EXPLORATORY RESEARCH

# Final Project Results Report

| | |
|---|---|
| Deliverable ID: | D1.2 |
| Dissemination Level: | PU |
| Project Acronym: | NOSTROMO |
| Grant: | 892517 |
| Call: | H2020-SESAR-2019-2 |
| Topic: | SESAR-ER4-26-2019 |
| Consortium Coordinator: | CRIDA |
| Edition date: | 04 November 2022 |
| Edition: | 00.02.00 |
| Template Edition: | 02.00.05 |

EUROPEAN PARTNERSHIP

Co-funded by the European Union

# Authoring & Approval

## Authors of the document

| Name / Beneficiary | Position / Title | Date |
| --- | --- | --- |
| M. Cano / CRIDA | Project Coordinator | 12/09/2022 |
| D. Mocholí / NOMMON | WP2 – WP6 Leader | 06/10/2022 |
| F. Camara / DTU | WP3 Leader | 06/10/2022 |
| J. Kuljanin / UPC | WP4 Leader | 06/10/2022 |
| G. Gurtner / UoW | WP5 Leader | 06/10/2022 |
| A. Perillo / CRIDA | WP7 Leader | 06/10/2022 |
| J. López / CRIDA | WP8 Leader | 06/10/2022 |

## Reviewers internal to the project

| Name / Beneficiary | Position / Title | Date |
| --- | --- | --- |
| F. Antunes / DTU | Project Member | 03/11/2022 |
| D. Mocholí / NOMMON | Project Member | 07/10/2022 |
| J. Kuljanin / UPC | Project Member | 07/10/2022 |
| G. Gurtner / UoW | Project Member | 07/10/2022 |
| Z. Tibichte / ISA | Project Member | 07/10/2022 |
| A. Perillo / CRIDA | Project Member | 07/10/2022 |
| J. López / CRIDA | Quality Manager | 07/10/2022 |
| R. Herranz / NOMMON | Project Member | 03/11/2022 |

## Reviewers external to the project

| Name / Beneficiary | Position / Title | Date |
| --- | --- | --- |
| José Manuel Cordero / CRIDA - ENAIRE | PJ19.04 Project Coordinator | 27/10/2022 |

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

| Name / Beneficiary | Position / Title | Date |
| --- | --- | --- |
| M. Cano / CRIDA | Project Coordinator | 04/11/2022 |

## Rejected By - Representatives of beneficiaries involved in the project

| Name and/or Beneficiary | Position / Title | Date |
| --- | --- | --- |
| N/A | | |

## Document History

| Edition | Date | Status | Name / Beneficiary | Justification |
|---|---|---|---|---|
| 00.00.01 | 12/09/2022 | Draft | CRIDA | Initial Draft |
| 00.00.02 | 04/10/2022 | Draft | CRIDA | Version for partners review and approval |
| 00.01.00 | 07/10/2022 | Final Draft | CRIDA | Version for SJU delivery |
| 00.01.01 | 02/11/2022 | Revised Draft | CRIDA | Update with SJU and PJ19.04 comments.<br>Version for partners' review. |
| 00.02.00 | 04/11/2022 | Final Draft | CRIDA | Version for SJU re-submission |

# NOSTROMO

NEXT-GENERATION OPEN-SOURCE TOOLS FOR ATM PERFORMANCE MODELLING AND OPTIMISATION

This document is part of a project that has received funding from the SESAR3 Joint Undertaking under grant agreement No 892517 under European Union's Horizon 2020 research and innovation programme.

## Abstract

The main objective of the NOSTROMO project has been to develop, demonstrate and evaluate an innovative modelling approach for the rigorous and comprehensive assessment of the performance impact of future ATM concepts and solutions at ECAC network level. This approach brings together the ability of bottom-up microscopic models to capture emergent behaviour and interdependencies between different solutions with the level of tractability and interpretability required to effectively support decision-making.

This report provides a summary of NOSTROMO accomplishments and contributions to the SESAR Programme. It gathers technical lessons learned and concludes proposing further developments to facilitate the use of the NOSTROMO methodology in the future SESAR 3 Programme.

# Table of Contents



# List of Tables

## List of Figures

# 1 Executive Summary

NOSTROMO is a research project funded by the SESAR Joint Undertaking within the European Unions's Horizon 2020 research and innovation programme under grant agreement No. 892517. Particulary NOSTROMO addresses the Research Topic SESAR-ER4-26-2019 'ATM Validation for a Digitalised ATM'. NOSTROMO is an applied research project aiming at achieving TRL2.

NOSTROMO was led by CRIDA, the R&D+I Centre of the Spanish Air Navigation Service Provider (ENAIRE), with participation from Nommon Solutions and Technologies SL (NOMMON), the University of Westminster LBG (UoW), Universitat Politecnica de Catalunya (UPC), Danmarks Tekniske Universitet (DTU) and ISA Software (ISA).

In recent years, the question of how to assess the performance impact of new ATM solutions at a system-wide level has arisen. To address this challenge, NOSTROMO project aimed to develop new approaches to ATM performance modelling able to **reconcile model transparency, computational tractability and ease of use with the necessary sophistication required for a realistic representation of the ATM system.**

The project was organised around five case studies and followed an incremental approach based on three iterations. The results obtained during the development and execution of the different case studies served to refine the proposed methodology in an iterative manner.

As **main achievements** of the project, the following items can be highlighted:

- Development of an **active learning metamodeling methodology** in the context of the SESAR Performance Framework aiming at reducing the computational burden often associated with fast-time simulation-based studies. This methodology is not expected to substitute the traditional simulation tools but instead to complement the current state of practice of ATM performance assessment.

- Successful **investigation of the feasibility** of such methodology using two state-of-art ATM simulators, namely Mercury and FLITAN, through a proof-of-concept framework. The various experiments point to a prediction performance 1,000 times faster and an accuracy error of around 11% with respect to the simulation tools under study.

- Development of a **prototypal API** enabling the employment of the proposed methodology and the integration of ATM simulators, which may support a future common SESAR integrated simulation and metamodelling platform for performance assessment, visualisation, and decision support.

- Proposal of some **guidelines for future simulation models** to be developed in the scope of SESAR, ensuring, to the maximum extent possible, compatibility with the NOSTROMO architecture.

- Development of an **interactive dashboard** that facilitates the understanding, analysis, and communication of the ATM performance metamodel results.

- **Cost benefit analysis** of the application of the metamodeling methodology proposed by NOSTROMO.



EUROPEAN PARTNERSHIP

Co-funded by the European Union

The final results of the project with real case studies and the most complex Solutions selected during the project showed that the metamodeling approach followed by NOSTROMO provides results very close to the simulator with much less computational time. The NOSTROMO metamodels allow a deeper assessment of a solution, amplifying the exploration of the simulation input and output behaviour space and helping to identify patterns and trends.

Considering the achievements and conclusions, at the moment of the development of this deliverable, the work performed in NOSTROMO and SIMBAD projects has been taken as candidate to be considered in the SESAR3 activities related to the Master Plan and Performance Assessment, through two SESAR 3 proposals for the IR1 Call, AMPLE3 and PEARL respectively. In particular, the NOSTROMO approach has been proposed to be used in the development of Optimised Deployment Scenarios in AMPLE3, while the performance dashboard to be developed in PEARL will build on the visualisation tools developed by NOSTROMO, among other relevant inputs.

For the transition of the NOSTROMO Methodology to the SESAR3 IR Programme a number of improvements will be required, such as the use of more active learning strategies, the development of APIs for other ATM simulators and the integration of multiple simulators in a single metamodel.



EUROPEAN PARTNERSHIP

Co-funded by the European Union

# 2 Project Overview

## 2.1 Operational/Technical Context

### Problem addressed by the project

The Air Traffic Management (ATM) system is composed of a myriad of elements that interact with each other, including interdependent policies and regulations, stakeholders, technologies and market conditions. These interactions give rise to a number of properties characteristic of **complex adaptive systems**, such as non-linearity, emergence and adaptation, which make the ATM system intrinsically difficult to model. One of the most challenging modelling problems is the assessment of the performance impact of new solutions at a system-wide level.

In recent years, the question of how to assess the performance impact of new ATM solutions at a system wide level has arisen. The development of methodologies to **evaluate the impact of new ATM concepts and technologies on high-level, system wide Key Performance Indicators (KPIs)** has been a long-time objective of the ATM research community. Low-level validation activities based on fast-time simulation, human-in-the-loop (HITL) simulation, shadow-mode trials and live trials provide accurate estimates of the performance of a certain solution in a given operational environment; however, implementing such validation approaches for different combinations of solutions at a network-wide scale is infeasible, or at least prohibitive in terms of both cost and time. It is therefore necessary to resort to performance models that consolidate the results of low-level validation experiments conducted for different solutions at a local level and estimate the integrated impact of such solutions at network level.

To address this challenge, the NOSTROMO project aimed to develop new approaches to ATM performance modelling able to **reconcile model transparency, computational tractability and ease of use with the necessary sophistication required for a realistic representation of the ATM system.**

### Technical Context

Innovative developments in the fields of **artificial intelligence and data science** have opened new opportunities to overcome the traditional limitations of large-scale, bottom-up or microscopic simulation models, such as the limited number of scenarios assessed or higher model complexity and computational costs.

NOSTROMO addressed the construction of metamodels as a means to minimise the need for simulation runs and allow a more efficient exploration of the simulation input-output space, based on a machine learning paradigm called **Active Learning**. Active Learning is a special case of supervised machine learning consisting of an oracle (i.e., an instance label provider, which in our case will be a simulator) and an iterative sampling scheme that allows the algorithm to choose the data points from which it learns. The general idea is to actively select the most informative data points, as few as possible, in order to simultaneously boost the model training efficiency and its prediction performance.

NOSTROMO also covered the use of **interactive visualisation and visual analytics** as a means to facilitate the analysis, interpretation and communication of the results of the metamodels and ultimately support evidence-based decision making.

## 2.2 Project Scope and Objectives

### 2.2.1 Project Objectives

In the context previously described in §2.1, the goal of NOSTROMO was to develop, demonstrate and evaluate an innovative modelling approach for the rigorous and comprehensive assessment of the performance impact of future ATM concepts and solutions at ECAC network level. This approach brings together the ability of bottom-up microscopic models to capture emergent behaviour and interdependencies between different solutions with the level of tractability and interpretability required to effectively support decision-making.

To achieve this, the project pursued the following objectives.

#### Objective#1: Metamodelling Methodology

Develop a **methodology for the construction of ATM performance metamodels** that approximate the behaviour of computationally expensive simulation models so as to allow a systematic and efficient exploration of the model input-output space and a robust handling of the uncertainty associated with the model predictions, by exploiting recent advances in the field of active learning.

#### Objective#2: Metamodels Implementation

**Implement and validate the proposed metamodelling methodology** by developing metamodels of different state-of-the-art microsimulation tools able to reproduce ATM performance at ECAC level.

#### Objective#3: Visualisation

Develop a set of **visualisation and visual analytics tools that facilitate the analysis, interpretation and communication of the results** of the new performance metamodels.

#### Objective#4: Evaluation and Cost Benefit Analysis[1]

**Demonstrate and evaluate the maturity of the NOSTROMO approach and the capabilities of the newly developed toolset** through a set of case studies addressing the performance assessment of SESAR Solutions at ECAC level. The case studies shall cover a variety of ATM phases, solutions and KPAs/KPIs sufficiently heterogeneous to allow a comprehensive **cost benefits analysis**, with the aim to analyse the added value and the limitations of the NOSTROMO approach and evaluate the appropriateness of its transition to SESAR IR.

---

[1] Please note that this objective was slightly updated from the Grant Agreement. A benchmarking against the performance modelling methodologies currently in use was initially planned, but, during the 1st NOSTROMO workshop, the replacement by a cost benefit analysis of the NOSTROMO methodology application was agreed. But it should be noted that these two activities have a common objective: to demonstrate that this methodology is more effective and economically efficient.

### 2.2.2 NOSTROMO Approach

The NOSTROMO project followed an iterative and incremental approach, by evaluating and refining the metamodelling methodology in an iterative manner in the light of the results obtained during the development and execution of the different case studies. The project was performed in a three cycle-approach:

- A **first preliminary iteration** was just a single **test case**, which was executed using a fast prototype that was lean but functional, to verify the setup from beginning to end and to evaluate the proposed architecture.

  This first iteration, conceived as a "case zero", took advantage of the micromodels already developed in previous ER projects, with the aim of providing initial evidences on the technical feasibility of the methodology defined by the project rather than on the operational aspects that were addressed in the following iterations.

- The **second iteration** was the first attempt to evaluate the case studies defined in the project to demonstrate and assess the applicability of the proposed methodology to support performance assessment at ECAC level. These case studies covered isolated operational concepts developed in different selected SESAR Solutions modelled by two simulators, Mercury and FLITAN.

- The **third and final iteration** was carried out to refine the metamodeling methodology and obtain conclusions on the expected benefits and potential drawback of the new modelling approach for the ATM performance assessment. The use cases evaluated in this iteration covered the integration of different SESAR Solutions modelled by the same simulator, Mercury or FLITAN, producing multiple outputs from multiple inputs.

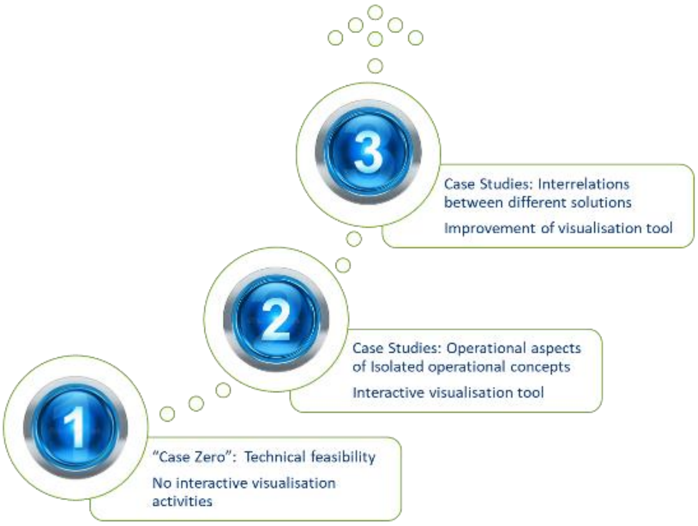


**Figure 1: NOSTROMO Incremental Approach**

### 2.2.3 Work Breakdown Structure

The NOSTROMO project was structured into eight Work Packages (WPs), as presented in Figure 2. The technical WPs of the project are described below, including their link to the project objectives:

- The metamodeling methodology and the enabling metamodeling toolset were developed in WP3 (OBJ#1: Metamodelling Methodology);
- The architecture required to apply the metamodeling methodology was defined in WP3 (OBJ#2: Metamodels Implementation);
- The data repository gathering all the datasets required by the project was deployed in WP2 (OBJ#2: Metamodels Implementation)
- The case studies were selected and defined in WP4 (OBJ#4: Evaluation and Cost Benefit Analysis);
- The ATM performance metamodels were constructed in WP5, including the adaptation and calibration of the simulators (OBJ#4: Evaluation and Cost Benefit Analysis);
- The interactive dashboard was developed in WP6 (OBJ#3: Visualisation);
- The ATM performance metamodels were executed and assess in WP7 (OBJ#4: Evaluation and Cost Benefit Analysis);
- The cost benefit analysis was performed in WP7 (OBJ#4: Evaluation and Cost Benefit Analysis).

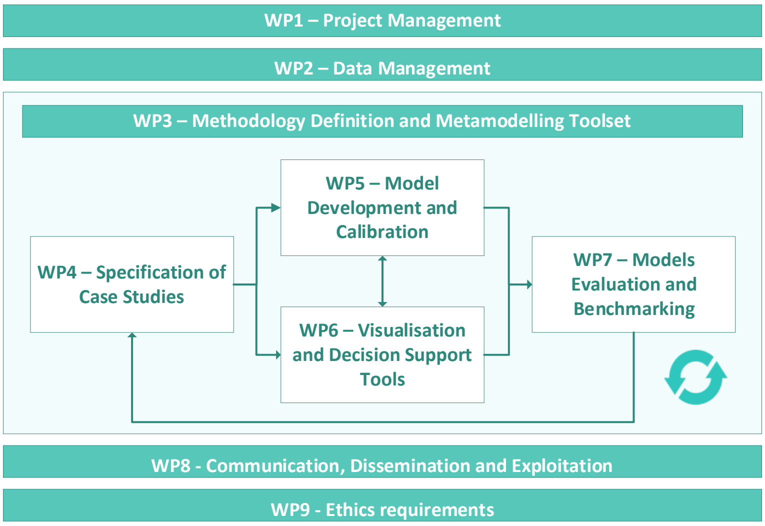


Figure 2: Project Work Breakdown

## 2.3 Work Performed

This section details the main work performed during NOSTROMO lifecycle related to each of the project objectives described in §2.2.1.

### 2.3.1 Metamodelling Methodology

The overall objective of WP3 was to develop a methodology capable of enhancing existing fast-time simulation-based ATM performance assessment studies. Modelling techniques such as fast-time simulation are a common de facto approach to evaluate the performance of current ATM systems or planned ones. These techniques allow researchers to investigate, test and propose a wide range of designs and alternative solutions within a virtual environment (i.e., a computer model), which would be otherwise practically infeasible to conduct in the real-world system. Despite their clear advantages, simulation models are, more often than not, likely to become computationally expensive to run, thereby curbing the potential of their exploratory nature and the reach of their insights. With this concrete problem in mind, WP3 proposed a methodology integrating two well-established techniques: active learning and simulation metamodeling. Whereas the latter aims at approximating the underlying simulation models through fast, explicit and transparent models, the former tries to ensure that such approximation is conducted most efficiently, i.e., requiring as few simulation results (or runs) as possible.

Besides the development and proposal of the methodology itself, WP3 also explored its feasibility in the context of simulation reproducing ATM performance at the ECAC level by investigating the impact of new SESAR Solutions sustained on a proof-of-concept framework. The exploratory studies conducted with Mercury and FLITAN simulators, albeit not in any way definitive, showed promising results for the field, representing an important contribution to the continuous and ongoing development of the SESAR Performance Framework. The joint use of active learning and simulation metamodels effectively constitutes a promising auxiliary tool able to complement and improve upon the current state of practice of simulation-based ATM performance assessment studies and related. It is worthwhile mentioning that it was never WP3's intent, nor would it be possible, to entirely dismiss the simulation models after the corresponding metamodels are obtained. On the contrary, the ultimate goal of the proposed methodology is to deploy the active learning metamodels alongside the simulators themselves within a bundled modelling solution operated and supervised by domain knowledge experts and practitioners. We strongly believe that the adoption of NOSTROMO's methodology, can, in all likelihood, pave the way for a novel and innovative ATM performance assessment framework, greatly enhancing current fast-time simulation practices and associated studies.

### 2.3.2 Metamodels Implementation

The work of WP2 focused on data management activities, as detailed in the Data Management Plan (DMP) [2], and on the creation of a data repository where all the information necessary for the development of the project and produced by the project was stored.

The developed DMP established the mechanisms for sharing, verification, curation, preservation, reuse and further exploitation of the data used by the NOSTROMO project. In particular, the NOSTROMO DMP included information on:

- the data collected, processed and generated,
- the handling of research data during and after the end of the project,
- the methodologies and standards applied within the project,
- whether data will be shared/made open access, and

- how data will be curated and preserved (including after the end of the project).

WP2 also set up the NOSTROMO data repository. This repository was aimed at providing the consortium members with secure, efficient and reliable access to the project datasets. To choose the appropriate software technology, a series of requirements were developed: web graphic user interface (GUI), metadata support, API for system integration, apps for functionality extension, user management and database and file backup. According to these needs, the chosen storage software technology for the repository was Nextcloud.

In addition, WP3 developed an API enabling the user to interact with the earlier-mentioned simulators from an active learning metamodeling perspective. In the context of NOSTROMO, this interface was mostly tailored to be exploited for visualisation/decision support (WP6) and evaluation/benchmarking (WP7) purposes. Both the API and the underlying modelling scripts were totally implemented using free and open-source technologies. Despite being accessible worldwide, the tool is currently limited to consortium members and interested partners. NOSTROMO’s vision for its current architecture, which comprises the proposed methodology delivered within a fully operational API, is that it is capable of integrating a broad spectrum of ATM simulation models developed within the scope of SESAR, not only facilitating their access and usage but also their individual exploration from an active learning metamodeling lens. A more mature version of such architecture should support a future common ATM performance assessment platform with which several simulators can be studied through a single decision support tool. Furthermore, from the gained experience and lessons learnt, WP3 also drew some guidelines for ensuring that future simulation models are more easily integrated with NOSTROMO’s proposed architecture.

### 2.3.3 Visualisation

Metamodels aim to remove computational barriers to perform a complete and efficient exploration of the input-output space defined by complex simulation models. This exploration, however, is ultimately linked to a decision-making process where computational tractability is a necessary but not sufficient condition. The way in which the results of the model are presented is crucial, so that they can be clearly analysed in order to make correct decisions.

The work in WP6 was focused on the development of a dashboard equipped with a set of interactive visualisation tools that allow the user to analyse the outputs of the metamodels and explore trade-offs between KPIs with the ultimate purpose of supporting different types of decision-making process related to performance management.

WP6 had a set of defined goals that have been achieved during the development of the tool. These objectives are the following:

- Collect requirements from different stakeholders regarding the visualisation and analysis of the results of the new ATM performance metamodels.
- Define an approach that allows complete extensibility and personalisation of the tool.
- Integrate the API developed on the WP3 with the visualisation tool.
- Design a set of meaningful and compelling visualisations that facilitate the understanding, analysis, and communication of the simulator results.

### 2.3.4 Evaluation and Cost Benefit Analysis

The goal of WP4 was to select and specify the case studies that were later used to evaluate the feasibility and applicability of the metamodeling methodology developed in NOSTROMO. The preliminary set of Case Studies for the second iteration of the project (D4.1 [8]) included a variety of solutions, ATM phases and KPAs/KPIs sufficiently heterogeneous to capture the specific added value of the proposed metamodeling methodology. WP4 also defined the methodological framework to specify the case studies that were selected as well as the research questions that were assessed in them.

The final specification of Case Studies for the third iteration of the project (D4.2 [9]) included the refinement of the solutions considering the results from the previous iteration. Whereas the preliminary specification was more focused on the understanding and development of the single concepts, the final specification of case studies exploited additional mechanisms that could be incorporated in a particular concept as well as their potential combinations with other solutions.

WP5 focused on the implementation of solutions within the two simulators Mercury and FLITAN, their testing and calibration, and the construction of the corresponding metamodels. After the decision (from WP4) on the solutions that would be implemented in each simulator, WP5 lead the implementation for their respective simulators. The first iteration of the implementation was meant as a testing run, in order to have simple but reliable results, while the second and third iterations focused on solutions selected for their suitability to the micromodels and their future potential for SESAR. The micromodels were calibrated based on historical data prepared specifically for the project.

On top of the solution implementation, WP5 developed additional interface layers that allowed the metamodels to encode the micromodels input, gather the micromodels output, and query the models during the learning process. The metamodels were then developed, based on Gaussian processes, testing different learning flavours to explore the learning capabilities of the metamodels on the micromodels, in particular in terms of computational time and accuracy.

The metamodels, once trained, were used to produce results that were analysed in WP7. The validation process in WP7 was made using datasets completely different than the used for training these metamodels. The process followed to measure the performance and get the different metrics was performed several times, obtaining an average value of the metamodel outputs to be compared with the simulator's ones.

Finally, WP7 carried out a cost benefit analysis of the metamodeling methodology proposed by NOSTROMO that included all the costs identified within its implementation and development as well as the expected benefits and disadvantages.

## 2.4 Key Project Results

This section details the main project results traced against project objectives described in §2.2.1.

### 2.4.1 Metamodelling Methodology

#### Simulators

ATM is a complex system where improvements such as SESAR Solutions are constantly proposed to enhance efficiency, capacity, resilience, and sustainability, inter alia. Extensive studies are needed to assess the feasibility of the concepts and their potential benefits. Simulators are frequently used to perform these assessments, often across different levels of Solution maturity.

In brief, a simulator is a software program designed to reproduce behaviour likely to occur in the existing (ATM) system. The design of simulators requires efficient and effective computational models for data representation, analysis and visualisation. Various simulator types are available to analysts based on the required level of the investigation and the maturity of the proposed concept:

1. Fast-time simulators (FTS),
2. Human-in-the-loop simulators (HITL),
3. Real-time simulators (RTS).

## Metamodels

By definition, **a metamodel is a model of a model**. Although the term itself is relatively vague, having different meanings and interpretations across the fields where it is used (see [26][27][28], for other SESAR-related metamodels), we solely focus on simulation metamodels [29][30][31], that is to say, models specially designed to reproduce the behaviour of simulation models (e.g. simulators). If a simulation model corresponds to an abstraction of a particular real-world system or phenomena, a metamodel can be regarded as an abstraction of the simulation model itself, as depicted in Figure 1. We may use the terms 'simulation metamodel' and **'metamodel'** interchangeably; also, we refer to the process of designing and building it as **'metamodelling'**.

Formally, a simulation metamodel is any type of model that can be used to deduce the unknown input-output mapping inherently defined by the simulation model, essentially serving as a **surrogate or proxy** with respect to the associated simulator. Although simulation models are simplified representations of the real-world system, they can still be, and often are, complex and detailed enough to yield significant inconveniences in practice. The most common shortcoming is their tendency to exhibit expensive simulation runs. Furthermore, the size and range of the input variable space can make it difficult to efficiently study and explore the behaviour of computer simulations as a whole, even with current computing technologies.

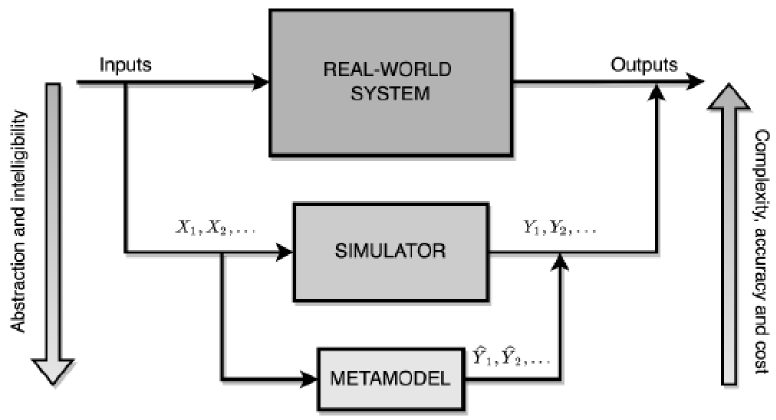


**Figure 3: Relationship between real-world system under study, simulator and metamodel**

Simulation metamodels can then be employed to minimise the computational drawbacks posed by exhausting and time-consuming simulation runs by jointly exploiting their approximate nature, functional simplicity, and fast computing. Being approximations of the underlying simulation functions, the metamodels' design and general performance can achieve balanced trade-offs between computational speed and controlled accuracy loss, depending on their ultimate objectives. Another feature of metamodels is that their respective **functional structures are generally known** and analytically defined, as opposed to those of most simulators. It is worthwhile noting that, although the average arbitrary simulator is oftentimes comprised of a plethora of internal analytic expressions and logical relationships, it can be externally treated as a single 'black-box' function with no clear mathematical formula. Nevertheless, an 'emergent behaviour', resulting from its inner interactions and dynamics that evolve over time, can be directly observed. Metamodels aim at mimicking precisely this output behaviour, as a function of the simulation inputs.

Illustrated in Figure 4 is one of the simplest metamodelling scenarios consisting of a simulator with two input variables and one output, along with a simple linear regression model in the role of the metamodel. Here, the metamodeling assumption is that the unknown function f represented by the simulator, and consequently its single output, can be reasonably well approximated by a linear combination of its two simulation inputs plus a normally distributed noise term. Naturally, the three parameters of this linear function have to be estimated using some data generated by the simulator itself. This process is typically termed the **'training'** of the model within a machine learning context. In our particular case, this is the process through which the metamodel learns to fit itself to the observed simulation data, ultimately aiming at approximating the simulator's output behaviour.

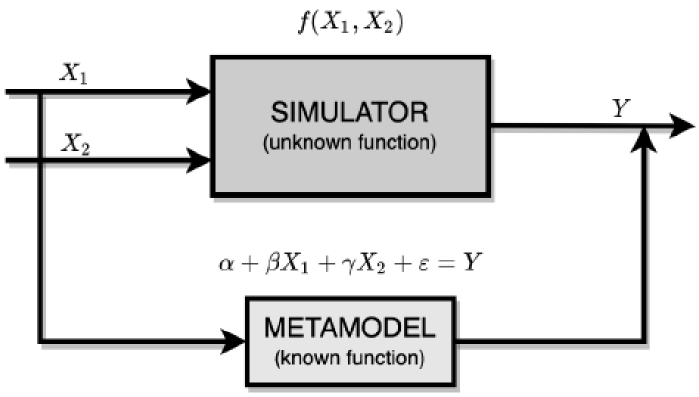


**Figure 4: A simple linear regression model acting as a simulation metamodel**

Despite requiring an initial and unavoidable computational effort, both for sampling the data from the simulator and then for training, the metamodelling approach relies on the fact that most metamodels are (and should be by default) computationally fast, provided that their parameters are already estimated. At this point, if the metamodel represents a fairly good approximation of the simulator, it can thus be employed as a proxy replacement to attain a more efficient exploration of the latter's behaviour. This exploration is conducted by means of predicting the output values for a set (typically a rather large one) of combinations of input values that have not been simulated. Hence, through a surrogate metamodel, exploration by proxy can effectively **bypass the need for new simulation runs** with a minor and controlled accuracy loss and instead generate predictions for unobserved input combinations. Figure 5 summarises two important types of data sets used in metamodeling, namely, the training set to which the metamodel is fitted and the prediction set used to explore the simulation input by proxy and the corresponding output behaviour.

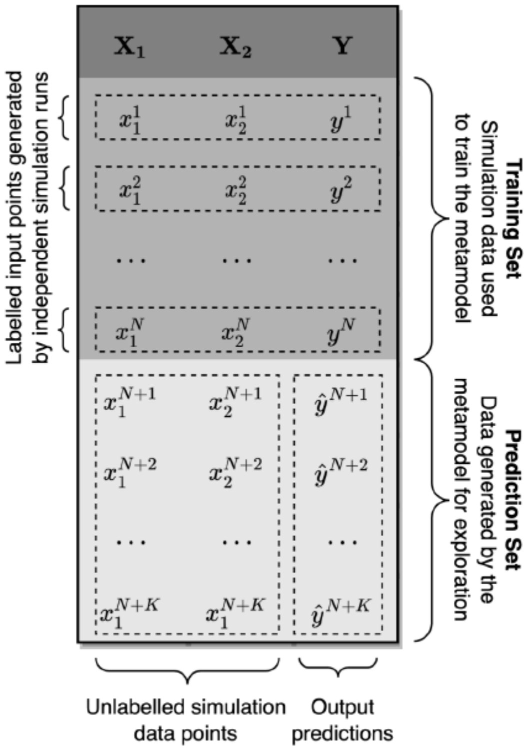


**Figure 5: Basic ingredient of training and prediction sets within the same linear metamodel example**

Typically, the training set is called a labelled set, whereas the prediction set in the absence of the predicted values is an unlabelled set. The term 'label', which is commonly associated with classification problems, is here adopted to refer to any value lying in the range of the simulation output space. The exploration process encompasses the prediction of labels which otherwise would have to be generated through simulation, thereby consuming more computational resources and time.

In the context of the NOSTROMO project, a more complex and **powerful family of metamodels is being employed, namely, the Gaussian process (GP)** modelling framework [32][33]. Indeed, GPs have been widely studied and used as simulation metamodels across many different fields, corresponding to the de facto default approach in most metamodelling settings. Besides their flexible non-parametric and highly non-linear characteristics, GPs provide a native Bayesian inference system that allows them to handle and quantify their own predictions' uncertainty and the variability naturally present in the data.

## Active Learning

Previously, it was mentioned that metamodels need to be trained with the simulation data so that they can serve as approximations for the simulator at hand. On the other hand, we recognise that simulation results might be computationally expensive and cumbersome to generate in a systematic matter, which ultimately constitutes one of the core issues that metamodelling aims to address.

In this context, i.e., in modelling settings where labelled data is difficult to obtain, active learning [34][35] emerges as a powerful learning paradigm that enhances metamodels and underlying algorithms to attain a high predictive performance using as little data as possible. This is achieved by an iterative scheme that, in its simplest form, sequentially selects the most informative input data points to be run through the simulator, eventually adding them to the current training set for model refitting. Here, the informativeness of an arbitrary single unlabelled point is measured as a function of its potential relative contribution to improving the metamodel's performance. In other words, if a data point is associated with a strong information index, then it is more likely to pose a greater performance boost than otherwise. Several information criteria can be adopted, but their reference is out of the scope of this document.

In sum, active learning generally seeks to optimize and, essentially, accelerate the metamodel's learning curve by avoiding redundancy in the training set, simultaneously making training more efficient and saving significant computational resources, simulation run time, and workload in the process.

Metamodels and active learning are conceptually intertwined and somewhat complementary in practical terms since both generally aim at reducing computational costs. Whereas the metamodels' contribution to this goal lies in providing a parsimonious approximation serving as simulation replacement, **active learning focuses on delivering an efficient training process**. Overall, with the natural combination of the two approaches, more insights concerning the simulator's behaviour are obtained with fewer data, i.e., at a reduced computational cost to the maximum possible extent.

Further information about the NOSTROMO's Methodology can be found in D3.4 [7].

### 2.4.2 Metamodels Implementation

Within NOSTROMO, the best of both worlds are integrated into a **single auxiliary framework** with the objective of complementing and improving, through **active learning and metamodelling**, the current state of the art for assessing the simulation-supported design and performance impacts of SESAR Solutions on ATM systems. Figure 6 depicts an overview of this architecture's main elements along with its process flow.

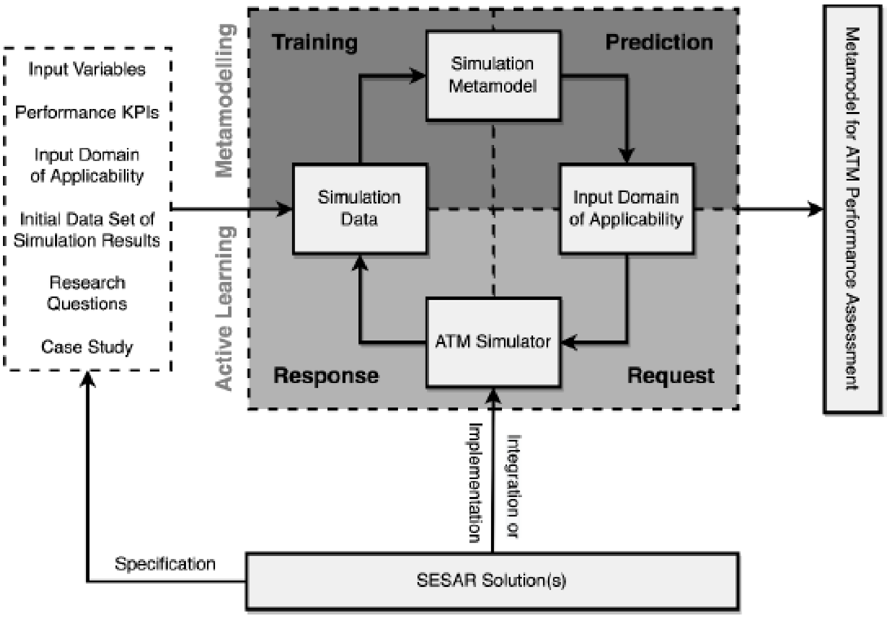


**Figure 6: Overview of the NOSTROMO's Active learning-based metamodelling architecture**

Overall, the ultimate goal of this approach is to assist ATM researchers, modellers and practitioners with an auxiliary tool to study the input-output behaviour of simulation models in a more insightful, systematic and computationally efficient fashion. The underlying metamodelling process includes the fulfilment of several prerequisites. First of all, and rather obviously, it cannot advance without a clear selection of the SESAR Solution(s) to be assessed. In addition, these **Solutions must be jointly integrated or implemented** *a priori* into the ATM simulator, upon which metamodelling is then performed.

The simulation model should be capable of encoding the SESAR Solutions into a specific set of input variables and KPIs (output variables) designed for performance assessments. These are the same variables that are eventually used by the metamodel to approximate the simulator's inherent function. Besides the simulation variables of interest, the metamodelling process itself also requires the definition of the simulation input region where it should be employed. Here, this region is deemed the input domain of applicability, essentially encompassing the value ranges within which metamodelling is conducted. Furthermore, due to the iterative nature of the approach, an Initial (training) set, comprised of previously generated simulation results, must be fed to the metamodel so that the process can be initiated. Finally, it is of utmost importance to keep in mind the research questions to be answered by the metamodel and the case study to be analysed in terms of performance impacts.

Finally, alternating between the metamodeling and the active learning phases, the integrated approach is composed of four elementary steps:

1. **Training**: the metamodel is fitted to the simulation data;

2. **Prediction:** the fitted metamodel is used to predict over the simulation input domain of applicability;

3. **Request:** based on some acquisition criteria, new input data points (unlabelled) are selected to be run by the simulator;

4. **Response:** the simulator provides new simulation output results corresponding to the points from step 3, which are then added to the current training set.

Steps 1-4 are repeated cyclically until a stopping criterion is satisfied. This criterion can be defined, for example, as a function of the metamodel's performance, such as accuracy and error-based metrics, or simply the number of iterations to be performed with respect to the available time, budget and resources. The active learning metamodelling process eventually provides a trained metamodel designed to help answer the posed research questions and assess the performance impact of the previously selected SESAR Solutions.

It is important to always bear in mind the approximative nature of the metamodel which calls for careful handling of the trade-off between speed, accuracy and computational budget. This balance should constantly be monitored and adjusted whenever required, to ensure the metamodeling's ultimate objectives are attained. This means that, if the finally obtained metamodel is fit for purpose, it can be reintroduced in the active learning metamodelling cycle to allow its parameters to be reestimated. Consequently, and on a similar note, it is equally crucial to recognize and identify the performance threshold from which the mere addition of new training points will not significantly improve the ability of the metamodel to approximate the simulation results. In those cases, and especially from a metamodelling perspective, requesting more simulation results might prove to be a waste of computational resources.

Further information about the implementation of NOSTROMO Methodology can be found in D3.4 [7].

### 2.4.3 Visualisation

The NOSTROMO dashboard was not intended to be a static platform that simply displays results, but rather an interactive platform where the user can personalise their experience. In this way, the development process prioritised factors such as ease of use, interactivity and extensibility. The visualisation platform was totally agnostic to the data and metamodels available in the NOSTROMO API. In this way, if the inputs or outputs of that metamodel changed, the platform continued to be functional. Similarly, the dashboard is also fully extendable to explore the metamodel of a different simulator. The tool also offers a report generation capability allowing the user to download the visualisations explored.

The NOSTROMO dashboard is a web-based platform where the user can visualise the impact on performance of alternative decisions in a simple but rigorous way, allowing a comparative assessment. It communicates dynamically with the metamodel to be investigated through the NOSTROMO API to allow a simple and fast exploration of the simulator's input-output space. The methodology chosen for the representation of results is the following: (i) the user can define combinations of inputs of interest through a series of tool selectors; (ii) this information is communicated to the metamodel so that it makes the corresponding prediction; (iii) the metamodel outputs are sent back to the dashboard to be visualised in the chosen plots. Given the computational speed of metamodels, this combined approach enables visual exploration in real time.

Further information about the NOSTROMO dashboard can be found in D6.2 [13].


**Figure 7: NOSTROMO Interactive Dashboard**

### 2.4.4 Evaluation and Cost Benefit Analysis

The development and evaluation of the ATM Performance metamodels for the selected case studies showed that the metamodelling approach followed by NOSTROMO provide results very close to the simulator with much less computational time, allowing deeper assessment of a solution, amplifying the exploration of the simulation input and output behaviour space, and helping to identify patterns and trends.

It was demonstrated that the creation of a single metamodel that combines different Solutions is possible and provides strong results as long as both Solutions were previously modelled in the same simulator. Both FLITAN and Mercury metamodels tackled two different solutions during the final iteration without increasing the computational effort in a meaningful way.

Better results were obtained with a deeper active learning process, increasing the computational time of training the metamodel. However, as it is stated in the cost benefit analysis, the training cost can be expensive, but it has to be performed only once. With the metamodel trained, the predictions can be obtained as fast as thousands of results per second. It was proved that it costs a lot more to run Mercury or FLITAN X times than to create metamodels with these simulators and run their metamodel X times.

#### 2.4.4.1 Applicability of metamodeling approach

Following pros and cons related to the applicability of the metamodeling approach can be extracted from the validation of the ATM performance metamodels developed in the second and third iterations of the project.

##### Pros of using the metamodeling approach

As stated previously, a metamodel is a functional approximator for a simulator, mostly employed in the context of computationally expensive and/or complex computer experiments. Characteristics include **functional simplicity, computational speed, and general intelligibility**. Contrary to simulation models, metamodels are explicitly defined by known analytic mathematical formulas, which contributes to an enhanced understanding of the dynamics and associations between the simulation inputs and the outputs of interest. Furthermore, the exploration of the simulation input space, and corresponding output behaviour, is greatly improved. Whilst conducted by surrogate approximation, this exploration allows for fast and efficient identification of patterns and general trends, and it is even able to correct itself via active learning whenever the metamodel's performance starts to drop below unacceptable levels.

In essence, metamodels act as proxy replacements for simulators. In the context of the SESAR Performance Framework, they are specially conceived to focus on the input/output variables that specifically encode the Solutions under study. Due to their computational speed and their ability to predict 'in bulk', it makes it easy to run up to thousands of combinations of input values in a manner of seconds.

Another important feature of metamodels is their portability. In principle, a fully trained metamodel should be easily executable and relatively dependency-free across different machines and operating systems, especially when compared to the average ATM simulator. For similar reasons, and by adopting current cloud deployment technologies, it should be fairly straightforward to make metamodels worldwide available via Application Programming Interfaces (APIs).

##### Cons of using the metamodeling approach

It is always important to be mindful of the unescapable approximate nature of simulation metamodels, which naturally come at the cost of reduced accuracy and detail. Consequently, metamodels should be regarded as auxiliary tools that complement the conventional simulation-based analyses rather than completely substitute them. The NOSTROMO's approach embraces this shortcoming of simulation metamodelling while simultaneously building upon it to exploit its advantages and, in a sense, aiming at deploying metamodels side-by-side with their corresponding simulation models, never actually discarding the latter, in a bundled modelling package. While the metamodel aims at reducing the exploratory redundancy by seeking the most informative and distinct input data points, the simulation model ensures, by providing labelled data whenever necessary, that this exploration process is maintained close enough to the simulation data distribution.

In practice, another important aspect of metamodelling, especially when coupled with active learning strategies, is that it does **not represent a universal and plug-and-play approach**. Depending on the characteristics and design details of the simulator in question, the construction of corresponding metamodels might require more or less implementation effort. This is particularly true for those cases when the input/output data require some sort of transformation or encoding, for example, from categorical to numerical values or when the simulator runs over multiple data log files. Eventually, **each metamodel is highly tailored for the specific ATM simulator**, SESAR Solutions and case studies under study.

Furthermore, metamodels, like most data-driven models, do also suffer from the so-called 'curse of dimensionality' phenomenon [36][37][38], which refers to the problem of exponential data sparsity in high-dimensional spaces. Whilst metamodeling is a useful approach, it is unwise to consider that a metamodel can entirely approximate the simulation model as a whole. This would be rather impractical, if not virtually impossible, for most simulation approaches with numerous input variables, as the number of required simulation runs would be too high, inevitably rendering the metamodeling itself unattractive and computationally unable to meet its goals. As dimensionality increases, the number of required data points to attain reasonably good model performance increases exponentially. In the context of metamodelling, a data point corresponds to an input-output tuple generated from a single simulation run. Consequently, the need for more data points yields higher computational demand, which is what metamodelling aims at reducing ultimately. Instead, the domain of applicability, or experimental region, in which the metamodel should be a valid approximation, should be established first [30]. In essence, this simplification can be regarded as restricting the metamodel training to a bounded subset contained within the sparse simulation input-output space, unlikely considering all the input variables at once, thereby reducing its dimensionality to more manageable and intelligible sizes. From this, another important question emerges: even if the metamodel is able to encompass, from a modelling perspective, a great number of input variables, should it? We believe that a metamodel should be maintained as simple as possible with enough input variables so that we can have the best of the two worlds: simplicity, intelligibility and transparency on the one hand and reasonably good description and approximation of the underlying problem under study.

#### 2.4.4.2 Cost Benefit Analysis

As mentioned in §2.3.4, a cost benefit analysis of the application of NOSTROMO approach was performed. The main expected benefits and disadvantages obtained in this analysis are collected bellow (further information can be found in D7.2 [15]).

**Table 1: Benefits of NOSTROMO approach**

| Benefits | | Comments |
|---|---|---|
| Benefits of NOSTROMO versus the present | Execution of more solid PARs (not based on estimations and extrapolations because of a reduced amount of outputs/lack of resources for all a big number of simulations) and of higher quality. | A quantification of feasibility and computational cost is needed. |
| | NOSTROMO can offer approximations / simulations for the entire ECAC area (very high cost). | |
| Benefits of active learning | Active learning generally seeks to optimize and, essentially, accelerate the metamodel's learning curve by avoiding redundancy in the training set, simultaneously making training more efficient and saving significant computational resources, simulation run time, and workload in the process. | |
| Benefits of the dashboard | The dashboard allows metamodels output analysis and the relationships exploration between the KPIs to facilitate the decision-making process. | |
| Benefits over a simulation model | Optimization of computational time/speed: it has been quantified that there are computational savings with respect to the simulator once the metamodel has been created. | It should be studied/ quantified, e.g., for a specific simulator and solution, how many simulation model runs are equated with how many metamodel runs. |
| | Generic savings: ex: it costs more to run Mercury X times than to create metamodel with Mercury and run metamodel X times. | |
| | Functionally structures are generally known. Contributes to an enhanced understanding of the dynamics and associations between the simulation inputs and the outputs. | |
| | Portability. In principle, a fully-trained metamodel should be easily executable and relatively dependency-free across different machines and operating systems, especially when compared to the average ATM simulator. | |
| | For similar reasons, and by adopting current cloud deployment technologies, it should be fairly straightforward to make metamodels worldwide available via Application Programming Interfaces (APIs). | |
| | One can use the metamodels in two different ways: as auxiliary tools to the conventional simulation-based analysis for exploring the input data and finding the most useful cases to be simulated by the simulation model or, as the main tool for simulation taking into account its benefits versus the simulation model. | |

**Table 2: Disadvantages of NOSTROMO approach**

| Disadvantages |
|---|
| Possible reduced detail and accuracy compared to the full simulator. |
| It is unwise to consider that a metamodel can entirely approximate the simulation model as a whole, this would be impractical, if not, virtually impossible, as the number of simulation runs would be too high. The domain of applicability should be established first restricting the metamodel training to a bounded subset contained within the input-output space unlikely considering all the input variables at once. |
| A metamodel cannot be created for multiple simulators. Nevertheless, if it is evaluated that two simulators are logically and technically integrable, the metamodel is able to approximate the inputs/outputs of the set. The functions and relationship between the two simulation models is a black box. |

## 2.5 Technical Deliverables

**Table 3: Project Deliverables**

| Reference | Title | Delivery Date[2] | Dissemination Level[3] |
|---|---|---|---|
| Description | | | |
| D1.1 | Project Management Plan | 19/10/2020 | Confidential |
| This document described the Project Management Plan (PMP) of NOSTROMO project according to the guidelines described in the Project Handbook of SESAR 2020 Exploratory Research Call. NOSTROMO PMP describes how the project management processes were executed during the project lifecycle. It also sets up the project overview and scope, as well as the project organisation and structure. | | | |
| D1.2 | Final Project Results Report | 04/11/2022 | Public |
| The present document provides a summary of NOSTROMO achievements and contributions to the SESSAR Programme. It gathers technical lessons learned and concludes proposing further steps for the future SESAR3 Programme. | | | |
| D2.1 | Data Management Plan | 01/12/2020 | Confidential |
| This document describes the data management life cycle for the data collected, processed and generated by the NOSTROMO project. Through this document, the project aimed to ensure that all the research data were findable, accessible, interoperable and reusable (FAIR) as well as that ethics and data security aspects were properly addressed. | | | |
| D2.2 | NOSTROMO Data Repository | 23/12/2020 | Confidential |
| This document described the specifications and the implementation of the NOSTROMO Data Repository. Relevant screenshots were included to illustrate the appearance and functioning of the data repository. | | | |
| D3.1 | Preliminary Metamodelling Methodology | 01/10/2020 | Public |

[2] Delivery data of latest edition

[3] Public or Confidential

<table>
<tr><th>Reference</th><th>Title</th><th>Delivery Date[2]</th><th>Dissemination Level[3]</th></tr>
<tr><th colspan="4">Description</th></tr>
<tr><td colspan="4">This document described a preliminary version of the metamodeling methodology employed within this project. It provided brief descriptions of two core concepts that compose the base structure of the proposed methodology, namely, active learning and simulation metamodeling itself.</td></tr>
<tr><td>D3.2</td><td>Requirements Specification</td><td>08/12/2020</td><td>Public</td></tr>
<tr><td colspan="4">This deliverable specified the simulation metamodeling framework's main technical and modeling requirements. Several considerations were drawn concerning the active learning strategy used in the methodology, which eventually required a constant link between the metamodel, the simulation model, and the data repository.</td></tr>
<tr><td>D3.3</td><td>NOSTROMO Framework API + Associated Documentation</td><td>28/05/2021</td><td>Public</td></tr>
<tr><td colspan="4">This deliverable defined the general guidelines and requirements for the development and implementation of the metamodels' Application Programming Interface (API). For each case study, this API should enable the modeller to select the simulation model to be used, specify the relevant input variables and corresponding ranges, and choose a set of relevant KPIs (simulation outputs) to be analysed, as well as the initial simulation data. Eventually, the API should return the resulting metamodelling results to be exploited for visualisation purposes.</td></tr>
<tr><td>D3.4</td><td>Final Metamodelling Methodology</td><td>30/09/2022</td><td>Public</td></tr>
<tr><td colspan="4">This document was the last deliverable of NOSTROMO's WP3. It compiled the work conducted along the lifespan of the work package in the form of revisited versions of earlier deliverables, namely, D3.1 Preliminary Metamodeling Methodology, D3.2 Metamodels Requirements Specification, and D3.3 NOSTROMO Framework API + Associated Documentation.<br>The final proposed methodology did not diverge significantly from that introduced in D3.1 nor from the refinements carried out during the 2nd iteration of the project.</td></tr>
<tr><td>D4.1</td><td>Preliminary Specification of Case Studies</td><td>06/01/2021</td><td>Public</td></tr>
<tr><td colspan="4">This deliverable provided the methodological framework which enabled to specify the case studies that were used to demonstrate and evaluate the maturity of the NOSTROMO approach as well as the capabilities of the methodology and the tools developed in the project. The preliminary set of case studies included a variety of solutions, ATM phases and KPAs/KPIs sufficiently heterogeneous to allow a comprehensive analysis of the added value and the limitations of the NOSTROMO approach.</td></tr>
<tr><td>D4.2</td><td>Final Specification of Case Studies</td><td>30/06/2022</td><td>Public</td></tr>
<tr><td colspan="4">This deliverable provided the final specification of the case studies that were used in the third (last) iteration to demonstrate and evaluate the maturity of the NOSTROMO approach as well as the capabilities of the methodology defined and the micro-simulators and tools developed in the project.</td></tr>
<tr><td>D5.1</td><td>ATM Performance Metamodels – Preliminary Release</td><td>12/04/2022</td><td>Public</td></tr>
<tr><td colspan="4">This deliverable presented the results obtained with the meta-modelling process presented in D3.1 and D3.2 applied to the two micromodels (or simulators), Mercury and FLITAN, themselves implementing concepts from four SESAR solutions, PJ01.01, PJ07.02, PJ08-01, and PJ02.08.</td></tr>
<tr><td>D5.2</td><td>ATM Performance Metamodels – Final Release</td><td>04/10/2022</td><td>Public</td></tr>
<tr><td colspan="4">This deliverable presented the third iteration of the development of the two micromodels Flitan and Mercury and the results obtained with them with the active learning process, as described in the deliverables D3.X.</td></tr>
<tr><td>D6.1</td><td>NOSTROMO Interactive Dashboard – Preliminary Release</td><td>22/12/2021</td><td>Public</td></tr>
<tr><td colspan="4">This deliverable was the first of a set of two phases deliverables and encompassed the methodology, architecture and component structure that was used to develop the tool foreseen in the WP6 of the NOSTROMO project.</td></tr>
<tr><td>D6.2</td><td>NOSTROMO Interactive Dashboard – Final Release</td><td>30/09/2022</td><td>Public</td></tr>
<tr><td colspan="4">This document was the second of a set of two-phase deliverables. It described the approach taken to develop the tool, which involved an integration between the generation of visualisations and the API developed and specified in the WP3. Therefore, this document compiled all the design documentation of the visualisation tool as well as the approach followed.</td></tr>
</table>

| Reference | Title | Delivery Date[2] | Dissemination Level[3] |
|---|---|---|---|
| **Description** | | | |
| D7.1 | Preliminary Assessment of the NOSTROMO Performance Evaluation toolset | 04/08/2022 | Public |
| This deliverable presented the validation of the metamodeling approach implementation during the second iteration of the project. During this iteration, the first real case studies were tested with the two simulators (Mercury and FLITAN) implementing concepts from SESAR Solutions.<br>In the previous (i.e. first) iteration of the project, the technical feasibility of the approach was tested, without using the case studies defined in D4.1. The results were presented also in the Annex of this document. | | | |
| D7.2 | Evaluation of the NOSTROMO Performance Evaluation Toolset and Implementation Guidelines | 04/11/2022 | Public |
| This document was the last technical deliverable of the project. It summarised the validation of the third and final iteration of the NOSTROMO methodology. In the document was explained the validation methodology follow to evaluate the metamodels and its performance in terms of computational time and accuracy.<br>In this document it was included also the cost benefit analysis for the implementation of NOSTROMO's methodology for any other simulations and Solutions. | | | |
| D8.1 | Dissemination, Exploitation and Communication Plan | 16/09/2020 | Public |
| This document described the different dissemination, exploitation and communication activities that were planned to be undertaken by NOSTROMO partners. These activities were identified to ensure the proper usability and exploitation of NOSTROMO results and achievements. | | | |
| D8.2 | Dissemination, Exploitation and Communication Report | 04/11/2022 | Public |
| This document covered the results from the dissemination, exploitation and communication actions performed during the NOSTROMO lifecycle. | | | |
| D9.1 | H – Requirement No. 1 | 16/09/2020 | Confidential |
| This document presented how the project intended to satisfy the ethics requirements related to the organisation of events for dissemination and communication purposes that included the participation of people external to the project. | | | |

# 3 Links to SESAR Programme

## 3.1 Contribution to the ATM Master Plan

An important element of SESAR Programme, as the technological pillar of the Single European Sky (SES) initiative, is to bring about improvements, as measured through specific KPIs, and as implemented by a series of so-called SESAR 'Solutions'. These 'Solutions' are new or improved operational procedures or technologies, designed to meet operational and performance improvements described in the European ATM Master Plan.

Central to performance assessment in SESAR is its Performance Framework, and this is supported, in part, by EATMA – the European Air Traffic Management Architecture. This is the common architecture framework for SESAR, its means of integrating operational and technical content developments. In these various SESAR contexts, the term 'meta model' is not used extensively, and typically describes, at a high level, logical entity relationships, e.g., for performance data and as an architecture mapping and database model. Whilst different definitions of metamodelling indeed prevail in different scientific contexts, usually referring to some form of abstractions of complexity, NOSTROMO has applied simulation metamodels, as defined in §2.4.1, in the performance assessment of SESAR, identifying their benefits and limitations.

Different SESAR Solutions variously deploy different simulations to demonstrate their expected performance contributions across the ICAO set of eleven KPAs, using a number of specific KPIs defined in the Performance Framework. Indeed, the corresponding projects are compelled to assess performance expectations as part of the SESAR programme. This brings challenges in terms of computational effort, simulation consistency, assessing KPI interdependencies and general integration.

NOSTROMO does not **set out to build or specify a single, integrated metamodel** for different SESAR Solutions or simulators. Nor does it aim to generalise all such simulations. Each simulation metamodel is a modelling proxy for, or simplified abstraction of, a specific (Solution, or combination of Solutions) simulation model. Whilst not replacing these simulations, simulation metamodelling is a **powerful complementary tool**, improving the state of the art for performance assessment, for example in terms of delivering computational efficiency and driving enhanced standardisation.

## 3.2 Maturity Assessment

The activities proposed by NOSTROMO ranged from basic research to applied research, encompassing from TRL1 to TRL2. The first stage of the project conducted an exploratory analysis of metamodeling techniques, such as active learning (TRL1). Subsequently, these techniques were applied to a number of case studies in order to analyse, evaluate and describe the characteristics of the newly developed technology (TRL2).



EUROPEAN PARTNERSHIP

Co-funded by
the European Union

**Table 4: ER Fund / AO Research Maturity Assessment**

| ID | Criteria | Satisfaction | Rationale - Link to deliverables - Comments |
|---|---|---|---|
| OPS.ER.1 | Has a potential new idea or concept been identified that employs a new scientific fact/principle? | Achieved | NOSTROMO Metamodelling Methodology described in D3.4 [7].<br><br>NOSTROMO Interactive Dashboard described in D6.2 [13]. |
| OPS.ER.2 | Have the basic scientific principles underpinning the idea/concept been identified? | Achieved | Basic scientific principles are described in D3.4 [7] and D6.2 [13]. |
| OPS.ER.3 | Does the analysis of the "state of the art" show that the new concept/Idea/technology fills a need? | Achieved | Problem addressed is described in §2.1 of the present document. |
| OPS.ER.4 | Has the new concept or technology been described with sufficient detail? Does it describe a potentially useful new capability for the ATM system? | Achieved | NOSTROMO Metamodelling Methodology described in D3.4 [7].<br><br>NOSTROMO Interactive Dashboard described in D6.2 [13].<br><br>NOSTROMO Framework API [6]. |
| OPS.ER.5 | Are the relevant stakeholders and their expectations identified? | Achieved | NOSTROMO Grant Agreement [41] |
| OPS.ER.6 | Are there potential (sub)operating environments identified where, if deployed, the concept would bring performance benefits? | N/A | N/A (Technological Solution) |
| SYS.ER.1 | Has the potential impact of the concept/idea on the target architecture been identified and described? | Achieved | NOSTROMO Metamodelling Methodology described in D3.4 [7].<br><br>NOSTROMO Interactive Dashboard described in D6.2 [13]. |



EUROPEAN PARTNERSHIP

Co-funded by the European Union

| ID | Criteria | Satisfaction | Rationale - Link to deliverables - Comments |
|---|---|---|---|
| | | | NOSTROMO Framework API [6]. |
| SYS.ER.2 | Have automation needs e.g. tools required to support the concept/idea been identified and described? | N/A | N/A (Technological Solution) |
| SYS.ER.3 | Have initial functional requirements been documented? | Achieved | Requirements Specification included in D3.2 [5]. |
| PER.ER1 | Has a feasibility study been performed to confirm the potential feasibility and usefulness of the new concept/idea/Technology being identified? | Achieved | Description of selected cases studies in D4.1 [8]and D4.2 [9].<br><br>Description of ATM Performance Metamodels in D5.1 [10] and D5.2 [11].<br><br>Evaluation of ATM Performance Metamodels in D7.1 [14] and D7.2 [15] |
| PER.ER.2 | Is there a document analysis and description of the benefit and costs mechanisms and associated Influence Factors? | N/A | N/A (Technological Solution) |
| PER.ER.3 | Has an initial cost/benefit assessment been produced? | Partially Achieved | Cost Benefit Analysis available in D7.2 [15]<br><br>It should be noted that CBA developed in WP7 is not fully in line with expectations with regards to its structure, objectives and content. In particular, a reference scenario should be added. |

| ID | Criteria | Satisfaction | Rationale - Link to deliverables - Comments |
|---|---|---|---|
| PER.ER.4 | Have the conceptual safety benefits and risks been identified? | N/A | N/A (Technological Solution) |
| PER.ER.5 | Have the conceptual security risks and benefits been identified? | N/A | N/A (Technological Solution) |
| PER.ER.6 | Have the conceptual environmental impacts been identified? | N/A | N/A (Technological Solution) |
| PER.ER.7 | Have the conceptual Human Performance aspects been identified? | N/A | N/A (Technological Solution) |
| VAL.ER.1 | Are the relevant R&D needs identified and documented?<br><br>Note: R&D needs state major questions and open issues to be addressed during the development, verification and validation of a SESAR Solution. They justify the need to continue research on a given SESAR Solution once Exploratory Research activities have been completed, and the definition of validation exercises and validation objectives in following maturity phases. | Achieved | NOSTROMO Grant Agreement [41] |
| TRA.ER.1 | Are the recommendations proposed for completing V1 (TRL-2)? | Achieved | Further enhancements are detailed in §4.3 of the present document. |

# 4 Conclusion and Lessons Learned

## 4.1 Conclusions

According to metamodelling approach defined by NOSTROMO, a metamodel is a fast and computationally efficient approximator of a simulator's behaviour. As such, it allows the surrogate exploration of the input-output simulation space with much less computationally hassle. Given a Solution, or multiple ones, already integrated and implemented within the simulation model in question, and a concrete case study, the metamodel is able to run multiple input combination values and predict their corresponding output values in a relatively short amount of time (especially when compared with the simulator's average runtimes), consequently bypassing exhausting and systematic simulations runs. Due to approximation, accuracy is sacrificed to the detriment of faster speeds and exploratory abilities. Besides model approximation and exploration (prediction) of the simulation input-output mapping, metamodels can also be used for optimization support, sensitivity analysis, and verification and validation [30][36][39][40].

The interactive dashboard built for NOSTROMO was developed as a web application, so that it is accessible from multiple devices by different users concurrently. The system architecture is distributed between the user interface itself (frontend) running on the user device and the components executing the system logic (backend) running on a server.

The final results of the project with real case studies and the most complex Solutions selected during the project showed that the metamodeling approach followed by NOSTROMO provides results very close to the simulator with much less computational time, as stated in D7.2 [15], allowing a deeper assessment of a solution, amplifying the exploration of the simulation input and output behaviour space, helping to identify patterns and trends.

## 4.2 Technical Lessons Learned

The main pros and cons related to the application of the NOSTROMO metamodelling approach have been discussed in §2.4.4.1. As summary, it is important to note that metamodels are limited by their intrinsic approximative nature and one-to-one relationship with respect to the simulator being approximated. NOSTROMO's metamodelling framework is Solution-oriented, i.e., an individual metamodel is produced per SESAR Solution and the corresponding simulator implementing it. A combination of Solutions can also be considered, provided they are jointly integrated into the same simulation model. Therefore, metamodels cannot combine Solutions for themselves, nor they can generalize across different simulation models. Instead, generalization is only conducted across the input space of the same simulator and for a given SESAR Solution. Similarly, metamodels cannot be used for extrapolation purposes for other case studies and sets of input/output variables that have not been considered during their design and training in the first place.

The metamodel treats the simulation model as a black box that transforms an input space of values into an output space of values, regardless of the numerical computations that run underneath it. Some simulation models are naturally more straightforward to metamodel than others, mostly depending on, but not limited to, the complexity of their input-output relationship, the number and type of variables, and the size or range of the variables' values spaces. Note, however, that whereas the approach is theoretically applicable to any simulator, the obtained metamodels are simulator-specific, thus not generalizable.

Other lessons learnt can be also derived in relation to the application of the NOSTROMO metamodelling approach to ATM performance assessment:

- There is no universal/plug-and-play or unique metamodeling solution. Each simulation model, case study, and their modelling objectives and research questions often require individual treatment and methodological tuning.
- The process of active learning metamodeling is relatively exploratory. Several parameters, such as the initial data set, stopping criteria, acquisition function and family of metamodels (Gaussian Processes, Neural Networks, etc.) have to be tested before setting them.
- Simulation models with non-numerical input and output variables/parameters require additional steps prior to the application of the methodology itself, mostly encompassing data conversion and encoding, as well as collection and merging/fusion. The latter is particularly relevant when the simulation data is scattered across multiple log files.
- The design and implementation diversity of the available simulation models in the context of SESAR may hinder their compatibility with the current NOSTROMO Architecture (methodology + API). While it is true that the developed API constitutes only a proof-of-concept seed of what can become a common SESAR metamodeling platform and novel paradigm for the field, current and future simulation models should be enhanced with their own individual APIs. This should greatly improve the adoption of and integration with the broader and future vision for the proposed architecture.

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

The results and conclusions of the application of metamodelling approach proposed by NOSTROMO and the evaluation activities were used to develop a consolidated set of guidelines and applicability methodology for the integration of the NOSTROMO models and tools into the SESAR3 IR programme.

To maximise its transferability to SESAR3 IR programme, NOSTROMO conducted a joint research effort with SESAR ER SIMBAD project. SIMBAD has also addressed the use of machine learning techniques to improve ATM simulation and performance assessment, working with EUROCONTROL’s R-NEST simulation tool, which will be one of the main simulation tools for performance assessment in SESAR3. In this joint research effort, NOSTROMO’s metamodelling methodology was used to develop a metamodel of R-NEST, demonstrating the potential benefits of this approach in SESAR3 performance assessment activities. At the moment of the development of this deliverable, the work performed in NOSTROMO and SIMBAD projects has been taken as candidate to be taken up in the SESAR3 activities related to the Master Plan and Performance Assessment, through two SESAR 3 proposals for the IR1 Call, AMPLE3 and PEARL respectively. In particular, the NOSTROMO approach has been proposed to be involved in the development of Optimised Deployment Scenarios in AMPLE3. Additionally, the performance dashboard to be developed in PEARL will build on the visualisation tools developed by NOSTROMO, among other relevant inputs.

In this context, the following enhancements of the NOSTROMO approach should be considered:

- Addition of more active learning strategies and metamodel forms, allowing the user to select which ones to use, therefore making the API more flexible and complete;
- Future ATM simulators to be integrated should develop their own APIs, this will facilitate the communication between NOSTROMO's API and the simulators to be studied;
- Integration of multiple simulators in a single metamodel should be addressed. It might be feasible to design a metamodel that integrates multiple simulators insofar that the simulators in question are technically and meaningfully integrable in the first place. However, that the integration capabilities lie mostly on the integrability of the underlying simulators and not on the metamodelling approach per se. The lack of standardisation and use of different technologies and programming platforms between simulators might represent a real hindrance in practice.

  Any metamodel is agnostic to the design and implementation details of the simulation model it aims to approximate. If the simulator in question already integrates the different Solutions by encoding them into specific input/output variables, then the metamodelling procedure follows naturally. Therefore, distinct Solutions can be integrated with the metamodeling approach as long as they are first integrated and implemented in the simulator of interest. The metamodel aims to mimic the latter and not the former. Metamodels are not designed to model Solutions per se. Instead, they model the simulation input-output relationships representing Solutions which are encoded via simulation variables and related KPIs previously embedded into the simulator for that effect.

  In the simplest case, a unique metamodel is required per simulator and Solution. However, suppose the Solutions are implemented into a single simulator. In that case, the metamodel should be able to handle them, as they technically correspond to the addition of new input-output variables to the metamodelling dimension space.

  In practice, it might be feasible to design a metamodel that integrates multiple simulators insofar that the simulators in question are technically and meaningfully integrable in the first place. Notice, however, that the integration capabilities lie mostly on the integrability of the underlying simulators and not on the metamodelling approach per se. In this case, the metamodel will regard the final integrated simulators simply as a novel simulator.

  Figure 8 and Figure 9 illustrate two possible simplified ways of integrating two simulators, A and B, with posterior metamodelling in mind, each one individually implementing its own homonymous Solution. The first situation (see Figure 6) comprises the case where the output of one simulator serves directly as the input to the other. Here, we observe that the metamodel approximates the black box whose simulation variables are the inputs of Simulator A and the outputs from Simulator B, being completely agnostic to this serialised integration.

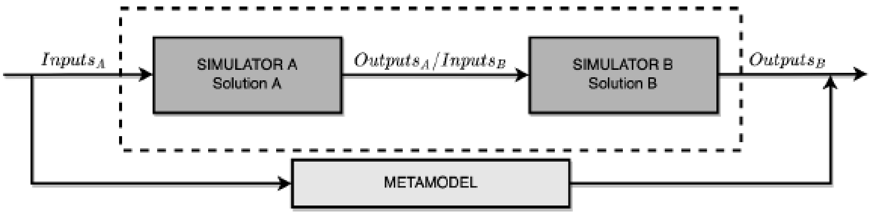


**Figure 8: Serial integration of two simulators implementing different Solutions and corresponding metamodel**

Alternatively, the two simulators can be integrated via parallelisation, as depicted in Figure 9. In this situation, both simulators can have their own input spaces and share a set of common variables. The integration itself is performed by somehow individually combining the outputs generated by each simulator in a meaningful and useful way. This integration is independent of and occurs *a priori* to the metamodel's building process. Once again, the metamodel only considers the newly formed simulator's input and output variables, ignoring its inner simulation subcomponents.

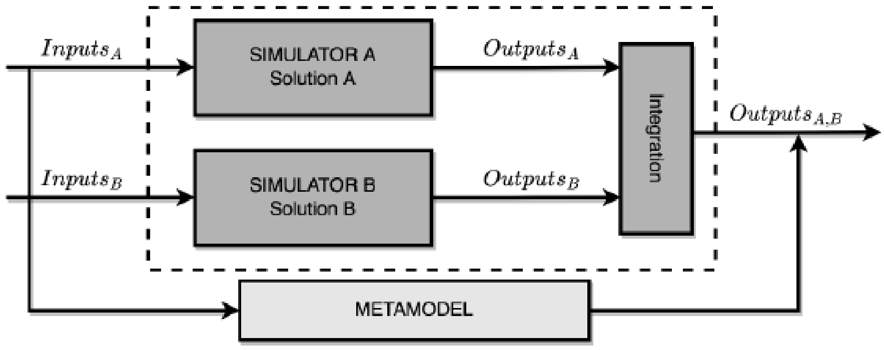


**Figure 9: Parallel integration of two simulators implementing different Solutions and corresponding metamodel**

For all intents and purposes, the simulator resulting from the integration of independent simulators is a new simulator (depicted with dashed lines in both figures) from a metamodelling perspective. The metamodel takes no part in this integration process which is only limited by the ability and utility of combining the simulators in question and, consequently, the associated Solutions. If an arbitrary set of simulators can be integrated in a reasonable and meaningful manner, then a metamodel can be used to approximate the resulting integrated simulator as it is regarded as a new simulation model. Therefore, building a metamodel that encompasses multiple simulators makes sense only if the simulators themselves can be integrated *a priori* and run as a whole. The combination of Solutions, and thus its corresponding metamodelling, is heavily dependent on the success of incorporating them into a single simulator, via integration, as seen before, or from scratch. The lack of standardisation and use of different technologies and programming platforms between simulators might represent a real hindrance in practice. Moreover, even if it is technically possible, one should always firstly investigate if the combination of certain Solutions does indeed make sense from theoretical, research and practical perspectives.

# 5 References

## 5.1 Project Deliverables

[1] NOSTROMO, D1.1 Project Management Plan, 00.01.00, 11 Feb 2021

[2] NOSTROMO, D2.1 Data Management Plan, 00.01.00, 18 Feb 2021

[3] NOSTROMO, D2.2 NOSTROMO Data Repository, 00.01.00, 18 Dec 2021

[4] NOSTROMO, D3.1 Preliminary Metamodelling Methodology, 00.01.00, 20 Nov 2020

[5] NOSTROMO, D3.2 Requirements Specification, 00.01.00, 18 Feb 2021

[6] NOSTROMO, D3.3 NOSTROMO Framework API – Associated documentation, 00.01.00, 19 Aug 2021

[7] NOSTROMO, D3.4 Final Metamodelling Methodology, 00.01.00, 30 Sep 2022

[8] NOSTROMO, D4.1 Preliminary Specification of Case Studies, 00.01.00, 18 Feb 2021

[9] NOSTROMO, D4.2 Final Specification of Case Studies, 00.01.00, 18 Aug 2022

[10] NOSTROMO, D5.1 ATM Performance Metamodels – Preliminary Release, 00.01.00, 28 Apr 2022

[11] NOSTROMO, D5.2 ATM Performance Metamodels – Final Release, 00.01.00, 4 Oct 2022

[12] NOSTROMO, D6.1 NOSTROMO Interactive Dashboard - Preliminary Release, 00.01.00, 11 Feb 2022

[13] NOSTROMO, D6.2 NOSTROMO Interactive Dashboard - Preliminary Release, 00.01.00, 30 Sep 2022

[14] NOSTROMO, D7.1 Preliminary Assessment of the NOSTROMO Performance Evaluation toolset, 00.01.00, 18 Aug 2022

[15] NOSTROMO, D7.2 Evaluation of the NOSTROMO Performance Evaluation Toolset and Implementation Guidelines, 00.01.00, 05 Oct 2022

[16] NOSTROMO, D8.1 Dissemination, Exploitation and Communication Plan, 00.01.00, 13 Jan 2021

[17] NOSTROMO. D8.2 Dissemination, Exploitation and Communication Plan, 00.01.00, 07 Oct 2022

[18] NOSTROMO, D9.1 H – Requirement No. 1, 00.01.00, 13 Jan 2021

## 5.2 Project Publications

[19]Christoffer Riis, Francisco Antunes, Gérald Gurtner, Francisco Cámara, Luis Delgado and Carlos Azevedo (2021) Active Learning Metamodels for ATM Simulation Modeling. SESAR Innovation Days 2021

[20]Christoffer Riis, Francisco Antunes, Gérald Gurtner, Francisco Pereira, Luis Delgado and Carlos Azevedo (2022) Active Learning for Air Traffic Management Simulation Metamodeling. Transportation Research Board Annual Meeting 2022

[21]Christoffer Riis, Francisco Antunes, Francisco Cámara (2022) Bayesian Active Learning with Fully Bayesian Gaussian Processes. Neural Information Processing Systems 2022

[22]Christoffer Riis, Francisco Antunes, Francisco Cámara (2023) Opening the Simulation Box: Towards Explainable Metamodels for Enhanced Air Traffic Management Performance Assessment. . Transportation Research Board Annual Meeting 2023

[23]Christoffer Riis, Francisco Antunes, Francisco Cámara, Tatjana Bolic, Gérald Gurtner (2022) Explainable Metamodels for ATM Performance Assessment. SESAR Innovation Days 2022

[24]Christoffer Riis, Francisco Antunes, Francisco Cámara (2022) Active Learning Metamodelling for R-NEST. SESAR Innovation Days 2022

[25]NOSTROMO Project Website https://nostromo-h2020.eu/

## 5.3 Other

[26] PJ19 CI, PJ190.5, May 2017, European ATM Architecture (EATMA) Guidance Material Version 9.0, Deliverable 5.1.

[27] [2] PACAS, December 2017, Modelling Language Metamodel, Deliverable D3.3.

[28] [3] PJ 19 CI, Content Integration, November 2019, SESAR2020 Performance Framework, Deliverable D4.7.

[29] [4] L. W. Friedman, The simulation metamodel. Springer Science & Business Media, 2012.

[30] [5] Jack PC Kleijnen and Robert G Sargent. A methodology for fitting and validating metamodels in simulation. European Journal of Operational Research, 120(1):14–29, 2000.

[31] [6] Gramacy, R. B. (2020). Surrogates: Gaussian process modeling, design, and optimization for the applied sciences. Chapman and Hall/CRC.

[32] [7] Williams, C. K., & Rasmussen, C. E. (2006). Gaussian Processes for Machine Learning. Cambridge, MA: MIT Press.

[33] [8] Bishop, C. M., & Nasrabadi, N. M. (2006). Pattern recognition and machine learning (Vol. 4, No. 4, p. 738). New York: Springer.

[34] [9] Burr Settles. Active learning literature survey. Computer Sciences Technical Report 1648, University of Wisconsin–Madison, 2009.

[35] [10] Xizhao Wang and Junhai Zhai. Learning with Uncertainty. CRC Press, 2016.

[36] [11] Wang, G. G., & Shan, S. (2007). Review of metamodeling techniques in support of engineering design optimization.

[37] [12] Köppen, M. (2000, September). The curse of dimensionality. In 5th online world conference on soft computing in industrial applications (WSC5) (Vol. 1, pp. 4-8).

[38] [13] Kuo, F. Y., & Sloan, I. H. (2005). Lifting the curse of dimensionality. Notices of the AMS, 52(11), 1320-1328.

[39] [14] Kleijnen, J. P. (2009). Kriging metamodeling in simulation: A review. European journal of operational research, 192(3), 707-716.

[40] [15] Kleijnen, J. P. (2017). Regression and Kriging metamodels with their experimental designs in simulation: a review. European Journal of Operational Research, 256(1), 1-16.

[41] NOSTROMO Grant Agreement 892517

# Appendix A

## A.1 Acronyms and Terminology

| Term | Definition |
|---|---|
| ATM | Air Traffic Management |
| CBA | Cost Benefit Analysis |
| DAC | Dynamic Airspace Configuration |
| ECAC | European Civil Aviation Conference |
| E-MAN | Extended Arrival Manager |
| GP | Gaussian Process |
| HITL | Human in the Loop |
| IR | Industrial Research |
| KPA | Key Performance Area |
| KPI | Key Performance Indicator |
| NOSTROMO | Next-Generation Open-Source Tools For Atm Performance Modelling And Optimisation |
| S3JU | SESAR3 Joint Undertaking (Agency of the European Commission) |
| SESAR | Single European Sky ATM Research Programme |
| SIMBAD | Combining Simulation Models and Big Data Analytics for ATM Performance Analysis |
| TRL | Technology Readiness Level |
| WP | Work Package |

**Table 5: Acronyms and technology**