traffic Samples *This document describes the new traffic pattern classifier and the proposed methodology for the selection of a representative set of traffic samples. It is available in the project webpage.*	04/10/2022	Public

[1] Delivery data of latest edition

[2] Public or Confidential

D5.1	Active Learning Metamodelling *This document describes the metamodelling methodologies and querying strategies developed, evaluating their performance in a test environment. It is available in the project webpage.*	14/09/2022	Public
D6.1	Evaluation of the SIMBAD performance modelling framework and implementation guidelines *This document describes the results of the evaluation of the SIMBAD methodologies performed through the case studies and the guidelines on how to apply the new methods and tools to other simulation tools. It is available in the project website.*	24/11/2022	Public
D7.1	Project Website *This document describes the public website of the project, including all the communication and dissemination material produced by the project.*	21/04/2021	Public
D7.2	Combining Simulation Models and Big Data Analytics for ATM Performance Analysis: Lessons Learnt from the SIMBAD Project and Way Forward *White paper providing a high-level view of the main results and conclusions of the project. It is available in the project website.*	16/12/2022	Public
D8.1	H - Requirement No. 2 *This document complements D1.2 Data Management Plan by detailing the mechanisms and procedures put in place by TRANSIT to ensure full compliance with the applicable EU and national laws on data protection.*	29/07/2021	Confidential

3 Links to SESAR Programme

3.1 Contribution to the ATM Master Plan

The new techniques developed by SIMBAD will allow a more comprehensive and accurate assessment of the performance impact of new ATM solutions/concepts, thus having a positive impact across all the KPAs considered in the SESAR performance ambitions (Environment, Capacity, Cost Efficiency, Operational Efficiency, Safety and Security). Additionally, the enhancement of existing trajectory prediction models and the extraction of air traffic patterns will allow both the Network Manager and ANSPs to make more accurate forecasts of the network conditions. This is expected to allow a better management and a more efficient use of airspace resources, which will positively impact on five of the areas specified in SESAR performance ambitions: capacity, cost-efficiency, operational efficiency, environment and safety. Improved traffic forecasting will diminish the number of unnecessary ATFM measures, thus reducing the level of delay, while ANSPs will be able to better adapt their resources to the actual demand, avoiding the personnel costs associated to demand overestimation. Moreover, better matching of airspace capacity with traffic demand will make flight trajectories more efficient. Finally, an improved management of ANSP resources will also have a positive impact on safety, by reducing the risk of overloads.

The SIMBAD project has delivered one solution, whose contribution is discussed below.

SIMBAD Performance Modelling Framework

The SIMBAD solution consists in an ATM performance modelling framework aimed at enhancing the capabilities of large-scale ATM microsimulation models to effectively support performance evaluation at network level. The framework enables the comprehensive exploration and understanding of the performance impact of ATM concepts/solutions under different traffic conditions.

The framework comprises: 1) a model for the estimation of hidden variables related to airspace user behaviour that are necessary inputs for the ATM microsimulation models (e.g., cost index and landing weight); 2) a methodology to identify representative traffic patterns at different scales (ECAC, ANSP, ACC and airport) for each particular problem under study; and 3) a metamodelling framework that enables the approximation of the results of a microsimulation model to facilitate a more efficient exploration of its input-output space.

The SIMBAD performance modelling framework will facilitate a more comprehensive, accurate, and efficient assessment of the performance impact of new ATM solutions/concepts, thus having a positive impact across all the SESAR KPAs and ultimately on the achievement of the ATM Master Plan performance ambitions.

Table 5 Maturity of the solution delivered by the project

Solution	Maturity at project start	Maturity at project end
SIMBAD Performance Modelling Framework	TRL0	TRL1

3.2 Maturity Assessment

Table 6 Research Maturity Assessment of SIMBAD Performance Modelling Framework

ID	Criteria	Satisfaction	Rationale - Link to deliverables - Comments
TRL-1.1	Has the ATM problem/challenge/need(s) that innovation would contribute to solve been identified? Where does the problem lie? Has the ATM problem/challenge/need(s) been quantified that justify the research done?	Achieved	The problem to be solved is the difficulty of the practical application of complex simulation models to strategic ATM performance assessment and decision-making. There are several factors: hidden variables (such as cost index or landing weight), computational complexity, and interpretability issues.
TRL-1.2	Have the solutions (concepts/capabilities/methodologies) under research been defined and described?	Achieved	• Three different methods, namely modelling of trajectories and estimation of hidden variables (cost index and payload mass), traffic pattern classification, and active learning metamodeling, were defined to facilitate hidden variable estimation and increase the tractability and interpretability of large-scale, complex, fast-time ATM simulation models. • The performance modelling framework proposed by SIMBAD aimed at enhancing the capabilities of large-scale ATM microsimulation models to effectively support performance evaluation at network level is described and validated in D6.1.
TRL-1.3	Have assumptions applicable for the innovative concept/technology been documented?	Achieved	Hypothesis and assumptions to implement the three different methods are defined in deliverables D3.1, D4.1 and D5.1, respectively, as well as in the Validation Plan. Some examples are: • the assumption that relevant hidden variables related to airspace users (such as the cost index or the landing weight) can be accurately estimated given M1 flight plans and weather forecast, or • the assumption that a clustering-based approach can be used to identify the different traffic patterns and behaviours present in a

			certain region and to select a set of traffic samples that effectively represents those traffic patterns to allow a more efficient performance assessment of new concepts and solutions.
TRL-1.4	Have the research hypothesis been formulated and documented?	Achieved	The list of hypotheses and modelling assumptions has been formulated and is well documented in the Validation Plan and in deliverables D3.1, D4.1, and D5.1. These hypotheses include, among others, the use of Gaussian processes to model the behaviour of the microsimulation tools or the use of reinforcement learning techniques for trajectory prediction.
TRL-1.5	Do the obtained results from the fundamental research activities suggest innovative solutions (e.g., concepts/methodologies/capabilities? What are these new concepts/methodologies/capabilities? Can they be technically implemented?	Achieved	SIMBAD proposes a solution in the form of a performance modelling framework that provides three main capabilities: • Estimation of hidden variables related to AUs and prediction of trajectories. • Identification of traffic patterns and selection of representative traffic samples. • Definition of active learning metamodels able to efficiently explore the input-output space of ATM microsimulation models. These capabilities have been technically implemented through the components of the SIMBAD Performance Modelling Framework described in D3.1, D4.1, and D5.1.
TRL-1.6	Have the potential strengths and benefits of the solution identified and assessed?	Achieved	The strengths and benefits of each method have been identified and are included in deliverables D3.1, D4.1, D5.1, and D6.1. These are: (i) the accurate estimations of the hidden variables provided by the considered machine learning models; (ii) the high quality and trustful traffic patterns obtained with the methodology for the identification of traffic patterns and selection of traffic samples, and (iii) the metamodels developed

			provide very close results to the simulation tools with much less computational time.
TRL-1.7	Have the potential limitations, weaknesses and constraints of the solution under research been identified and assessed?	Achieved	The limitations, weaknesses and constraints of each method have been identified and are included in deliverables D3.1, D4.1, D5.1, and D6.1. In particular, some of these limitations and weaknesses are: (i) trajectory prediction models sometimes produce unflyable trajectories; (ii) the selection of the traffic features for the identification of traffic patterns should be guided or conducted by well-experienced and skilled analysts who have substantial knowledge in the ATM domain; and (iii) the development of metamodels requires special skills of the analysts who conduct the whole process due to its complexity.
TRL-1.8	Do fundamental research results show contribution to the Programme strategic objectives e.g., performance ambitions identified at the ATM MP Level?	Achieved	The new techniques developed by SIMBAD allow a more comprehensive and accurate assessment of the performance impact of new ATM solutions/concepts, thus having the potential to positive impact across all the SESAR KPAs and contributing to the achievement of the ATM Master Plan ambitions.
TRL-1.9	Have stakeholders been identified, consulted and involved in the assessment of the results? Has their feedback been documented in project deliverables? Have stakeholders shown their interest on the proposed solution?	Partial	The Advisory Board is composed by 15 experts: 10 ATM experts, which include representatives from airlines, air navigation service providers (ANSPs), SESAR PJ19.04 project and EUROCONTROL, and 5 academic experts, including at least one expert in each of the core research areas involved in WP3, WP4, and WP5, namely reinforcement learning, traffic patterns, metamodelling, and visual analytics. The ATM experts were involved in the definition of the case studies in a first workshop. Their feedback is gathered in deliverable D2.1.

			The academia experts validated the machine learning models developed in WP3, WP4 and WP5 in a second workshop. Their feedback is gathered in deliverables D3.1, D4.1, and D5.1. Their feedback will be collected in a final workshop for fine-tuning of the final recommendations and guidelines.
TRL-1.10	Have initial scientific observations been communicated and disseminated (e.g., technical reports/journals/conference papers)?	Achieved	The Communication and Dissemination Plan includes a number of communication and dissemination activities that have been carried out during the project, including conference and journal papers that will be made open-accessible. The results included in the different deliverables were presented at several conferences and workshops. As part of these communication activities, a poster about the project was presented in the SIDs2021, and three papers describing the results obtained in WP3, WP4 and WP5 were submitted and accepted for the SIDs2022. The project will continue the communication and dissemination activities after the project close-out. In particular, the final workshop will take place in January 2023.
TRL-1.11	Are recommendations for further scientific research documented?	Achieved	A set of lessons learnt and way forward will be provided in the deliverable D7.2. Also, a set of lessons learnt and future research is included in Section 4 of this document. Some of the lessons learnt of the project are: • the selection of the most convenient machine learning method for the estimation of hidden variables highly depends on the specific hidden variable;

			• the variables selected and their granularity is of paramount importance to find representative traffic patterns, as well as the data preparation process; • more labelled data points and representative days should be taken in order for the a metamodels to be able to generalize beyond the set of days for which it was trained.

4 Conclusion and Lessons Learned

4.1 Conclusions

The SIMBAD project has delivered a performance modelling framework able to provide state-of-the-art ATM microsimulation models with the level of reliability, tractability and interpretability required to effectively support performance evaluation at ECAC level. The designed models aim to overcome the three main problems that these complex simulation models present when applying to strategic ATM performance assessment and decision-making: hidden variables, computational complexity, and interpretability issues.

To do that, the project has developed three different components, that address the three problems previously mentioned:

- A set of machine learning techniques for the estimation of hidden variables from historical air traffic data that allow a more accurate modelling of aircraft trajectories. In particular, the models defined estimates two hidden variables related to Airspace Users, the CI and the PL. Also, reinforcement learning models for trajectory prediction were developed. These algorithms allow the trajectory modelling according to AUs preferences and behaviour. Moreover, the trajectory prediction and the estimation of hidden variables for these trajectories support the prediction of trajectory KPIs. These results show the achievement of objective O1 of the project (reported in Section 2.2).

- A clustering-based methodology for the identification of traffic patterns and the selection of representative traffic demand samples at different geographical scales for the performance assessment of a given SESAR solution. This methodology groups the days of the year into traffic patterns and from this classification, select a set of representative traffic samples. It was validated through two SESAR solutions, FR and DCB. Representative traffic patterns were obtained for four different geographical scales, namely, ECAC, ANSP, ACC, and airport. These results show the achievement of objective O2 of the project (reported in Section 2.2).

- The definition of performance metamodels able to provide a computationally efficient approximation of the input-output function defined by complex microsimulation models. Three metamodels for two state-of-the-art ATM microsimulation tools, DYNAMO and R-NEST, were developed. For each simulation tool, the metamodels were defined in terms of a different SESAR Solution, namely, FR for the DYNAMO metamodel and DCB for the R-NEST metamodels. The metamodels were trained using the active learning technique. These results show the achievement of objective O3 of the project (reported in Section 2.2).

Each component was developed and validated independently. However, they were developed in such a way that they can work independently or jointly. To demonstrate the potential of the combination of these components, as part of the metamodelling developments, the metamodelling methodology was combined with the methodology for traffic samples selection to generate a metamodel that, being trained with a set of presentative days of an AIRAC, is able to estimate for the rest of the days of that AIRAC. This combined approach leads to very promising results.

The performance of the different components as well as the developed tools and algorithms have been evaluated through a set of case studies based on two SESAR solutions, FR and DCB. This evaluation corresponds to objective O4 of the project (reported in Section 2.2).

The evaluation of the hidden variable estimation models and trajectory prediction models clearly indicate that the application of different machine learning methods leads to a satisfactory performance in terms of accuracy when compared to the brute force (i.e., DYNAMO simulation). In particular, the two hidden variables considered (CI and PL) were estimated with great accuracy. Also, the approach proposed for the estimation of trajectory KPIs led to quite accurate estimations, compared as well to the brute force approach. Nevertheless, despite the good accuracy of the trajectory prediction models, during the evaluation process some unacceptable predicted trajectories were detected, these inaccuracies could lead to unflyable proposals.

Regarding the methodology for the identification of traffic patterns and the selection of representative traffic demand samples, the evaluation exercises involving ATM experts showed that it provides high quality and trustful results for the two SESAR solutions. Moreover, in some specific cases, the tested techniques outperform the experts' judgement, as it reveals the subtle traffic patterns which can be easily omitted by the experts due to their small shares in the total traffic sample. Moreover, they provide a reliable approach for calculating relevant KPIs. In this way, the results demonstrate that a more guided and intelligent selection of representative traffic samples may bring additional benefits to the analysis of ATM performance problems.

Finally, the evaluation of the metamodels shows that they can imitate the behaviour of the microsimulators, providing very close results with much less computational time. These results are quite accurate in the case of DYNAMO, while they show a larger error in the case of R-NEST, however this can be explained by the reduced training set due to time constraints in manual training of R-NEST. These metamodels enable one to explore the two SESAR solutions more deeply, amplifying the exploration of the simulation input and output behaviour space, helping to identify patterns and trends and guiding the simulation analysis in a more efficient way, facilitating the analysis of ATM researchers. The results of the metamodels indicate the performance gain of employing a metamodeling strategy compared to a traditional approach based on the exhausting simulations.

Finally, as the complexity of ATM concepts and solutions is becoming extensively high and multifaceted, the application of novel approaches will foster their performance assessment in more efficient way. The results of the SIMBAD project present a good starting point for further exploration of these techniques in an attempt to improve the performance assessment methodology within SESAR framework.

4.2 Technical Lessons Learned

After the validation and evaluation of the SIMBAD solution through the specific case studies, we can draw the following main technical lessons:

- Splitting the trajectories into the different flight phases and training separate models for each phase shows high potential to increase the accuracy of imitating flown trajectories, improve the trajectory modelling.
- The selection of the most convenient machine learning method for the estimation of hidden variables highly depends on the specific hidden variable. This reinforces the message that

philosophy of "one type fits all" cannot be applied in this field calling for careful considerations of the obtained results and a number of "trial and error" iterations.

- The outcomes obtained from the validation of the methodology for the identification of traffic patterns and the selection of representative traffic samples for both case studies highlight the importance of conducting a thorough data preparation process tailored to the specific operational use case, starting from the selection of variables and features and concluding with the generation of the dataset that will be used as input for the clustering algorithms.

- The variables selected and their granularity is of paramount importance to find representative traffic patterns. If the selected features are not fully appropriate for the specific concept under study or some important variables are not included in the analysis, the interpretation of results may be misleading and not fully reliable. Thus, it is of high importance that the analysis is conducted by well-experienced and skilled analysts who have substantial knowledge in the ATM domain. In addition to the selection of traffic features, the appropriate granularity should also be specified.

- More labelled points should be needed to train the metamodels. Despite the remarkable results obtained with the combination of the metamodelling and the selection of representative samples methodologies, the predictive performance of the design metamodel was not as good as expected. The reason is that, due to time restrictions, not enough labelled data was available during the active learning training process. This could be also overcome taking the day that represents the biggest potential deviation of each cluster (i.e., the day with the lowest silhouette score) as well.

- Taking advantage of the experience of previous SESAR projects such as TRANSIT or IMHOTEP, the first workshop in which the development of the project was discussed was divided into two different events, one in which the end users (Airlines, ANSPs, …) were invited to discuss the case studies and related operational concepts of the project, and another one in which academic experts were invited to discuss and validate the technical developments of the different WPs. Although it is important that every member of the advisory board has a basic understanding of both parts, this way of working has proved to be much more efficient and effective, maximizing the feedback received from each stakeholder.

4.3 Plan for next R&D phase (Next steps)

From the work started at SIMBAD, several lines of research can be distinguished that can be grouped depending on the technological maturity of these new developments.

- On the one hand, it is proposed to promote the developments initiated in SIMBAD to **higher TRLs** through, for example, **participation or collaboration with the SESAR Industrial Research projects**, some examples are:

    - Application of the machine learning models developed for the estimation of hidden variables to improve the work carried out in the PJ18-W1 project on this same topic.

    - Improvement of the combined metamodelling and traffic pattern selection approach. Potentially this continuation could take place within the context of the new SESAR3 call. In particular, these activities have been proposed within the PEARL and AMPLE3 project proposals inside the Transversal Activities (IR1-WA1) area of the call.

- On the other hand, in parallel, SIMBAD defines other more **exploratory lines of research** that could complement the solution but could not be done in the context of the project due to limited time and resources.
    - Development of a new clustering classification refining the set of traffic features used for the clustering analysis, including the hourly entry counts into the geographical area instead of the number of flights per day.
    - Improvement of the active learning scheme used for the metamodel training through the investigation of new faster and more efficient algorithms.

5 References

5.1 Project Deliverables

- D1.1 Project Management Plan, Edition 02.01.00, July 2021.
- D1.2 Data Management Plan, Edition 01.02.00, October 2022.
- D1.3 Final Project Report, Edition 00.01.00, November 2022.
- D2.1 Specification of Case Studies, Edition 01.00.00, January 2022, http://www.nommon-files.es/simbad/SIMBAD_D2.1_Specification_of_Case_Studies_v01.00.00.pdf
- D2.2 SIMBAD Data Repository, Edition 00.01.00, July 2021.
- D3.1 Data-Driven Methods for Trajectory Modelling, Edition 01.00.00, September 2022, http://www.nommon-files.es/simbad/SIMBAD_D3.1_Data_Driven_Methods_for_Trajectory_Modelling_v01.00.00.pdf
- D4.1 Methodologies and Algorithms for the Selection of Representative Traffic Samples, Edition 01.00.00, October 2022, http://www.nommon-files.es/simbad/SIMBAD_D4.1_Methodologies_and_Alg_for_the_Selection_of_Representative_Traffic_Samples_v01.00.00.pdf
- D5.1 Active Learning Metamodelling, Edition 01.00.00, September 2022, http://www.nommon-files.es/simbad/SIMBAD_D5.1_AL_metamodelling_v01.00.00.pdf
- D6.1 Evaluation of the SIMBAD performance modelling framework and implementation guidelines, Edition 00.01.00, November 2022.
- D7.1 Project Website, Edition 02.00.00, July 2021.
- D7.2 Combining Simulation Models and Big Data Analytics for ATM Performance Analysis: Lessons Learnt from the SIMBAD Project and Way Forward, in preparation.
- D8.1 H - POPD - Requirement No. 2, Edition 00.01.00, July 2021.

5.2 Project Publications

- Andrienko, N., Andrienko, G., Adilova, L., & Wrobel, S. (2022). Visual analytics for human-centered machine learning. IEEE Computer Graphics and Applications, 42(1), 123-133, https://openaccess.city.ac.uk/id/eprint/27550/1/
- Vouros, G., Tranos, T., Blekas, K., Santipantakis, G., Melgosa, M., & Prats, X. (2022). Data-driven estimation of flights' hidden parameters. In 12th SESAR Innovation Days (SIDs), https://whova.com/xems/whova_backend/get_event_s3_file_api/?event_id=sesar_202212&file_url=https://d1keuthy5s86c8.cloudfront.net/static/ems/upload/files/1670002505_fifsc

_SIDs_2022_paper_54_final.pdf&eventkey=4c68ac23dcb73ce4b5be87ca1b5b0ccaed6016dfcefe0c4b6af300caca061224

- Sánchez-Cauce, R., Mocholí, D., Cantú Ros, O. G., Herranz, R., Rodríguez, R., Tello, F., & Fabio, A. (2022). Identification of Traffic Patterns and Selection of Representative Traffic Samples for the Assessment of ATM Performance Problems. In 12th SESAR Innovation Days (SIDs), https://whova.com/xems/whova_backend/get_event_s3_file_api/?event_id=sesar_202212&file_url=https://d1keuthy5s86c8.cloudfront.net/static/ems/upload/files/1670001508_imeun_SIDs_2022_paper_89_final.pdf&eventkey=4c68ac23dcb73ce4b5be87ca1b5b0ccaed6016dfcefe0c4b6af300caca061224

- Sánchez-Cauce, R., Riis, C., Antunes, F., Mocholí, D., Cantu Ros, O. G., Câmara Pereira, F., Herranz, R., & Lima Azevedo, C. (2022). Active Learning Metamodelling for R-NEST. In 12th SESAR Innovation Days (SIDs), https://whova.com/xems/whova_backend/get_event_s3_file_api/?event_id=sesar_202212&file_url=https://d1keuthy5s86c8.cloudfront.net/static/ems/upload/files/1670001800_dyxyj_SIDs_2022_paper_91_final.pdf&eventkey=4c68ac23dcb73ce4b5be87ca1b5b0ccaed6016dfcefe0c4b6af300caca061224

- Spatharis, C., et al. (2022). Modelling Flight Trajectories with Multi-modal Generative Adversarial Imitation Learning. Submitted to the Intl. Journal Expert Systems with Applications, Elsevier.

- SIMBAD Project website: https://www.simbad-h2020.eu/

5.3 Other

- Grant Agreement No 894241 - SIMBAD, Annex 1 Description of the Action.

- SIMBAD Consortium Agreement, Issue 1, October 2020.

Annex 1. Experimental Plan

1. Overview

The practical application of complex simulation models to strategic ATM performance assessment and decision-making is hindered by several factors: (i) non-observable variables; (ii) computational complexity; and (iii) interpretability issues.

SIMBAD aims to explore how recent advances in machine learning can help overcome these problems, by developing data-driven simulation techniques that facilitate hidden variable estimation and increase the tractability and interpretability of large-scale, complex, fast-time ATM simulation models.

The specific objectives of the project include:

1. to explore the use of machine learning techniques for the estimation of hidden variables from historical air traffic data;
2. to develop new machine learning algorithms for the classification of traffic patterns that enable the selection of a sufficiently representative set of simulation scenarios;
3. to investigate the use of active learning metamodelling to facilitate a more efficient exploration of the input-output space of complex simulation models; and
4. to demonstrate and evaluate the newly developed methods and tools through a set of case studies in which the proposed techniques are integrated with existing, state-of-the-art ATM simulation tools and used to analyse a variety of ATM performance problems.

Objectives 1, 2 and 3 are addressed through the use of machine learning algorithms. The different machine learning algorithms obtained need to be evaluated and validated, in order to ensure its reliability. For that, three validation processes are developed during the project, one per objective:

- The evaluation and validation of data-driven methods for trajectory prediction. This task compares the performance of the two methods developed, machine learning for efficient model calibration based on historical data and reinforcement learning, in a set of specific evaluation scenarios.
- The assessment and comparison of traffic pattern classification techniques. This task evaluates and validates the different clustering and classification methods developed in terms of accuracy, sensitivity and specificity.
- The evaluation and validation of the metamodelling methodology and the querying strategies. This task evaluates the active learning metamodels definition and their predictive performance. This validation activity is supported by the exploration of the solution space of the metamodel through the use of visual analytics tools.

2. General approach

This section describes the methodology which will be followed to accomplish the three validation activities.

2.1 Data-driven methods for the estimation of hidden variables and trajectory models

SIMBAD explores how data-driven techniques can be used to estimate hidden variables of mechanistic trajectory models, and compare this approach with pure data-driven trajectory prediction models, in order to derive conclusions on the most adequate solution for each specific application.

The data-driven methods for both the estimation of hidden variables and trajectory prediction provide quantitative results. Specifically, errors in prediction by means of the root mean squared error (RMSE), the mean absolute error (MAE) and the mean absolute percentage error (MAPE) are provided, while for trajectories the cross-track (CTE) and the along-track trajectory (ATE) prediction errors in 4D, together with the estimated time of arrival (ETA) error, are also provided.

Models are validated using folds of training and testing trajectories. Regarding the hidden variables, the errors in estimating these variables are provided for the test trajectories, i.e., in trajectories where the actual values of hidden variables are known. On the other hand, regarding the trajectory modelling, the predictive error with respect to the test trajectories is also provided, which relates to the specific conditions considered as input for the prediction (origin destination (OD) pair and weather conditions).

In addition to the above, Key Performance Indicators (KPIs) for trajectories, as estimated by DYNAMO, are evaluated and compared in two specific cases:

1. When the trajectory (e.g., M1 flight plan or flown trajectory) is given and the machine learning module provides an estimation of the hidden variables for this trajectory. In this case, DYNAMO (optimization) takes as input the actual trajectory and the estimated hidden variables, and predicts the specific trajectory KPIs.
2. When the trajectory is not known, but it is predicted, and the machine learning module provides an estimation of the hidden variables for the predicted trajectory. In this case, DYNAMO (optimization) takes as input the predicted trajectory and the estimated hidden variables, and predicts the specific trajectory KPIs.

Specifically, the trajectory KPIs considered are gate-to-gate flight time, average fuel burnt per flight, and flown distance.

Finally, results on estimated KPIs from these two cases are compared between them and with the KPIs obtained when both the flown trajectory and the actual hidden parameters are known, in order to estimate the feasibility of using hidden variables' estimators and trajectory predictors in the pipeline, thus evaluating the potential to simulate the predicted behaviour of airspace users.

For further details on the planning, methodology and outcomes of the validation activities for the estimation of hidden variables and trajectory prediction methods, we encourage the reader to read D3.1 Data-Driven Methods for Trajectory Modelling.

2.2 Multiscale traffic pattern classification

SIMBAD explores how different clustering and classification techniques can be used to identify different traffic patterns and network conditions at different spatial and temporal scales and classify operational days into one of these detected patterns. As a result, a set of representative demand patterns will be generated at different scales, allowing the optimal selection of a sufficiently representative set of traffic scenarios for each particular problem under study.

This section describes the methodology proposed for the identification of representative traffic samples. In particular, the approach proposed and that is validated within the SIMBAD project is the following:

1. Definition of relevant scenarios and traffic features

This step includes:

- the understating of the operational concept to be assessed,
- the identification of the potential operational scenarios to be analysed,
- a preliminary identification, if possible, of the ATM performance questions that are relevant, and
- the selection of the performance indicators and traffic features that are representative of the operational concept under analysis.

Once the first step has been completed, and for each one of the scenarios identified:

2. Data management

The data management process includes both the data exploration phase and all the data pre-processing steps.

The data exploration phase is intended to provide a preliminary understanding of the kind of data available and to gain insights of the potential traffic patterns that can be obtained based, mainly, on the analysis of the temporal evolution of the different traffic features selected.

Regarding the data pre-processing phase, even though it will vary depending on the analyst, the following steps are recommended to be followed, as minimum:

- Data cleaning, in order to remove potential data noise (variables with zero variance, outliers, etc.) This step might include a variable normalisation process.
- Correlation analysis and feature selection, aiming at identifying the relation between the variables available in the dataset to avoid variables containing similar or redundant information.
- Dimensionality reduction, in order to reduce the number of variables to use as input for the identification of traffic patterns with clustering techniques.

3. Application of algorithms for traffic patterns identification

Once the data has been cleaned and the input data, i.e., the final variables selected, have been obtained, the set of clustering algorithms that are more appropriate for the specific problem under analysis should be applied (visual analytics clustering techniques or classical clustering approaches).

As part of the clustering process, the number of optimal clusters needs to be determined. It should be noted that the data exploration results should be used to support the establishment of the number of clusters together with the results obtained by applying state-of-the-art algorithms available for this purpose.

4. Technical evaluation of results

This step is devoted to the analysis of the quality or goodness of the clusters resulting from the application of the clustering algorithms. To this end, several state-of-the-art assessment scores can be used to evaluate the inter-cluster and intra-cluster similarity (e.g., silhouette score, Calinski Harabasz, or Davies Bouldin). These metrics allow the assessment of the performance of the different algorithms in terms of accuracy, sensitivity and specificity.

5. Interpretation of results

To close the methodology, an operational interpretation of the clusters and traffic patterns identified needs to be carried out.

The interpretation of results might be supported with several statistics of the variables that conform each one of the clusters identified. Temporal attributes are also of high relevance in most of the cases.

To complete this step, additional information might be required (e.g., special events that have occurred during the time period covered by the data set).

6. Selection of the traffic sample

Finally, and considering the clusters obtained and that they have been previously validated and interpreted, the set of the most representative days for each one of the clusters has to be selected in order to conform the traffic sample that will be used to assess the performance of the concept under study. To achieve this step, it is proposed to select, if data permits, two days per cluster (the most representative and the less similar one). The representativeness of the days to be chosen can be assessed according to the similarity metrics computed when evaluating the clusters from a purely technical performance point of view (e.g., days with the highest and minimum silhouette score).

The overall methodology is depicted in Figure 2.

Figure 2 Overall methodology for WP4.

This methodology provides both a qualitative and quantitative assessment of the clustering results obtained.

For further details on the planning, methodology and outcomes of the validation activities of the traffic pattern classification, we encourage the reader to read D4.1 Methodologies and Algorithms for the Selection of Representative Traffic Samples.

2.3 Application of active learning to air traffic simulation

Due to its multidimensionality nature, the ex-ante assessment of the performance impact of new ATM concepts and solutions at network-wide level is not often approachable through analytic methods. In these cases, microsimulation models are usually the only feasible and reliable alternative. However, when embedded with enough detail, ATM microsimulation models can be computationally expensive, especially when applied at the scale of the full European network, which in practice limits the number of input-output points that can be explored. This makes it difficult to find the combination of inputs that optimises a certain output (e.g., the optimal mix of new solutions to maximise a certain

performance objective function), as well as to detect possible conditions leading to unacceptable performance degradation. SIMBAD explores how to overcome these shortcomings through the use of active learning metamodelling.

The evaluation and validation of the active learning metamodelling methodology consists in training, and evaluating the metamodels defined.

The first step to build the metamodel is the specification of the model inputs and outputs. These are directly related with the use case under study. In the case of SIMBAD we have selected two use cases to test and validate the methodology, these are Free-Routing (FR) for the metamodel of DYNAMO and Demand and Capacity Balancing (DCB) for the metamodels of R-NEST. This way, the inputs are a number of variables defining different possible implementations of the case study, while the outputs are a list of KPIs related to the relevant SESAR solution. A simplified testing scenario is defined in order to develop and test the metamodels with the corresponding simulation tool, R-NEST or DYNAMO, for each case study. For that, the range of values of the input variables that are parameters of the simulation tools needs to be specified.

The metamodels are trained with the active learning (AL) technique using the following AL framework:

- The labelled training set, L, contains 5 points.
- The rest of the training input points belongs to the unlabelled set, U.
- The machine learning model, M, is a Gaussian process (GP) regressor.
- The oracle, O, is the simulation model (DYNAMO or R-NEST).
- The query function, Q, is to label the instance whose prediction products the highest variance.
- The stopping criterion is that the 10% of the input space is labelled and the total variance of the test set is reduced by 80% of the initial total variance of this set in each of the five last iterations.

This way, the GP model is initially trained with an initial set of labelled points, and then the query function is used to select the following points from the unlabelled set to query the microsimulation model and progressively refine the metamodel until the stopping criterion is met.

To implement the AL process and train the metamodels, the input data is split into three different sets:

- training set: this set is used to train the model. The labelled and unlabelled sets are generated from this set;
- test set: this set is used by the stopping criterion;
- validation set: this set is used to assess the predictive performance of the metamodel once the training process has finished. This set is labelled and the true labels are compared with the predicted on.

The evaluation of the metamodel predictive performance and the active learning scheme is done by means of two error metrics, RMSE and MAPE. During the training process, the predictive error over the test set is computed to monitor the improvements among iterations. Finally, the predictive performance of the trained metamodel is assessed on the separated validation set to analyse its ability to generalise to new data.

The SIMBAD simulation metamodels provide quantitative results, such as the fuel efficiency of the network in the region considered.

For further details on the planning, methodology and outcomes of the metamodel validation activities, we encourage the reader to read D5.1 Active Learning Metamodelling.

3. Validation approach

This section discusses the validation approach to be followed for each of the three components of the SIMBAD solution presented in Section 2.

3.1 Data-driven methods for trajectory prediction evaluation and validation

3.1.1 Validation objectives

The trajectory prediction validation process encompasses two activities:

- the validation of the hidden variable estimation models and
- the validation of the trajectory prediction.

These activities are sequential, as the estimated hidden variables are used for trajectory modelling. One validation objective has been identified for each activity and, for each validation objective, two success criteria have been defined:

- OBJ.01: The hidden variable estimation models reach a good predictive performance and support the accurate prediction of flights' KPIs.
  - SC-OBJ.01-01: Hidden variables are estimated with sufficiently low prediction error (using MAE) given a trajectory (flight plan or flown trajectory) and weather forecast. Prediction errors are judged as sufficiently low by aviation domain experts. Specifically, the MAE must be at most two times the discretization interval of the hidden variable values provided in the training data: 4% for cost index (CI) (i.e., 4 units in [0,100]) and 40% for payload mass (PL) (i.e., 0.4 units in [0,1]).
  - SC-OBJ.01-02: Estimated hidden variables for known trajectories support the prediction of trajectory KPIs using standard prediction error metrics (MAE and RMSE). Prediction errors are judged as sufficient by aviation domain experts. Specifically, the MAE must be at most 4% for fuel, distance, and gate-to-gate time.
- OBJ.02: The trajectory prediction models reach a good predictive performance and support the accurate prediction of flights' KPIs.
  - SC-OBJ.02-01: Trajectories in 4D and patterns of trajectory evolution (in space and time) chosen by AUs are predicted with sufficiently low prediction error in all dimensions (using RMSE, as well as CTE and ATE, and error in ETA), given initial conditions and weather forecast. Prediction errors are judged as sufficient according to errors reported by state-of-the-art methods and the EUROCONTROL Specification for Trajectory Prediction [1]. In any case, prediction errors must not exceed 10% (i.e., 0.1) in 3D with respect to the flight distance and with respect to the flight time.
  - SC-OBJ.02-02: Trajectory predictions and the estimated hidden variables for these trajectories support the prediction of trajectory KPIs with sufficiently low prediction errors using standard prediction error metrics (MAE and RMSE), compared to the KPIs estimated for known trajectories (according to SC-OBJ.01-02). Prediction errors are judged as sufficient according to state-of-the-art methods. Specifically, the MAE must be at most 4%

for fuel compared to that provided by mechanistic models, and the same holds for distance and gate-to-gate time.

3.1.2 Validation exercises

3.1.2.1 Validation of the hidden variable estimation models

Validation objectives and success criteria

The associated validation objective is OBJ.01 and the associated validation criteria are SC-OBJ.01-01 and SC-OBJ.01-02.

Validation assumptions

The research hypotheses for the hidden variable estimation are that they can be estimated with low prediction error given valid M1 flight plans and weather forecast, and that estimated hidden variables for known trajectories can improve the computation of trajectory KPIs.

Validation planning

The input (independent) variables of the hidden variable estimation models are spatial and temporal information concerning aircraft positioning and weather variables (temperature, pressure, wind components, derivatives of true/calibrated/indicated airspeed and geometric altitude). The output (dependent) variables are the hidden variables to estimate, in this case, PL and CI. These estimations are deterministic.

For each of the machine learning methods identified, a sufficient number of independent models is trained using different training sets. For this step, accurate weather forecasts and valid training data provided by DYNAMO are assumed. Additionally, models are defined for a specific OD pair. Next, the models are tested in a sufficient number of different (per model) testing cases to measure accuracy of estimations. Statistical measures concerning accuracy are also provided.

Solution scenarios comprise a set of trajectories for a specific OD pair in different weather conditions, under the existing operational setting, together with estimated hidden variables. The reference scenarios comprise historical trajectories (flight plans or flown trajectories) for a specific OD pair enriched with DYNAMO variables and operated in different weather conditions, associated with the hidden variables to be estimated. Estimated variables are compared to those included in the reference scenarios to prove the accuracy of the models.

The models obtained could be trained with different OD pairs to generalise beyond a single pair and test the accuracy of model's prediction in several pairs different from those in which it has been trained.

At least two iterations of the training cycle will be implemented, the first one produces preliminary results and the second one, final results.

This validation activity is carried out by UPRC, with the support of UPC and Fraunhofer. Table 7 shows the list of risks that may affect this validation activity and their mitigation actions.

The results of this validation activity can be found in Sections 5 and 7.1 of D3.1 Data-Driven Methods for Trajectory Modelling.

Table 7 List of risks and mitigation actions

Risk description	Impact	Prob.	Risk management (mitigation / contingency plan)
New data-driven methods and explored techniques are not relevant for applications	High	Low	This risk is largely mitigated by the continuous dialogue and interaction between WP2, WP3, WP4 and WP5, to ensure that the case studies address questions that are relevant both from the ATM point of view and from the point of view of the applicability of the new methods developed by SIMBAD
The computational resources required for some of the new machine learning algorithms exceed the resources available for the project	High	Low	- The algorithms are developed in an iterative manner and tested in a set of scenarios of increasing complexity - The Nommon, FHFR, CRIDA, UPRC and UPC personnel involved in the project have a wide experience in the design and implementation of scalable big data solutions
Inability to collect sufficient data	High	Low	- The data needs and the relevant sources were identified at the beginning of the project, so that data collection tasks could be launched as early as possible - The participation of CRIDA is of great help to ensure access to detailed, high quality operational data
Delays in software implementation	High	Medium	The Consortium members involved in software development implemented software quality management systems to ensure a robust management of the whole software life cycle. The use of agile development methodologies allows the obtention of valuable results soon that are further developed in successive iterations
Data preparation for machine learning methods takes time and creates delays in the pipeline	Medium	High	- To minimise the impact of the situation, the machine learning algorithms were trained, validated and tested incrementally, considering all batches that have been shared from partners up to a specific point. This allows the consortium to report on results and continue with further updates - The consortium used advanced methods for data preparation and pre-processing, according to their extended experience to handle large volumes of data

3.1.2.2 Validation of the of the trajectory prediction

Validation objectives and success criteria

The associated validation objective is OBJ.02 and the associated validation criteria are SC-OBJ.02-01 and SC-OBJ.02-02.

Validation assumptions

The development of the trajectory prediction models is based on the hypotheses that trajectories can be successfully predicted in 4D in cases where airspace users choose different patterns of trajectory evolutions (in space and time) given initial conditions and weather forecast, and relying on the fact that estimated hidden variables for known trajectories can advance the computation of trajectory KPIs.

Validation planning

The models for the trajectory prediction take as input (independent) variables the OD pair, the aircraft type, and the weather forecast, and output the predicted trajectory (spatial-temporal states in 4D). It is important to note that accurate weather forecasts are assumed and models are restricted to specific OD pairs, although this constraint is trying to be alleviated.

Trajectory modelling involves uncertainties and provides the probability that a predicted trajectory follows a specific pattern. This is an inherent feature of the methods used, which model patterns as latent variables and learn the probability of assigning specific values to these variables.

In the same way that for the validation of the hidden variable estimation models (see Section 0), several independent models for each machine learning method selected are trained using different training sets. The resulting models are tested in a large number of different (per model) testing cases to measure accuracy of predictions. Statistical measures concerning accuracy are provided.

Solution scenarios comprise the OD pair with different weather conditions, under the existing operational setting. Based on these, trajectory models provide solutions that assess the potential of a trajectory to follow a specific pattern and evolve in a specific way in 3D.

The reference scenarios comprise the set of trajectories (flight plans or flown trajectories) for the specific OD pair for different weather conditions, under the existing operational setting. These are the trajectories to model and against which the predicted trajectories are compared.

As for the hidden variable estimation, at least two iterations of the training cycle are implemented, in order to obtain preliminary results and final results.

This validation activity is carried out by UPRC, with the support of UPC and Fraunhofer. The list of risks that may affect this validation activity and their mitigation actions are the same as those for previous activity (Section 0), see Table 7.

The results of this validation activity can be found in Sections 6 and 7.2 of D3.1 Data-Driven Methods for Trajectory Modelling.

3.2 Traffic pattern classification techniques assessment and comparison

3.2.1 Validation objectives

The qualitative and quantitative assessment of the traffic patterns identification framework encompasses two activities:

- Validation of the traffic patterns identification algorithms.
- Validation of the traffic patterns identified as a mean to select the most appropriate traffic samples for the performance assessment of ATM concepts.

One validation objective has been identified for both activities, as both of them are developed simultaneously, and five success criteria have been defined to assess the validation objective:

- OBJ.01: The clustering algorithms generate a rich enough set of clusters able to capture all possible relevant traffic patterns casuistic, where at the same time all days belonging to the same cluster present equivalent, in terms of performance, traffic patterns. This is validated independently for both DCB and FR SESAR solutions.
    - SC-OBJ.01-01: The data features selected as input for the clustering algorithms by applying feature engineering techniques are coherent with the list of KPIs considered as relevant for the use case based on SIMBAD's Advisory Board feedback.
    - SC-OBJ.01-02: The inertia of the clusters identified by the different algorithms is judged as sufficient by the technical developers according to state-of-the-art quality metrics.

- SC-OBJ.01-03: The Calinski-Harabasz score of the clusters identified by the different algorithms is judged as sufficient by the technical developers according to state-of-the-art quality metrics.
- SC-OBJ.01-04: The Silhouette Coefficient of the clusters identified by the different algorithms is judged as sufficient by the technical developers according to state-of-the-art quality metrics.
- SC-OBJ.01-05: The Davies-Bouldin Index of the clusters identified by the different algorithms is judged as sufficient by the technical developers according to state-of-the-art quality metrics.

In all the cases, the technical developers judge as sufficient the state-of-the-art quality metrics when, within the metric range, the results obtained indicate a good performance considering the minimum and maximum possible values.

3.2.2 Validation assumptions

The validation assumptions are that a clustering-based approach can be used to identify the different traffic patterns and behaviours present in a certain region, and to select a set of traffic samples that effectively represents those traffic patterns to allow a more efficient performance assessment of new concepts and solutions.

3.2.3 Validation exercises

Two validation exercises were performed, one to evaluate the validation objective for each use case (DCB and FR). For each case, the methodology introduced in Section 2.2 was applied independently, and the performance of the clustering algorithms is validated in each case. Also, for each use case, the methodology is applied at different geographical scales: ECAC, ANSP, ACC, and airport.

Next, each validation exercise is described.

3.2.3.1 Validation of the clustering results for DCB

Validation objectives and success criteria

The associated validation objective is OBJ.01 and the associated validation criteria are SC-OBJ.01-01, SC-OBJ.01-02, SC-OBJ.01-03, SC-OBJ.01-04, SC-OBJ.01-05.

Validation planning

The methodology introduced in Section 2.2 is applied for this case as follows.

1. Definition of relevant scenarios and traffic features

All the variables identified with the support of the Advisory Board for each geographical scale (ECAC, ANSP, ACC, and airport) are used for the clustering algorithms. For the ECAC and ANSP scales, these variables are: number of instrumental flight rules (IFR) movements, average departure delay per flight, average minutes of airport and en-route ATFM delay per flight attributable to air navigation service (ANS), MET and non-ANS reasons, additional distance flown. For the ACC level, only the number of IFR movements and the ATFM delay variables were computed. Finally, for the airport level, only the number of IFR movements and the airport ATFM delay variables were considered. In turn, for each variable, the following statistics are computed: sum, average, standard deviation, variance, maximum,

minimum, mode, percentiles 25, 75, 90 and 95, median, skewness, and kurtosis. All this information is computed for the year 2019.

2. Data management

Three datasets with different characteristics are generated to compare the performance of the clustering algorithms.

All the features with variance equal to zero are removed from the initial dataset and the remaining ones are normalized. This process provides the first dataset with cleaned data.

Next, a filter is applied to remove the correlated features, yielding the second dataset with non-correlated data.

Finally, a non-supervised dimensionality reduction technique known as Feature Agglomeration is applied to the non-correlated dataset. After this process, the third dataset with dimensionality reduction is generated.

3. Application of algorithms for traffic patterns identification

Two clustering algorithms are tested: k-means and hierarchical clustering. In particular, the agglomerative version of the hierarchical clustering is used.

4. Technical evaluation of results

To obtain the optimal number of clusters for each dataset, both algorithms are run for different number of clusters and for each one the silhouette score is computed to measure the quality of the clusters. In order to complement the silhouette score, the elbow method is also applied for the k-means algorithm. The number of clusters is selected looking for a trade-off between these two metrics.

Once the number of clusters for each dataset has been decided, two additional metrics are computed to compare the quality of the clustering classification for each dataset. These metrics are the Davies Bouldin score and the Calinski Harabasz score. This provides the final clustering classification.

5. Interpretation of results

Once the final classification is selected, the clusters obtained are depicted in a calendar plot in which each day of the year is colored by cluster belonging. The temporal distribution of the clusters is analysed by means of two auxiliary variables:

- Weekend: binary variable indicating whether each day is a weekend day (Weekend = 1) or not (Weekend = 0), and
- IATAWinterSeason: binary variable indicating the IATA season each day belongs to. The IATA summer season starts on the last Saturday of March and ends on the last Saturday of September. For winter days, IATAWinterSeason = 1, while for summer days, IATAWinterSeason = 0.

Also, the results are interpreted by means of the distribution within each cluster of the variables used to perform the clustering analysis, as well as the ones not used.

6. Selection of the traffic sample

Two representative days for each cluster are selected based on the silhouette score:

- the day with the highest silhouette value, that will ensure that the cluster is well represented in the final selection, and
- the day with the lowest silhouette value, that will ensure that the sample also considers potential deviations from the cluster centroid.

This validation activity is carried out by CRIDA with the support of Fraunhofer. The list of risks that may affect this validation activity and their mitigation actions are shown in Table 7.

The results of this validation activity can be found in Section 4.3.3 of D4.1 Methodologies and Algorithms for the Selection of Representative Traffic Samples.

3.2.3.2 Validation of the clustering results for FR

Validation objectives and success criteria

The associated validation objective is OBJ.01 and the associated validation criteria are SC-OBJ.01-01, SC-OBJ.01-02, SC-OBJ.01-03, SC-OBJ.01-04, SC-OBJ.01-05.

Validation planning

The methodology introduced in Section 2.2 is applied for this case as follows.

1. Definition of relevant scenarios and traffic features

Using the variables identified for FR with the support of the Advisory Board, five different scenarios, i.e., combinations of variables, are defined to find the traffic patterns at ECAC and ANSP levels:

- All the FR variables: average flown distance per flight, gate-to-gate flight time, actual average fuel burnt per flight, route charges, average of difference between flown trajectories and flight plans, and average minutes of en-route ATFM delay per flight attributable to ANS, MET or non-ANS reasons.
- Efficiency variables: average flown distance per flight, gate-to-gate flight time, actual average fuel burnt per flight, and route charges.
- Predictability variable: average of difference between flown trajectories and flight plans.
- Regulation variables: average minutes of en-route ATFM delay per flight attributable to ANS, MET, or non-ANS regulation causes.
- Efficiency variables and MET regulation variables.

In the case of the ACC level, as some of the variables cannot be computed, only one scenario with the following variables is considered: entry-to-exit time in the ACC (adaptation of the gate-to-gate time), average of differences between flown trajectories and flight plans in the ACC, and the en-route ATFM delay variables. As this SESAR solution does not affect the airports, the airport scale is not considered.

Finally, as for the DCB use case, the following statistics are computed for each variable: sum, average, standard deviation, variance, maximum, minimum, mode, percentiles 25, 75, 90 and 95, median, skewness, and kurtosis. This information is computed for the year 2019.

2. Data management

For each scenario considered, the following data pre-processing is applied to generate four datasets.

In first place, the features with variance smaller than 0.001 are removed and the remaining features were scaled. Next, a correlation analysis was applied. This analysis revealed that the efficiency variables are highly correlated, so, two datasets were considered in parallel, one containing all the features and another one without the features with correlation coefficient bigger than 0.98.

Finally, principal component analysis (PCA) is applied to both datasets (the one with all the variables and the one without the correlated ones) in order to reduce the dimension of the data, while keeping at the same time the original datasets without PCA. The four resulting datasets are:

1. The dataset with all the features with variance bigger than 0.001 scaled,
2. The dataset without the correlated features,
3. The dataset with all the features with variance bigger than 0.001 and PCA, and
4. The dataset without the correlated features and PCA.

3. Application of algorithms for traffic patterns identification

As for the DCB use case, two clustering algorithms are tested: k-means and the agglomerative version of the hierarchical clustering.

4. Technical evaluation of results

The same process described to evaluate the clustering results in the DCB case is applied here. The optimal number of clusters for each dataset is selected looking for a trade-off between the silhouette score and the elbow method. Next, the quality of the clustering classification for each dataset is compared by means of the Davies Bouldin score and the Calinski Harabasz score. This provides the final clustering classification.

5. Interpretation of results

Once the final classification is selected, the clusters are interpreted in the same way as described for the DCB use case. A calendar plot is used to depict the clusters along the year and two auxiliary binary variables, Weekend and IATAWinterSeason, are used to analyse the temporal distribution of each cluster. Also, the results are interpreted by means of the distribution within each cluster of the variables used to perform the clustering analysis, as well as the ones not used. This analysis provides a characterisation of each cluster.

6. Selection of the traffic sample

Finally, two representative days for each cluster are selected following the same criterion as for the DCB use case: the day with the highest silhouette value and the day with the lowest silhouette value.

This validation activity is carried out by Nommon with the support of Fraunhofer. The list of risks that may affect this validation activity and their mitigation actions are shown in Table 7.

The results of this validation activity can be found in Section 4.3.2 of D4.1 Methodologies and Algorithms for the Selection of Representative Traffic Samples.

3.3 Metamodelling methodology and querying strategies evaluation and validation

3.3.1 Validation objectives

The validation activities of the simulation metamodels are focused on the training of these models using the active learning (AL) technique. One objective has been identified to validate this activity, together with a success criterion to assess its fulfillment:

- OBJ.01: The trained metamodel using the AL scheme reaches a good predictive performance.
    - SC-OBJ.01-01: The predictive error of the metamodel on a separated validation set is less than 15% (using the MAPE metric).

3.3.2 Validation assumptions

The research hypothesis for the development of the active learning metamodelling is that this methodology can be used to translate a complex simulation model into a performance metamodel able to provide a computationally efficient approximation of the input-output function defined by complex ATM microsimulation models, improving computational tractability and interpretability of results.

3.3.3 Validation exercises

For each metamodel defined, one exercise was carried out to validate the metamodel's predictive performance. For that, the predictive error of the trained metamodel is assessed by comparing two scenarios using the RMSE and the MAPE metrics:

- a solution scenario, containing the KPIs estimated by the metamodel, and
- a reference scenario (i.e., the validation set), containing the KPIs provided by the simulation models.

Next, we describe them.

3.3.3.1 Validation of the DYNAMO metamodel

Validation objectives and success criteria

The associated validation objective is OBJ.01 and the associated validation criterion is SC-OBJ.01-01. In order to validate this objective for each metamodel, we proceed as follows.

Validation planning

The metamodel is implemented in the Free Route Airspace Italy (FRAIT). The input variables of the metamodels and their ranges of values are:

- FRAIT variable: binary variable taking value 0 when the airspace configuration used for the DYNAMO simulations is the one corresponding to the 10th November, 2016; and value 1 when it is the one corresponding to the 22nd June, 2017.
- fuel cost: numerical variable that takes values in the interval (0, 3.5] in 0.01 steps.

The rest of the parameters of DYNAMO are fixed. The same set of flights is used for the simulations in both scenarios, the set of flights crossing the FRAIT region on the 22nd June, 2017.

The data split for this metamodel is the following:

- Training set: 85% of the input data.
- Test set: 7.5% of the input data.
- Validation set: 7.5% of the input data.

This validation activity is carried out by Nommon with the support of UPC. The list of risks that may affect this validation activity and their mitigation actions are shown in Table 7.

3.3.3.2 Validation of the R-NEST initial metamodel

Validation objectives and success criteria

The associated validation objective is OBJ.01 and the associated validation criterion is SC-OBJ.01-01. In order to validate this objective for each metamodel, we proceed as follows.

Validation planning

The metamodel is implemented in the Bordeaux ACC, specifically on the lower and east cluster (LFBBCTAE) for the day 5th July 2019.

The input variables of the metamodel and their ranges of values are:

- configurations: numerical variable taking values in the interval [0, 300] in steps of 10.
- sectors: numerical variable taking values in the interval [0, 300] in steps of 10. This variable is always greater than or equal to the configurations variable.

The data split for this metamodel is the following:

- Training set: 70% of the input data.
- Test set: 15% of the input data.
- Validation set: 15% of the input data.

This validation activity is carried out by Nommon. The list of risks that may affect this validation activity and their mitigation actions are shown in Table 7.

3.3.3.3 Validation of the R-NEST extended metamodel

Validation objectives and success criteria

The associated validation objective is OBJ.01 and the associated validation criterion is SC-OBJ.01-01. In order to validate this objective for each metamodel, we proceed as follows.

Validation assumptions

The methodology for the selection of traffic samples (see Section 2.2) can be used to find a representative set of days of an AIRAC to train a metamodel, and this metamodel is able to generalise to the rest of the days of the AIRAC.

Validation planning

The metamodel is implemented in the Bordeaux ACC, specifically on the lower and east cluster (LFBBCTAE) for the 7º AIRAC of 2019 (from 20th June to 17th July). The set of representative days selected are:

- 10/07/2019
- 30/06/2019
- 29/06/2019
- 15/07/2019
- 20/06/2019
- 02/07/2019

The input variables of the metamodel and their ranges of values are:

- configurations: numerical variable taking values in the interval [0, 300] in steps of 10.
- Sectors: numerical variable taking values in the interval [0, 300] in steps of 10. This variable is always greater than or equal to the configurations variable.

The data split for this metamodel is the following:

- Training set: 70% of the input data.
- Test set: 15% of the input data.
- Validation set: 15% of the input data.

An additional validation set corresponding to other days different than the six representatives selected is considered. The purpose of this dataset is to analyse and validate the ability of the metamodel to generalise to other days different from the ones used to train it. This dataset contains 28 points.

This validation activity is carried out by Nommon. The list of risks that may affect this validation activity and their mitigations actions are shown in Table 7.

3.3.3.4 NOSTROMO metamodelling API

As explained in D5.1, the NOSTROMO metamodelling API is finally used to train the defined metamodels, and the AL framework defined in SIMBAD is used to generate the labelled datasets. For the metamodels validation using this API, the following validation activity is defined.

Validation objectives and success criteria

The associated validation objective is OBJ.01 and the associated validation criterion is SC-OBJ.01-01. In order to validate this objective for each metamodel, we proceed as follows.

Validation planning

The AL framework used by the API is:

- The labelled training set, L, contains 3 points in the case of the DYNAMO metamodel and the initial R-NEST metamodel, and to 10 in the case of the extended R-NEST metamodel.
- The rest of the training input points belongs to the unlabelled set, U.
- The machine learning model, M, is a GP regressor.
- The oracle, O, is the simulation model (DYNAMO or R-NEST).
- As query function, two options are available: to label the instance with highest predictive variance and to randomly select the instance to be labelled.
- The stopping criterion is a predefined number of iterations. In all cases, the number of iterations is the difference between the size of the training set and the size of the initial labelled set, in order to use all the labelled data available.

The data split used to train each metamodel is:

- DYNAMO metamodel: 25 points for the training set, 10 points for the test set, and 12 points for the validation set,
- first R-NEST metamodel: 45 points for the training set, 13 points for the test set, and 15 points for the validation set, and
- second R-NEST metamodel: 65 points for the training set, 15 points for the test set, 13 points for the validation set with the set of the six representative days, and 28 labelled points for the validation set with dates different from the six representative days.

The metamodels are trained using both query functions and the one that produces the best predictive results for the test set is finally used. In order to analyse the stability of the metamodel, the training process is repeated 100 times. Then, the predictive performance of the trained metamodels is assessed on the validation set. The predictive error of the metamodels is measured by means of the RMSE and MAPE metrics.

This validation activity is carried out by Nommon, with the support of the NOSTROMO team. The list of risks that may affect this validation activity and their mitigation actions are shown in Table 7.

The results of this validation activity can be found in Section 8 of D5.1 Active Learning Metamodelling.

4. Data and software collection and generation

4.1 Purpose of the data collection/generation

The goal of the SIMBAD data collection is to provide access to all the different types of data required to support the proper development of the project. This may include, for instance, flight plans, delay data, and weather information. However, the specific case studies defined for the different incremental steps of the SIMBAD project will determine the exact datasets to be collected.

These data are analysed and combined during the project execution to feed the machine learning models and the two microsimulation ATM models used within SIMBAD; R-NEST and DYNAMO. The ATM simulation models outputs allow the computation of relevant indicators and parameters which are used to train the built metamodels.

Both the generated data and the collected data which need to be shared between two or more Consortium Members is stored and shared through the SIMBAD Data Repository. Public results are published in the SIMBAD project website.

4.2 Relation to the objectives of the project

Data is one of the cornerstones of the SIMBAD project, whose main goal is to develop and evaluate a set of machine learning approaches aimed at providing state of-the-art ATM microsimulation models with the level of reliability, tractability and interpretability required to effectively support performance evaluation at ECAC level.

As stated in Section 1, the specific objectives of the project are:

1. to explore the use of machine learning techniques for the estimation of hidden variables from historical air traffic data;

2. to develop new machine learning algorithms for the classification of traffic patterns that enable the selection of a sufficiently representative set of simulation scenarios;
3. to investigate the use of active learning metamodelling to facilitate a more efficient exploration of the input output space of complex simulation models; and
4. to demonstrate and evaluate the newly developed methods and tools through a set of case studies in which the proposed techniques will be integrated with existing, state-of-the-art ATM simulation tools and used to analyse a variety of ATM performance problems.

All these objectives are deeply related with the collection and generation of data. The implementation and training of the machine learning models as well as the implementation, development and validation of the metamodels require the collection of a series of datasets (e.g., flight schedules, weather data, and delay data). These data feed also the ATM microsimulation tools. In turn, these models generate new predicted data which need to be stored.

4.3 Features of the collected data

The following table summarise the datasets identified for the development of the SIMBAD machine learning models (objectives 1, 2, and 3).

Dataset	Description	Category	Origin	Format	Expected size	Objective
COLLECTED DATA						
DDR2	Flight plan (routes) pool, ATFM regulation pool (via statistical analysis), turnaround time (via statistical analysis), taxi-out time (via statistical analysis)	Air transport data	ECTL online portal	csv	6GB per month	1, 2, 3
ADS-B	Aircraft radar tracker. Contains both aircraft position and flight information	Air transport data	Open Sky Database	txt	6 GB per month	1
GFSANL	Wind, temperature, pressure and a wide range of meteorological indicators	Meteorological data	NOAA	GRIB2	60 MB per 6 hours	1
EUMETSAT	Convective weather/storms	Meteorological data	AEMET	NetCDF	700 MB	2
reanalysis-era5-single-levels	Wind, temperature, pressure and a wide range of meteorological indicators	Meteorological data	ECMWF	GRIB2	60 MB*	1

Dataset	Description	Category	Origin	Format	Expected size	Objective
BADA	Aircraft performance data	Air transport data	ECTL	txt, csv and xml	<1MB per aircraft model	1, 3
METAR	Wind strength and direction, temperature, pressure, visibility, etc.	Meteorological data	IOWA mesonet	csv	3 MB per year per airport	1, 3

*Depending on the features needed.

4.4 Features of the generated data

This section describes the data generated when addressing each project objective.

4.4.1 Data generated for objective 1

Table 8, Table 9, and Table 10 show the different data generated for the estimation of hidden variables, the trajectory modelling, and the estimation of flights' KPIs, respectively.

Table 8 Data generated for the estimation of hidden variables

Dataset	Description	Format
DYNAMO_FP	Dataset with 78 columns (features) containing the DYNAMO simulated trajectories (flight plans) of flights which connect Charles de Gaulle (LFPG) and Istanbul Ataturk airports (LTBA), for all the combinations of the CI and PL	csv
DYNAMO_FP (69)	Dataset containing the trajectories of the DYNAMO_FP dataset with 69 features estimated by exploiting the trajectory variables	csv
DYNAMO_FP (51)	Dataset containing the trajectories of the DYNAMO_FP dataset with 51 features estimated from the pre-processing of the trajectories, considering only weather variables associated to spatiotemporal trajectory points and the patio-temporal variables at each trajectory point	csv
Enriched predicted trajectories	Dataset containing the trajectories provided by the trajectory prediction algorithms with 51 features estimated from the pre-processing of the trajectories	csv

Dataset	Description	Format
Estimated hidden variables for DYNAMO FP trajectories	Dataset containing estimated hidden variables for the trajectories contained in the DYNAMO_FP dataset. There is one file for each of the methods tested (SVR, Lasso, KRR, GBM, DNN) and for each of the datasets (DYNAMO_FP(69) and DYNAMO_FP(51)) used. Each file comprises 5 columns with flight id, the true hidden variables and the estimated hidden variables	csv
Estimated hidden variables for predicted trajectories	Dataset containing estimated hidden variables for the trajectories contained in the enriched predicted trajectories dataset. There is one file for each of the methods tested (SVR, Lasso, KRR, GBM, DNN). Each file comprises 3 columns with flight id, and the estimated hidden variables	csv

Table 9 Data generated for the trajectory modelling

Dataset	Description	Format
Processed ALLFT+	Dataset containing flights connecting Charles de Gaulle airport (LFPG) and Istanbul Ataturk airport (LTBA) with a flight state reported as final (i.e., reported values in the record are either “FI” or “TE”). All trajectory points are enriched with the corresponding weather variables as well as with a variable specifying the estimated phase of the flight, resulting into 10 columns per trajectory point	csv
Predicted trajectories	Dataset containing the predicted trajectories with 5 columns specifying flight ID and spatio-temporal variables at each trajectory point	csv

Table 10 Data generated for the estimation of flights’ KPIs

Dataset	Description	Format
ALLFT_FLIGHTID	Dataset with 5 columns containing the KPIs for a specific ALLFT+ flight	csv
ALLFT_FLIGHTID	Dataset with 27 columns containing the KPI for each segment of an ALLFT+ flight	so6
ALLFT_ALLFLIGHTS	Dataset with 5 columns containing the KPIs for each ALLFT+ flight	csv
PREDICTED_FLIGHTID	Dataset with 5 columns containing the KPIs for a specific predicted flight	csv
PREDICTED_FLIGHTID	Dataset with 27 columns containing the KPI for each segment of a predicted flight	so6

Dataset	Description	Format
PREDICTED_ALLFLIGHTS	Dataset with 5 columns containing the KPIs for each predicted flight	csv
DYNAMO_FPFLIGHTID	Dataset with 5 columns containing the KPIs for a specific DYNAMO_FP flight	csv

4.4.2 Data generated for objective 2

Table 11, Table 12, Table 13, and Table 14 show the different data generated for the identification of traffic patterns and selection of representative traffic samples for each geographical scale. As can be seen, for each scale eight datasets were generated:

- Two input datasets containing the computed KPIs.
- Three datasets for each case study (FR and DCB): the clustering classification, the statistical analysis of each cluster, and the selection of representative days.

At airport level, only 5 datasets are generated as the FR case study does not include this geographical scale.

Table 11 Data generated for the traffic patterns at ECAC level

Dataset	Description	Format
Input data - KPIs (non-regulations)	Dataset with 102 columns containing all the KPIs and corresponding statistics for every day of 2019 at ECAC level	csv
Input data - KPIs (regulations)	Dataset with 93 columns containing all the KPIs and corresponding statistics related to regulations for every day of 2019 at ECAC level	csv
Clustering data - FR	Dataset indicating the cluster each day belongs to for each scenario of the FR case study at ECAC level	csv
Statistical analysis - FR	Dataset containing the statistical analysis of each cluster obtained for the FR case study at ECAC level	csv
Representative days - FR	Dataset containing the representative days for each cluster and scenario of the FR case study at ECAC level	csv
Clustering data - DCB	Dataset indicating the cluster each day belongs to for the DCB case study at ECAC level	csv
Statistical analysis - DCB	Dataset containing the statistical analysis of each cluster obtained for the DCB case study at ECAC level	csv
Representative days - DCB	Dataset containing the representative days for each cluster of the DCB case study at ECAC level	csv

Table 12 Data generated for the traffic patterns at ANSP level

Dataset	Description	Format
Input data - KPIs (non-regulations)	Dataset with 103 columns containing all the KPIs and corresponding statistics for every day of 2019 at ANSP level	csv
Input data - KPIs (regulations)	Dataset with 93 columns containing all the KPIs and corresponding statistics related to regulations for every day of 2019 at ANSP level	csv
Clustering data - FR	Dataset indicating the cluster each day belongs to for each scenario of the FR case study at ANSP level	csv
Statistical analysis - FR	Dataset containing the statistical analysis of each cluster obtained for the FR case study at ANSP level	csv
Representative days - FR	Dataset containing the representative days for each cluster and scenario of the FR case study at ANSP level	csv
Clustering data - DCB	Dataset indicating the cluster each day belongs to for the DCB case study at ANSP level	csv
Statistical analysis - DCB	Dataset containing the statistical analysis of each cluster obtained for the DCB case study at ANSP level	csv
Representative days - DCB	Dataset containing the representative days for each cluster of the DCB case study at ANSP level	csv

Table 13 Data generated for the traffic patterns at ACC level

Dataset	Description	Format
Input data - KPIs (non-regulations)	Dataset with 31 columns containing all the KPIs and corresponding statistics for every day of 2019 at ACC level	csv
Input data - KPIs (regulations)	Dataset with 51 columns containing all the KPIs and corresponding statistics related to regulations for every day of 2019 at ACC level	csv
Clustering data - FR	Dataset indicating the cluster each day belongs to for each scenario of the FR case study at ACC level	csv
Statistical analysis - FR	Dataset containing the statistical analysis of each cluster obtained for the FR case study at ACC level	csv
Representative days - FR	Dataset containing the representative days for each cluster and scenario of the FR case study at ACC level	csv
Clustering data - DCB	Dataset indicating the cluster each day belongs to for the DCB case study at ACC level	csv

Dataset	Description	Format
Statistical analysis - DCB	Dataset containing the statistical analysis of each cluster obtained for the DCB case study at ACC level	csv
Representative days - DCB	Dataset containing the representative days for each cluster of the DCB case study at ACC level	csv

Table 14 Data generated for the traffic patterns at airport level

Dataset	Description	Format
Input data - KPIs (non-regulations)	Dataset with 18 columns containing all the KPIs and corresponding statistics for every day of 2019 at airport level	csv
Input data - KPIs (regulations)	Dataset with 49 columns containing all the KPIs and corresponding statistics related to regulations for every day of 2019 at airport level	csv
Clustering data - DCB	Dataset indicating the cluster each day belongs to for the DCB case study at airport level	csv
Statistical analysis - DCB	Dataset containing the statistical analysis of each cluster obtained for the DCB case study at airport level	csv
Representative days - DCB	Dataset containing the representative days for each cluster of the DCB case study at airport level	csv

4.4.3 Data generated for objective 3

Table 15, Table 16, and Table 17 show the different data generated for the DYNAMO metamodel, the first R-NEST metamodel, and the extended R-NEST metamodel, respectively. As can be seen, for each metamodel three datasets are generated:

- Input data: dataset containing the complete input space of the metamodel.
- Labelled data: dataset containing the labelled data generated along the active learning (AL) training process.
- Predicted data: dataset containing the predictions produced by the trained metamodel and their standard deviation.

Table 15 Data generated for the DYNAMO metamodel

Dataset	Description	Format
DYNAMO data	Dataset with 79 columns (features) containing the DYNAMO simulated trajectories (flight plans) of the flights crossing the FRAIT region for a specific fuel cost and airspace configuration	csv
Input data	Dataset with two columns, fuel cost and FRA, containing all the possible combinations of the input variables of the metamodel	csv

Dataset	Description	Format
Labelled data	Dataset with three columns, fuel cost and FRA (input variables), and FEFF1 (output variable), containing the labelled set generated with the AL process	csv
Predicted data	Dataset with four columns, fuel cost and FRA (input variables), and predicted FEFF1 and predictive standard deviation (output variables), containing the predictions and the predictive standard deviations of the metamodel	csv

Table 16 Data generated for the first R-NEST metamodel

Dataset	Description	Format
Input data	Dataset with two columns, configurations and sectors, containing all the possible combinations of the input variables of the metamodel	csv
Labelled data	Dataset with four columns, configurations and sectors (input variables), and PUN1 and CEF2 (output variable), containing the labelled set generated with the AL process	csv
Predicted data	Dataset with six columns, configurations and sectors (input variables), and predicted PUN1, predicted CEF2, and predictive standard deviation for each variable (output variables), containing the predictions and the predictive standard deviations of the metamodel	csv

Table 17 Data generated for the extended R-NEST metamodel

Dataset	Description	Format
Input data	Dataset with 26 columns, configurations, sectors, and (24) hourly entry counts, containing all the possible combinations of the input variables of the metamodel for the six representative days	csv
Labelled data	Dataset with 28 columns, configurations, sectors, and hourly entry counts (input variables), PUN1 and CEF2 (output variable), containing the labelled set generated with the AL process	csv
Predicted data	Dataset with 30 columns, configurations, sectors, and hourly entry counts (input variables), predicted PUN1, predicted CEF2, and predictive standard deviation for each variable (output variables), containing the predictions and the predictive standard deviations of the metamodel	csv

4.4.4 Data generated for objective 4

As this objective deals with the evaluation activities of the project, the data generated for this objective are the same as the ones described in Sections 4.4.1, 4.4.2, and 4.4.3.

5. Research coordination and development

5.1 Research data management

The project aims to ensure that all the research data is findable, accessible, interoperable and reusable (FAIR) as well as that ethics and data security aspects are properly addressed.

The mechanisms for sharing, verification, curation, preservation, reuse and further exploitation of the data used by the SIMBAD project are established in the Data Management Plan (DMP). The SIMBAD DMP includes information on:

- what data is collected, processed and generated,
- the handling of research data during and after the end of the project,
- which methodologies and standards are applied,
- whether data are shared/made open access, and
- how data are curated and preserved (including after the end of the project).

All the data is documented and stored in such a manner that they are searchable efficiently. This philosophy will increase the potential reuse of these data, both inside and beyond the project. The effective application of this principle requires the implementation of a consistent and meaningful meta-information for each dataset. To that end, SIMBAD adopted conventions for metadata creation, naming, searching and control version management as stated in the DMP.

The data collected by the project may be re-used by third parties only if allowed by the data owner. Under request, SIMBAD will provide the metadata to ease the identification of the datasets.

The re-use of the data produced by the project will be subject to the general principles for dissemination and transfer of results set in the SIMBAD Consortium Agreement. In particular, all the data produced will be shared for re-use unless:

- the protection of one Consortium Member's results or background would be adversely affected;
- legitimate academic or commercial interests of one Consortium Member in relation to the results or background would be significantly harmed.

The SIMBAD Consortium shall ensure that no privacy or data protection rights are violated and that data management procedures comply with all relevant national and EU legislation. In particular, the Consortium ensures full compliance with the European General Data Protection Regulation (Regulation (EU) 2016/679). The Consortium shall implement the necessary technical and organisational measures to ensure data security and prevent tampering, loss, or unauthorised access. Further details on ethical aspects are included in deliverable D8.1 H – POPD – Requirement No. 2.

Personal data (names, organisations, nationality, e-mail, phone, etc.) are collected when contacting external experts to ask for information and inputs. The use of personal data is restricted to contact purposes (for participation in workshops, surveys, interviews, etc.) and for providing feedback (e.g.,

for communicating that a deliverable is publicly available). These data is stored in the collaboration space of SIMBAD and access to the information is allowed only to the members of the project team involved in the organisation of the events.

5.2 Research data dissemination

The data is made open accessible in different ways according to their confidentiality:

- Data which is only for the use of the Consortium is accessible through the SIMBAD Data Repository or the SIMBAD Information System for Consortium members, but not for general public.
- Data generated by the project and catalogued as public is accessible through the SIMBAD website.
- Data which is only for the use of specific Consortium member(s) is not accessible for any other member of the Consortium or the general public.

6. References

[1] EUROCONTROL Specification of Trajectory Prediction, September 2017, available at: https://www.eurocontrol.int/publication/eurocontrol-specification-trajectory-prediction

Appendix A

A.1 Acronyms and Terminology

Table 18 Acronyms and technology

Term	Definition
AL	Active Learning
ANS	Air Navigation Service
ATCo	Air Traffic Controller
ATE	Along-track Trajectory Error
ATM	Air Traffic Management
AU	Airspace User
CI	Cost Index
CTE	Cross-Track Error
DCB	Demand and Capacity Balancing
ETA	Estimated Time of Arrival
FR	Free-Routing
FRAIT	Free Route Airspace Italy
GP	Gaussian Process
IFR	Instrumental Flight Rules
KPA	Key Performance Area
KPI	Key Performance Indicator
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
OD	Origin Destination
OS	Opening Scheme
PCA	Principal Component Analysis
PL	Payload mass
RMSE	Root Mean Squared Error

SESAR	Single European Sky ATM Research Programme
S3JU	SESAR3 Joint Undertaking (Agency of the European Commission)