Modeling Tail-Specific Performance Using Historical Flight Data and Machine-Learning Techniques

Master of Science Thesis

María del Pozo Domínguez

Modeling Tail-Specific Performance Using Historical Flight Data and Machine-Learning Techniques

Master of Science Thesis

by

María del Pozo Domínguez

in partial fulfillment of the requirements for the degree of

Master of Science
in Aerospace Engineering

at the Delft University of Technology,
to be defended on March 30, 2020.

MSc. Track:	Flight Performance and Propulsion	
Student number:	4748514	
Project duration:	May, 2019 – March, 2020	
Thesis supervisors:	Dr. J. López Leonés,	BR&T-E
	Ir. P.C. Roling,	TU Delft
Thesis committee:	Prof.dr.ir. L.L.M. Veldhuis,	TU Delft
	Dr. M.A. Mitici,	TU Delft

This thesis is confidential and cannot be made public until March 30, 2022

Preface

This project is *the last rung of the ladder*. It is the final requirement to obtain the degree of Master of Science in Aerospace Engineering. This project describes two different methodologies to generate tail-specific aircraft performance models based on historical flight data using machine-learning techniques. The goal is to develop performance models that accurately mimic the real behavior of the aircraft, in order to ensure more realistic results in pre-, in- and post-flight applications. This project was done in collaboration with Boeing Research & Technology - Europe, provider of the data required to carry out the research work.

This thesis is the culmination of years of persistence and dedication. It is the end of my time as a student of TU Delft and as an intern of Boeing. Both experiences have been demanding, but very rewarding and enriching. They have allowed me to grow personally and professionally, and they have helped me to know myself better and to meet great people. Even though some phases of my life end here, this also means the beginning of new adventures and challenges.

Many people have helped me to be where I am today. First, I would like to thank my daily supervisors: Paul Roling and Javier López Leonés. Without their time and effort this project would not have been possible. I highly appreciate their advice, trust and continuous support. I would like to thank my managers: Miguel Vilaplana and Enrique Casado, for their encouragement and recognition. I also want to express my gratitude to all my colleagues at Boeing, specially to Rubén, Manuel, Antonio and Mevlüt (a.k.a. *The Fuel Team*), Adrián and Alejandro. Thank you for daily accompanying me in this journey.

Lastly, deep gratitude also needs to be expressed to my family: mum, dad and Alicia, and to Antonio, thank you for always believe in me. I have had some ups and downs, and I could not have made it through without your love. You are the most special people in my life.

María del Pozo Domínguez
Delft, March 2020

Abstract

Aircraft performance has always been a focus of attention in aviation. The work of aircraft designers, certifying agencies, aircraft operators, and air traffic controllers relies on aircraft performance models. Current aircraft performance models are based on performance data of brand-new aircraft, independent of airline configuration and customizations. Nonetheless, over time aircraft suffer structure, engine and aerodynamic deterioration, as well as maintenance actions. These factors, which vary with tail number, make aircraft performance deviate from the theoretical and create the need for aircraft performance monitoring, and ultimately for aircraft performance tailoring.

This thesis proposes two novel approaches to develop up-to-date, tail-specific performance models. These approaches are based on the use of historical flight data, namely Quick Access Recorder (QAR) data, and machine-learning techniques. First, a methodology is designed to calibrate Base of Aircraft DAta (BADA), a widely consolidated physics-based performance model. As a result, more accurate performance models are generated, maintaining the same applicability over the entire flight envelope and during all phases of flight as BADA nominal models. The second approach is purely data-driven, and in contrast to the first approach, describes aircraft performance without modeling the underlying physics. The resulting models proved to be more accurate than BADA nominal models. Additionally, they provided insights on the parameters that have a significant impact on aircraft performance. This research project is the first to provide a comparison between physics-based and non-physics-based performance tailoring approaches. Despite the differences in accuracy achieved with both methods, each one has its own advantages and disadvantages. What approach to follow is determined first by the application, and secondly by the demanded level of accuracy.

Contents

List of Figures

List of Tables

1
Introduction

On December 1903, the Wright Brothers were the first to achieve sustained, controlled flight in a heavier-than-air aircraft. More than a century has passed since that attainment. However, the subject of flight performance arose only after the disciplines of aerodynamics, aircraft structure and aircraft propulsion were studied and consolidated [1]. Performance is understood as the ability of an aircraft to accomplish a specific maneuver or operation. Aircraft performance is a multidisciplinary subject, result of aerodynamic, gravitational and propulsive forces acting on the aircraft. Apart from the physical laws governing these forces, other factors such as aircraft weight, runway configuration and atmospheric conditions affect aircraft performance [2].

Aircraft performance has been a focus of attention in aviation for decades. The work of aircraft designers, certifying agencies, aircraft operators and air traffic controllers relies on aircraft performance in one way or another. Therefore, the need of a tool that mimics aircraft performance becomes apparent. Several models exist whose goal is to capture aircraft performance. The most widely accepted is Base of Aircraft DAta (BADA), developed and maintained by EUROCONTROL. BADA has been traditionally used for Air Traffic Management (ATM) applications, simulations, environmental impact studies or airspace redesign. BADA performance models are expressed in the form of polynomials, and each polynomial depends on a set of coefficients. As it will be explained in detail later in this report, BADA provides an average aircraft performance for each aircraft type [3], independent of airline configurations and customizations. Even though BADA performance models represent average aircraft performance with high fidelity, they fail to accurately model the performance of a specific tail number. Unfortunately, for some applications type is not specific enough to classify an aircraft, and an average aircraft performance is not sufficiently accurate.

Over time aircraft suffer structure, engine and aerodynamic deterioration, as well as maintenance actions. These factors make aircraft performance vary from the theoretical and create the need for aircraft performance monitoring, and ultimately for tailored performance modeling, since performance deviations differ from tail to tail depending on their age, user-configurable specifications or maintenance history. The higher computing capabilities, higher quality, data and more sophisticated analysis techniques that exist nowadays, encourage the development of tail-specific performance models.

The purpose of this thesis is to create tail-specific performance models based on historical Quick Access Recorder (QAR) data, using machine-learning techniques. The ultimate goal is to enrich pre-, in- and post-flight applications like flight planning, aircraft performance monitoring, fuel analytics, air traffic simulation, aircraft conflict detection, etc. This research goal is sought through two different approaches: by calibrating physical models like BADA and by developing purely data-driven, non-physical models.

For the first approach, a methodology has been developed to calibrate BADA performance models based on the use of historical flight data and machine-learning techniques. The methodology includes all the necessary steps to identify BADA coefficients from QAR data, grouped in: data ingestion, data preparation, tailoring process and model evaluation. The data ingestion includes the loading of QAR data and the generation of synthetic performance data. These are two important steps, since the the quality of the resulting calibrated models will rely on the coverage, precision and granularity of the reference performance data. After the required data are gathered, they must be preprocessed. QAR data commonly contain outliers and noise due to measurement errors and inaccuracies, for which the methodology includes cleaning and filtering techniques. In addition, data do not come in a format and shape that machine-learning algorithms can handle directly, so the required data preparation is also defined. The tailoring process consists of multiple regression models, through which the sets of BADA coefficients that best describe the aircraft represented in the QAR data are obtained. The tailoring stage focuses on adjusting BADA actions models, responsible for the computation of the forces acting on the aircraft, and hence responsible for its motion. Actions are classified in three main categories: gravitational, aerodynamical and propulsive. They include four forces acting on the aircraft: weight, lift, drag and thrust, together with the derivative of weight over time. In total, seven different models are considered and calibrated. Once the calibrated coefficients have been found, the model is validated to guarantee that it does not only provide accurate results, but also that it is valid over the entire flight envelope and during all phases of flight.

For the second approach, a methodology has been developed to create purely data-driven, non-physical models based on historical flight data and machine-learning techniques. The methodology considers all the necessary steps to generate tree-based regression ensembles from QAR data. The main steps to follow are: data ingestion, data preparation, feature selection, model selection, model training and model evaluation. The first step includes the loading of QAR data. After loaded, QAR data must be cleaned and filtered, in order to minimize the impact that outliers and noise might have on the developed performance models. The preparation phase also includes the appropriate split of the datasets, specific to the tackled problem and key to prevent overfitting. Given that not all the parameters available in the QAR data contribute to predict aircraft performance, the designed methodology includes a process of feature selection, to be launched before the construction of the decision trees. Once the optimum set of input variables is determined, the best algorithm and hyperparameters are selected following a random-search strategy. After datasets are loaded and prepared, irrelevant features are discarded, and the best model is selected, this is trained and evaluated with a different set of flights. This is done to assess the ability of the tailored performance models to robustly generalize lessons learned to unseen flight data.

This report is structured as followed. Chapter 2 discusses the current state-of-the-art on aircraft performance modeling and tailoring. It also provides an introductory overview of machine-learning methods, as well as some examples of application in aircraft operation in general, and specifically in aircraft performance modeling. Findings of this literature review served as starting point to formulate and address the research questions of this project. These main research questions, defined to fill the research gaps identified in the literature study, as well as the research objectives, are presented in Chapter 3. Chapter 4 is devoted to the development of tail-specific hybrid performance models, based on BADA performance models but adjusted using historical QAR data. This chapter provides a description of the QAR datasets used in the project, a comparison between real and BADA average performance, a detailed description of the developed methodology, and the results of the multiple performed calibrations. Chapter 5 covers the development of purely data-driven, non-physical performance models, which predict aircraft performance without modeling the underlying physics. This chapter discusses the motivations of this part of the project, and introduces the main machine-learning algorithms suitable to model aircraft performance. Nonetheless, the chapter is focused on the description of the developed methodology and the analysis of the multiple created non-physical models. The work presented in Chapters 4 and 5 answers the main research questions, and meaningful conclusions can be drawn from the obtained results. These conclusions, together with several recommendations for future research projects, are presented in Chapter 6.

2 State of the Art

Aircraft performance plays a key role in aviation, being a determinant factor in aircraft design, certification and operation. With regard to the scope of this thesis, aircraft performance is of special interest to airlines, since it dictates aircraft operation strategies. Airlines rely on aircraft performance monitoring activities to ensure optimal use of their resources and maximize profit, for which models that mimic aircraft performance are necessary.

The purpose of this chapter is to gain a wide overview of the current consolidated aircraft performance models, as well as of the approaches to improve these models and develop new ones. In Section 2.1 an extensive literature review on aircraft performance modeling is provided, with special emphasis on BADA performance models. This section also includes alternative approaches to model aircraft performance, based on numerical, statistical and machine-learning methods. Section 2.2 addresses the need for aircraft performance tailoring and elaborates on the existing approaches to generate tail-specific models. Given the importance of machine learning on the topic, Section 2.3 is devoted to this discipline.

2.1. Modeling Aircraft Performance

The increase in air traffic makes the airspace denser every day. The International Air Transport Association (IATA) forecasts that passenger numbers could double by 2037, reaching 8.2 billion [4]. To maintain the level of safety, the ATM community needs to face the challenge of coping with traffic increase. ATM applications like air traffic and trajectory simulators rely on the estimation of aircraft performance.

Aircraft performance describes the ability of an aircraft to operate efficiently while accomplishing a specific purpose. Examples of aircraft performance problems are the determination of take-off and landing distances, rates and angles of climb and descent, ceiling, cruising speed, payload or fuel consumption. The disciplines involved in aircraft performance are aerodynamics, aircraft structures and aircraft propulsion. Nonetheless, aircraft performance problems cannot be answered by these disciplines individually [1].

Aircraft performance is key in aircraft design and certification, aircraft comparison, aircraft operation, mission assessment, and even in the investigation of causes of aircraft accidents. In terms of aircraft operation, aircraft performance drives pre-flight, in-flight and post-flight activities. An example of pre-flight activity is flight plan generation, based on aircraft trajectory simulations and fuel calculations. Airlines continuously monitor the aircraft of their fleet to ensure optimal performance in their daily operations. Typically performance levels available in Aircraft Flight Manuals (AFMs), which are provided by the manufacturers, are used. These theoretical performance levels, also called book levels, represent average performance of a fleet with brand-new airframe and engines. However, over time aircraft suffer deterioration that causes performance variations [5]. Performance monitoring is defined by AIRBUS as the procedure through which aircraft data are gathered in order to determine the real performance level of an aircraft with respect to the theoretical level [6]. In order to ensure optimal operation, it is important that airlines detect anomalies in their fleet and fix the problems as soon as possible. E. Chu, *et al.* [7] propose an approach to detect anomalies from aircraft cruise data. They use historical data from a fleet of aircraft of the same type to build a data-driven model of the average operation of the fleet. Later, they detect anomalies by considering them deviations from the models.

Every activity involving aircraft performance relies on aircraft performance models. Accuracy and reliability of results will to a large extent depend on how well the performance models used mimic the real behavior of the airplane. An aircraft performance model is a representation of the real physics of an airplane, of how it interacts with the real world. The main goal of a performance model is to provide a realistic, accurate and complete description of how an aircraft behaves under certain flight conditions. Aircraft performance models can be presented in the form of figures, tables or equations. According to A. Nuic *et al.* [3], a good performance model should be able to support accurate calculations regarding the geometric, kinematic and kinetic aspects of the aircraft behavior. In addition, it should be applicable to a wide set of aircraft types, and valid over the entire flight envelope and during all phases of flight. And last but not least, it should stay within reasonable complexity, maintenance and computing efforts.

2.1.1. Types of Aircraft Performance Models

Flight performance can be modeled with different level of granularity depending on its intended utilization. Whereas studies regarding aircraft control use six degree-of-freedom models, applications for air transportation generally use less detailed models. Aircraft performance models used for this purpose can be classified according to different criteria. In relation to the foundation on which they are built, two types of models can be distinguished: physics-based models and machine-learning-based models. At the same time, physics-based models can be divided in kinetic and kinematic models. Kinetic models represent aircraft forces, while kinematic approaches model the aircraft flight path characteristics without modeling the underlying physics [3]. These models are generally built on ideal assumptions, which reduces their complexity but also their accuracy. An alternative methodology to represent aircraft performance consists of applying machine-learning techniques. In the case of machine-learning-based models, weak or even no assumptions are made, as discussed in Section 2.1.4.2 and later in Section 2.2.3. More specific information about machine learning, and more of its applications in aircraft operation is given in Section 2.3.

In order to develop a model capable of representing aircraft performance, reference data are needed. As a consequence, aircraft performance models can also be classified according to the data source used in their creation: flight-test data, manufacturer data, Quick Access Recorder (QAR) data, surveillance data, etc. Data gathered during flight tests are used by aircraft manufacturers to develop aircraft performance engineering programs in-house. These software, based on a kinetic approach, allow the generation of high quality aircraft performance data for the complete range of operating conditions. Two examples of such programs are INFLT/REPORT Boeing Performance Software and the PEP Airbus Performance Engineering Program [8]. With flight-test data and these programs, manufacturers create aircraft performance models in the form of figures and tables, which later are included in their flight manuals. It is not practical to use manufacturers models directly in aircraft operation applications, due to their dimension, low computational speed and company's Intellectual Property (IP). Alternatively, performance data generated by manufacturers serves as the starting point for other performance models like BADA, described in Section 2.1.2. As well as other models developed from in-flight data recorded with sensors and surveillance data.

2.1.2. Base of Aircraft DAta (BADA) Overview

BADA is an aircraft performance model developed and maintained by EUROCONTROL. As it can be seen in Figure 2.1, BADA consists of two models: an Aircraft Performance Model (APM) and an Airline Procedure Model (ARPM). BADA APM is based on a mass-varying kinetic approach. It represents the aircraft as a point and models the underlying forces that cause aircraft motion. In addition, BADA provides information on nominal aircraft operations thanks to the ARPM. The way an aircraft is operated depends on the specific airspace procedures and operating policies of local airlines, information contained in this second model of BADA. Even though the ARPM is important for different simulation and modeling tools used in ATM, the main focus of this work is the APM.

BADA APM encloses four different sub-models named actions, motion, operations and limitations, as described in [3] and [8]. The actions model is in charge of computing the forces responsible for the aircraft motion. These forces can be distributed in three categories or actions: aerodynamic (lift and drag), gravitational (weight) and propulsive (thrust). Fuel consumption is modeled as a function of thrust.

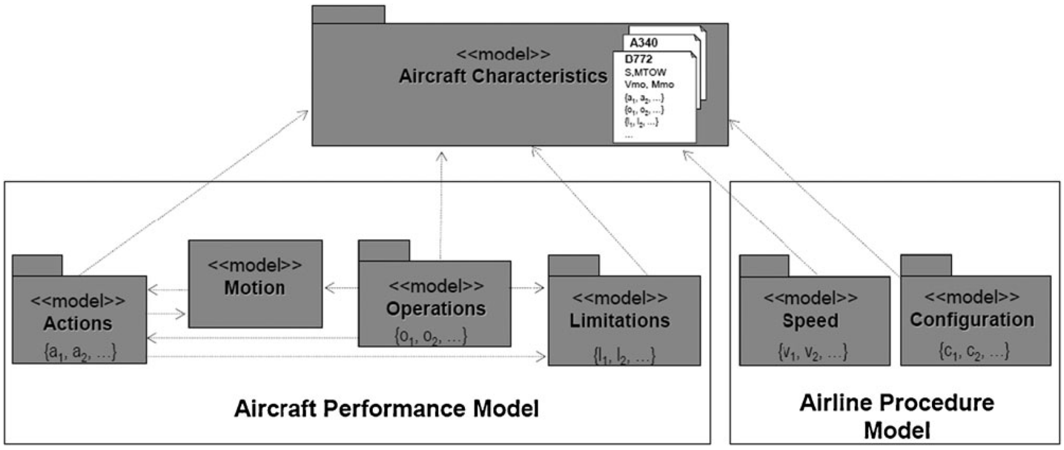


Figure 2.1: Structure of BADA model [3]

The motion model is a Total-Energy Model (TEM). It relates the geometrical, kinematic and kinetic features of the aircraft motion by equating the rate of work done by the forces acting on the airplane to the rate of increase in potential and kinetic energy, as shown in equation (2.1):

$$(T - D)v = W\dot{h} + mv\dot{v} \tag{2.1}$$

where v is the true airspeed (TAS), h is altitude, $\dot{h}$ is vertical speed and m is aircraft mass. This model allows the calculation of aircraft performances and trajectories. Aircraft trajectories are computed by joining the solutions of a sequence of motion problems.

Even though any possible aircraft motion is considered by the previous models, different ways of operating an aircraft lead to different trajectories. To take this into account BADA APM includes the operations model. This model is responsible of restricting the performance solution according to the method used to operate the aircraft (e.g. flying at constant Mach number). Therefore, the combination of actions, motion and operations sub-models is necessary to solve performance problems.

Finally, the boundary conditions of the problem are given by the limitations model, which aims to restrict aircraft behavior to guarantee safe operation of the vehicle and avoid excessive degradation. BADA considers four types of limitations: geometrical, which includes maximum altitudes; kinematic, referring to speed limitations; dynamic, consisting of, for example, maximum weight; and environmental.

One BADA model exists for every aircraft family, which has common aircraft type and version (e.g. all the Boeing 737-800 with winglets and engines CFM56-7B27). In order to generate each BADA model particular technical specifications (e.g. model, engine, dimensions, nominal operating speeds, aerodynamic configurations, etc.) and performance characteristics are required. In the past, AFMs were the main source of reference performance data. However, nowadays manufactures count with performance software capable of generating high quality reference data (e.g. INFLT/REPORT, PEP). Each aircraft model in BADA is characterized by a set of polynomials and coefficients. Coefficients represent the aircraft characteristics and are used by the APM sub-models to model the action, motion, operation and limitations of the particular aircraft. These coefficients aim to achieve the best fit between calculated and reference performance data. In other words, they aim to be the set of coefficients that best describe the aircraft family under consideration. Therefore, aircraft family discretization highly depends on performance reference data. The quality of BADA models rely on the coverage of the entire flight envelope, precision and granularity of the reference performance data, coming from performance software.

Currently, two families of the BADA APM exist: BADA family 3 and BADA family 4. Both of them follow the same modeling approach and have a similar structure. The first dates back to the early 90s, when availability and quality of reference data, computing capabilities and target applications were not the same as today. As a consequence, BADA 3 assures accurate results only for the range of nominal operating conditions, where reference data was concentrated. Nonetheless, it cannot guarantee the same level of accuracy across the complete flight envelope. In addition, BADA 3 does not consider compressibility effects, important at high speeds and altitudes. The latest revision of BADA 3 found models 166 aircraft types from original aircraft data, but it is able to simulate the performance of 272 extra aircraft types, considering them as equivalent in performance [9]. The need for an updated performance model arose as a result of BADA 3 main disadvantage: the inability to accurately model performance across the complete flight envelope. An important research work was performed by EUROCONTROL in cooperation with Boeing Research & Technology - Europe (BR&T-E) to investigate how and to what extend enhancements to BADA 3 were possible [8] [10]. As a result BADA 4 was developed.

The new version of BADA represents aircraft performance over the entire envelope, covering all phases of flight. BADA 4 provides improved expressions for drag, thrust and fuel consumption, applicable to all types of engines and to non-clean configurations. Drag is computed by means of the drag coefficient, which is modeled as a function of lift coefficient, Mach number and the position of high-lift devices, landing gear and speed brakes. Thrust is computed by means of the thrust coefficient, modeled as a function of Mach number and throttle position. Throttle positions for maximum climb, maximum cruise and low idle are provided by engine ratings, which are predefined power settings that pilots select. The high accuracy of drag and thrust models over the entire flight envelope enables the computation of trajectories flown at very dissimilar speeds. In addition, BADA 4 calculates fuel consumption by means of the fuel coefficient, which at the same time depends on the thrust coefficient. BADA 4 incorporates a new sub-model, the Earth model [10], which accounts for the interaction between the aircraft motion and the environment. This sub-model contains an atmospheric and a weather model. BADA 4 also differs from BADA 3 in the operations model, which in this case includes more complex aircraft operations, such as economy cruise (ECON) operation. More on complex flight instructions and operating regimes included in BADA 4 is explained in [11]. All these upgrades culminated in BADA 4 reducing BADA 3 errors, as A. Nuic *et al.* show in one of their conference papers [8]. A comparison of both BADA families in terms of Key Performance Indicators (KPIs), mean of evaluation of an aircraft performance model in the context of its use in ATM, can be seen in Table 2.1. From which it can be concluded that thanks to the high accuracy, effectiveness and applicability of BADA 4 this model can be use in several aviation applications.

Table 2.1: Comparison of BADA 3 and BADA 4 using KPIs

KPI	Definition	BADA 3	BADA 4
Capability	Defined by the performance aspects that the model can provide	Both provide the same aspects	
Realism	Extent to which a model captures true physical dependencies underlying performance aspects	Limited by low quality reference data	Uses high quality reference data
Complexity	Defined by the number of parameters provided	$<$ 20 coefficients [11]	$>$ 50 coefficients [11]
Goodness-of-Fit	Extent to which a model fits the reference data set	RMSE in vertical speed $<$ 100 fpm and fuel flow error $<$ 5% [11]	RMSE in vertical speed $<$ 70 fpm and fuel flow error $<<$ 5% [11]
Accuracy	Extent to which a model represents the validation data set	Validity domain: nominal flight envelope	Validity domain: entire flight envelope
Maintainability	Amount of resources necessary for a model to be kept updated	Both require same amount of maintenance resources	
Applicability	Defined by the area of application of each capability provided	438 aircraft types [9]	Approx. 350 aircraft types [12]

Many applications rely on aircraft performance models and their accuracy. Therefore, the accuracy of the models need to be evaluated. In the case of BADA models, EUROCONTROL has used manufacturer performance data to evaluate the correctness and precision of their models. Regularly, EUROCONTROL Experimental Center (EEC) has been submitting technical and scientific reports, including the denominated ”Model accuracy summary report for the Base of Aircraft Data (BADA)”. In addition, further accuracy evaluations have been performed which used flight data recordings. For example, in [13] the accuracy of BADA 3 model is evaluated by comparing calculated fuel flow and total fuel consumption with QAR data recorded by an airline. Analyses concluded in a good agreement between calculated fuel flow and flight data in the cruise phase. Even though accuracy reduced in climb and descend, the influence on the error for the whole flight was not significant, since the cruise phase is the longest. The error between BADA 3 model and the analyzed flight data was always within $\pm$ 5%.

One example of application in which BADA pays a key role is trajectory optimization, which consists on determining the values of the flight parameters that minimize or maximize trip cost, trip fuel, noise or emissions, for example. The influence of such flight parameters on the optimization criteria is dictated by aircraft performance. The applicability of BADA 4 to trajectory optimization was evaluated by V. Mouillet *et al.* in [14]. To perform the study they use Boeing's reference data and compute the reference optimization results with the Boeing Performance Software (BPS). These results are compared with BADA 4 optimization solutions. It is demonstrated that BADA 4 is suitable for cruise speed optimizations in ATM simulations and for environmental impact assessments. However, it may not be accurate enough for on-board applications and business and economic studies.

Over all, the development of BADA model was a success for the aviation community. BADA has demonstrated to be useful specially for research purposes. However, it models average performance of brand-new aircraft and differentiates performance only according to aircraft type. This leaves a gap that must be filled for the utilization of BADA in certain applications.

2.1.3. Other Aircraft Performance Models

The manufacturers' performance models and BADA are not the only models that can be implemented in ATM systems. Hereinafter, this chapter presents alternatives to these aircraft performance models.

The oldest aircraft performance model developed for ATM applications is the so-called look-up tables model [15]. This model is based on a kinematic approach and it only models average performance, without taking meteorological conditions into consideration. The source data required is minimal and it can be taken from radar track recordings. Due to the simplicity of the model, the accuracy it provides is low, but so are its processing requirements.

In the early 90s, the ATC Research Group at the National Aeronautics and Space Administration (NASA) designed a system for the automatic management and control of arrival traffic, named Center-TRACON Automation System (CTAS) [16] [17]. CTAS operates thanks to a data base consisting of aircraft performance models, airline preferred operational procedures and real-time wind measurements. The CTAS en-route trajectory predictor relies on a kinetic performance model derived from aircraft manufacturer's performance data, including aircraft specific airframe drag and engine thrust data. The aerodynamics model database contains lift- and drag-coefficient data, and the engine model database provides thrust and fuel-consumption data. NASA cooperated with Boeing so that CTAS researches were allowed to use Boeing aircraft performance simulation software (e.g. INFLT) to improve their performance model. It was not until 2012 that NASA tried to integrate BADA in CTAS. M. Abramson *et al.* [18] demonstrated that BADA can improve the accuracy of some CTAS predictions, especially in climb. They also showed that BADA can enhance CTAS performance when dealing with small and regional aircraft types, but apart from that, differences between BADA and CTAS performance models did not appear significant.

General Aircraft Modeling Environment (GAME) is another kinematic performance model developed by EUROCONTROL. GAME is completely based on a parametric approach. In their analysis of performance modeling for ATM applications, A. Suchkov *et al.* [15] define GAME as an upgrading extension of EROCOA/PARZOC, a model developed also by EUROCONTROL in the 70s. Similar to BADA, mathematical functions in GAME are the same for all aircraft types, but their coefficients are equal only for a given airframe-engine combination. GAME and BADA have similar data requirements. They both need flight data in clean and non-clean configurations to allow performance calculations for all phases of flight. For this purpose, flight-profile data are used. In the development of both models, source data come directly or indirectly from aircraft manufacturers, with the IP legal issues, confidentiality, licensing and other commercial interests that entails.

Aircraft performance models can be developed with the goal of being successful in multiple applications, or to fulfill a particular need in a specific application. The manufacturers' performance models, BADA, the look-up table model, CTAS and GAME were designed to be extensively applicable. Nonetheless, different research groups and institutes are working on distinct tools, which may demand a particular performance model with specific capabilities, constraints and assumptions. As a result, multiple aircraft performance models have been developed by different research groups.

In 2013, the Control & Simulation group of the Delft University of Technology introduced BlueSky, a free and open source real-time air traffic simulator. Air traffic simulators rely on aircraft performance models to gain knowledge about the different aircraft procedures. When BlueSky was first developed it used BADA models. However, BADA is owned by EUROCONTROL so its utilization conflicted with the goal of proposing an air traffic simulator free of restrictions and licenses. Therefore, TU Delft had to develop its own aircraft performance model. The first approach to face this challenge consisted on the development of a kinetic Flight Dynamics Model (FDM) based on empirical methods, able to calculate performance parameters from a limited set of input data. The methodology proposed by I. Metz *et al.* [19] starts by taking type-specific aircraft and engine parameters from literature when available. Later, these parameters are stored in a database separately, so that variations in performance caused by different airframe-engine combinations are taken into account. From aircraft parameters, drag coefficient is computed as the sum of parasite and induced drag, both calculated before running the simulation. The value of the drag coefficient is constantly updated using the current lift during the simulation. This model calculates the actual thrust with the TEM, and it finds fuel consumption from thrust and Thrust-Specific Fuel Consumption (TSFC). To validate this methodology I. Metz *et al.* compared computed drag polars with drag polars in literature. Comparisons demonstrated that the model serves as a base for the prediction of aircraft performance within BlueSky, but they showed room for improvement.

Next, a method was developed by T.W. Gloudemans [20] to identify aircraft performance parameters using Automatic Dependent Surveillance-Broadcast (ADS-B) and other open sources of data, like manufacturers' websites and the International Civil Aviation Organization (ICAO) Engine Emissions Databank. In particular, the ultimate goal of the investigation was to estimate lift and drag coefficients. Results demonstrated that the method underestimates drag, possibly because of assumptions made in the development, like neglecting wind effects or simplifying data filtering techniques.

Also in Delft (the Netherlands), J. Sun *et. al* [21] [22] proposed the utilization of ADS-B data, data mining methods and statistical analyses to create a kinematic parametric performance model called WRAP. In this work, three continuous probability functions (Normal, Gamma and Beta) are used to describe performance parameters, from which the most appropriate is selected. One of the main disadvantages of WRAP is that some variables, like cruise altitude, cannot be properly described by a continuous distribution. For example, step climb operations are hardly modeled with this approach. In WRAP model each performance parameter is described by an optimal, a minimum and a maximum value. The first is computed with Maximum Likelihood Estimation (MLE), and the other two by setting confidence intervals. The dataset used to develop WRAP counts with flights which took place all around the world. Thus, researches believe that the distributions of performance parameters obtained reflect to some extend performance characteristics inherent to the different areas of the globe considered. In addition, this work is of interest to the writer because it considers that there exists an interdependence between parameters within each flight phase, and it includes a correlation matrix for each combination of variables. In order to validate WRAP, BADA and the Eurocontrol Aircraft Performance Database were used. Comparisons with these models showed akin results. J. Sun *et. al* have also worked in the estimation of mass and fuel flow based on ADS-B data [23].

In 2010, researchers from the Technical University of Dresden developed an analytic performance jet aircraft model called Enhanced Jet Performance Model (EJPM), with the goal of improving trajectory prediction methods. According to M. Kaiser *et al.* [24], the use of purely analytical methods provides a more reliable representation of the real physics of the aircraft. EJPM takes as input data radar information, weather data and specific attributes of the airframe and engine necessary for performance analyses. Most of these features are obtained from manufacturers. However, design values provided by manufacturers can change with time, so EJPM allows the use of FDR data to register and adjust deviations. EJPM is based on a kinetic approach and it includes four main modules: for speed determination, for calculating lift and drag, for computing thrust, and another one reserved for high precision fuel flow calculations, build with the aid of BADA 3. To test the model real-time data of standard airline operations collected from cockpit instrumentation during cruise was used. The utilization of the EJPM demonstrated an increase in accuracy to predict fundamental performance parameters, and proved that the model can be applied in fuel flow computations. For example, the error between real and computed true airspeed was found to be less than 0.1%. Errors in the drag polar remained within 1% and regarding fuel flow, BADA 3 errors were reduced from 5% to 1.8% after introducing missing dependencies as compressibility effects. On the other hand, the accuracy of the thrust model could not be validated because flight data was recorded during steady cruise flight phases.

Also in Dresden and also with the ultimate goal of optimizing trajectories, but 5 years later, J. Rosenow *et al.* [25] developed the Compromised Aircraft performance model with Limited Accuracy (COALA). This model is based on physical functions only, except for the drag polar, which is modeled with BADA 4 if the aircraft model exists in the database or with BADA 3 otherwise. As a consequence, the resulting trajectories are only the physically possible trajectories. COALA combines the impact of aircraft-specific aerodynamics and the influence of 3D weather information, provided by the National Weather Service NOAA. It describes flight performance in terms of TAS, thrust, fuel flow, forces of acceleration, time of flight and emissions. With COALA an optimum trajectory with respect to either minimum fuel burnt, minimum time flown or minimum emissions can be estimated. Trajectory optimization is realized via the cruising altitude and a speed correction factor, used to manipulate true airspeed and climb gradient. Since COALA is based on BADA to some extent, it inherits its disadvantages.

In order to develop a robust Conflict Detection and Conflict Resolution (CD&CR) tool, the Center for Transport Studies at Imperial College London derived a model to accurately predict trajectories [26] [27]. At a broader level, the Trajectory Prediction (TP) model consists of characterization of the aircraft intent, modeling of the aircraft dynamics and environmental conditions, and modeling of the Flight Management System (FMS) to emulate operational procedures. TP model is a kinetic model based on BADA APM and the analysis of real flight trajectory data recordings, wind and temperature data, and aircraft intent information. TP model uses BADA APM to represent drag and mass evolution, but follows a different methodology to model lift. Particularly, a relation between angle of attack and lift coefficient is derived from the analysis of flight data of a specific commercial aircraft for a real trajectory. As shown in [27], there is a good agreement between predicted and true trajectories. Results show that for trajectory prediction the enhanced model is more accurate than BADA [28].

2.1.4. Alternative Approaches to Aircraft Performance Modeling

As in the case of EJPM, COALA or the TP model, there are more examples showing that researchers not always develop a completely new performance model, but use an existing model as baseline and upgrade it to meet their specific needs. On the other hand, the development of non-physical, data-driven performance models has also been pursued. Alternative approaches to aircraft performance modeling found in literature have been classified according to the main technique used: numerical and statistical methods or machine-learning algorithms.

2.1.4.1. Numerical and Statistical Methods

Numerical methods are mathematical techniques used for calculating approximate solutions of mathematical problems that are difficult to solve analytically. Statistics is a branch of mathematics that aims to solve problems with the aid of collected data. Two examples of statistical methods are regression and stochastic modeling. Both numerical and statistical methods have been used in alternative approaches to aircraft performance modeling.

Aircraft performance is not only present in pre-flight applications, it plays a key role in post-flight analyses. In order to detect performance anomalies, E. Chu *et al.* [7] developed an average, data-driven model to trace the nominal operation of the vehicles of a fleet using historical data. Specifically, they computed the linear and rotational accelerations as a function of some performance parameters. To find the relation between performance variables and accelerations a regression problem was solved by applying the method of least squares. To make the model realistic some noise, consequence of unconsidered flight conditions and turbulence effects, was added. To detect anomalies the residuals of the model were computed and a threshold was established. This is another example of a performance model developed with a specific purpose.

With the goal of providing realistic aircraft behaviors in air traffic simulations, EUROCONTROL developed a methodology to improve BADA ARPM by considering specific aircraft operation parameters identified from historical data [29]. To carry out the proposed analysis data had to be collected first. In their study the dataset consisted of flight plan information, radar track data and weather data. After gathering all the data, it was processed and additional parameters such as rate of climb/descent (ROCD), calibrated airspeed and true airspeed were computed. The dataset was divided in flight categories for which similar aircraft behavior was observed, and operational parameters for each category were found using statics. This work resulted in a customized aircraft behavior model. To validate to what extent this model improves BADA ARPM performance, simulated flight profiles were compared with average real trajectories. S. Guillet *et al.* [29] demonstrated that the analysis of flight recordings can bring valuable information that can improve the accuracy of aircraft modeling.

Not all researches are looking for a more accurate aircraft performance model. Some have tried to improve the estimation of the input variables of the models, which is important specially when using the model in ground-based applications. For example, BADA models use a standard mass and thrust profile per aircraft type, which may lead to inaccurate results. Multiple research projects focus on the estimation of aircraft mass. Aircraft initial mass is an important source of uncertainty in trajectory prediction. It determines mass evolution, and consequently fuel consumption and trip cost. Accurate aircraft initial mass is required to provide a good estimation of fuel consumption and passenger and cargo capacity, two of the main concerns of airlines. Several studies proposing methods to estimate aircraft mass can be found in literature. Fang He *et al.* [30] developed a method to accurately estimate the mass of each flight using QAR data. They reformulated the flight dynamics equations and solved them with an improved structured nonlinear total least-squares method using Monte Carlo experiments. The main advantages of their methodology is that it does not depend on thrust and does not require information regarding aerodynamic coefficients, but it only considers steady level flight. R. Alligier *et al.* [31] improved the accuracy of trajectory prediction of climbing flights by inferring mass and thrust from past observations. First, mass is estimated using a least-squares approximation and several points of the trajectory. Mass is adjusted so that the modeled power fits the energy rate observed on past points. Then, the thrust setting law is computed iteratively using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method, coupled to the mass estimation model. Following their methodology trajectory prediction accuracy improves up to 50% when compared with BADA.

2.1.4.2. Machine-learning Algorithms

As shown earlier in this section, there exist several aircraft performance models that distinguish performance mainly based on aircraft type. Even though some of them include weather effects, it is believed that other factors like operator, airport or take-off weight also influence performance. Multiple approaches exist that suggest to consider flight attributes to model aircraft performance. They rely on machine-learning techniques to predict aircraft performance. This chapter provides an overview of these methods. More information on machine learning is provided in Section 2.3, which includes a general description of the discipline, as well as some other examples of application.

In the 33^{rd} Digital Avionics Systems Conference (DASC), M. Hrastovec *et al.* [32] presented a machine-learning model to predict the aircraft performance that best represent each flight, using large amounts of data. Theirs is a non-physical model, since it does not explicitly describe any physical relation. Instead, they use collected data enriched with non-physical attributes such as airline, departure, destination and flight date and time. Then, they group flights based on similar attributes. Their main source of data are radar data recordings, accompanied by weather information and flight plans. All three sources of data must be properly correlated, the right flight attributes must be coupled with the corresponding radar data. As explained in [33], M. Hrastovec *et al.* assume that flights with similar flight plans also flight similarly. In all data analytics problems, data can contain errors and even missing values, so a preprocessing phase is necessary to assess data consistency. However, in the methodology firstly proposed [32], authors do not include cleaning of inconsistent data, since they believe these outliers are not representative in their datasets and thus, do not significantly influence results. First, they estimate airspeed using radar data and numeric weather prediction models, to later segment flights in different phases depending if aircraft is ascending, flying at constant altitude or descending. They propose to gather as many variables as possible, and let the machine-learning algorithm decide what attributes are more important to describe aircraft performance. Performance models based on machine learning and historical data are dynamic models, which means that they require an updating mechanism to keep the database, and thus the performance model, up-to-date. In this regard, M. Hrastovec *et al.* [32] propose to every night process the flights from the previous day.

In their next proposal [33], the goal is to utilize machine learning to provide BADA input parameters and make predictions closer to real trajectories. By means of the predicted input parameters BADA default parameters can be substituted. Their methodology includes three phases: data acquisition, data preprocessing and performance prediction using machine learning, as illustrated in Figure 2.2. Regarding the machine-learning technique used for prediction, the selected algorithm was the k-Nearest Neighbor (k-NN), where k is the number of considered *neighbors*. In this particular case, the distance between two different neighbors (or flights) is given by the number of non-matching attributes. The quality of the prediction greatly depends on the quality of the set of k nearest *neighbors*, so it is essential to find the set of attributes that cause two flights to exhibit close aircraft performances, which can be done manually or automatically. Feature selection can be done once globally for all predictions (manual feature selection), or dynamically for each prediction (automatic feature selection). Not to complicate the model even more, M. Hrastovec *et al.* assumed flight attributes to be independent from each other. The accuracy of the method is computed by comparing calculated and actual performance. Comparisons of results between using BADA, using average performances calculated from the database of historical data, and using the machine-learning method were performed. As a result, it can be concluded that BADA nominal values give the worst results because the model is made for a general average case. Even though the machine-learning model does not improve accuracy drastically, it makes good predictions fast enough to be used in real-time applications.

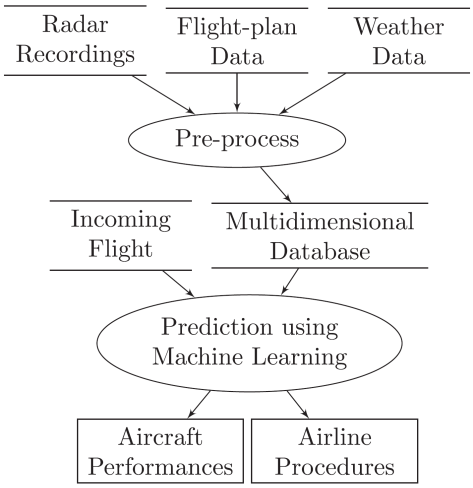


Figure 2.2: Schema of the methodology proposed by M. Hrastovec *et al.* [33] to predict aircraft performance using machine learning

As a continuation to their previous work [31], R. Alligier *et al.* developed a machine-learning regression model to estimate aircraft mass for aircraft climb prediction [34]. Again, the purpose is to substitute standard BADA input parameters, rather than to enhance BADA. In this problem, aircraft mass is the variable to predict (target variable), but actual mass is not available in the dataset used to train the model, which consists on radar and weather data. To overcome this problem, actual mass is estimated by adjusting it so that the modeled power fits the observed energy rate in the best possible manner, as proposed by the authors in other publications [31]. Five different machine-learning regression algorithms were tested: Multiple Linear Regression (MLR), Ridge regression, Principal Component Regression (PCR), single layer neural network and Gradient Boosting Machine (GBM). The latter algorithm provides the best results, reducing BADA error from 30 %, except for one aircraft type, for which BADA approximation and machine-learning estimation culminated in similar errors. In order to be as close as possible to the operational context, the default BADA speed profile was first used. However, this speed profile is not always close to the actual one. Aircraft climb at constant calibrated airspeed until the transition altitude is reached. Afterwards, they ascend at constant Mach number. BADA correlates one pair of calibrated airspeed and Mach number values to each aircraft type, but these values depend on some unconsidered parameters, such as Cost Index (CI). To solve this problem, the authors trained another GBM algorithm to predict airspeed during climb, specifically calibrated airspeed and Mach number [35]. This upgrade allows improvements from 36 % with respect to BADA, and an error reduction of 79 % for the aircraft type that was not improved in the previous approach. A combination of both approaches decreased the error by at least 45 %.

Apart from aircraft aerodynamic characteristics or aircraft mass, engines play an important role in aircraft performance. Fuel consumption is strongly related to direct operating costs of airlines. Hence, fuel savings is a mayor driver in air transportation. Fuel flow rate (mass of fuel injected into the engine per unit time) is a key indicator of engine performance, so a correct modeling of such parameter is important to assess engine performance, as well as to estimate aircraft emissions. Existing performance models have shown insufficient capabilities in fuel consumption modeling, specially when high level of accuracy is required. The ICAO Aircraft Emissions Databank is sometimes used. This database includes values of fuel flow rate corresponding to different aircraft engines. However, these values not always match actual fuel flow rate, since they are obtained through ground tests done at sea level static conditions on uninstalled engines.

One possible approach to represent engine performance is to model fuel flow rate using historical data. This exercise has been found multiple times in literature. In [36], Y. S. Chatu *et al.* provide a methodology on how to use machine-learning techniques and operational FDR data of several aircraft types to model fuel flow rate. Two different regression algorithms are described: the Classification and Regression Trees (CART) and the Least Squares Boosting (LSB) algorithm. The first is commonly known as decision tree algorithm. The later is an ensemble method, which combines weak prediction models (e.g. a decision tree) to create a more powerful model. Before implementing any algorithm, input features are selected based on their importance in the prediction. Input features are normalized with reference values. Then, the dataset is divided in train, test and validation sets. The training set is used to build the model, the validation set is used to select the model and the test set is used to evaluate the final model. The mean error and the percentage coverage are the metrics used for model validation. As expected, the LSB algorithm is more accurate than the CART, but it also requires higher computational times. An approach that diverges from the conventional is followed to improve model accuracy . Decision trees break-up the predictor space into different areas in which all points show homogeneity in fuel flow rates values. Therefore, instead of training a single model for the entire dataset, different algorithms are fitted in each region. By comparing fuel flow rate predictions with those given by the ICAO Databank and BADA models, it can be concluded that both machine-learning models represent fuel flow rate more accurately than the ICAO Databank and BADA models.

As a continuation to [36], Y. S. *et al.* followed a Gaussian Process Regression (GPR) approach to model aircraft engine fuel flow rate [37]. GPR is a regression algorithm in which the underlying regression function follows a Gaussian distribution. It is a Bayesian approach, so it is characterized by its ability to directly give a predictive distribution for new data. In this work, the dataset available is formed by FDR data. However, input variables are restricted to those accessible by ground-based surveillance systems. For training, aircraft gross mass available in the FDR data are used, but such parameter is not accessible on-ground, so in order to make a prediction for a new flight its value must be estimated. The developed algorithm estimate both instantaneous aircraft gross mass and fuel flow rate, assuming that mass reduction is only due to fuel consumption, and assuming the values of both parameters to be known at take-off. Estimated fuel flow rate is used to estimate gross mass, and estimated gross mass is used to estimate fuel flow rate again. Hence, one input variable of the regression problem has uncertain values. This time, apart from mean error and percentage coverage, the normalized root mean squared prediction error (NRMSPE) is used. Performance metrics reveal that the GPR model is less accurate in descent, which could be a result of uncertainty propagation from take-off, all the way throughout the flight until the descent phase. However, it could also be a result of the typical operational variations in aircraft descent. Nevertheless, the GPR model improves BADA performance in all the flight phases considered in the experiment, reducing mean error up to 50 %.

To estimate fuel consumption and detect engine abnormalities, Zhen Pan *et al.* [38] developed a fuel estimation model based on a Back Propagation (BP) neural network, using flight data. The network consists of three layers: the input, hide and output layer, as illustrated in Figure 2.3. The training data points propagate from the input layer to the hidden layer, and later to the output layer. This process is called forward propagation. The output obtained is compared to the expected value, if they are different the error is propagated backwards through the hidden layer until it reaches the input layer. The relative contribution of each layer to the error is computed. By doing so the hidden layer learns to minimize the output error. The outputs of the neural network are fuel flow of the left and right engines. The product of fuel flow and flight time gives an approximate total fuel consumption. In this approach the influence of weather parameters, longitudinal and vertical acceleration, and turning rotation on fuel consumption is taken into account, which improves model accuracy. As in most of the approaches to model aircraft performance, Zhen Pan *et al.* segmented flight data according to flight phases and trained a different model for each phase. To do so, they analyzed the values of variables like height or flight speed. Aircraft perform differently in each flight phase, so flight segmentation is important. The more accurate flights are segmented the more faithful the model is.

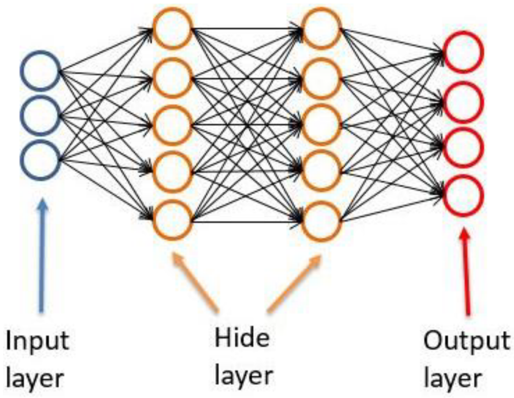


Figure 2.3: Schematic diagram of the neural network proposed by Zhen Pan *et al.* [38] to estimate aircraft fuel flow

With respect to flight segmentation, the Control & Simulation group of the TU Delft proposed the use of various machine-learning and fuzzy-logic methods to estimate flight phases using solely ADS-B data [39] [40]. Clustering techniques are used to extract continuous flights and divide them in phases. Clustering is an unsupervised-learning process that groups data based on the difference of features. For this application, the cluster algorithm must be able to adapt the number of clusters depending on the data, since different flights can have different number of segments. The two algorithms proposed based on this requirement are DBSCAN (density-based spatial clustering of applications with noise) and BIRCH (balanced iterative reducing and clustering using hierarchies). The first separates data into areas of high and low density, eliminating noise data that is at a lower density than the clusters. The latter builds a Characteristic Feature (CF) tree, in which each leaf node is a sub-cluster. First, the goal was to cluster scattered flight data into sets of continuous trajectories, for which both clustering methods performed similarly. However, DBSCAN proved to be more efficient and practical. Then, fuzzy logic is utilized to identify different flight phases. Fuzzy logic considers concepts where no precise definition of criteria for classification exist. The variables considered are the altitude, airspeed and ROC. One of the main issues in flight segmentation is data noise, since variables such as speed and ROC can exhibit large variations. J. Sun *et al.* [39] suggested to filter data before it is processed with fuzzy logic. In addition, they divided data into time windows to reduce the steps necessary for segment identification. The steps followed in their approach can be seen in Figure 2.4. As explained in [40], to validate the quality of the phase recognition the number of phase transitions and the feasibility of these transitions are checked. This method to label continuous trajectory and flight phases was used by TU Delft to build the aircraft performance models for the open-source BlueSky simulator.

Throughout this chapter multiple methods to address aircraft performance have been presented. Aircraft performance modeling has been a concern for the aviation community since last century, and it still is nowadays. Every approach presented so far focus on performance of aircraft types, assuming that every aircraft belonging to a particular family (e.g. all the Boeing 737-800 with winglets and engines CFM56-7B27) behave similarly. However, in reality, there could be significant differences in the performances of two aircraft belonging to the same family.

2.2. Tailoring Aircraft Performance

Ideally aircraft performance models would demonstrate good agreement with the true performance of the aircraft. Unfortunately, in reality this is rarely the case. Most models represent average aircraft performance, depending on the aircraft type. As a consequence, some researchers found necessary to calibrate aircraft performance models based on aircraft tail number, also called registration number.

This section contains a brief introduction of the concept of aircraft performance tailoring, explaining what is it and why is it necessary. Afterwards, it presents the efforts found in literature regarding aircraft model calibration according to tail number, and it gives an analysis of the different approaches and methodologies.

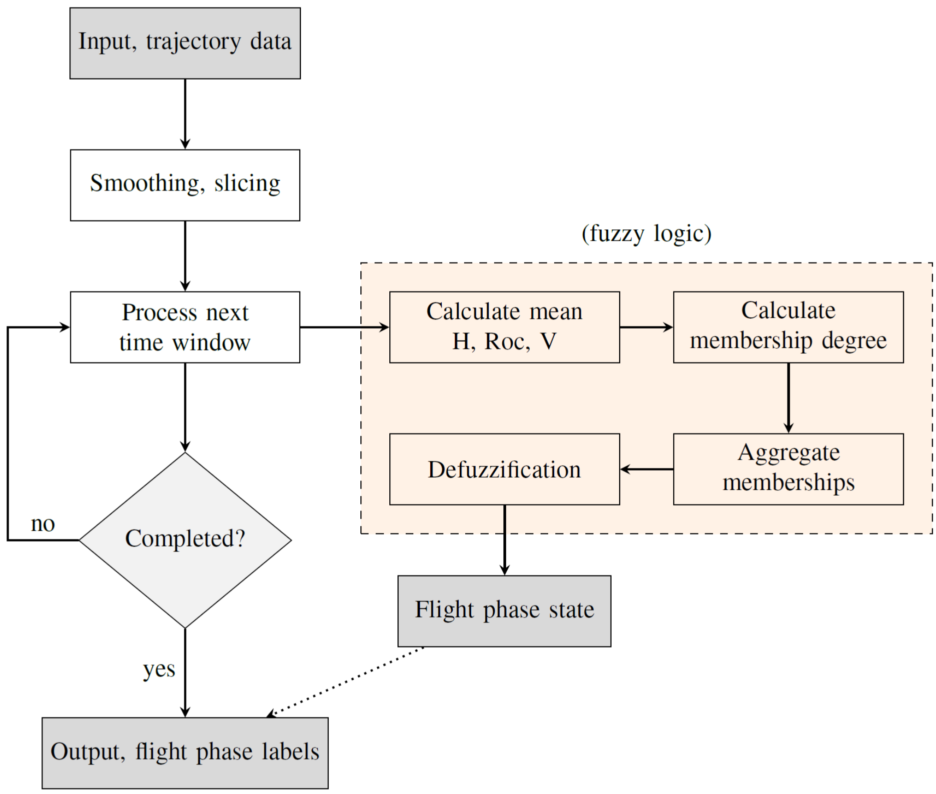


Figure 2.4: Flow diagram of flight segmentation process using fuzzy logic [39]

2.2.1. Introduction to Aircraft Performance Tailoring

Even though the development of aircraft-type-specific performance models has been a success for the aviation community, individual aircraft have their unique flight performance due to aging, user-configurable specifications, maintenance variations, etc. The monitoring practices performed by airlines are an example of how important is to know the performance of particular tail numbers. Airlines use performance factors to adjust the FMS aircraft performance level to the actual performance capability of the aircraft. In other words, performance factors are used to express aircraft efficiency loss. Using the same performance factors for all aircraft would penalize most of the fleet. Therefore, airlines typically compute performance factors by analyzing real flown data, and tailor them to each individual aircraft with the aid of aircraft performance monitoring software. This is important since performance deviations vary from tail to tail depending on several factors. Like for example, degradation in performance of aircraft systems with age, additional drag due to unclean surfaces, modifications consequence of maintenance actions, or operational history.

Tailored performance models are useful to reflect aircraft-specific performance degradation. BADA not only represents the average performance characteristics of all the aircraft of one type, it also reflect the performance of a brand-new aircraft, ignoring performance degradation as time passes. Degradation must be taken into account, since fuel consumption for the same flight route is higher the more deteriorated the aircraft is. Aircraft performance deterioration can be due to airframe deterioration and engine performance degradation. Airframe deterioration can be caused by deformed aerodynamic surfaces, flaking paint, repairing patches and seals missing, among others. On the other hand, engine deterioration can be caused by fan blade leading edge erosion, dirt contamination, fretting corrosion or blended blades reparations, for example. Studies show that generally engine deterioration accounts for 80 % of the total performance degradation, leaving the remaining 20 % to airframe deterioration [41] [42]. Aircraft degradation is already considered in flight plan generation, but considering it in tail-assignment processes would allow airlines to allocate aircraft to flights such that costs could be minimized. In [43] aircraft degradation is modeled as performance factors. However, it also exists the possibility to integrate deterioration effects in the performance models. Which is what this chapter addresses.

By *Aircraft Performance Tailoring* the writer refers to the action of either adjusting existing aircraft performance models to particular tail numbers, or developing new models making a distinction according to tail number. In any case, the resulting model must represent the performance of a specific aircraft more accurately than an average performance model. The performance models introduced in Section 2.1, such as BADA APM, assume the same performance for all the aircraft of the same type. However, in reality different aircraft exhibit different performance characteristics, and these evolve differently with time. This assumption creates uncertainty in the knowledge of the performance of a specific aircraft.

At least three approaches can be followed to develop an adapted performance model that accurately represent the performance of individual aircraft. Two of them involve the "recycling" of existing parametric models like BADA. One approach consists of maintaining the established polynomials, and finding the new coefficients that best represent the performance of the aircraft under consideration. Alternatively, the effort can involve a change in the given polynomials. These models are called *hybrid performance models* hereinafter in this report, because they are based on a physical model (so preserve the physical relations to some extent), but are adapted using historical data. In fact, this is one of the main advantages of hybrid performance models. Since they are somehow based on known physical laws, explanation and justification of results is more interpretable. In addition, because they use historical data to fine-tune physical models they allow to include the effect of external factors such as weather, origin and destination airports, etc. The main disadvantage of hybrid performance models is that they inherit physics-based models assumptions and limitations.

The third and last approach included in this chapter deals with the generation of *non-physical models* derived solely from historical data, using machine-learning techniques and without modeling the underlying physics. The main advantage of these data-driven performance models is that they are built on weak or even without assumptions, which increases accuracy. In addition, machine learning allows to make predictions fast and cheap. In short, these methods are generally quick and robust. Furthermore, machine-learning-based models allow the use of non-physical variables in the development of the model, and therefore, consider factors that may affect aircraft performance from the beginning. However, machine-learning systems are often considered "black boxes" because of the difficulty to explain what functions are used by the algorithm to make predictions.

In both cases, a limited amount of flight data covering areas of the flight envelope away from typical operation conditions is a challenge to confront when willing to develop a performance model valid for the entire envelope. In the case of hybrid approaches, corrections to the physics-based model can only be made for the range of the flight envelope where enough historical data exist. Similarly, non-physical models are only able to learn aircraft performance corresponding to the flight conditions contained in the dataset. A possible solution to this issue is the generation of artificial data to complement historical data. The hybrid and non-physical approaches towards aircraft performance tailoring found in the literature are detailed below.

2.2.2. Hybrid Tailored Aircraft Performance Models

For certain applications, model calibration is necessary to tune available performance models to obtain a better agreement with real performance. Model calibration is "*the process of adjusting numerical or physical modeling parameters in the computational model for the purpose of improving agreement with experimental data"* [44].

T. G. Puranik *et al.* [45] addressed the calibration and validation of a performance model developed for general aviation aircraft. In their work, the shape of the performance model is maintained and the parameters are optimized to increase accuracy. They propose a two-level approach to aircraft performance calibration. The first layer uses publicly available data, specifically, the Pilot Operating Handbook (POH) provided by the manufacturer. The used performance model is an empirical model in the form of curves. In order to parameterize the curves, they use several factors corresponding to engine, propeller and aerodynamic characteristics. For example, engine de-rate factor, scaling of propeller chord and scaling of parasite-drag factor for clean configuration, respectively. Calibration factors are the variables used to systematically modify performance models. They can be added or multiplied to the curves. In the first level, a multi-objective optimization algorithm is used to minimize the error between predicted and real performance. Even though this calibration level increases the accuracy of the initial model, the data contained in the POH correspond to brand-new aircraft under ideal conditions, so further actions must be taken to truly represent real performance. For this purpose, the second level of the calibration process uses flight recorded data. This time, Specific Total Energy

Rate (STER) is used to calibrate the model, since it can be calculated using both; flight data and the basic performance model. In this level, first an evolutionary optimization algorithm is chosen. Then, a genetic algorithm is applied to ensure that the best calibrated model is obtained. The resulting hybrid performance model has a root mean square of residuals lower than 1, being the residuals the difference between actual and predicted performances. As expected, the model calibrated with actual recorded data is more accurate than the model calibrated with manufacturers' performance data only. Even though this work provides a methodology to calibrate a performance model using flight data, the used model was developed for general aviation aircraft. The interest of this literature study is commercial aviation.

There exist some studies in literature that focus on commercial aircraft and use BADA as reference performance model. For example, E. Casado *et al.* [46] proposed a methodology to evaluate the impact of the aircraft performance uncertainty consequence of the use of BADA 4 models on trajectory-prediction accuracy. By now, it is already known that BADA actions model consists of a series of polynomials and a set of coefficients that determine the nominal performance of brand-new aircraft. Performance degradation can be attained by modifying the polynomials. In the methodology proposed in [46], only the independent coefficients of the polynomials are modified for simplicity. To find the independent coefficient that best represent actual performance, a Monte Carlo experiment is used to randomly generate aircraft performance models with different independent coefficients, within defined bounds. The analysis of results showed that any possible prediction was contained in a plane defined by the trajectories given by the nominal and a set of degraded aircraft performance models. This allows to compute the characteristic analytical formula of the plane. On the one hand, the methodology is applicable to any performance models based on polynomials. On the other hand, it only modifies the independent coefficient and it was only applied to those flight segments where the aircraft was flying with clean configuration. This explains why the resulting hybrid model provides a good agreement between predicted and actual cruise, but shows big differences in climb and descent.

The work presented above gave rise to a patent [47], in which a method for modeling aircraft performance including degradation is provided. Such accomplishment is achieved by introducing degradation coefficients in a nominal performance model. The idea is to improve the fidelity of aircraft performance models by utilizing recorded flight data to identify deviations between the actual performance of the aircraft and the nominal performance values, given by the nominal performance model for that aircraft type. The methodology proposed in the patent is illustrated in Figure 2.5. First, the aircraft performance data must be gathered, which include recorded trajectory data, weather data, aircraft configuration and aircraft intent. Then, the recorded and computed trajectories (TRJ_R and TRJ_C, respectively) are obtained. Degradation coefficients are optimized such that the difference between computed and real trajectories is below a predetermined threshold (ε). Once this criterion is satisfied, the degraded aircraft performance model is considered the enhanced aircraft performance model. Apart from the recorded flight data, another input to the process is the aircraft performance degradation model, which is the selected criterion used to change the aircraft performance model so that degradation is represented. A possible criterion is to modify the independent term of BADA polynomials based on theoretical hypothesis, like the different influence of engine and aircraft degradation to total performance deterioration, approximately 80% and 20%, respectively [41] [42]. As stated in the patent, the proposed methodology can be used to generate a hybrid performance model for aircraft of the same type, for the same type and airline, and if desired, for the same aircraft. If more than one set of recorded flight data is available for the same aircraft, the most updated information should be used.

As it was mentioned before, one advantage of this methodology is its possible application to any performance model based on polynomials and coefficients. In addition, the resulting enhanced aircraft performance is adjustable through the criterion chosen to define the aircraft performance degradation model. This means that if new knowledge regarding aircraft performance degradation becomes available, the aircraft performance degradation model could be updated, and the methodology would still be valid. Even though the methodology would be applicable to tailor performance model to specific aircraft, E. Casado believes that this is not the optimal used of the method. Instead, he finds more interesting to use it to obtain the model that best fits the performance of a fleet of aircraft of the same type and of the same airline. Therefore, aircraft performance tailoring was not further explored. Likely, the proposed method could be complemented with more sophisticated techniques by using Artificial Intelligence (AI).

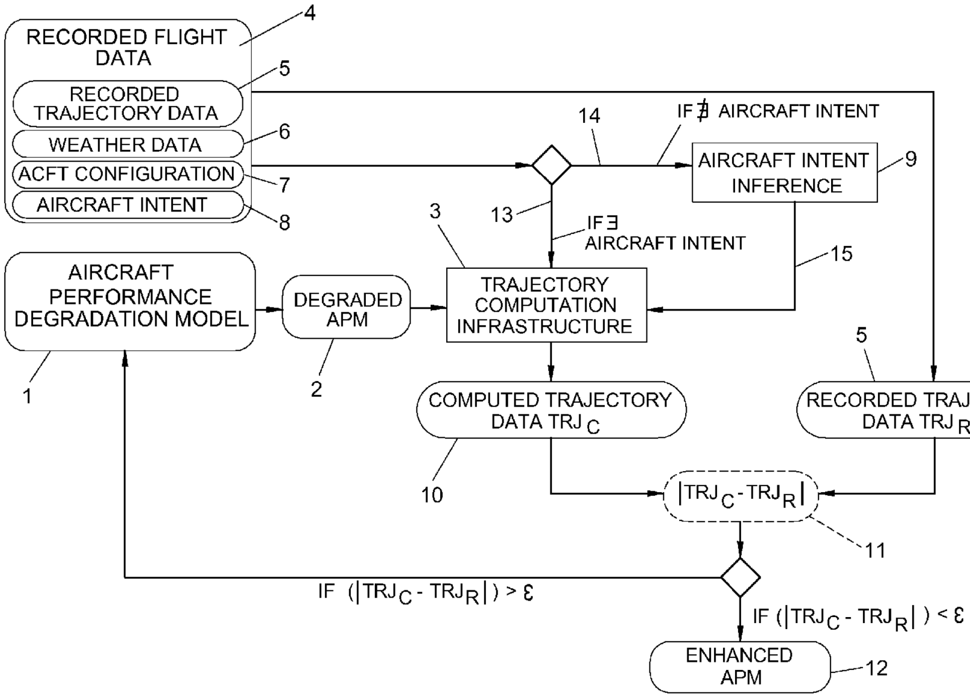


Figure 2.5: Flow diagram of the methodology proposed in [47] to model aircraft performance considering degradation

2.2.3. Non-physical Tailored Aircraft Performance Models

The use of AI techniques, specifically machine-learning techniques, in aircraft performance tailoring has been explored by the Aerospace Research Center (ARC) of the Istanbul Technical University (ITU), with the ultimate goal of improving aircraft trajectory prediction. It is known that aircraft mass and speed are important parameters in aircraft trajectory prediction. The estimation of initial mass is essential in predicting future trajectory of the aircraft, as demonstrated by the multiple studies devoted to this matter [23] [30] [31] [34], and so is the fuel consumption model, since it updates aircraft mass with time. Research activities performed by the ARC of ITU aim to improve the fuel consumption model provided by BADA 4, aircraft performance model used as benchmark. Researchers believe that high-fidelity models can be achieved by using actual recorded flight data and aircraft communications data. The main source of recorded flight data is the QAR, and communication data can be obtained from the Aircraft Communications Addressing and Reporting System (ACARS).

In [48], an approach to improve BADA fuel consumption model using deep learning and QAR data is presented. QAR data are used to train and validate the deep-learning model, as well as to feed the BADA fuel consumption model, used as a benchmark to evaluate the developed (non-physical) performance model by comparing the pair of fuel consumption values achieved. To solve the regression problem, an Artificial Neural Network (ANN) is designed. The model uses 8 variables as inputs, processes them through 4 hidden layers and gives the fuel consumption as output. A detailed representation of the model can be seen in Figure 2.6. As in all machine-learning problems, selection of input variables (or features) is crucial. In this model, 2 out of the 8 features correspond to the throttle levers, which directly represent thrust. As expected, thrust information proved to be important in the learning of the fuel consumption. The deep-learning model illustrated in Figure 2.6 is capable of estimating a fuel flow very close to the real value, which BADA 4 cannot do.

Even though the throttle lever positions are available in their set of QAR data, such information is not known beforehand. Hence, the proposed model can be used to estimate fuel consumption in in-flight use cases and post-flight applications, for example in fuel analytics. Nonetheless, it is not valid to predict fuel consumption of future flights or ”what-if” simulations. In addition, only 400 flights are used to train and validate the deep-learning model. Those flights correspond to two different long-haul routes. However, [48] does not specify if the routes were flown by the same specific aircraft, by aircraft of the same type or none of the two. It is therefore assumed that the resulting model would describe the performance of the aircraft represented in the dataset. Regardless this assumption and compared

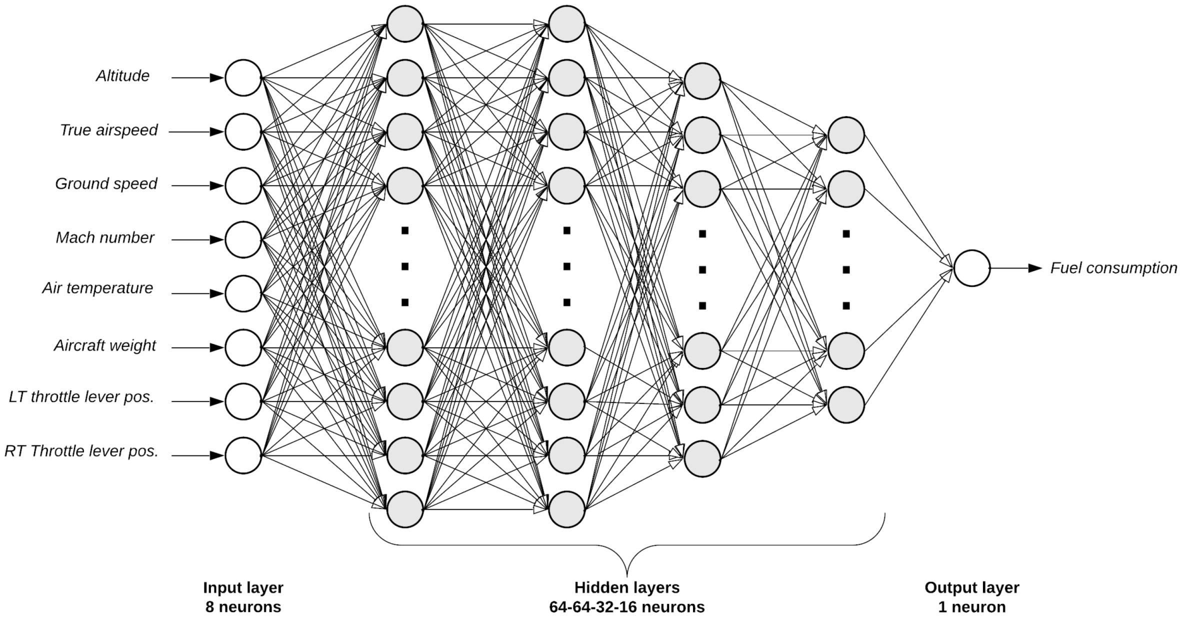


Figure 2.6: Detailed model architecture proposed by A. Alizadeh *et al.* in [48] to estimate fuel consumption from QAR data

to the amount of actual flown data available today, 400 flights may not be constitute a representative sample.

Further studies have been performed at the ARC of ITU to pursue a solution to these issues. In March 2019, M. Uzun *et al.* presented their last achievements in building data-driven models to represent actual fuel flow [49]. This time more than 1000 flights are used. This study uses QAR data and filed flight plans. With such information it aims to model the actual fuel flow of specific aircraft for each principal flight phase: climb, cruise and descent. For which they suggest to segment flights into different phases first of all. This classification problem is solved by using unsupervised machine-learning methods, in particular density-based clustering algorithms. A different sub-model is developed for each flight phase. In [49], three different deep neural network models to calculate tail-number specific correction factors to BADA models are proposed. More accurate performance calculations can be achieved by implementing such correction factors into the nominal performance model.

The essence of the first method proposed corresponds to the model presented in [48], a model that estimates total fuel flow only using variables available in the QAR data. Hence, the first model is a non-physical data-driven model. Comparing this model with real flight data it was found that precise estimations of fuel flow, with mean errors lower than 0.15 %, could be achieved. However, this model can only be used when throttle information is accessible. This restriction is addressed in the second model proposed.

To avoid the need to use the throttle information contained in the QAR data, the second model reconstructs the thrust force, using BADA to compute drag, and the QAR data to obtain the rest of the parameters required. Hence, the second model is a hybrid model. Once the thrust force is known, the throttle information can be inferred from the thrust coefficient, for which BADA model is used again. After the throttle information has been modeled, it can be used as an input feature to the neural network model explained earlier. This second approach gives similar results as the first approach for climb and descent phases, but perform worse during cruise. Since this method is based on BADA to some extent, it inherits its drawbacks. In this case, the erroneous modeling of the throttle information during cruise. Nonetheless, comparing this second model with real flight data it was found that estimations of fuel flow with mean errors lower than 0.7 % could be obtained for all flight phases.

The third method is again a hybrid model, in which throttle positions are estimated and then fed into the first fuel flow estimator. In this case, the model estimates two variables; the thrust lever of both engines, so a more complex neural network is trained. A general diagram showing the architecture of this neural network is shown in Figure 2.7. One of the input features chosen to estimate the thrust levers is the thrust force, computed with the aid of BADA. This is why the third model is again a hybrid model.

Comparing this third model with real flight data it was found that estimations of fuel flow with mean errors lower than 0.3 % could be obtained for all flight phases. Once an accurate estimation of the fuel flow is attained, a correction factor tailored to specific aircraft can be computed and multiplied to BADA fuel consumption model in future trajectory predictions. One conclusion driven from this exercise is that flight segmentation is a necessary step in performance modeling, since the values of the obtained correction factors vary with flight phase.

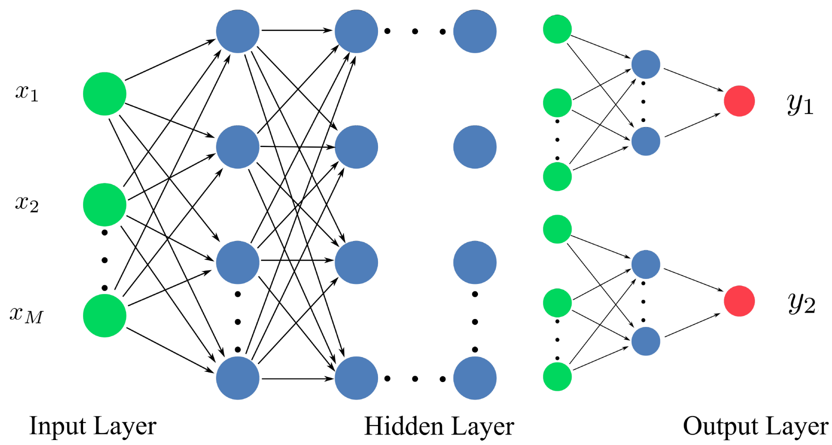


Figure 2.7: General architecture of the neural network model proposed by M. Uzun *et al.* in [49] to estimate throttle lever of the two engines of the aircraft

Although the results of the study are promising, the methodology gives room for improvement. For example, in the third provided approach, the selected input features allow to estimate throttle levers with an acceptable level of accuracy. Like it was demonstrated in [48], throttle levers and fuel consumption are strongly correlated. As a consequence, it may be possible to directly estimate fuel flow from those input features. This would eliminate the need of a two-layer performance model and would reduce computing time and memory. Additionally, the input features of the models could be selected automatically from every variable available in the QAR data and filed flight plans, by analyzing their influence in the estimated parameter. In [48], the feature selection process is not specified, and in [49], the input features selected to estimate fuel flow are the ones used by BADA (except from the throttle levers). However, having more than 1000 parameters recorded by the QAR gives the possibility of finding determinant parameters that have not been considered before, and that may have an impact on aircraft performance. Last but not least, remark that the second and third approaches, which are the only ones valid for pre-flight applications, are based on BADA drag model. Nonetheless, this model has not previously been tailored to specific aircraft tail numbers, so the accuracy of results is constrained by BADA main drawback; it models average performance of a brad-new aircraft representing an aircraft type. Despite these opportunities of improvement, the work done by the ARC of ITU sets the basis for other approaches to update performance models using machine learning.

This chapter has summarized the state-of-the-art of aircraft performance tailoring. The studies described above show the importance of having a model that mimic real performance. They also demonstrate the viability of the creation of a methodology to tailor aircraft performance. On the other hand, apart from the different assumptions made in each approach, none of the found methodologies mention the incorporation of an updating mechanism. When pursuing tail-specific performance models to account for performance deterioration it is critical that models are dynamic. For which methodologies must include some kind of update mechanism, to allow the models to change through time as the aircraft components wear, for example. To the knowledge of the writer, no complete methodology has been provided so far, and no further work is available regarding the adjustment of existing performance models or the creation of data-driven models based on tail number.

2.3. Machine Learning

As it has been demonstrated throughout this literature review, AI, and specifically machine learning, is a extremely useful tool in aircraft performance modeling. This section is devoted to this discipline. And its goal is to provide extra knowledge for a better understanding of the methods involving machine learning that were mentioned in Sections 2.1 and 2.2.

This chapter contains general information about machine learning and its methodology, as well as some additional examples of application in aircraft operation. At the beginning, it explains the main concepts involved in machine learning and it presents its utility nowadays. Then, the main types of algorithms are explained. This section ends with some machine-learning models developed for aircraft operation applications, and with a reflection regarding what specific techniques could be applied in future aircraft performance tailoring efforts.

2.3.1. General Concepts

If we define AI as the science of mimicking human abilities, machine learning is the branch of AI that trains a machine how to learn. Indeed, machine learning is based on the idea that computers can learn from data and make decisions in real-time with minimal human intervention. Thanks to machine learning we benefit from email spam filters, reliable web-search engines and ambitious chess-playing programs, among others.

Nowadays, we have large amounts of structured and unstructured data which can be analyzed. Machine learning allows solving problems by capturing the knowledge in all these data to make predictions. Machine-learning algorithms consider a dataset with a specific number of samples and later try to predict properties of unknown data. Every data point constitutes one row of the dataset, it has its own features or attributes (independent variables, X), and it might have a label or target (dependent variable, y). A very interesting characteristic of machine learning is its capacity to evolve and adapt when new observations are added to the initial dataset.

Every machine-learning system is formed by four main phases [50]: preprocessing, learning, evaluation and prediction, as illustrated in Figure 2.8.

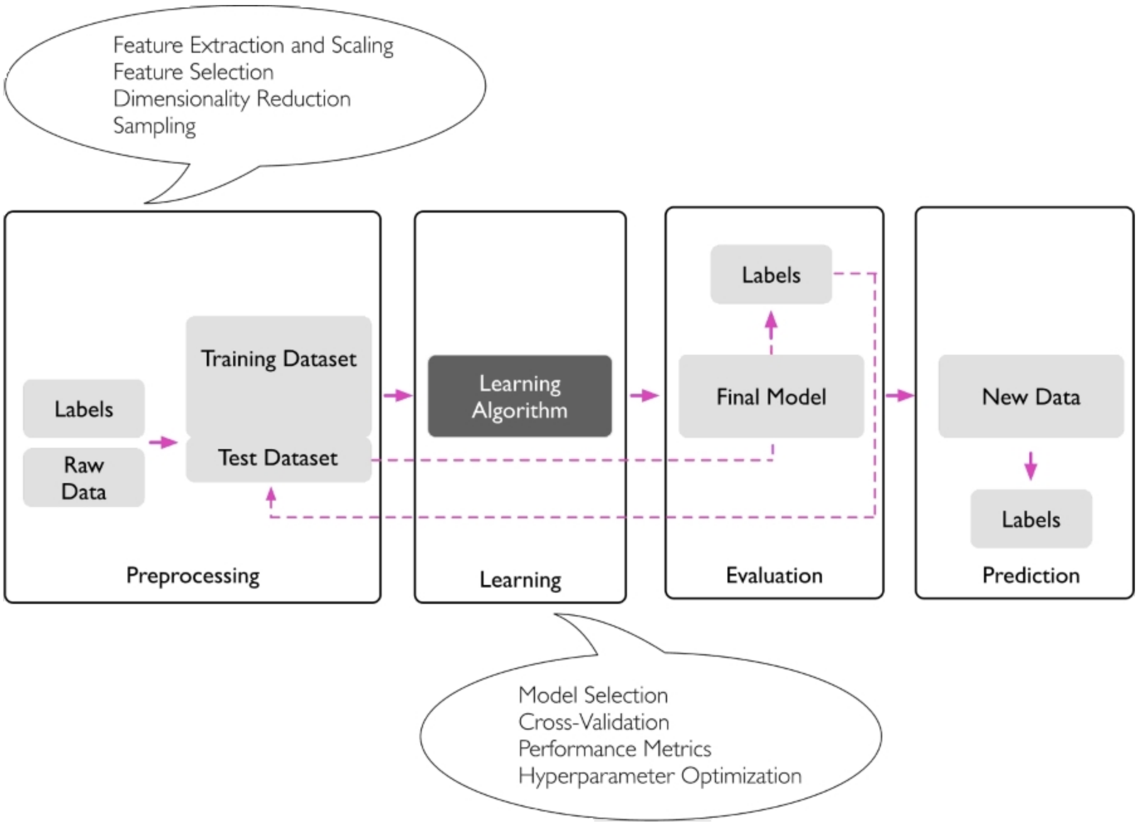


Figure 2.8: Typical workflow for using machine learning in predictive modeling [50]

2.3.1.1. Preprocessing

Data rarely come in a format and shape that a machine-learning algorithm can directly handle. Thus, data preprocessing is the first, and one of the most crucial steps in the creation of a machine-learning models. This phase starts by gathering (if not gathered yet) and cleaning data. As W. Koehrsen mentioned in one of his articles [51], about 80 % of the time spent in data analysis is retrieving and preparing data, so this first step must not be underestimated.

Often, data samples contain missing values that cannot be handled by a computer. Taking care of those missing values is necessary before continuing the analysis. Entries with missing values can be eliminated, but if no data wants to be lost, missing values can be estimated through interpolation. Another issue that can be encountered is related to the types of variables. Datasets are mostly formed by both continuous and categorical variables. Continuous variables are those that have an infinite number of possible values. These variables are not a problem for machine-learning algorithms. On the other hand, categorical or discrete variables are those that only can take a certain number of values (generally in *string* format). Each value represents a different category or group. The problem with categorical data is that most machine-learning algorithms cannot work with it directly, and require all input variables to be continuous. Within categorical data, it is possible to further distinguish between ordinal and nominal variables. In the first, values can be sorted (e.g. t-shirt sizes) but in the latter they cannot (e.g. t-shirt color). It is important to make this distinction because methods to handle categorical data depends on this differentiation. The values of ordinal variables can be transformed to integers, provided that the order of the labels and the distance between them are maintained. If the same approach is applied to nominal categorical data, machines would assume that a color is ”larger” or ”smaller” than other based on the mapped integers. To avoid this problem, one-hot encoding is generally used. This method creates one dummy feature for each unique value of the nominal variable and fills them with 0s and 1s accordingly.

Once data have been transformed to a format and shape understandable by a machine, the dataset is randomly subdivided in training and test set. The training set is used to train and optimize the model, and the test set is used at the end to validate the final model. It is important that the intersection of the training and test sets is empty for the algorithm to appropriately learn. In other words, none data point must be contained in both subsets. The same idea can be applied to human learning. For example, if a student is given some practice problems and their solutions to prepare for an exam, studying them will help him to do better. However, if he is given the exam problems he would memorize the solution, without properly learning the underlying principles [52].

Many machine-learning methods require the selected features to be in a similar scale to properly make predictions. Generally, feature scaling is necessary, and it can be done by means of normalization or standardization. If some of the features are highly correlated it means that one of them is duplicating information and therefore, it is not needed. Afterwards, the features meaningful for the predictions must be found. By applying dimensionality reduction techniques, the magnitude of the problem decreases, and so do required storage space and computing time. Feature selection and feature extraction are the main types of dimensionality reduction techniques. The first selects a subset of meaningful features, the latter constructs a new essential feature by extracting information from other variables. A well-known feature selection algorithm is Sequential Backward Selection (SBS), which reduces dimensionality based on a minimum detriment in performance of the algorithm to improve computational, and even predictive, efficiency.

2.3.1.2. Learning, Evaluation and Prediction

After the data set has been prepared it is time to choose the predicting model. It is important to select, train and compare several different algorithms to choose the one that gives more accurate predictions. Once the algorithm has been selected it should be fine-tuned, since its default hyperparameters may not give the most accurate predictions. The process by which the optimal values of the hyperparameters are selected is called model selection. The most common model-selection technique is the k-fold cross-validation, which requires to split the data into three subsets: a training, a validation and a test set. The first is used to fit the different models and the second evaluates the performance of each one, so it is used to select the final model. The test set, which has not seen the data during the training and model selection, is used to evaluate the performance of the chosen model. The working principle of the k-fold cross-validation technique can be seen in Figure 2.9. In k-fold cross-validation the training set of each different model is randomly splitted into *k* folds. *k*-1 folds are used for training and the remaining one is used for performance evaluation. This process is repeated *k* times, and then, the average performance of each different model is computed. This is illustrated in Figure 2.10. Finally, performance metrics of each model are compared to find the optimum hyperparameters.

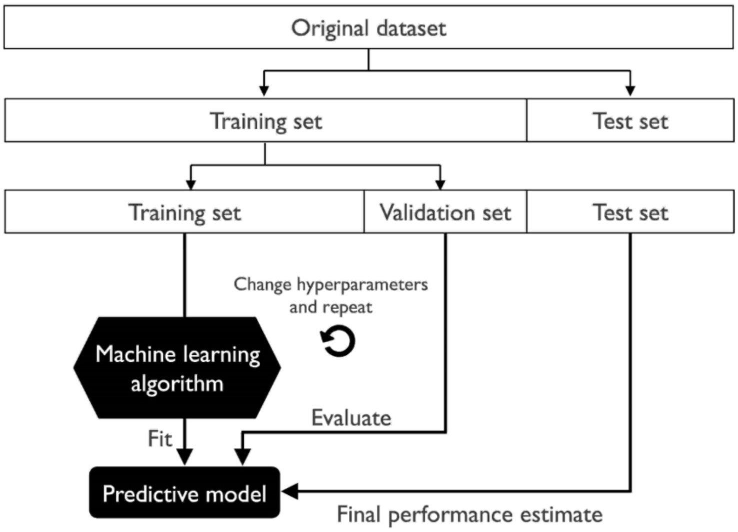


Figure 2.9: Typical workflow of k-fold cross-validation for model selection [50]

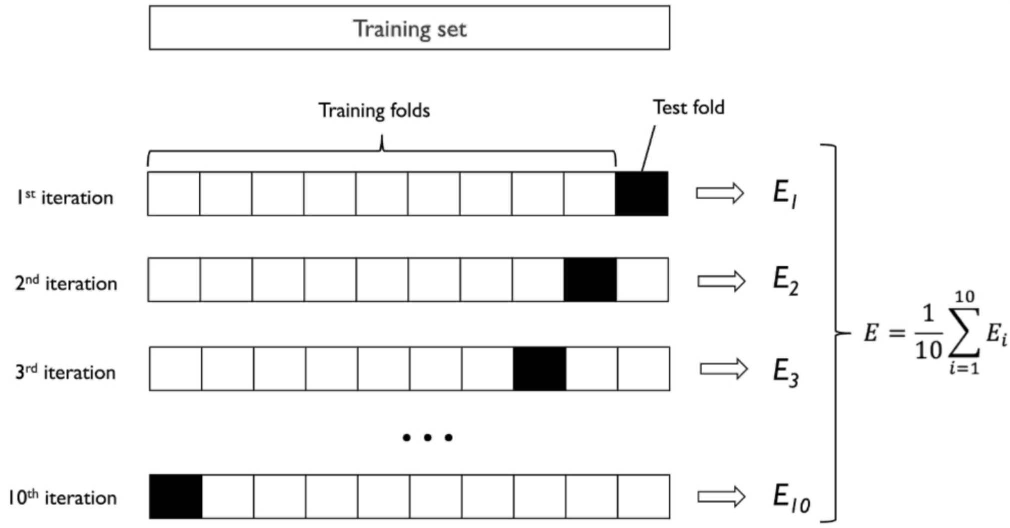


Figure 2.10: Concept behind k-fold cross-validation with k = 10 [50]

Once the final model has been selected it needs to be retrained using the entire training set. The reason for fitting the model to the whole training set after k-fold cross-validation is that training the model with more samples generally result in a more accurate and robust model. In order to obtain a final performance estimate, the test data that has not been seen by the model yet is used. To evaluate the performance of the machine-learning system the so-called performance metrics are used. Some examples of performance metrics are the Mean Absolute Error (MAE), the Mean Squared Error (MSE) or the Root Mean Square Error (RMSE). In addition, confidence and prediction intervals can be used to quantify the uncertainty of the model and of a particular prediction, respectively.

If after the evaluation phase the model proves to have an adequate level of accuracy, it can be used to make predictions on new data. If this is not the case, it is necessary to take a step backwards to analyze the problem, which would be either in the data or in the algorithm, or in both.

Several challenges are faced when developing a machine-learning system [53]. Those related with data can be due to insufficient quality or non-representative training data. If training data are scarce or full of outliers and noise, the system would hardly learn the underlying principles and would not be able to make accurate predictions. As a popular saying among machine-learning experts says: *garbage in, garbage out*. With respect to an incorrect performance of the algorithms, issues can be caused by the choice of irrelevant features, overfitting or underfitting. Overfitting occurs when the model performs well on training data, but cannot generalize. Which means that the model has learned patterns that take into account training data noise, and that cannot be extrapolated to new data. On the other hand, underfitting takes place when the model is too simple and cannot learn the underlying structure of the training data. In conclusion, developing a machine-learning system is not straightforward, going back to the preprocessing phase and iterating may be necessary.

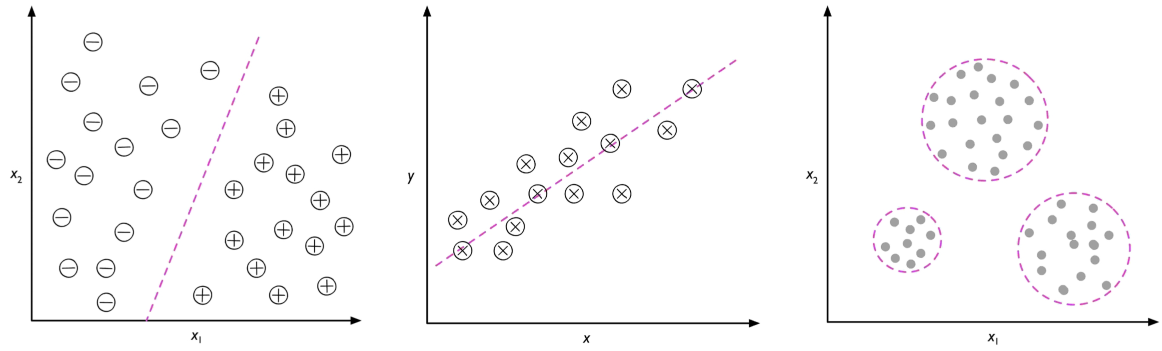
The methodology explained above includes the main steps to follow in the development of machine-learning models, but each problem is different. Some of them may require more intermediate steps, and some others may not demand all the steps described above. However, Figure 2.8 serves as a baseline for machine-learning model development.

2.3.2. Types of Algorithms

One advantage of machine learning is its capability of solving problems of different nature. The types of learning problems that can be addressed with machine learning are [50]:

- **Supervised learning**, in which the problem involves labeled data and a target variable to be predicted. At the same time, supervised-learning problems can be divided into:
  - **Classification**, used when samples belong to different classes and the goal is to predict the class of new, unknown data. In the example of email spam filtering, classification algorithms and a dataset of emails adequately marked as spam or not-spam can be used to predict to what category a new email belongs. The concept of classification problems is depicted in Figure 2.11a. Naive Bayes classifiers, Support Vector Machines (SVMs), K-Nearest Neighbor (k-NN) algorithms and decision trees are some examples of classification methods.
  - **Regression**, used when the goal is to predict the continuous value of the target variable corresponding to new, unknown data. Figure 2.11b shows the idea behind linear regressors. One example of regression problem is the prediction of the score of a student in an exam. Assuming that there is a relationship between grade and time spent studying, it is possible to use a regression algorithm to predict the exam grades of new students based on their study time. The most popular regression algorithms include linear and polynomial regressors, neural networks and regression trees.
- **Unsupervised learning**, in which data are unlabeled and no target variable is present. In this case, the goal of machine learning is to extract meaningful information from the data without guidance of known outcome variables. For example, by discovering groups of similar attributes within the data (also called clustering) or determining the data distribution (also known as density estimation).
  - **Clustering**, unsupervised classification technique that organizes observations into meaningful subgroups, called clusters, based on their degree of similarity, as illustrated in Figure 2.11c. For example, clustering allows market analysts to divide potential customers in groups based on their interests. Within the clustering methods, K-means clustering and hierarchical clustering standout [54].
  - **Dimensionality reduction**, subfield of unsupervised learning in charge of decreasing data dimensions to avoid issues related to limited storage space and computational time. Dimensionality reduction removes noise from data and hence, enhance predictive performance. The most popular dimensionality reduction algorithm is Principal Component Analysis (PCA).
- **Reinforcement learning**, whose goal is to develop a system or agent that improves its performance by interacting with the environment. This kind of learning involves a reward function, which measures the efficiency of the actions taken by the agent. The interaction with the environment allows the agent to improve its actions through trial-and-error or deliberative planning approaches. Reinforcement learning can also be considered a sub-class of supervised learning. A typical example of reinforcement learning are chess games, in which the agent decides its moves based on the status of the board. In this context, the board is the environment and a win at the end of the game is the reward.

A part from the capability of handling supervised, unsupervised and reinforcement algorithms, machine learning sustains ensemble methods, which combine different models into a single one that



(a) Classification (b) Regression (c) Clustering

Figure 2.11: Representation of classification, linear regression and clustering algorithms [50]

makes better predictions than each individual model alone. A lot of ensemble algorithms are based on plurality voting, meaning that the prediction given is the one that ”received more votes”. In other words, the one that was given by the higher number of models. Two examples of ensemble methods are random forest and boosting algorithms.

2.3.3. Machine Learning Applied to Aircraft Operation

Apart from the efforts that apply machine-learning techniques to model aircraft performance, mentioned in Sections 2.1.4.2 and 2.2.3 of this literature review, machine learning has been applied in aircraft operation with different purposes.

Aircraft departure and arrival management, conflict detection or flight planning depend on accurate trajectory predictions. As it has been mentioned throughout this report, some approaches to trajectory prediction are based on aircraft performance models like CTAS [16] [17], EJPM [24], TP model [26] [27] [28] or BADA [46] [47] [49]. However, in the literature there also are approaches to trajectory prediction based solely on machine learning. These approaches do not rely on a explicit modeling of the aircraft performance, procedures or airspace.

In [55], A. M. P. de Leege *et al.* propose a method to predict trajectory by solving a supervised-learning regression problem, through the use of surveillance and meteorological data, without explicitly modeling aircraft performance. Their ultimate goal is to make arrival time predictions and to determine the required initial spacing between aircraft for safe Continuous Descent Operations (CDOs). The input variables of the machine-learning model are the aircraft state parameters at the beginning of the CDO and the meteorological conditions. The output is the predicted time over some fixed points along the route. Three different regression techniques were considered: Generalized Linear Models (GLMs), SVMs and ANNs. The first technique is used to predict values that are supposed to be exponentially distributed. The second constructs a hyperplane, or set of hyperplanes, that separates the training set based on the labels and makes predictions accordingly. Lastly, ANNs build connections between units (or neurons) distributed in different layers to make predictions. From among the three techniques, GLMs are easier to implement and interpret. One interesting approach taken in this work is the way the impact of the input variables on the trajectory prediction performance is assessed. To make predictions four variable sets are used: aircraft parameters, meteorological information, altitude and wind information. First, predictions were made using aircraft parameters only, then the rest of the sets of variables were included one by one. Results proved that performance improves when new input variables are added to the model.

Also through the use of an ANNs and also for trajectory prediction, Xiangmin Guan *et al.* [56] aimed to develop a model capable of predicting 4D trajectories more accurately than the methods based on physical aircraft performance models do. Their model is trained using historical flight data and the current aircraft state information (*x,y,z,t*) is taken as input to predict the next point of the state information (*x’,y’,z’,t’*). A second model is presented in the paper which considers also aircraft velocity. Again taking more variables into consideration increases the accuracy of the prediction.

Another possible utilization of machine learning in aircraft operation is for the prediction of Estimated Time of Arrival (ETA). Punctuality of flights is affected by several factors. Some of them are controllable by the pilot, like flight level or aircraft speed, and some are not, like weather conditions and airport congestion. Traditional approaches to generate ETA predictions rely on aircraft performance models, but these do not always consider all influencing factors. To solve complex enigmas like this one, machine learning plays an important role. C. S. Kern *et al.* [57] presented a method to enhance aircraft ETA predictions with tree-based ensemble models, using information about the flight, weather and air traffic. In particular, utilizing random forests. These algorithms can be used for both regression and classification problems. They are ensembles of decision trees which can handle both numerical and categorical variables. In addition, they can rate features depending on their importance in the prediction of the output, easing the feature selection process. However, they tend to suffer from overfitting. When working with random trees it is recommendable to make a trade-off between the number of trees used and the computational cost associated to it. Another important advice given in [57] is that when using weather data to feed a machine-learning model it is important to consider a significant number of samples representative of each weather condition. Otherwise, the machine-learning model cannot adequately learn.

Again to predict ETA accurately, Zhengyi Wang *et al.* [58] developed a hybrid machine-learning model with the ultimate goal of improving trajectory prediction in the Terminal Manoeuvring Area (TMA). The method they propose contain a clustering-based preprocessing model that after the data has been cleaned, filtered and reduced, clusters the trajectories contained in the dataset into several patterns. The second part of the model consists of a Multi-Cells Neural Network (MCNN)-based prediction model. In which each cell is trained with an associated cluster of trajectories. In other words, each cluster of trajectories has its own prediction model. Therefore, when a prediction on new data needs to be made the new trajectory is first allocated to one of the clusters using a tree-based classifier, so that the correct prediction model can be used. The approach followed is depicted in Figure 2.12. They demonstrated that the addition of the clustering-based preprocessing model to the prediction model decreases prediction errors.

As becomes obvious, machine learning can be really helpful in addressing nowadays aviation challenges. The studies mentioned above are examples that show the usefulness of machine learning to answer problems that otherwise are puzzles hard to solve. An extensive knowledge on machine learning is a must for this research project. Since it involves the creation of create machine-learning systems, using datasets of various shape and format, applying different techniques and algorithms, and within reasonable memory and computing requirements.

2.4. Conclusion

This literature review has shown that aircraft performance plays a key role in aeronautics, being a determinant factor in aircraft design, certification and operation. Aircraft performance is of special interest to airlines, since it dictates aircraft operation strategies. Aircraft performance monitoring activities are carried out by airlines to ensure optimal use of their resources and maximize profit. In order to study the performance of their fleet, airlines rely on aircraft performance models. The modeling of aircraft performance was found to be of interest for the entire aircraft operation community, shown by the large number of research and development efforts devoted to it. Aircraft performance modeling is present in activities such as post-flight analytics, trajectory prediction, flight plan generation, air traffic management, and aircraft conflict detection and resolution, among others.

As it has been shown in this literature study, existing aircraft performance models represent average performance of brand-new aircraft, and they differentiate performance only according to aircraft type. In addition, most models do not include external factors that affect aircraft performance, such as the weather conditions or the airspace flown. This is the case of BADA, the most widely used aircraft performance model nowadays. BADA demonstrated to be useful specially for research purposes, but it has some shortcomings that complicate its use outside a research framework, where aircraft type may not be specific enough to classify an aircraft and an average aircraft performance may not be sufficiently accurate.

Multiple research efforts were found in the literature whose main goal was either to enhance BADA capabilities or to develop new performance models, always seeking for more accurate results. To pursue these objectives different techniques were used, classifiable into numerical and machine-learning techniques. These efforts, as well as BADA, aimed to model average performance of aircraft types. However, aircraft performance is closely related to aging, maintenance variations, user-configurable specifications, specific aircraft degradation, etc. All these factors are tail-number dependent, and must

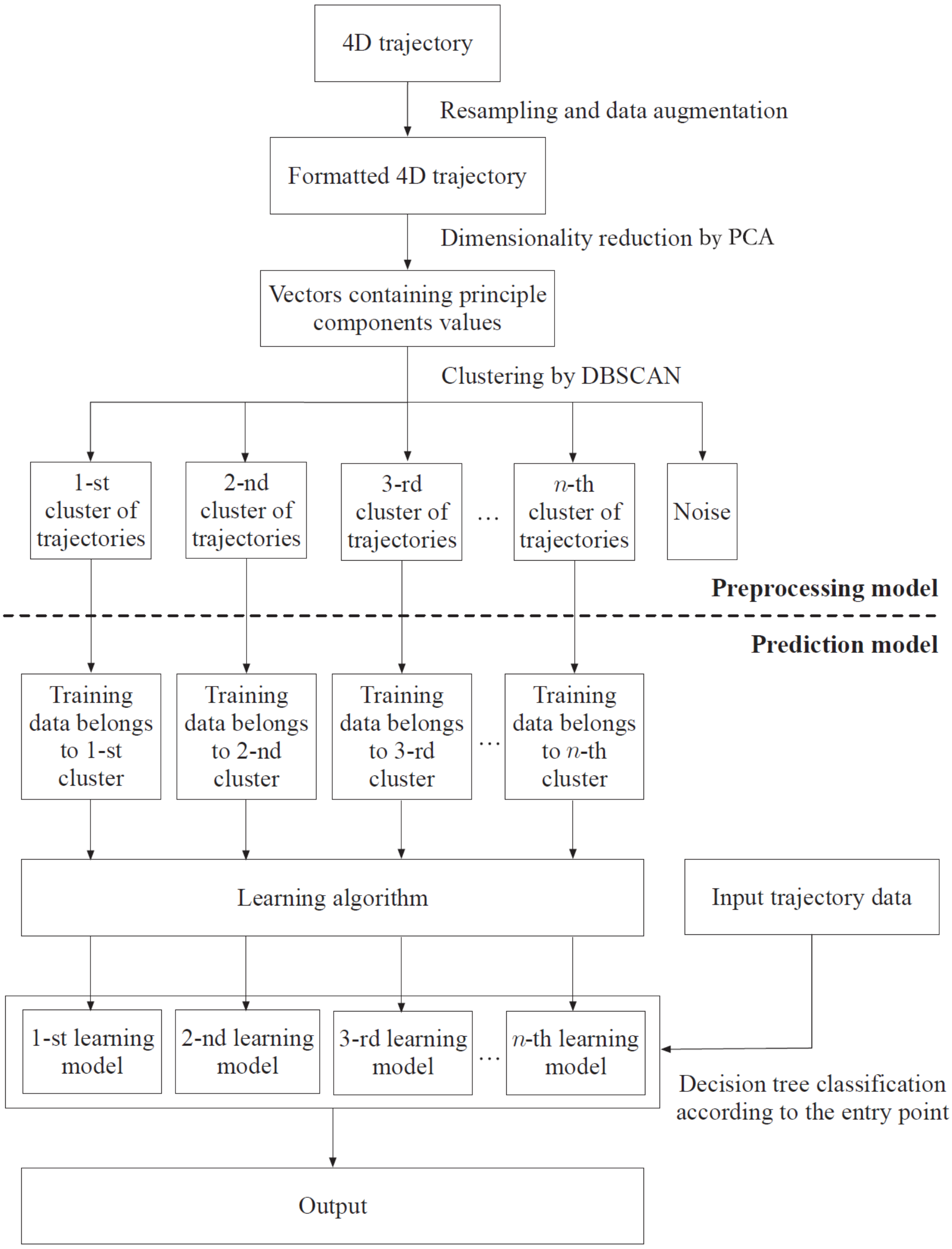


Figure 2.12: Approach proposed by Zhengyi Wang *et al.* in [58] to predict ETA using a hybrid machine-learning model

be considered in aircraft performance modeling. Otherwise, models do not demonstrate good enough agreement with the true performance of the aircraft. Finding exact solutions to certain performance problems, such as the estimation of fuel consumption, can be a real challenge. To face this challenge tailored performance models must be achieved. With regard to the development of tail-specific performance models, hybrid and non-physical approaches can be chosen. Both have their own advantages and disadvantages, but both seem suitable for the job. Even though some work towards aircraft performance tailoring has been done, there are research lines that have not been explored yet. In addition, no work proposing a complete methodology to develop tail-specific performance models is available in the literature so far.

Machine learning demonstrated to have a key role in aircraft performance modeling and tailoring. Two main learning problems can be addressed with machine learning: supervised and unsupervised learning. While supervised techniques such as neural networks can be applied to predict aircraft performance, unsupervised algorithms such as clustering methods can be used to segment flights in different phases. Aircraft perform differently depending on the phase of flight, so flight segmentation is key in aircraft performance modeling.

3
Research Definition

Based on the literature review presented in the previous chapter, there seem to be certain research lines that have not been explored yet, as well as some aspects that have not been considered in previously proposed methodologies. The identified research gaps, as well as the formulated research goal and questions, are detailed in this chapter.

3.1. Research Gap

The research gaps identified from the literature review are:

1. From current literature it can be concluded that models calibrated with actual flight data are more accurate than models calibrated with manufacturers' performance data only. In fact, the utilization of actual flight data is a requirement in aircraft performance tailoring. From this research study it can also be affirmed that machine learning is a great partner to model aircraft performance. Until now, two approaches have been followed to pursue the goal of modeling tail-specific performance. On one hand, the fine-tuning of physical models like BADA. In this respect, there is no methodology to correct BADA models using **machine-learning techniques**. On the other hand, the formulation of purely data-driven performance models. Therein, there is no methodology to develop entirely non-physical tail-specific performance models **applicable on-ground**, or usable to accurately predict aircraft performance in "what-if" scenarios accurately.

2. Even though some work has been done regarding aircraft performance tailoring with the aid of recorded flight data and machine-learning techniques, there currently is no methodology that gives the **preprocessing phase** the importance it deserves. Recorded flight data will likely contain missing values, noise and inconsistencies. Therefore, a clear procedure must be established to address all these issues. In addition, data will likely enclose a high number of variables, so a consistent feature selection process must be included in the methodology. As well as feature scaling and normalization practices, since the range of values of flight variables will vary widely and will be different from variable to variable. For example, one will find Mach numbers around 0.8 and flight altitudes around several thousand feet. Furthermore, one advantage of using machine learning in performance modeling is the consideration of the information provided by non-physical variables, for which further actions are needed. These variables will mostly be categorical, so they will have to be transformed appropriately.

3. A good performance model must be able to support accurate calculations during all phases of flight, and it should also be valid over the **entire flight envelope**. The use of recorded flight data in aircraft performance modeling has one major disadvantage as compared to the use of manufacturers' performance data gathered during flight testing: reference data will unlikely cover the entire flight envelope. The limited amount of flight data corresponding to flight conditions away from typical operation values is a challenge to confront when willing to develop a performance model valid for the entire envelope. There currently is no methodology that accounts for this dataset imbalance. Machine-learning systems require a dataset representative of the problem to

be solved. The model will not be able to capture the relationships that exist between input and output features if there is not enough representative data samples. One potential solution to this problem may be the generation of artificial data to populate the areas of the flight envelope that are underrepresented in the historical flight dataset.

4. The need for a tailored aircraft performance model arises from the inability of nominal performance models like BADA to account for performance degradation effects. Performance degradation may be due to airframe deterioration, engine degeneration or both, but in any case, it changes with time. Therefore, it is critical that tail-specific performance models are essentially dynamic. Nowadays, there is no methodology that includes an **updating mechanism**, which guarantees that the resulting model reflects changes of performance with time. In the case of models developed with the aid of machine learning it is important that the database remains updated. A sensitivity analysis on how often the algorithms must be retrained may be required.

After a methodology to model aircraft tail-specific performance taking into consideration these research gaps is created, it can be applied to pre-, in- and post-flight applications such as flight planning, aircraft performance monitoring, fuel analytics, air traffic simulation, aircraft conflict detection, etc. It might be interesting to implement the created methodology on a tool designed for any of these purposes, and to evaluate the outcomes.

3.2. Research Objective

The project goal, based on the identified research gaps, is to develop a complete methodology to model tail-specific performance by using historical flight data and machine-learning techniques, such that it can be applied in pre-, in- and post-flight applications. In order to achieve this objective, the following sub-goals will mainly be pursued:

- The creation of a hybrid approach to calibrate BADA to particular tail numbers, based on the application of machine-learning algorithms.
- The development of a purely data-driven approach capable of modeling the performance of specific tail numbers accurately, through the use of machine-learning techniques.
- The formulation of a preprocessing procedure required in order to use flight data in performance modeling, including the necessary preparation, cleaning, normalization, encoding, etc.
- The assessment of performance degradation with time, in order to establish an updating mechanism that guarantees accordance between modeled and real performance.

3.3. Research Question

To fill the research gaps and meet the project goal presented above, the following research question is proposed as the main research question for the thesis:

> *How can aircraft tail-specific performance be modeled by using historical recorded flight data together with machine-learning techniques, in order to enhance aircraft trajectory simulations by considering aircraft performance degradation?*

This research question can be broken into four different sub-questions:

1. **What data preparation is required in order to successfully use historical flight data and machine-learning techniques in aircraft performance tailoring activities?**
   A popular saying among data scientists says: *garbage in, garbage out*. Data preprocessing is key in the successful development of data-driven models, and this research is an exception. It is necessary to investigate if the flight data contains enough parameters to carry out the research plan. In other words, if the available flight data are representative of the problem or if it should be complemented with other data sources. For example, with artificially-generated performance data in order to cover the entire flight envelope.

2. **Given certain historical flight data of a tail number, how can tail-specific hybrid performance models be developed?**
   The main advantage of hybrid models is the transparency of their outcomes. Since they are based on the physical laws governing the motion of the aircraft, their results are easier to interpret. This part of the research project consists on designing a methodology to tailor BADA performance models. In particular, it is investigated how can machine-learning techniques be applied to identify tail-specific BADA coefficients. For which it is necessary to gain extensive knowledge of BADA performance models.

3. **Given certain historical flight data of a tail number, how can tail-specific purely-data driven performance models be developed?**
   The main advantage of non-physical, purely data-driven models is that they are built on weak or even without assumptions, and they can consider non-physical attributes that may influence aircraft performance. In order to develop tail-specific non-physical models, it is necessary to familiarize with the appropriate machine-learning methods, to guarantee that the more suitable algorithms are selected, as well as the optimum sets of hyperparameters and features.

4. **When should tail-specific performance models be updated in order to properly reflect variations of aircraft performance with time?**
   Once the factors that provoke inaccuracies in aircraft performance modeling have been identified, addressed and modeled, how their influence vary with time should be investigated. Tail-specific performance models must be dynamic to reflect aircraft performance degradation, maintenance actions, etc. An updating mechanism should be defined and included to the developed methodologies.

4

Hybrid Aircraft Performance Tailoring

The first part of the research on hands consists of a hybrid approach to tailor aircraft performance. The result of this approach are tailored models that are based on existing physical models, but that are adjusted using historical flight data. Specifically, a methodology to calibrate BADA 4 models (hereinafter referred to as BADA models) based on QAR data has been created. This chapter starts with a brief description of QAR data, and the datasets available to carry out this research work. Section 4.2 studies to what extend BADA models are able to realistically capture aircraft performance, and it sustains the motivation of the research. The core of the chapter is Section 4.3, which describes the developed methodology, including all the necessary steps to obtain accurate and robust performance models. Section 4.4 discusses the results obtained after running three different tailoring processes with different datasets. Finally, Section 4.5 brings together the main conclusions on the developed methodology and gotten results.

4.1. Quick Access Recorder (QAR) Data

Airlines have been collecting flight data for several years to monitor their fleet performance. Originally data retrieval was manually done by a dedicated staff member in the cockpit or by one of the pilots. However, this way of gathering data is costly and cannot be used in routine aircraft performance monitoring. For this purpose, the most common procedure is automatic recording of in-flight data, done by data recorders onboard the aircraft. Flight Data Recorders (FDRs) are known for being critical in accident and incident investigations. Due to the requirement that they must survive an accident, accessing FDRs is difficult. Instead, QARs are generally used in aircraft performance monitoring. These devices save data observations as snapshots of the aircraft condition multiple times per second and, in contrast to FDRs, are easily removable. They are designed to provide quick and easy access to raw flight data. After landing, QAR data collected during flight are compressed, encrypted and stored in a ground database. Later, QAR data are analyzed to study performance deviations from book levels.

QARs can record a large amount of aircraft flight parameters. Some parameters are mandatory and should be recorded regardless airplane complexity. Others are configured by the airline provided they are available in the ARINC network. QAR data contain parameters required to determine aircraft flight path, speed, attitude, engine power, configuration of lift and drag devices, etc. It can include information regarding very different aircraft systems, such as flight controls, landing gear, bleed system, hydraulics, navigation, communication and environmental controls. The vast amount of flight information recorded and the fast access to the data are two advantages of QARs. However, QAR data are owned by airlines and therefore, not easily accessible.

Since this research work was conducted in collaboration with Boeing Research & Technology - Europe, the access to three datasets of QAR data was granted. The first is composed by more than 6,500 flights of a narrow-body airplane. The flights available correspond to approximately four years of operation, and are short-haul flights. The second dataset includes over 13,000 flights flown by 88 narrow-body airplanes of the same airline. These flights took place in the same month of two consecutive years, and are short- and medium-haul flights. Last but not least, the third dataset used is formed by approximately 1,000 flights of a wide-body airplane. The flights of this dataset correspond to slightly more than one year of operation, and are long-haul flights.

Before beginning the tailoring of aircraft performance models, correlation analyses were carried out to pre-select the most relevant flight parameters among the high number of parameters available in QAR data. Specifically, the correlation of flight parameters with fuel flow was computed using:

- **Pearson correlation coefficient (R)**: parametric measure of linear correlation between two variables.
$$R = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}} \tag{4.1}$$
Where n is the number of observations, and X and Y are the variables between which correlation is computed.

- **Kendall tau correlation coefficient (τ)**: non-parametric measure of ordinal association between two variables. Kendall's correlation coefficient uses pairs of observations and determines the strength of association, based on concordance and discordance between the pairs. A pair of observations is concordant if they are ordered in the same way. If they are ordered differently they are considered discordant.
$$\tau = \frac{n_c - n_d}{\frac{n(n-1)}{2}} \tag{4.2}$$
Where n is the number of observations, and n_c and n_d are the number of concordant and discordant pairs, respectively.

- **Spearman rank coefficient (ρ)**: non-parametric measure of how well the relationship between two variables can be described by a monotonic function.
$$\rho = \frac{cov(rg_X, rg_Y)}{\sigma_{rg_X}\sigma_{rg_Y}} \tag{4.3}$$
Where $cov(rg_X, rg_Y)$ denotes the covariance of the rank variables, and σ_{rg_X} and σ_{rg_Y} are the standard deviations.

The main variables pre-selected after studying their correlation with fuel flow refer to the following parameters, ordered alphabetically: acceleration, aircraft mass, airspeed, altitude, control surfaces, flight angles, high-lift devices, landing gear, temperature, wind conditions, and obviously engine variables such as throttle levers, N1, N2, exhaust gas temperature, etc. Besides these flight parameters, some non-physical variables were considered important: flight date, flight number, departure and destination airports, etc. Correlation analyses reduced datasets to manageable sizes and hugely helped with the familiarization of QAR data.

4.2. BADA vs. Real Performance

As mentioned above, calibrating BADA nominal models is the focus of this part of the research. Information on BADA models can be found in Section 2.1.2 and Appendix A. The latter provides an overview of BADA 4 models; including their modeling principles and assumptions, their formulation, and a detailed description of the models of interest and how to use them. This section is dedicated to numerically address the accuracy of BADA models. To do so, fuel flow given by these models has been computed and compared with fuel flow recorded in QAR data.

Fuel flow is modeled by BADA differently depending on the engine rate that is being used to fly the aircraft. Thus, one can generalize that different BADA models must be used to calculate fuel flow in different flight phases. In the climb phase, the general thrust model (A.2.3) together with the general fuel consumption model (A.2.4) are required. In cruise, thrust is modeled by means of the drag model (A.2.2), to be used before the general fuel consumption model. In descent, it is sufficient to apply the idle fuel consumption model.

The following error metrics are used to evaluate the performance of BADA nominal models:

- **Mean Absolute Error (MAE)**: average of the absolute differences between prediction and actual observation, where all individual differences have equal weight.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |\hat{y}_i - y_i| \tag{4.4}$$

- **Mean Absolute Percentage Error (MAPE)**: average absolute percent error for each prediction minus observation, divided by observation.

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n} \frac{|\hat{y}_i - y_i|}{y_i} \tag{4.5}$$

- **Root Mean Squared Error (RMSE)**: square root of the average of squared differences between prediction and actual observation.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (\hat{y}_i - y_i)^2} \tag{4.6}$$

- **Pearson correlation or correlation coefficient (R)**: measure of linear correlation between two variables. It has a value between +1 and -1, where +1 is total positive correlation, 0 is no correlation, and -1 is total negative correlation.

$$R = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{\hat{y}}_i)(y_i - \bar{y}_i)}{\sqrt{\sum_{i=1}^{n} (\hat{y}_i - \bar{\hat{y}}_i)^2}\sqrt{\sum_{i=1}^{n} (y_i - \bar{y}_i)^2}} \tag{4.7}$$

QAR data corresponding to 235 flights of a single tail number have been used to determine to what extend BADA nominal models are able to capture the real performance of the aircraft. This airplane is a short- to medium-range twin-jet airliner, and the 235 flights considered represent 5 months of operation. After calculating fuel flow modeled by BADA and comparing it with fuel flow recorded in the QAR data, results shown in Table 4.1 and Figure 4.1 were obtained. Note that the values of x- and y-axes have been omitted in most of the figures hereinafter in the report, due to confidentiality reasons.

Table 4.1: Estimation errors in fuel flow of BADA nominal models in (a) Climb, (b) Cruise and (c) Descent

	MAE [kg/h]	MAPE [%]	RMSE [kg/h]	R [-]
Climb	190.47	4.40	231.88	0.986
Cruise	95.11	4.71	107.60	0.907
Descent	35.20	7.99	48.67	0.968

As it can be seen, the gap between modeled and actual fuel flow is not negligible. In terms of fuel consumption, BADA nominal models typically underestimate the amount of fuel mass burnt. For the QAR dataset considered in this analysis, the MAE between real and modeled fuel consumption per hour of flight is 84.55 kg/h, and the MAPE is 3.48 %. Figure 4.2 depicts the distribution of these error metrics over the dataset. For most flights the error in fuel consumption per hour of flight is between 50 and 120 kg/h, and there are a few flights for which this error is greater than 200 kg/h. Over the 5 months of operation of the studied tail number, the amount of fuel misestimated by BADA sums more than 15 tons, worth approximately 14,500 USD [59]. Clearly, BADA is not able to accurately represent the real behavior of the tail number under consideration. Hence, the use of BADA nominal models in pre-, in- and post-flight applications would lead the airline to erroneous planning, optimization and analytic activities.

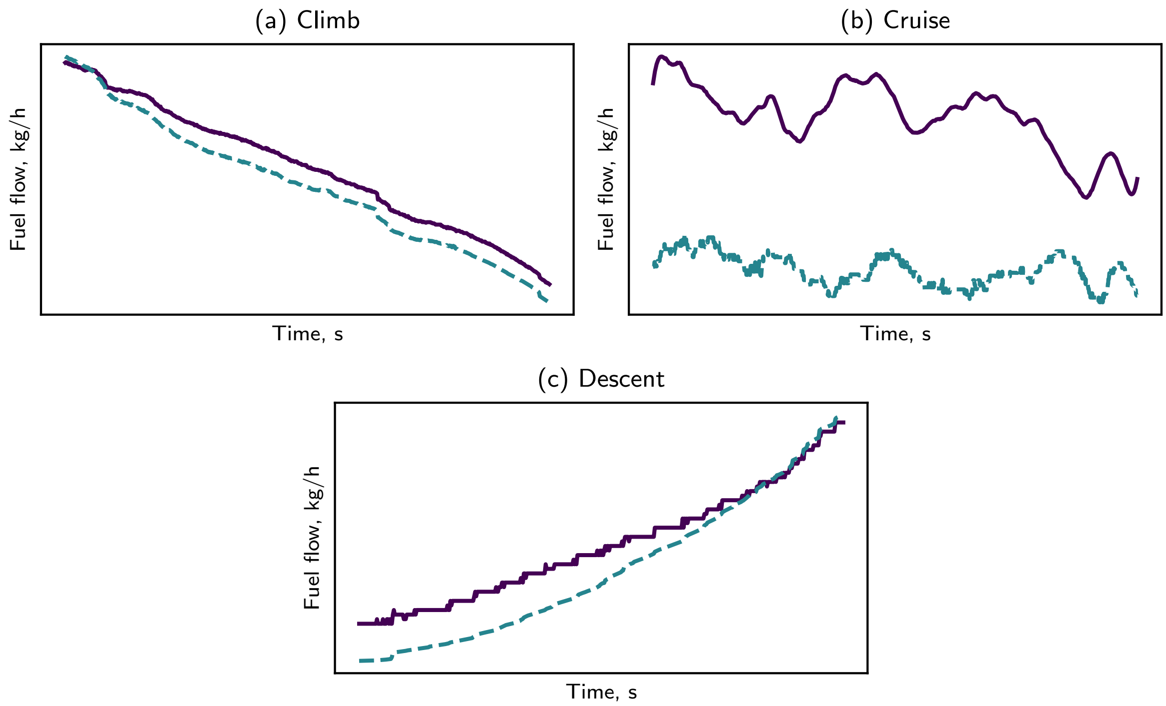


Figure 4.1: Evolution of fuel flow during a flight according to BADA nominal models (– – –) with respect to the actual fuel flow (——) in climb, cruise and descent

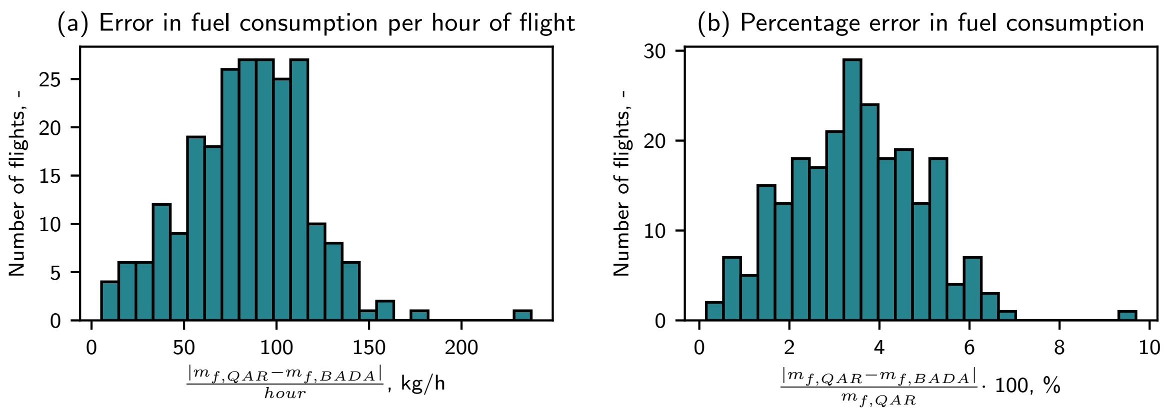


Figure 4.2: Error in fuel consumption per hour of flight and percentage error between the real fuel consumption and the fuel consumption given by BADA nominal models

4.3. Methodology

For this part of the research project, a methodology has been developed to identify all BADA coefficients using QAR data as reference performance data together with machine-learning techniques. It is anticipated that the calibrated models will provide more realistic results, thus enriching the applications for which aircraft type is not specific enough to classify an aircraft, and an average aircraft performance is not sufficiently accurate. This section includes a description of the developed methodology, which consists of four phases: data ingestion, data preparation, tailoring process and model evaluation. Figure 4.3 summarizes the main steps required to calibrate BADA nominal models, addressed one by one in this chapter.

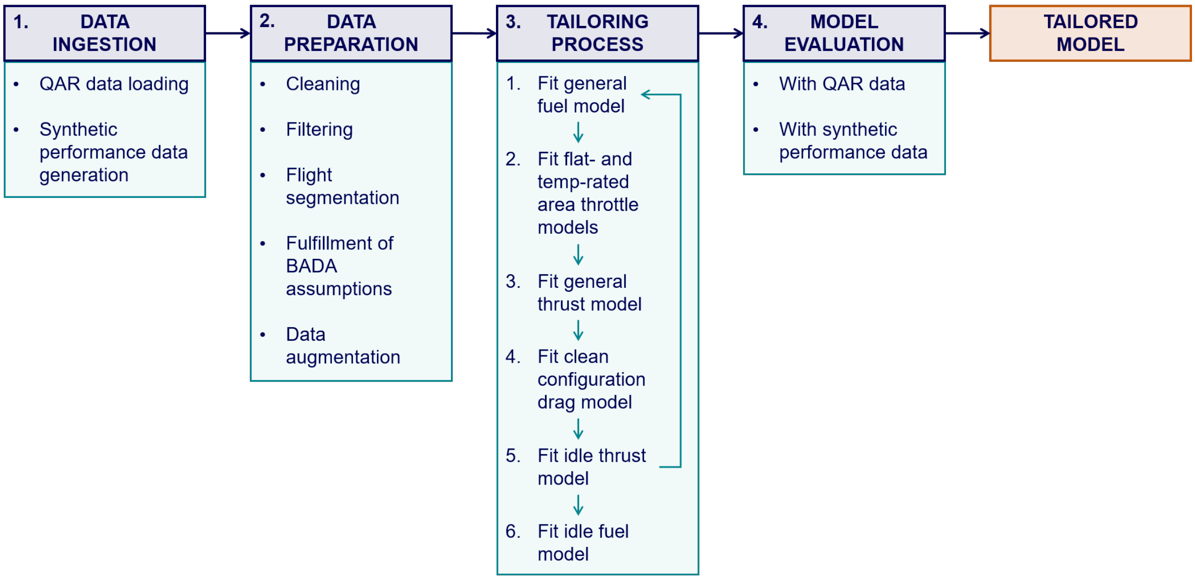


Figure 4.3: Schema of the methodology developed to calibrate BADA coefficients, including the main steps: data ingestion, data preparation, tailoring process and model evaluation, as well as some intermediate steps

4.3.1. Data Ingestion

Gathering data is the first step in every data-driven activity. It is key to create accurate predictive models, to successfully perform real-time optimizations and to obtain relevant conclusions from post-analyses. None of these activities are possible with low-quality or scarce data. In this case, the data ingestion phase includes the loading of QAR data, the main source of reference performance data, and the generation of synthetic performance data. These are two important steps, since the quality of the resulting calibrated models will rely on the coverage, precision and granularity of the reference performance data.

Typically, QARs record thousands of variables from which only dozens might be of interest for the problem addressed here. Therefore, it is desirable to study the available variables before loading the QAR data. As a starting point, hundreds of variables were preselected by computing the correlation coefficients between the features available in the QAR data and the target feature, which in this case is fuel flow. Then the final list was created, considering the variables required for the preparation phase, and the variables required for the tailoring process. The former are not fixed and depend on the dataset, and the latter are: pressure altitude, aircraft gross weight, total air temperature, Mach number, flight path angle, and obviously, fuel flow.

Airlines generally fly trajectories that correspond to an efficient operation of their fleet. As a consequence, QAR data tend to agglomerate in a narrow region of the complete flight envelope in which the aircraft can safely fly. Due to the lack of QAR data in those regions of the flight envelope, it is necessary to include synthetic flight points to the tailoring process. In addition, the utilization of synthetic data for validation purposes proved to be critical, since otherwise performance of calibrated models over the regions where there is not QAR data cannot be evaluated.

To create synthetic data one can use performance tables provided by the manufacturers, performance software or an aircraft performance model. In this research, BADA nominal models were utilized. BADA limitations model [60], which restricts the aircraft behavior to keep it between certain limits to ensure the safe operation of the aircraft, plays a key role in the generation of synthetic data. The limitations model is divided into five types of limitations: geometric, kinematic, buffet, dynamic and environmental. The geometric-limitations model provides the maximum geopotential pressure altitude for which the aircraft is certified. The kinematic-limitations model gives the maximum possible calibrated airspeed and Mach number. The buffet-limitations model computes the maximum lift coefficient at which the aircraft can operate safely based on the aerodynamic configuration. The dynamic-limitations model provides the minimum and maximum allowed weights, and lastly, the environmental-limitations model determines the maximum and minimum possible temperature deviations as a function of geopotential pressure altitude. Taking into consideration all these constraints, a dataset composed of synthetic flight points was generated.

4.3.2. Data Preparation

After the required data are gathered, it must be prepared for the tailoring process. QAR data commonly contains outliers and noise due to measurement errors and inaccuracies. To overcome this drawback, the developed methodology includes cleaning and filtering techniques. First, flights with erroneous signals are eliminated, typically those with inconsistent fuel flow information. Then, outliers in multiple variables are eliminated, based on the following criteria:

$$p_{outlier} < Q_1 - 1.5 \cdot IQR \quad (4.8)$$

$$p_{outlier} > Q_3 + 1.5 \cdot IQR \quad (4.9)$$

Where Q_1 and Q_3 are the first and third quartiles, and IQR is the interquartile range ($IQR = Q_3 - Q_1$). Savitzky-Golay filters are employed to smooth the noise present in QAR data. These filters smooth data without distorting the signal tendency, which is achieved by fitting subsets of contiguous data points with low-degree polynomials using linear least squares, process known by convolution. Figure 4.4 is an example of application of the explained techniques.

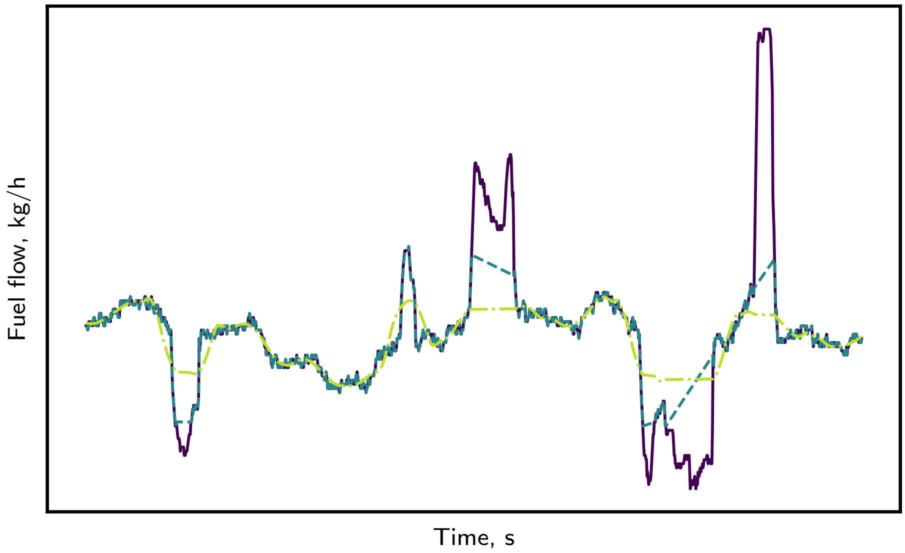


Figure 4.4: Evolution of fuel flow during a cruise segment of a flight, before any preprocessing (——), after deleting the outliers (– –) and after applying a Savitzky-Golay filter (–·–)

Before starting the tailoring process, it is necessary to segment flights appropriately, since different BADA models are applicable in climb, cruise and descent regimes. In addition, climb points are classified in flat-rated-area and temperature-rated-area points. As detailed in Section A.1, BADA models are built on multiple assumptions. To be faithful to BADA modeling principles, the preparation phase includes the necessary steps to guarantee that the calibration datasets fulfill as many assumptions as possible. These steps depend on the variables available in the QAR data. Currently, the calibration dataset is selected as follows:

- Only those flight points in which the aircraft is flying with clean configuration are considered. In other words, BADA drag coefficients in non-clean configuration are not calibrated.
- Only those flight points in which there is no derate and no assumed temperature are selected, to guarantee that those flight points at which the aircraft is flying with reduced thrust settings are not taken into account.
- Only those flight points in which no extra engine air bleeding is identified are used. Such identification is done based on bleed valves, anti-ice and packs-on information. More aspects could be considered if the necessary variables are available in the dataset.

- For simplification purposes, only those climb segments in which the aircraft is flying at MCMB are considered. In other words, BADA flat- and temperature-rated area throttle coefficients in MCRZ and MTKF engine ratings are not calibrated.
- Only those cruise segments in which the aircraft is flying at constant altitude are taken into account. If the aircraft flies at various flight levels during cruise the ascending segments are discarded.
- Only those flight segments in which the aircraft flies at constant airspeed are selected, so that thrust and drag forces are comparable ($T \approx D$).
- Only those flight points at which the aircraft flies at LIDL are used, since this is the only engine rating modeled by BADA for descent trajectories.

Last but not least, the preparation phase includes the augmentation of the QAR datasets with synthetic performance data. If the tailoring process is run with QAR data only, the resulting models are unstable when flight conditions approach the limits of the flight envelope, where QAR data are rarely found. The solution to reduce the instability of the models is to include synthetic points. The amount of synthetic points that should be added can defined as a percentage, with respect to the total amount of flight points in the QAR data. Among the added synthetic points, those points at the vertices of the flight envelope must be included. Furthermore, the set of added synthetic points must include points on the hyperplanes that define the flight envelope, to ensure that the behavior of the calibrated polynomials is constraint and valid over the complete flight envelope. Additionally, some of the added synthetic points should be uniformly-distributed over the regions of the flight envelope where there are not real flight points. By including synthetic points on areas that are not at all represented in the training QAR data, tailored models will represent aircraft performance at those flight conditions more realistically. As an example, Figure 4.5 depicts the distribution of climb points over the flight envelope after a QAR dataset was complemented with synthetic data. Specifically, the figure is a representation of a kernel-density estimate using Gaussian kernels, a way to estimate the probability density function (PDF) of multivariate data. Warm colors indicate higher density of flight points. Note that the color map is specific to each subplot.

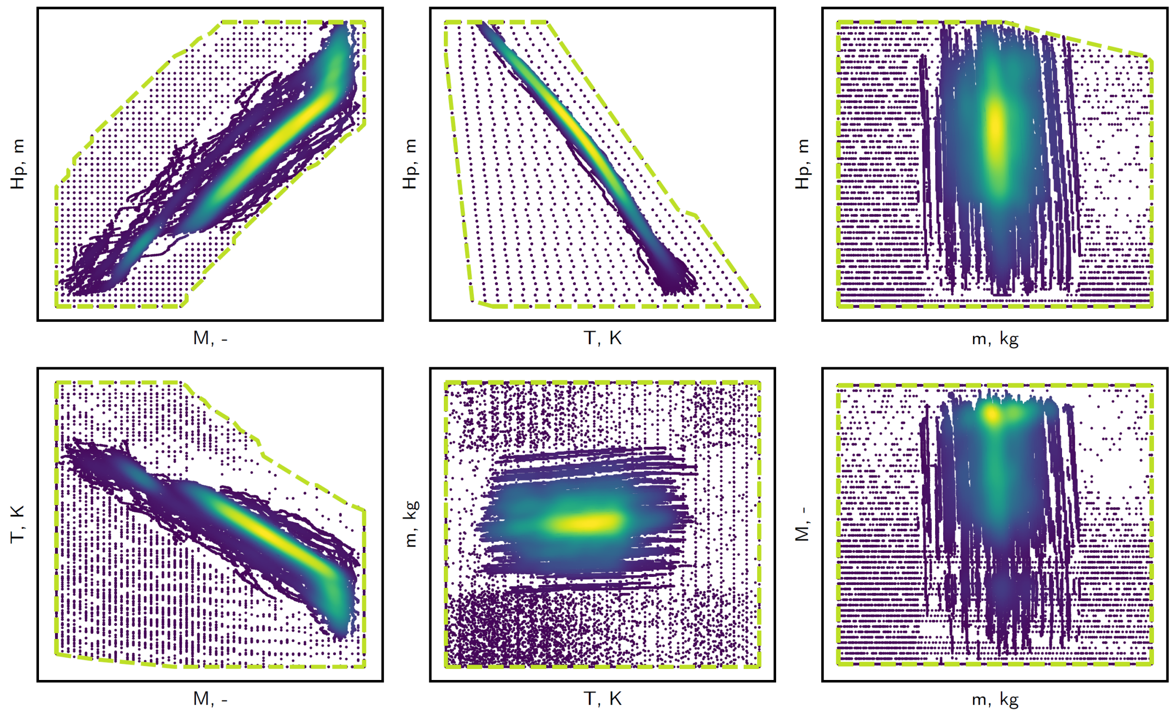


Figure 4.5: Kernel-density estimate of QAR climb data augmented with synthetic performance flight data, distributed over the operational envelope of the corresponding aircraft type (– –)

4.3.3. Tailoring Process

Once the necessary data are gathered and prepared, the tailoring process can be started. The purpose of the tailoring stage is to identify tail-specific BADA coefficients. In other words, the set of BADA coefficients that best describe the tail number represented in the QAR data. Limiting to turbofan models, considering clean configuration, MCMB ascents, cruise and LOW IDLE descent trajectories, seven different sets of coefficients are contemplated and adjusted:

1. General thrust coefficients (a_1 to a_{36})
2. MCMB Flat-rated area throttle coefficients (b_1 to b_{36})
3. MCMB Temperature-rated area throttle coefficients (c_1 to c_{45})
4. Clean configuration drag coefficients (d_1 to d_{15})
5. General fuel coefficients (f_1 to f_{25})
6. LIDL thrust coefficients (ti_1 to ti_{12})
7. LIDL fuel coefficients (fi_1 to fi_9)

The tailoring process consists of a fitting scheme through which dynamic information like thrust and drag forces are obtained from QAR data that contain kinematic information only. Specifically, reference performance data must include information regarding geopotential pressure altitude (from which ROCD is derived), fuel consumption, initial aircraft mass, speed profile and atmospheric conditions. These variables are related through BADA actions models, described in Section A.2, and BADA motion model, presented in Section A.3. In essence, the tailoring process aims to find those BADA coefficients that provide the best possible fit between real and modeled fuel flow and ROCD.

The identification of tail-specific coefficients is done by means of multivariate linear regression techniques, useful when various independent variables (or features) contribute to the dependent or target variable, and also when the regression coefficients are to be known. Multiple linear regressions have the following form:

$$Y_i = \beta_0 + \beta_1 x_i{}^{(1)} + \beta_2 x_i{}^{(2)} + ... + \beta_n x_i{}^{(n)} \tag{4.10}$$

Where Y_i is the estimate of the i^{th} component of the dependent variable, n is the number of independent variables, $x_i{}^j$ denotes the i^{th} of the j^{th} independent variable, and β_n are the regression coefficients.

The calibration model consists of seven multivariate linear regression models corresponding to the seven different sets of coefficients. Figure 4.6 represents each regression model with a different box. At the same time, each box is composed by three boxes. The upper gray box corresponds to the trajectory set used to train the regression model:

- CLIMB MCMB FLAT contains only the climb trajectories flown at MCMB rating below and at the kink point.
- CLIMB MCMB TEMP includes only the climb trajectories flown at MCMB rating above the kink point.
- CLIMB MCMB combines all the climb trajectories flown at MCMB rating, regardless the atmospheric conditions.
- CRUISE contains only those cruise segments at which altitude and airspeed is constant.
- DESCENT includes only the descent trajectories flown at LIDL rating.

The bottom boxes contain the BADA models involved in each regression process. The left yellow box encloses the models that are needed for the regression, and that are known from previous steps. The right blue box contains those BADA models whose coefficients are fitted through out the regression model. The nomenclature of the different BADA models to be calibrated is:

- MCMB FLAT refers to the MCMB rating model below and at the kink point.
- MCMB TEMP refers to the MCMB rating model above the kink point.

- GEN THRUST refers to the generalized thrust model.
- DRAG refers to the drag polar in clean configuration.
- GEN FUEL refers to the generalized fuel model.
- IDL THRUST refers to the low idle thrust model.
- IDL FUEL refers to the low idle fuel model.

The calibration of BADA models is done through an iterative process, which ends when the solution converges. In other words, when the minimum error between real and calibrated-modeled performance is achieved. To launch the process several sets of coefficients, marked with a red asterisk (*) in Figure 4.6, must be initialized. In this research work, coefficients of BADA nominal models are used as initial conditions. Figure 4.6 also represents the refeed of calibrated coefficients to the next iteration.

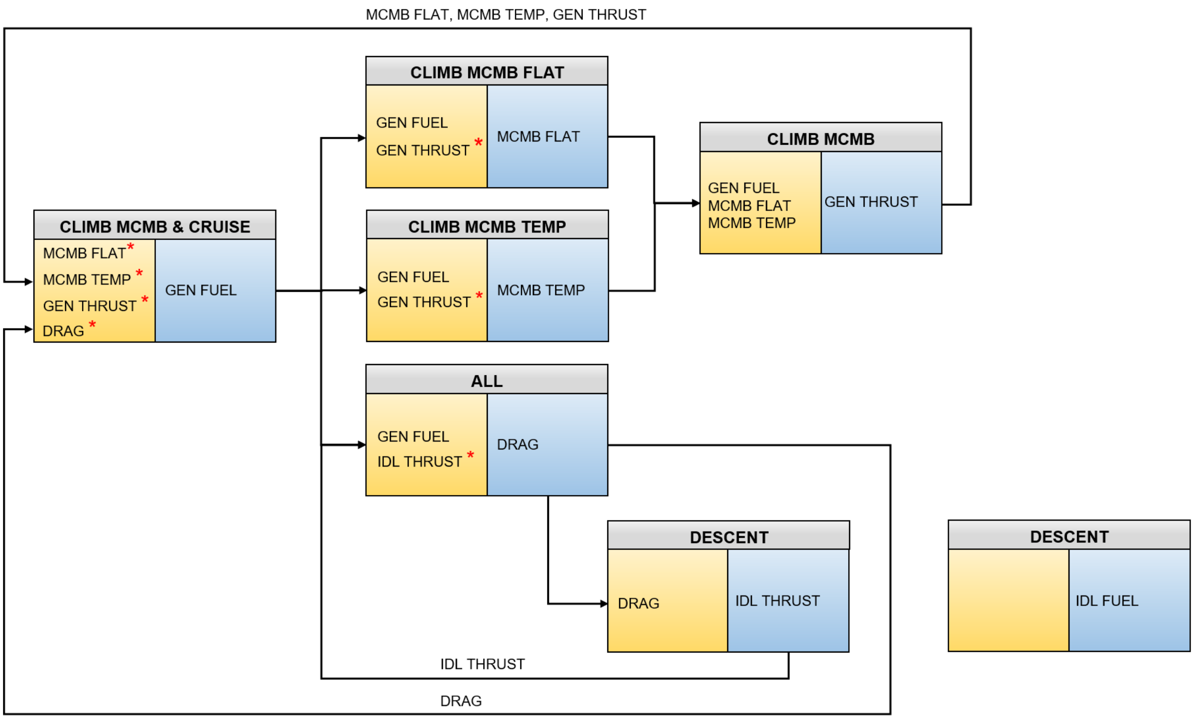


Figure 4.6: Schema of the tailoring process designed to find the sets of coefficients that best represent the reference performance data, including the types of trajectories and sets of coefficients involved in each multivariate regression model.

The tailoring scheme is summarized as follows. To start with the tailoring process, a first approximation of the generalized fuel model is calculated using climb and cruise flight points. Secondly, the flat-rated area throttle model is calculated using climb flight points under the kink point, assuming the previously-computed fuel coefficients as fixed. Similarly, the temperature-rated area model is identified using climb flight points above the kink point. Afterwards, the generalized thrust models is identified using all climb trajectories, and the flat- and temperature-rated are throttle models. In the next step, all flight points are used to find the drag polar from the generalized fuel model (previously-obtained) and the idle thrust model. The new drag model is then used to obtain a better approximation of the idle thrust model. After the first approximation of the six models has been calculated, the computed calibrated coefficients are used in a next iteration. The process is repeated until the difference between real performance (obtained from QAR data) and modeled performance (obtained with the calibrated models) is not increasing during a preselected number of iterations. Finally, the idle fuel coefficients are identified using only descent flight points.

Similarly, Figure 4.7 aims to graphically represent the tailoring process, this time in a table-like diagram. This figure also includes the seven multivariate regression models, represented with gray shadings, as well as all the intermediate steps that are performed with each set of trajectories. Once again, steps that demand a set of coefficients to be initialized are marked with a red asterisk (*). In this case, coefficients identified from a regression model and performance parameters computed from any of these coefficients are marked with a black asterisk (*). This diagram also counts with the refeed of calibrated coefficients to the next iteration.

As part of the developed methodology, criteria to stop the iterations of the tailoring process have been formulated. It is recommended to stop the process when the global MAE, computed as a weighted average of the MAE in all phases of flight, does not improve in a predefined number of iterations. By considering the weighted average MAE a trade-off between the possible improvement and deterioration in the model's accuracy for all flight phases is accomplished. If a specific requirement exists on calibration time, the methodology allows to stop the process after a predefined number of maximum iterations.

Two critical problems faced in the calibration of BADA; the instability and lack of robustness of the models over the entire flight envelope, are addressed not only in the data ingestion and data preparation phases, but also in the tailoring process. The problem of instability observed is inherent in BADA models, which are defined by high-degree polynomials that consist of a high number of terms. The number of coefficients, however, is not fixed and depends on the quantity and quality of the reference data with which the coefficients are identified, together with the modeler preferences. In many occasions, this results in simpler expressions where some coefficients are deactivated. It has been proven that the disregard of high-degree terms is beneficial for the stability of calibrated models. One important feature of the developed methodology, and of the tailoring process phase specifically, is that it allows to explicitly pre-select the coefficients to be adjusted. In other words, it allows the user to decide the degree of the polynomials to be calibrated. Based on documentation of BADA 4 models [61] [62], it is suggested to calibrate only those coefficients that are activated in BADA nominal models, this improves stability without notably degrading accuracy, and reduces calibration times. In short, the methodology can accommodate the requirements of different applications, which may demand different levels of accuracy, memory usage and computing times.

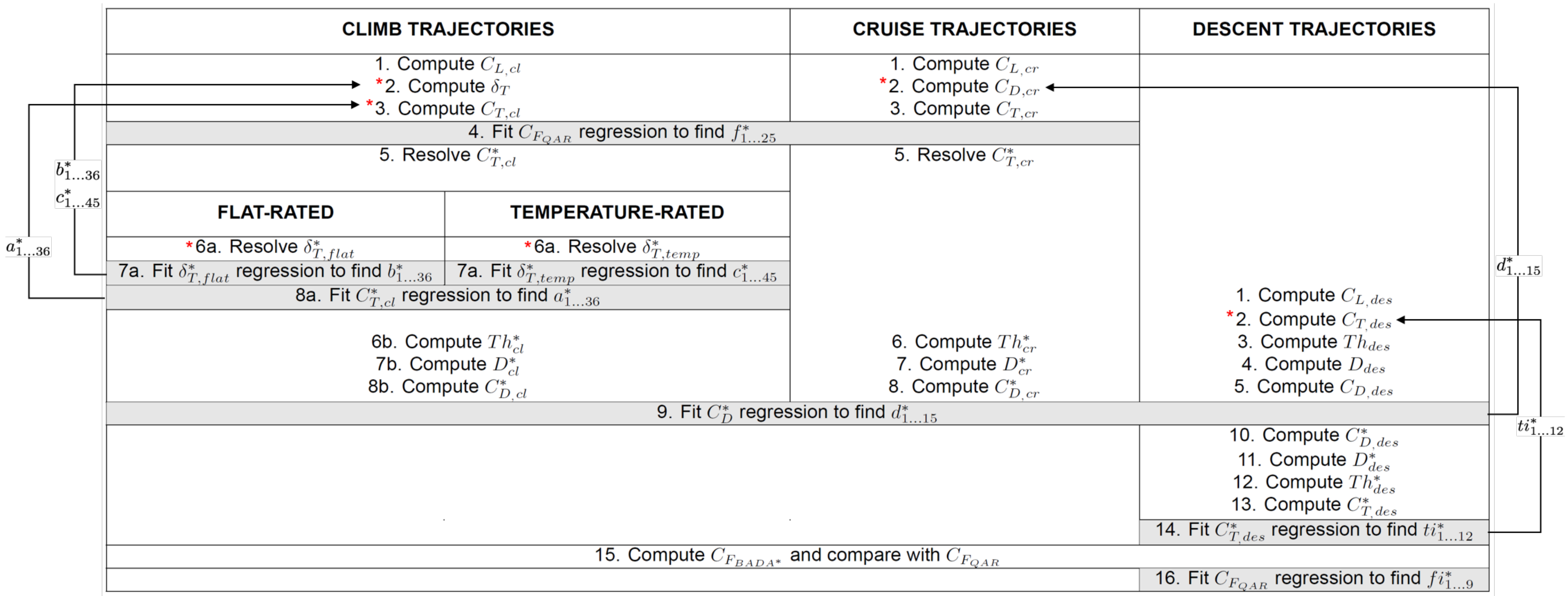


Figure 4.7: Schema of the tailoring process designed to find the sets of coefficients that best represent the reference performance data, including all the intermediate steps, the interconnection between steps and the multivariate linear regression problems to be solved.

4.3.4. Model Evaluation

Once the seven sets of calibrated coefficients have been found, models must be validated to guarantee that they provide more accurate results than BADA nominal models, and also that they are valid over the entire flight envelope and during all phases of flight. Thus, the proposed methodology includes two different ways to validate calibrated models: by means of a testing dataset and by means of synthetic data.

Firstly, in order to completely validate the methodology and the resulting calibrated models, it is necessary to evaluate the performance of calibrated models using QAR data that have not been used to train the regression models and identify the new sets of coefficients. A testing dataset, formed by flights of the same tail number characterized in the training dataset, is reserved to study the ability of calibrated models to generalize lessons learned to unseen flight data. The same preprocessing that is applied to the training dataset should now be applied to the testing dataset. Otherwise, differences in error metrics that might arise could be due to factors different than the incapacity to generalize, like for example data noise, which would difficult the evaluation of the models.

Then, the stability and robustness of calibrated models over the entire flight envelope and all phases of flight must be assessed. One of BADA's main advantages is its applicability over the entire operating envelope. It is expected that the calibration of BADA coefficients will improve the accuracy of BADA nominal models. However, calibrated models should also be applicable to every possible flight trajectory, regardless of the flight conditions. The synthetic dataset generated in the data ingestion should be used to address the performance of calibrated models over the entire envelope. For this research work, BADA nominal models were used to generate the synthetic dataset. As mentioned earlier, other sources of synthetic performance data could be used. One way to evaluate the behavior of calibrated models outside the regions of the flight envelope where QAR data are agglomerated is to compute the Absolute Percentage Error (APE) between fuel flow given by BADA nominal and BADA calibrated models, for each synthetic flight point. Models are validated if APEs below a predefined threshold are observed in all regions of the envelope, which would mean that models provide reasonable values of fuel flow regardless the flight conditions. Figure 4.8 is an example of this evaluation exercise. It represents the difference between BADA nominal and calibrated climb performances over the entire flight envelope of the aircraft under consideration. From this figure one can conclude that the calibrated climb models can be used to simulate any possible flight trajectory with guarantees. The same evaluation must be done for cruise and descent calibrated models. Note that in this case, it is proposed to use BADA nominal fuel flow as baseline, but any other performance parameter contained in the synthetic data, and estimated by calibrated models could be use.

4.4. Results and Discussion

Following the previously-described methodology, several calibrations with different datasets have been successfully performed and validated. This section includes results corresponding to three different calibrated models. Two of them are tail-specific models, meaning that they were derived using QAR data from one tail number only. On the other hand, the third calibrated model was generated with flight data from multiple tail numbers of the same aircraft type. The different error metrics shared in this section have been computed using testing datasets, in other words, datasets that were not used to run the tailoring processes.

4.4.1. Tail-specific BADA Calibrated Models

4.4.1.1. Narrow-body Aircraft

The first tail-specific calibration corresponds to a single-aisle, short- to medium-range twin-jet airliner. QAR data of 524 flights that took place between January and June 2016 were used to derive the calibrated models. A testing dataset conformed by 85 flights from December 2015 was utilized to validate the models and obtain the error metrics displayed below. Training and testing flights were chosen close in time, in order to reduce as much as possible the influence of performance degradation with time on the results.

As it can be seen in Table 4.2, the proposed methodology allows to reduce the error of estimated fuel flow with respect to BADA nominal models in the three main phases of flight: climb, cruise and descent. In average, calibrating BADA coefficients allows to reduce the error in fuel flow from 5.02 % to 1.49 %, considering all phases of flight. With respect to global MAE, this is reduced from 101.32 kg/h

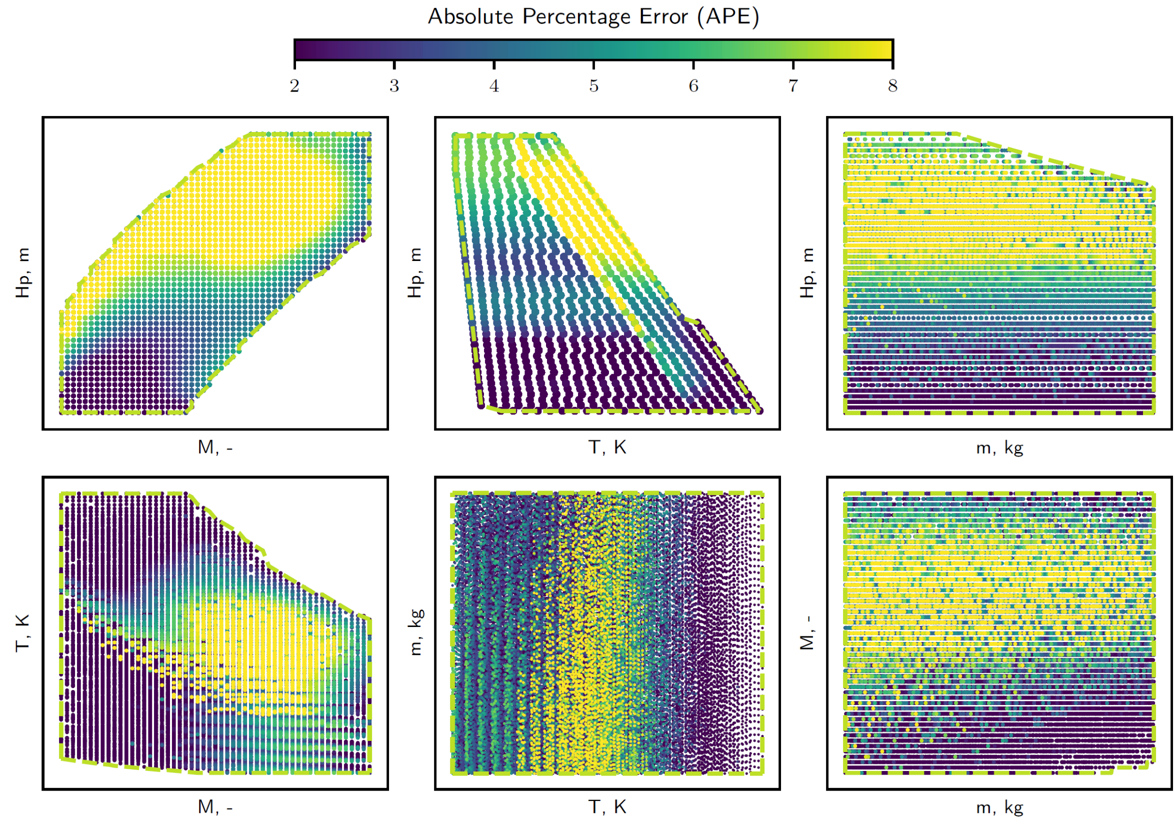


Figure 4.8: Absolute Percentage Error (APE) between fuel flow modeled by BADA nominal and BADA calibrated models in each synthetic flight point generated in the data ingestion phase of the methodology

to 30.10 kg/h, and the average RMSE is lowered from 124.27 kg/h to 43.18 kg/h. In terms of Pearson correlation, mention that the correlation of BADA nominal fuel flow with actual fuel flow is considerably high specially in climb and descent trajectories. Nonetheless, these correlations are further increased thanks to the developed methodology. For cruise trajectories, correlation slightly decreases.

Table 4.2: Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE) and Pearson correlation (R) of the fuel flow given by BADA nominal and BADA tail-specific calibrated models of a narrow-body aircraft with respect to the actual fuel flow

	MAE [kg/h]		MAPE [%]		RMSE [kg/h]		R [-]	
	BADA	Calibration	BADA	Calibration	BADA	Calibration	BADA	Calibration
Climb	176.40	20.99	3.99	0.42	196.43	28.72	0.997	1.000
Cruise	92.99	39.80	4.58	1.95	105.01	52.27	0.936	0.928
Descent	34.26	6.31	7.88	1.23	41.26	8.34	0.993	0.997

Figure 4.9 illustrates the evolution of fuel flow during a testing flight selected as an example, and confirms the improvement in fuel-flow modeling, which is more noticeable in climb and descent. In cruise, the calibration reduces the error, but fuel-flow tendency, well captured by BADA, is lost. It is believed that the shift between actual and BADA fuel-flow curves results in a flattened calibrated fuel flow. This de-synchronization may be due to the action of the engine fuel-flow controller, or even to recording errors. It is expected that the fix of this misalignment would result in better agreement with real performance also in the cruise phase.

Even though the most noticeable enhancement is observed in fuel-flow modeling, calibrated ROCD is realistic and consistent with actual and BADA nominal ROCDs. Figure 4.10 pictures the evolution of ROC and ROD over climb and descent of a flight selected as an example. Fuel flow is the parameter that mainly drives this performance tailoring approach, but it is important to highlight that the tailoring process adjust BADA fuel models without distorting thrust and drag models.

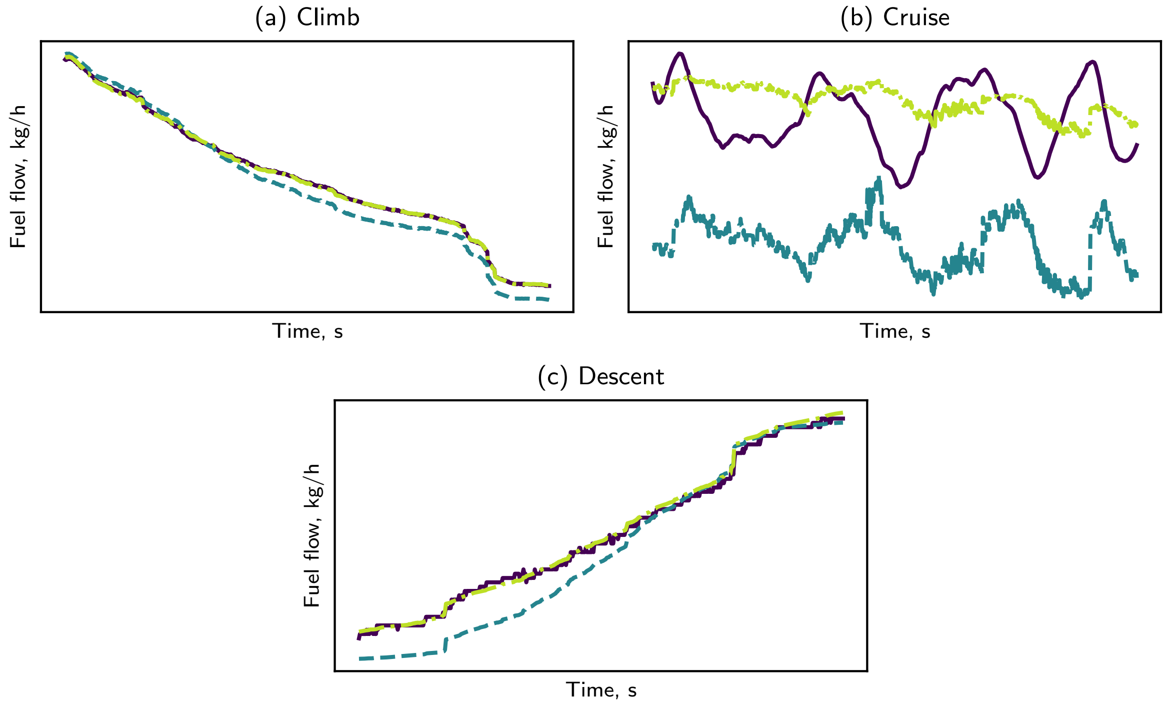


Figure 4.9: Evolution of fuel flow during a testing flight according to BADA nominal models(– –) and BADA tail-specific calibrated models of a narrow-body aircraft (–·–) with respect to the actual fuel flow (——) in climb, cruise and descent

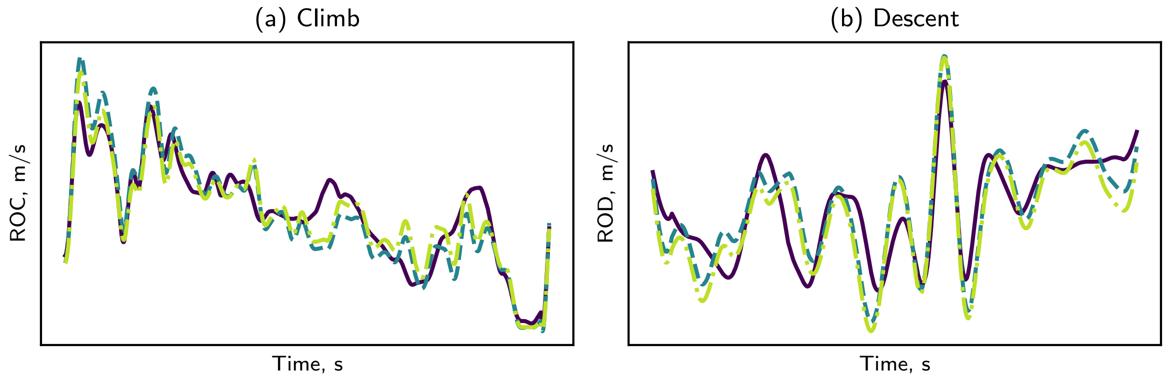


Figure 4.10: Evolution of ROCD during a testing flight according to BADA nominal models(– –) and BADA tail-specific calibrated models of a narrow-body aircraft (–·–) with respect to the ROCD (——) in climb and descent

Last but not least, errors in modeled fuel consumption have been computed and compared. This exercise helps to understand the advantages of using calibrated models in applications like fuel analytics or flight planing, for which the true importance is the accumulated error in fuel consumption. In this respect, and for the considered testing dataset, BADA nominal models estimate fuel consumption per hour of flight with a MAE of 80.78 kg/h, and a total MAPE of 3.33 %. On the other hand, the MAE between actual fuel consumption per hour of flight and the one given by calibrated models is 13.19 kg/h, and the total MAPE is 0.54 %. Figure 4.11 consists of 4 histograms that depict the distribution of these error metrics over the testing dataset. Histograms on the left (see Figure 4.11(a)) represent the distribution of MAE in fuel consumption per hour of flight given by BADA nominal models (top) and BADA calibrated models (bottom). In this regard it can be said 60% of the flights have an error between 70 kg/h and 120 kg/h, according to BADA. Thanks to the calibration of BADA coefficients, the error of most flights is reduced to less than 25 kg/h. Similar trends are observed in the histograms on the right (see Figure (4.11(b)), which represent MAPE in fuel consumption. Assuming that the 85 flights considered for this study include all the flights operated by the tail number during December 2015, the total amount of fuel misestimated by BADA sums more than 5 tons, worth approximately 5000 USD [59]. On the other hand, the total amount of fuel underestimated is around 800 kg, which worth less than 800 USD, if calibrated models are used.

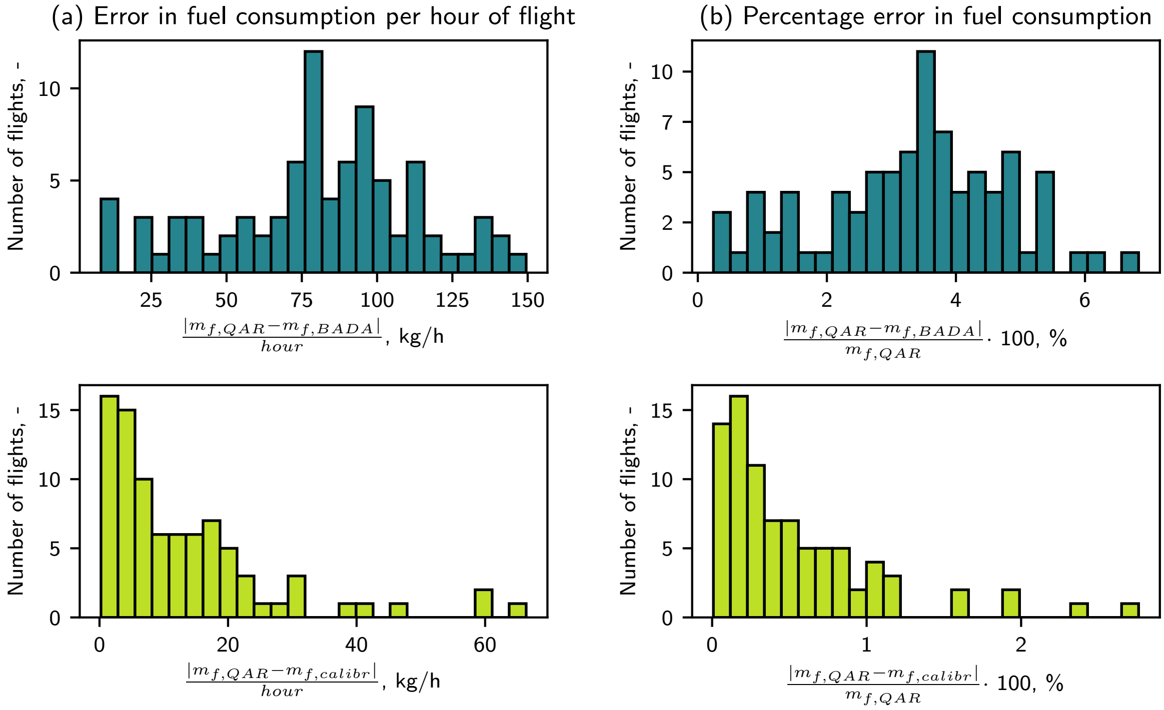


Figure 4.11: Error in fuel consumption per hour of flight and percentage error between the real fuel consumption and the fuel consumption given by BADA nominal models (■) and BADA tail-specific calibrated models of a narrow-body aircraft (■)

Based on results presented above, it can be affirmed that calibrating BADA coefficients using historical QAR data and machine-learning techniques significantly improves the accuracy of BADA nominal models. In order to support this conclusion and to validate the developed methodology, more calibrations were run, whose results are presented hereinafter in this section.

4.4.1.2. Wide-body Aircraft

The second tail-specific calibration corresponds to a twin-aisle, long-range twin-jet airliner. A very different aircraft type operated by a different airline was selected for validation purposes. To derive these calibrated models, QAR data of 661 flights that took place between January and October 2013 were used. To validate the models and obtain the error metrics displayed below, a testing dataset conformed by 126 flights from November and December 2013 was utilized.

As shown in Table 4.3, the methodology developed in this research work allows to improve BADA's estimation of fuel flow in the main phases of flight also for wide-body aircraft. Overall, it can be said that thanks to the developed methodology the error in fuel flow is reduced from 2.35 % to 1.76 % for an average flight of the tail number under consideration. In this case, as in the model calibration for the narrow-body aircraft, the biggest improvement occurs in the climb phase, in which the MAPE is reduced from 3.29 % to 0.48 %. Less noticeable improvements are observed in the cruise phase, in which BADA nominal models are able to model fuel flow more accurately than in the case of the narrow-body aircraft, with a MAPE of 2.19 % with respect to 4.58 %. Note that in the case of this tail number, the cruise phase is relatively longer compared to the previous tail number and to the other flight phases. Thus, its weight in the calculation of overall error metrics is greater. In the case of the descent phase, the MAPE of the calibrated models for the wide-body aircraft is reduced to more than half, from 5.32 % to 2.40 %. At the same time, it almost doubles the MAPE of the calibrated models for the narrow-body aircraft, it is 2.40 % with respect to 1.23 %. Figure 4.12 supports these statements. It shows that climb trajectories experience the most evident enhancement. It reveals why these calibrated models are less accurate than those for the narrow-body aircraft: the significant noise in the already-cleaned-and-filtered cruise-fuel-flow signal. Neither BADA nominal nor calibrated models are able to capture the peaks of the cruise fuel flow, only the mean slightly-decreasing tendency. Last but not least, it allows to understand why the MAPE of descent fuel flow is larger in this case: the tendency of fuel flow at the beginning of decent is not well-captured by BADA nominal models. This unexpected behavior was identified in the majority of flights used in the tailoring process, but the resulting models are not able to learn it and replicate it, which leads to believe that the limitation is inherent in the formulation of BADA models.

Table 4.3: Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE) and Pearson correlation (R) of the fuel flow given by BADA nominal and BADA tail-specific calibrated models of a wide-body aircraft with respect to the actual fuel flow

	MAE [kg/h]		MAPE [%]		RMSE [kg/h]		R [-]	
	BADA	Calibration	BADA	Calibration	BADA	Calibration	BADA	Calibration
Climb	581.67	83.93	3.29	0.48	625.71	118.36	0.999	0.999
Cruise	164.21	131.76	2.19	1.77	207.90	174.06	0.963	0.963
Descent	84.25	37.23	5.32	2.40	94.20	47.24	0.990	0.994

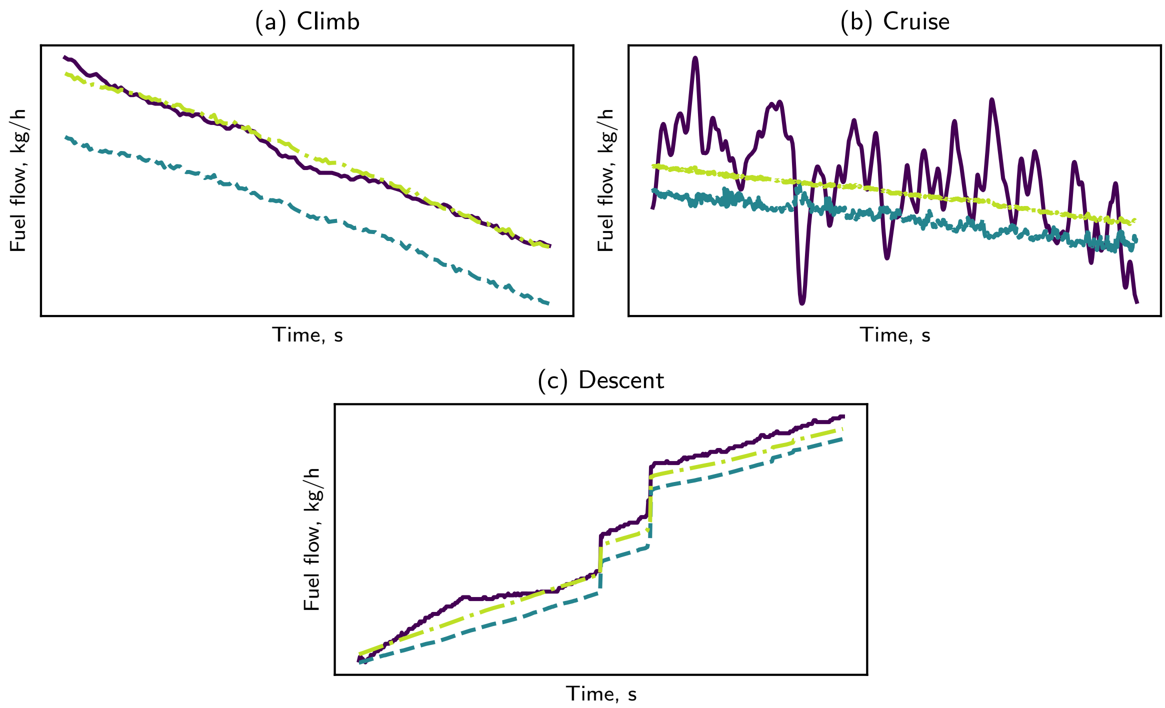


Figure 4.12: Evolution of fuel flow during a testing flight according to BADA nominal models (– –) and BADA tail-specific calibrated models of a wide-body aircraft (–·–) with respect to the actual fuel flow (——) in climb, cruise and descent

Figure 4.13 demonstrates that in the case of wide-body aircraft, calibrated models also give realistic and consistent ROCD. This proofs that the tailoring process is simultaneously modifying drag, thrust and fuel models, to enhance aircraft performance modeling.

Once again, errors in fuel consumption have been computed and compared to better understand the advantages of calibrating BADA nominal models in applications like fuel analytics or flight planning. In the case of the studied wide-body aircraft, and for the considered testing dataset, BADA nominal models estimate fuel consumption per hour of flight with a MAE of 119.48 kg/h and a total MAPE of 1.70%. Through the calibration, the MAE of fuel consumption per hour of flight is reduced to 52.39 kg/h and the total MAPE to 0.73 %. Figure 4.14 includes histograms that represent the distribution of these error metrics over the testing dataset. As it can be seen in Figure 4.14, the adjustment of BADA coefficients based on QAR data allows to reduce the error also in fuel consumption. For the set of

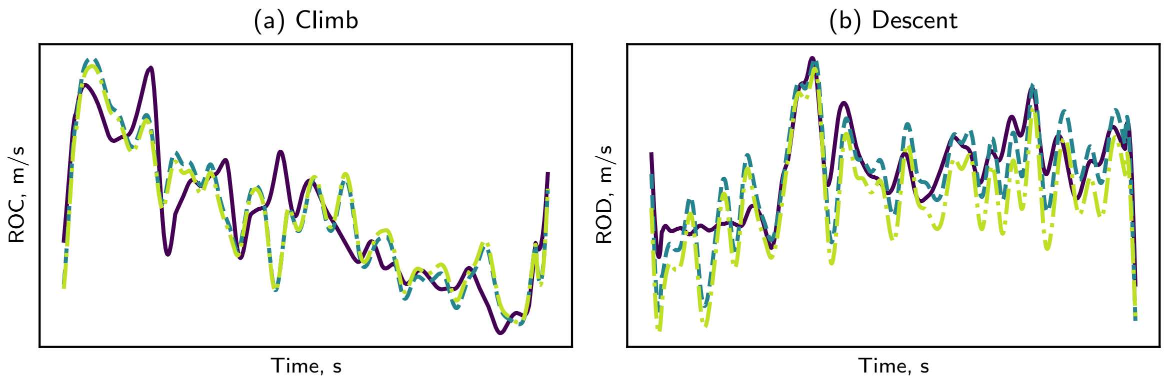


Figure 4.13: Evolution of ROCD during a testing flight according to BADA nominal models(– –) and BADA tail-specific calibrated models of a wide-body aircraft (–·–) with respect to the actual ROCD (——) in climb and descent

126 flights considered in this study, the total amount of fuel misestimated by BADA nominal models is approximately 67 tons, worth around 63000 USD [59]. On the other hand, the total amount of fuel misestimated by calibrated models is 28 tones, worth approximately 26000 USD [59].

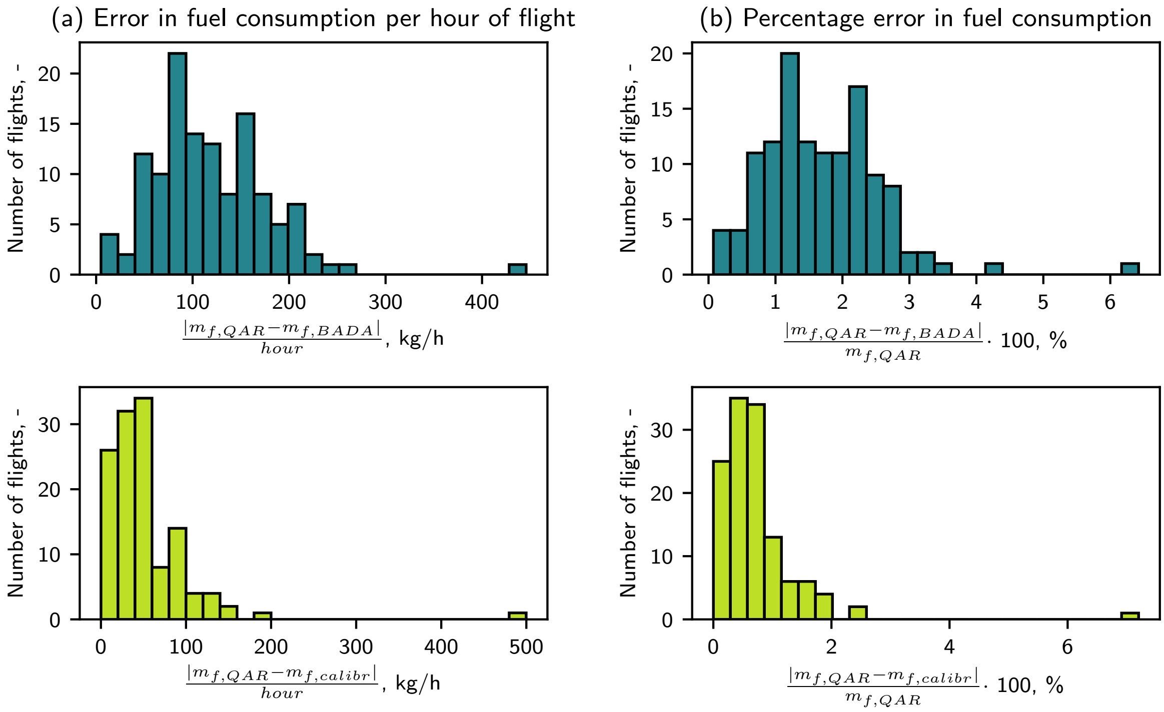


Figure 4.14: Error in fuel consumption per hour of flight and percentage error between the real fuel consumption and the fuel consumption given by BADA nominal models (■) and BADA tail-specific calibrated models of a wide-body aircraft (■)

In this section the developed methodology has been validated using a different tail number of a very different aircraft type. Despite the noise in the fuel-flow signal of the QAR data of the wide-body aircraft, this experiment corroborates that calibrating BADA coefficients using historical flight data and machine-learning techniques improves the accuracy of BADA nominal models. Therefore, enhancing all the applications based on them. In addition, it also demonstrates the importance of the quality of reference performance data, and its strong influence on results.

4.4.2. Aircraft-type-specific BADA Calibrated Models

4.4.2.1. Narrow-body Aircraft

So far, only tail-specific calibrations have been carried out. However, the developed methodology is not restricted to tail-specific modeling. It can be used to develop aircraft performance models corresponding to, apart from a tail number: an aircraft type, an aircraft type of an airline, an aircraft type flying similar routes, etc. To further validate the research work presented above and demonstrate that the methodology can be used at convenience, an aircraft-type-specific calibration has been accomplished. This calibration uses QAR data corresponding to 85 different tail numbers of the same aircraft type and airline. A total of 1926 flights that took place between the 1^{st} and 9^{th} of January 2017 were used to train the models, and 717 flights from the two following days were utilized to test them and gather the results shared below.

Error metrics of BADA nominal and BADA calibrated models with respect to the actual fuel flow recorded in the QAR data can be seen in Table 4.4. In this case, the calibration allows to reduce the average error in fuel flow from 6.82 % to 2.25 %, considering all phases of flight. In terms of MAE, the average global error is decreased from 160.35 kg/h to 52.01 kg/h, and the average global RMSE is lowered from 182.73 kg/h to 68.95 kg/h. Noting that the narrow-body aircraft considered in Section 4.4.1.1 and this aircraft type have similar characteristics, it is important to remark the different degree of accuracy with which the calibrations can mimic the real aircraft performance. In terms of MAPE, tail-specific models estimated fuel flow with an average error of 1.50 %, significantly lower compared to the error of aircraft-type-specific models. This reinforces the hypothesis that tail-specific modeling is necessary when high accuracy is needed.

Table 4.4: Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE) and Pearson correlation (R) of the fuel flow given by BADA nominal and BADA aircraft-type-specific calibrated models with respect to the actual fuel flow

	MAE [kg/h]		MAPE [%]		RMSE [kg/h]		R [-]	
	BADA	Calibration	BADA	Calibration	BADA	Calibration	BADA	Calibration
Climb	230.80	54.09	4.90	1.16	245.97	66.63	0.996	0.997
Cruise	159.74	56.72	6.83	2.46	177.34	73.53	0.785	0.823
Descent	48.09	13.45	10.03	2.70	52.51	17.41	0.969	0.979

The consequences of adjusting BADA coefficients to QAR data have once more been illustrated. In Figure 4.15, one can notice the better match between actual and calibrated fuel flow compared to BADA nominal fuel flow. This match is again stronger in climb and descent trajectories than in cruise, where again there seems to be a de-synchronization between actual and BADA nominal curves that results in a more-accurate but more-flattened calibrated fuel flow. As shown in Figure 4.16, aircraft-type-specific calibrated models lead to consistent and realistic ROCD. Which means that the accuracy of modeled fuel flow is increased and, at the same time, the rest of performance parameters represented by the model are physically coherent. In order to consider the developed methodology robust, BADA calibrated models must preserve the physical relations modeled by BADA nominal models.

Assuming that the tail numbers considered in this exercise conform all the airplanes of the aircraft type of the airline, the use of BADA nominal models in fuel-analytics applications would imply a under-estimation of more than 130 tones of fuel in two days of operation, worth approximately 125,000 USD [59]. By using calibrated models instead, these figures would reduce to a total of 30 tones and 29,000 USD [59]. Errors between real and modeled fuel consumption are plotted in Figure 4.17 in the form of histograms. It can be seen that while most of the flights accumulate a percentage error between 2 % and 8 % according to BADA, percentage error in most flights is less than 3 % thanks to the calibration. The difference between tail-specific and generic calibrated models is also apparent in terms of fuel consumption, for which MAPE is 0.54 % and 1.37 %, respectively.

In this section the developed methodology has once more been validated. Results obtained for this generic calibration demonstrate the flexibility to use the developed methodology to adjust BADA models at convenience, meaning that it can be used to develop tail-specific models or generic models, just by considering QAR data of several tail numbers together. As expected, generic models represent aircraft performance with less accuracy than tail-specific models.

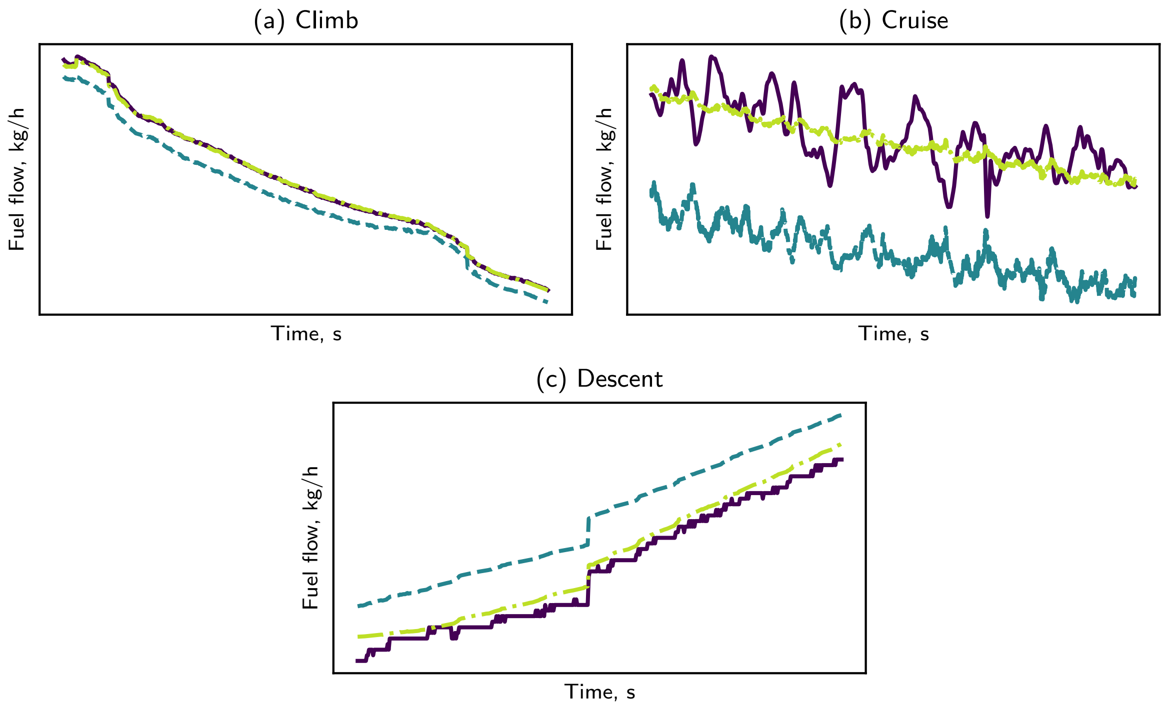


Figure 4.15: Evolution of fuel flow during a testing flight according to BADA nominal models(– –) and BADA aircraft-type-specific calibrated models (-·-·-) with respect to the actual fuel flow (——) in climb, cruise and descent

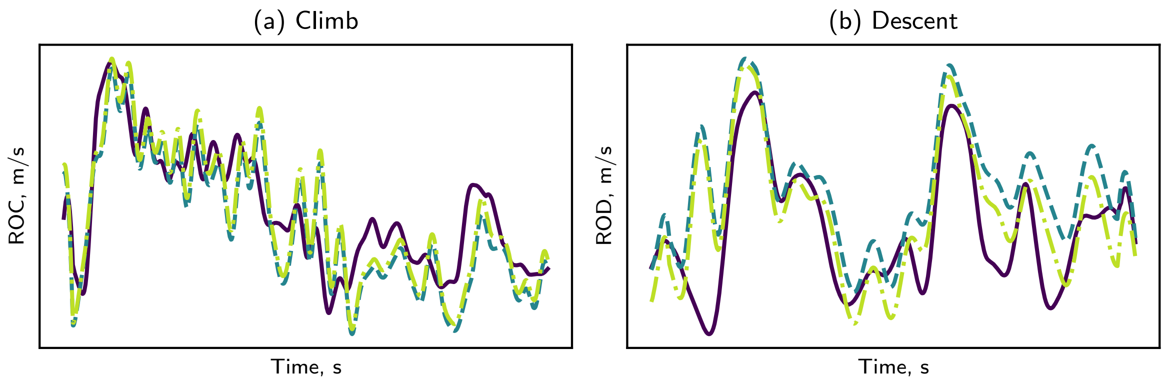


Figure 4.16: Evolution of ROCD during a testing flight according to BADA nominal models (– –) and BADA aircraft-type-specific calibrated models (-·-·-) with respect to the actual ROCD (——) in climb and descent

4.5. Conclusion

This chapter starts with a brief description of the QAR datasets available to develop the research work. Afterwards, it contains a comparison between BADA's estimations and real performance that demonstrate the need of a method to update BADA models. The core of this chapter is the methodology developed in this part of the research work, which includes all the necessary steps to calibrate BADA coefficients using QAR data and regression machine-learning algorithms. All these steps have been grouped in four phases: data ingestion, data preparation, tailoring process and model evaluation, described in Sections 4.3.1 - 4.3.4.

This chapter ends with the outcomes of following the proposed methodology, using three sets of QAR data corresponding to: a narrow-body tail number, a wide-body tail number and a narrow-body aircraft type. Based on the obtained results, it can be affirmed that the proposed methodology can be used to successfully calibrate BADA performance models based on real flight data. This can be done

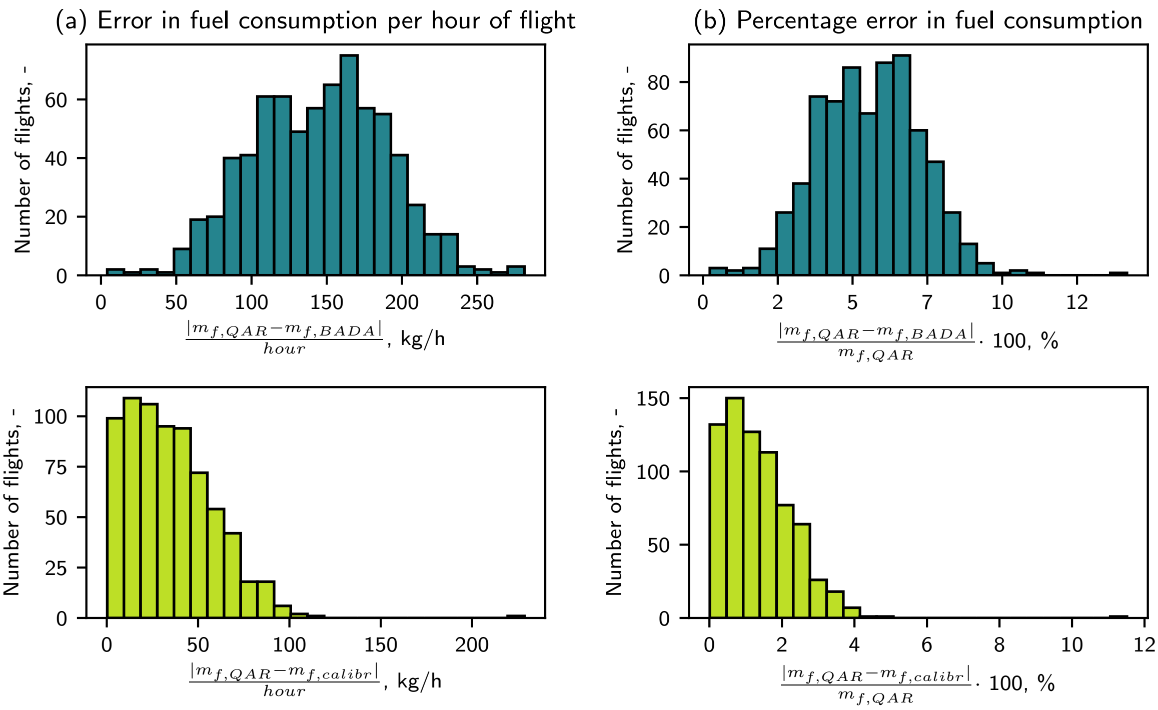


Figure 4.17: Error in fuel consumption per hour of flight and percentage error between the real fuel consumption and the fuel consumption given by BADA nominal models (■) and BADA aircraft-type-specific calibrated models of a narrow-body aircraft (■)

to model tail-specific performance, but also to create generic performance models, just by considering various tail numbers together. Calibrated models proved to be more accurate than BADA nominal models, but the quality of calibrated models highly depends on the coverage, precision and granularity of the reference flight data. Using the KPIs defined in Table 2.1, it can be said that calibrated models provide the same capabilities and level of complexity as BADA nominal models. With respect to the applicability of the models, BADA nominal models can be applied over the entire flight envelope of the aircraft type under consideration. By complementing real flight data with synthetic flight data, this key feature is also guaranteed by the calibrated models. In terms of accuracy, BADA nominal models predict fuel flow over an entire flight with a MAPE of 5.00 % in the case of the narrow-body airplane, 2.33 % in the case of the wide-body airplane, and 6.83 % in the case of the narrow-body fleet. With the calibrated models, these errors are reduced to 1.50 %, 1.78 % and 2.27 %, respectively. With respect to maintainability, BADA nominal models do not demand high maintenance resources. Nonetheless, in order to have calibrated models that accurately reflect the real performance of an aircraft, they need to be updated or re-calibrated. However, time and memory resources required to calibrate the models are reasonable and feasible.

In conclusion, a methodology to model aircraft performance using historical flight data together with machine-learning techniques has been developed. The purpose of the methodology is to identify all BADA coefficients using QAR data as reference performance data. As a result, the behavior of the aircraft characterized in the flight data is represented by so-called hybrid performance models, since they are based on physical models but are adapted using historical data. The developed methodology has proven to be consistent and complete, and calibrated models have demonstrated to improve the accuracy of BADA nominal models, while providing the same capabilities, applicability and level of complexity, all within feasible computing, maintainability and memory requirements. It is believed that the use of BADA calibrated models in pre-, in- and post-flight applications like fight planning, trajectory prediction, fuel analytics, air traffic simulation, or aircraft conflict detection, is desirable since it will ensure more realistic results. However, BADA models are built on several assumptions that are inherited by the calibrated models. As a continuation to the research work presented here, purely data-driven alternatives to BADA calibrated models have been developed.

5

Purely Data-driven Aircraft Performance Tailoring

In the previous chapter, a hybrid approach to tailor aircraft performance models has been described. This chapter presents an alternative concept, which consists on developing purely data-driven, non-physical models, rather than to calibrate existing, physical ones. The purpose of this effort is discussed in Section 5.1, which also introduces the tackled problem and some machine-learning algorithms available to address such type of problem. Section 5.2 details the methodology followed to develop the purely data-driven models, whose results can be seen in Section 5.3. A conclusion on the results and insights obtained is presented in Section 5.4.

5.1. Machine-learning Models

Multiple non-physical, purely data-driven models have been developed to model tail-specific performance. These models, in contrast to the hybrid performance models described in Chapter 4, are non-physical models that predict performance parameters without modeling the underlying physics. It is believed that purely data-driven models built with consolidated machine-learning algorithms might capture more complex relationships between input and output parameters. Additionally, they will allow to easily include more input parameters, and to study their importance on performance modeling. In short, the two purposes of the purely data-driven approach described hereinafter in this chapter are:

1. To study to what extent performance modeling can be enhanced by using non-parametric machine-learning regressors, compared to physical parametric models like BADA.

2. To assess the importance of various flight parameters on performance modeling, and analyze the impact on prediction accuracy.

In this part of the research project the problem of modeling aircraft performance from historical flight data is addressed with supervised learning. Having a set of input and output variables, supervised learning is used to find the mapping function that pairs inputs (or features) and outputs (or targets), with the ultimate goal of accurately predicting outputs of new input data. In this case, fuel flow is the output or target variable. The problem to be solved is then a regression problem, since fuel flow is a continuous variable.

There exist numerous types of regression algorithms. Simple and multivariate linear regressions are two of the most straightforward techniques, being the later one used to calibrate BADA nominal models. A variant of these techniques is polynomial regression, and another known method is Support Vector Regression (SVR) [63]. Nowadays, however, the most popular algorithms, because of their wide use across the literature, are: decision trees, ensembles and neural networks, they all can also be used to address classification problems.

Decision trees are non-parametric algorithms whose goal is to predict the value of the target variable by learning simple decision rules inferred from the data features. The main advantage of decision trees is their simplicity to understand and interpret, since trees can be visualized. Decision trees are considered white-box models that, unlike black-box models, can be explained by boolean logic. In addition, decision trees demand little data preparation. They can handle both numerical and categorical data and features do not need to be normalized. Decision trees are useful to address single- and multi-target problems. On the other hand, decision trees are susceptible to overfitting. If too complex trees are generated they may not generalize the data well. Decision trees can be unstable, meaning that small variations in the data can result in a completely different tree been created. Such instability is solved by using decision trees within an ensemble.

Ensembles methods use multiple learning algorithms to combine their predictions and improve their robustness and ability to generalize lessons learned to new unseen data. One can differentiate between two different families of ensembles methods: averaging (stacking, bagging) and boosting methods. The first builds base estimators independently and then average their predictions, and the latest constructs base estimators sequentially trying to reduce the bias in each iteration. The idea behind boosting ensembles is to generate a powerful ensemble from multiple weak models. Ensembles methods are suitable to address problems that would difficultly be solved with solely one decision tree. They can capture linear and non-linear relationships in the data. In addition, it is unlikely that they overfit. However, they are computationally expensive and if their structure is too complex they are more difficult to interpret than single decision trees.

Last but not least, neural networks are algorithms inspired by the structure and operation of the human neurons. Thus, these algorithms are built by neuron nodes connected together. Neurons are organized in one or multiple layers, used to progressively extract higher level features from the input data. Neural networks are known for being able to detect complex nonlinear relationships between features and target variables. The main disadvantage of these algorithms, besides their computational cost, is that they are black boxes that can hardly be interpreted and explained [64].

After making a trade-off between the different algorithms, it was decided that ensemble methods are the most suitable for this research effort. Ensembles can provide high level of accuracy and are less prone to overfit. Furthermore, they are not black boxes and can be deciphered. Apart from predicting the target variable they evaluate feature importance and rank variables based on their significance in the model.

5.2. Methodology

A methodology to build aircraft performance models based on QAR data and ensemble methods has been developed. Like the methodology to calibrate BADA models, it is expected that this approach leads to tail-specific models that mimic the real behavior of the aircraft more accurately than generic BADA nominal models. This approach will allow to compare BADA calibrated models with ensemble-based models, and to make a sensitivity analysis on how those models can be enhanced by including certain parameters neglected by BADA. This section is devoted to describe how the purely data-driven models have been created. The main steps followed are: data ingestion, data preparation, feature selection, model selection, model training and model evaluation, as it can be seen in Figure 5.1.

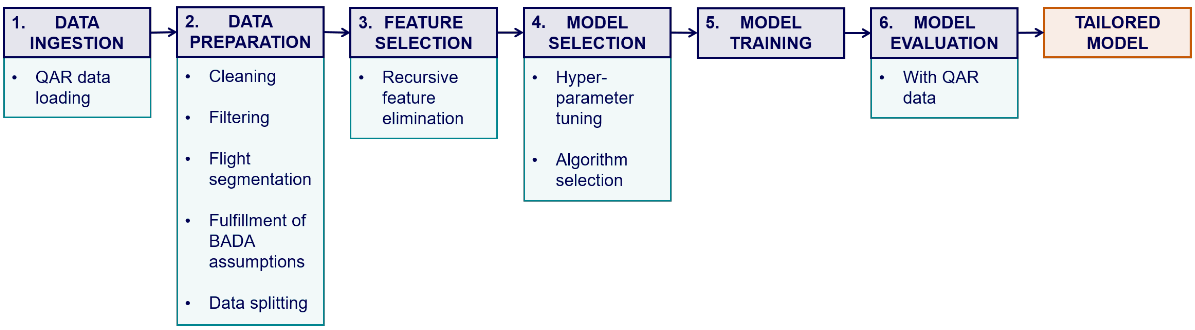


Figure 5.1: Schema of the methodology developed to generate purely data-driven aircraft performance models, including the main steps: data ingestion, data preparation, feature selection, model selection, model training and model evaluation

5.2.1. Data Ingestion

As in every data-driven process, the first step of this performance tailoring approach is to gather and load data. In this case, the data ingestion phase only includes the loading of QAR data. Synthetic data are not used in this approach to prevent biased results, conclusions and insights. The purpose of this approach is not to develop models applicable over the entire flight envelope, but to obtain meaningful insight from real performance, rather than from synthetic performance. This approach uses machine learning as a tool to study how ensemble methods can enhance performance modeling, and to analyze feature importance. However, it is believed that the concept of augmenting QAR with synthetic data is also interesting in the generation of non-physical models. The inclusion of synthetic performance data in the training of the models would strongly help the generalization to new unseen trajectories whose flight conditions deviate from the typically flown. The set of recommendations on what and how many synthetic points to include, which is defined in Section 4.3.2, could serve as baseline when non-physical models applicable over the entire flight envelope are to be generated.

This approach involves three different QAR datasets. The first dataset, of medium size, is used to define what input parameters give better results, to tune model hyperparameters and to select the final model. This dataset will be called *Dataset 1* from now on. The second and largest dataset is used to train the final model, and the third and smallest dataset to evaluate it. These datasets will be denominated *Dataset 2* and *Dataset 3*, respectively.

Once again, the parameters of interest are preselected before loading the QAR datasets. This preselection includes parameters required for the preparation phase, which are not fixed and depend on availability, and parameters used to build the models. These are: pressure altitude, aircraft gross weight, total air temperature, Mach number, true airspeed, ROCD, flight path angle, angle of attack, course, drift and track angles, longitudinal and vertical acceleration, control surfaces positions (elevators, ailerons and rudder), horizontal stabilizer position, N1 and N2, thrust levers, heading, ground speed, wind speed and direction, and obviously fuel flow. To limit the scope of the project only continuous and physical parameters were selected. However, one benefit of purely data-driven models is that they can consider non-physical parameters such as departure and destination airports, flight number or flight time. These are categorical variables that might provide valuable information, but that increases the dimensionality of the problem significantly.

5.2.2. Data Preparation

As in the case of the hybrid approach, the data processing phase starts with the elimination of flights with erroneous recordings, and continues with the cleaning and filtering of signals. Outliers are eliminated and noise is reduced by using Savitzky-Golay filters.

Aircraft performance varies from one flight phase to another. Prior to the generation of ensemble models it was decided to segment flights in climb, cruise and descent trajectories. One can expect different levels of accuracy and rankings of important variables depending on the considered flight phase. Since one of the goals of the purely data-driven approach is to study how can the accuracy of BADA calibrated models be further improved, the preparation phase includes the necessary steps to be faithful to several BADA modeling principles. Specifically:

- Only those flight points in which the aircraft is flying with clean configuration are considered. Flaps, slats and spoilers positions are not used by the model.
- Only those flight points in which there is no derate and no assumed temperature are selected, to guarantee that those flight points at which the aircraft is flying with reduced thrust settings are not taken into account.
- Only those flight points in which no additional engine air bleeding is identified are used. Such identification is done based on bleed valves, anti-ice and packs-on information. More criteria could be included if the necessary variables are available in the dataset.
- Only those climb segments in which the aircraft is flying at MCMB are considered.
- Only those cruise segments in which the aircraft is flying at constant altitude are taken into account. If the aircraft flies at various flight levels during cruise the ascending segments are discarded.

- Only those flight points at which the aircraft flies at LIDL are used.

These assumptions allow to keep the dimensionality of the problem under reasonable limits, and to guarantee that comparisons with BADA calibrated models are as unbiased as possible.

The last step of the preparation phase is data splitting, an important concept in data science key to prevent overfitting. Datasets are usually divided into training and test sets. Models learn on training data and are evaluated on test data. Sometimes data are splitted in training, test and validation. In this case, the validation set is held back from training the model and used to give an estimate of model performance while tuning hyperparameters, for example. The test set, also held back from training the model, is then used to evaluate the final model. In this project, data splitting varies depending on the dataset. Datasets 1 and 2 are splitted into train and validation with proportion of 7 to 3. Meaning that 70 % of the flight points are used to train the models and the other 30 % to give an estimate of the model accuracy on unseen points of the same flights. To do feature elimination and hyperparameter tuning, described in Sections 5.2.3 and 5.2.4, training points of Dataset 1 are further splitted in 5-folds. This approach is called k-fold cross-validation, technique based on the split of the training set into k smaller sets. The model is trained using k-1 folds and validated on the remaining fold. The final performance metrics are the average of all the metrics computed in the loop. To test how the final models would estimate performance of new flights Dataset 3 is entirely used. The proposed data splitting is illustrated in Figure 5.2.

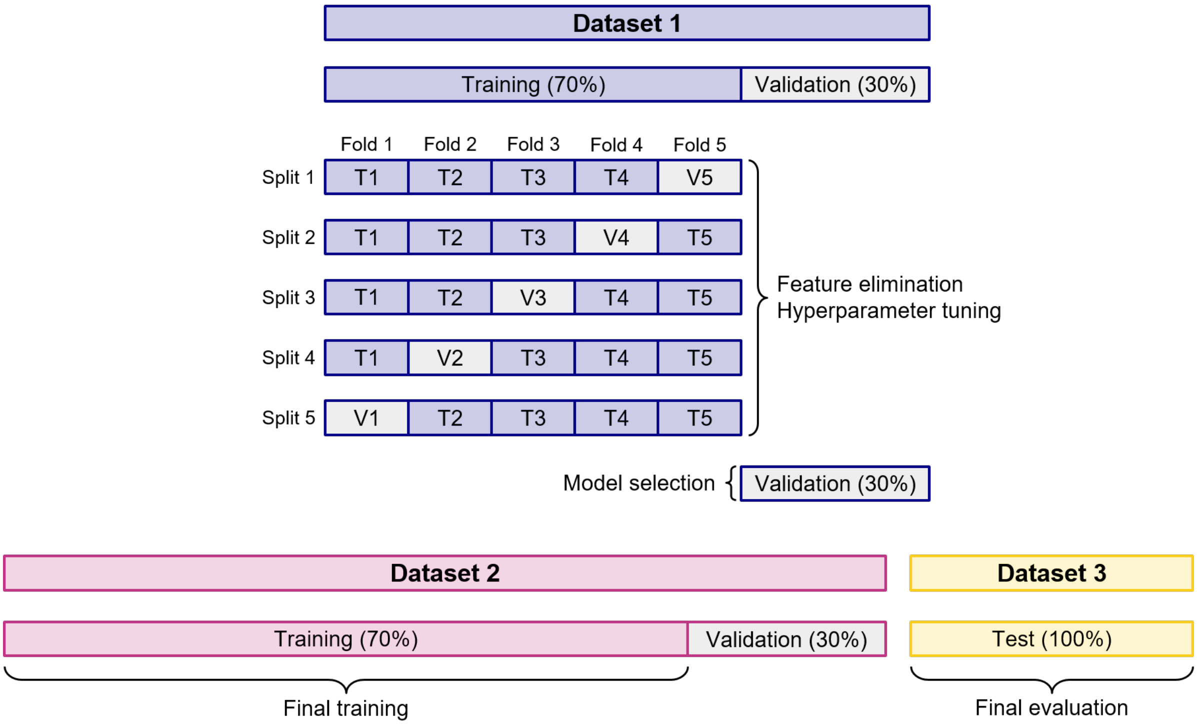


Figure 5.2: Schematic representation of the used datasets, their segregation in training, test and validation, and their connection with the different steps of the proposed machine-learning pipeline

5.2.3. Feature Selection

More is not always better. Not all preselected features might contribute to predict the target variable. In fact, removing features of low importance can improve accuracy while reducing complexity and proneness to overfitting. To find the optimum features for each model, the designed pipeline includes a process of feature selection, to be launched before the model construction. Among all the feature-selection methods available in literature, Recursive Feature Elimination (RFE) [65] was chosen. This method fits a model and removes the weakest variables until a certain number of features, previously specified, is attained. Unfortunately, the number of optimum input variables is not known beforehand, so RFE needs to be combined with k-fold cross-validation. As a result, different subsets of features are trained, evaluated and compared, to select the best subset of features.

5.2.4. Model Selection

Once the optimum set of features is determined, it is time to select the model. Model selection can have multiple meanings in the context of machine learning. On one hand, one can be interested in choosing the best possible hyperparameters for a specific machine-learning algorithm. Unlike other model parameters, which are derived via training, hyperparameters are set before fitting the model and govern the training process itself. On the other hand, one might want to select the best algorithm from a set of suitable methods. In this case both activities are combined: the best algorithm, together with its optimum hyperparameters, is selected from a set of eligible methods. Specifically, five different ensemble methods are considered:

1. **Adaptive boosting regressors**, meta-estimators that first fit a base regressor and later fit additional copies of the regressor but adjusting the weight of instances according to prediction error.

2. **Gradient boosting regressor**, similar to adaptive boosting regressors, but with the difference that they address the boosting problem as an optimization problem. They add weak learners to increase performance and build a strong model based on a loss function.

3. **Extreme gradient boosting regressor**, similar to gradient boosting regressors, but they stand out for penalizing trees and shrinking leaf nodes wisely. Additionally, they include an extra randomization parameter that reduces correlation between trees, which results in a better ensemble.

4. **Random forest regressor**, meta-estimators that fit several decision trees on different subsets of the training dataset and use averaging to improve accuracy and control overfitting.

5. **Extreme random forest regressor**, also called extra trees regressors, differ from random forest regressors because instead of computing the locally optimal split combination they select a random value for the split.

Prior to selecting the most accurate possible algorithm, the hyperparameters of each algorithm should be optimized. Hyperparameter tuning consists of testing different hyperparameter configurations to determine what are the values that maximize model's predictive accuracy. For more details on what hyperparameters are optimized for each regressor refer to Tables B.1 - B.5, in Appendix B.1. Any parameter search requires: an estimator, a parameter space, a method for sampling different configurations, a cross-validation scheme, and a score function. In this case, the five estimators previously listed are considered. Validation curves are plotted to define the parameter space, in other words, the list of parameters and range of values for each parameter and specified estimator. Validation curves plot the influence of a single hyperparameter on the training and validation scores, as shown in Figure 5.3, in which R^2 of a gradient boosting regressor is plotted as a function of max_depth. R^2, also called coefficient of determination [66], is the most common regression score function and represents the proportion of the variance in the dependent variable that is predictable from the independent variables.

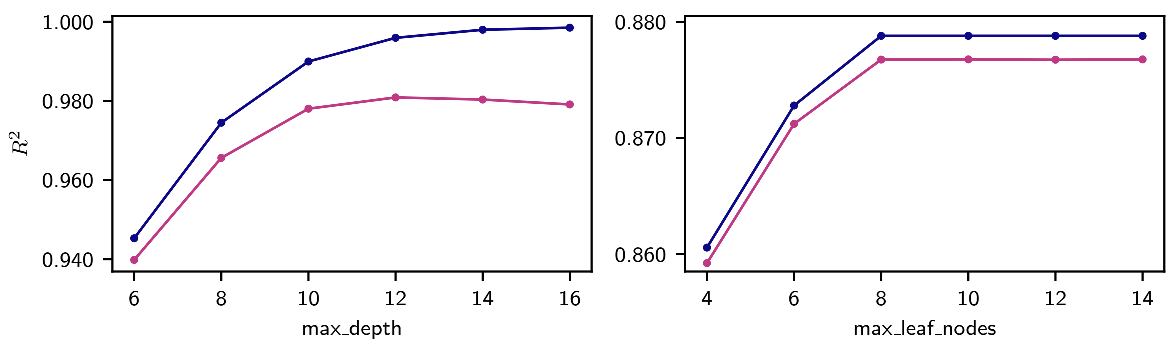


Figure 5.3: Training (——) and cross-validation (——) R^2 of a gradient boosting regressor as a function of hyperparameters max_depth and max_leaf_nodes

With respect the method for sampling different hyperparameter configurations, mention that there exist several search strategies. Two generic approaches were considered: grid search and random search. The first builds and evaluates a model for each combination of hyperparameters specified in the parameter space, which makes it computationally expensive. On the other hand, random search treats the parameter space as a statistical distribution for each hyperparameter, from which values are randomly sampled. This strategy is of special interest when one hyperparameter has significantly more influence on the prediction than the rest, which often occurs. In this case, grid search spends useless time exploring irrelevant hyperparameters, while random search can focus on finding the optimal values for the essential hyperparameters. For this reason, a random search strategy was finally implemented.

Once the best set of hyperparameters was found for each of the estimators, the models were evaluated on the validation set. The five regression models were compared based on their MAPE. The regressor with lower MAPE is chosen as the best predictor.

5.2.5. Model Evaluation

After datasets are gathered and prepared, irrelevant features are discarded, and the best models are selected, these must be evaluated. As in the evaluation of BADA calibrated models, four different error metrics are considered: MAE, MAPE, RMSE and R (see Equations 4.4, 4.5, 4.6, and 4.7, respectively). Models must be evaluated using data points that have not been used to train the models. This is done to understand what models are able to generalize to unseen new data. For this purpose, two sets of data points are reserved: a validation set corresponding to 30 % of Dataset 2, and Dataset 3. If only the first is used the error metrics obtained might be misleading. This is because even though the validation data are not strictly speaking the same as the training data, they belong to the same flights. To have a realistic idea of the performance of the models one should use QAR data from different flights, which simulates the practical use of the models.

5.3. Results and Discussion

Following the machine-learning pipeline described in the previous section, multiple non-physical models to predict fuel flow have been created and evaluated. In particular, a total of nine models per tail number, three models per flight phase. The first kind of model uses only 4 features, BADA input variables: pressure altitude, gross weight, total air temperature and Mach number. The second includes all the variables listed in Section 5.2.1, which come to a total of 25 features. Finally, the third type of model considers all these features apart from the engine variables (N1, N2 and thrust levers). For the sake of simplicity, these models will be called *ML 1*, *ML 2* and *ML 3* hereinafter in this report.

This section discusses the outcomes of 18 different purely data-driven performance models, corresponding to two different tail numbers: a narrow-boy aircraft and a wide-body aircraft. This section is devoted to analyze the performance of the models created. For more details on the outcomes of the feature and model selection processes refer to Appendices B.2 and B.3, respectively. The results presented here consist of error metrics in fuel flow prediction, including MAE, MAPE, RMSE and R. Rankings of features based on their importance on the prediction are also provided. Feature importance is calculated using Gini Importance, also called Mean Decrease in Impurity (MDI), which is computed as the decrease in node impurity weighted by the probability of reaching that node. In other words, MDI counts the times a feature is used to split a node, weighted by the number of samples it splits.

5.3.1. Tail-specific Purely Data-driven Models

5.3.1.1. Narrow-body Aircraft

QAR data from the same narrow-body aircraft considered in Section 4.4.1.1 are used to generate purely data-driven models. In this case, Dataset 1 is composed of 526 flights that took place between January and June 2016. Dataset 2 includes a total of 3442 flights from years 2016 and 2017, note that Dataset 1 is contained in Dataset 2. Dataset 3 consists of 86 flights that date from December 2015. Results shared in this section have been obtained using Dataset 3, to address the performance of the models on flights that have not been used for training.

Regarding *climb-fuel-flow modeling*, purely data-driven models are more accurate than physical and hybrid models, as it can be seen in Table 5.1. However, this difference is less than 5 kg/h in MAE, equivalent to 0.1 % in MAPE. As expected, the most accurate model is ML 2, mainly thanks to the engine variables. ML 3 performs slightly worse than ML 1, which means that the inclusion of additional non-engine variables is counterproductive. These variables are adding noise to the model, rather than providing useful information. Overall, it has been demonstrated that an ensemble model built only with BADA input variables is sufficient to estimate climb fuel flow with a high degree of accuracy. The main advantage of ML 1 is that it can used to simulate what-if scenarios, which makes it suitable for pre-flight applications. On the other hand, ML 2 and ML 3 demand input variables that might not be known *a priori*. By comparing ML 1 with BADA calibrated models, it can be said that ensemble methods are able to capture climb fuel flow more accurately than multivariate linear regressors. In panels (a-c) of Figure 5.4 it can be seen that the gap between BADA calibrated, ML 1, ML 2 and ML 3 is not apparent.

Table 5.1: Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE) and Pearson correlation (R) of climb fuel flow given by BADA nominal, BADA calibrated and purely data-driven models with respect to actual fuel flow, for the narrow-body airplane

	BADA	BADA Calibrated	ML 1	ML 2	ML 3
MAE [kg/h]	178.29	21.01	16.47	16.29	16.95
MAPE [%]	4.06	0.42	0.33	0.32	0.34
RMSE [kg/h]	198.02	28.81	23.15	22.92	24.14
R [-]	0.997	1.000	1.000	1.000	1.000

Table 5.2 provides a ranking of variables based on their importance on climb-fuel-flow modeling. Pressure altitude is the most important variable, with a weight of approximately 50 % in all purely data-driven models. Mach number and total air temperature are two other important flight parameters. They have similar weights, and their importance is between 13 % and 23 %, depending on the model. Fuel flow increases as air density increases, and vice versa. Air density changes with variations in pressure, temperature and humidity, and these flight parameters relate to speed of sound, and therefore, to Mach number. When selecting a fixed engine rating control law, as is the case, each engine rating provides a unique throttle position for given Mach number and atmospheric conditions, regardless aircraft mass. As a result, the impact of gross weight is neglected in all the purely data-driven models generated to model MCMB fuel flow. An engine variable is one of the most important features of ML 2 model. Both ML 2 and ML 3 models consider true airspeed, ground speed and elevator position when building up the decision trees. However high level of accuracy in MCMB-fuel-flow modeling can be achieved without considering wind and control-surface deflection, based on the results presented in Table 5.1.

Table 5.2: Ranking of feature importance in the prediction of climb fuel flow of the purely data-driven models, for the narrow-body airplane

ML 1		ML 2		ML 3	
Feature	%	Feature	%	Feature	%
Pressure altitude	56.42	Pressure altitude	44.76	Pressure altitude	45.11
Mach number	22.71	N1	15.36	Mach number	18.93
Total air temperature	20.85	Total air temperature	14.28	Total air temperature	18.68
Gross weight	0.02	Mach number	13.37	True airspeed	9.60
		True airspeed	6.10	Ground speed	3.45
		N2	2.10	Elevator position	3.05
		Ground speed	1.95	Others	< 1
		Elevator position	1.26		
		Others	< 1		

Cruise-fuel-flow modeling from QAR data is more challenging. As discussed in Section 4.4.1.1, cruise-fuel-flow modeling can be improved by calibrating BADA nominal models using QAR data as reference performance data. Even though mean errors are notably reduced, calibrated models are not able to capture the tendency of actual fuel flow. In the case of purely data-driven models, their performance hugely depends on the features used to build the decision trees. Regarding ML 1, mention that this model performs in average 0.21 % worse than BADA calibrated models. However, as illustrated in panel (d) of Figure 5.4, they are able to capture some tendency while BADA calibrated fuel flow is flattened. Once again, one can observe the misalignment or desynchronization between fuel flow and the rest of the considered variables. The addition of extra non-engine variables is beneficial in the case of cruise-fuel-flow modeling, proved by an improvement of 0.29 % of ML 3 with respect to ML 1. However, the tendency is again not captured. Undoubtedly, the most accurate model is ML 2, which not only provides the lowest mean error but also perfectly captures the tendency, as shown in 5.4(e). Fuel flow given by ML 1 is the least correlated to actual fuel flow, followed by fuel flow modeled by BADA calibrated, BADA nominal and ML 3, and ML 2, which has the R closest to 1. The consideration of engine variables is crucial to mimic the real cruise performance of the narrow-body aircraft.

Table 5.3: Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE) and Pearson correlation (R) of cruise fuel flow given by BADA nominal, BADA calibrated and purely data-driven models with respect to actual fuel flow, for the narrow-body airplane

	BADA	BADA Calibrated	ML 1	ML 2	ML 3
MAE [kg/h]	94.74	42.77	47.11	14.41	41.16
MAPE [%]	4.65	2.09	2.30	0.69	2.01
RMSE [kg/h]	108.55	57.50	61.84	21.90	54.77
R [-]	0.925	0.917	0.903	0.990	0.925

This time, the ranking of features based on their importance on the prediction is presented in Table 5.4. Gross weight is the most decisive feature to describe cruise performance. The required aerodynamic forces are directly related to aircraft mass, so this influences propulsive forces and therefore, fuel flow. Pressure altitude and total air temperature also proved to be important to model cruise fuel flow, mainly due to their relation with air density. It should be noted the low importance of Mach number in ML 2 and ML 3 (1.53 % and 3.27 %, respectively) compared to its importance in ML 1 (17.82 %). While it is true that Mach number provides information about speed of sound, covered by pressure altitude and total air temperature, and aircraft speed, included in ML 2 and ML 3. It is believed that the inclusion of engine information is the main responsible of the outstanding performance of ML 2. In fact, the importance of all engine variables (N1, N2 and thrust levers) is 32.78 %, greater than the importance of gross weight. Another remarkable parameter is angle of attack, with a weight of 8.49 % in ML 2 and 10.84 % in ML 3. Lift force is expressed as a function of speed, air density, wing area and angle of attack. Therefore, angle of attack influences aerodynamic forces, and consequently thrust force and fuel flow. Lift force is also affected by the position of the center of gravity. If it moves forward, the down-force on the horizontal tail required to trim the aircraft increases, and vice versa. Stabilizer position changes to counteract the variation of the longitudinal position of the center of gravity. Stabilizer deflection produces trim drag, additional drag that has an impact on thrust force if the aircraft has to fly at constant speed. Consequently, stabilizer position impacts cruise fuel flow and its consideration in the creation of aircraft performance models is beneficial. Other parameters such as longitudinal acceleration, heading, and wind speed and direction, also led to a more accurate prediction of cruise fuel flow.

Regarding *descent-fuel-flow modeling*, one more time, the major improvement is thanks to the use of historical flight data to develop the performance models. As shown in Table 5.5, BADA calibrated models are more accurate than all the purely data-driven models. The MAPE of BADA calibrated models is 0.03 % smaller than the MAPE of ML 2, 0.25 % smaller than the MAPE of ML 3, and 0.46 % smaller than the MAPE of ML 1. Panels (g-i) in Figure 5.4 show that actual fuel flow has a discrete behavior that purely data-driven models try to learn. This might be one of the reasons why BADA calibrated models are, in average, more accurate. A filter could have been applied to smooth actual fuel flow prior the generation of the purely data-driven models. However, it can be concluded that the precision of the sensors that record fuel flow is decisive at the achieved level of accuracy. The mean errors of the tail-specific models are in the order of the precision of the sensors, 16 lb/h ($\approx$ 7.25 kg/h). The ranking of features based on their importance in descent-fuel-flow modeling is presented in Table 5.6. As in climb-fuel-flow modeling, pressure altitude is the most determinant input variable and gross weight can be neglected when modeling LIDL fuel flow. Since the accuracy of the models is determined by the precision of the sensors, no additional insights can be extracted from the provided ranking of features.

Based on the results presented above, it can be affirmed that climb, cruise and descent fuel flow can be accurately modeled using QAR data and tree-based ensembles. These models improve the accuracy of BADA nominal models significantly, and depending on the input variables, they further improve the accuracy of BADA calibrated models. In order to support these conclusions and to validate

Table 5.4: Ranking of feature importance in the prediction of cruise fuel flow of the purely data-driven models, for the narrow-body airplane

ML 1		ML 2		ML 3	
Feature	%	Feature	%	Feature	%
Gross weight	38.27	Gross weight	21.33	Gross weight	34.44
Total air temperature	28.18	N2	20.14	Total air temperature	18.31
Mach number	17.82	Total air temperature	14.32	Angle of attack	10.84
Pressure altitude	15.73	Pressure altitude	13.63	Pressure altitude	10.49
		Angle of attack	8.49	True airspeed	8.45
		Thrust levers	6.51	Mach number	3.27
		N1	6.13	Stabilizer position	2.41
		True airspeed	4.29	Track angle	1.99
		Mach number	1.53	Heading	1.89
		Others	< 1	Longitudinal acceleration	1.67
				Ground speed	1.51
				Wind direction	1.19
				Wind speed	1.19
				Drift angle	1.18
				Course angle	1.16

Table 5.5: Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE) and Pearson correlation (R) of descent fuel flow given by BADA nominal, BADA calibrated and purely data-driven models with respect to actual fuel flow, for the narrow-body airplane

	BADA	BADA Calibrated	ML 1	ML 2	ML 3
MAE [kg/h]	33.16	6.47	8.73	6.54	7.77
MAPE [%]	7.62	1.26	1.72	1.29	1.51
RMSE [kg/h]	40.32	8.53	11.36	8.07	9.47
R [-]	0.993	0.997	0.996	0.999	0.998

Table 5.6: Ranking of feature importance in the prediction of descent fuel flow of the purely data-driven models, for the narrow-body airplane

ML 1		ML 2		ML 3	
Feature	%	Feature	%	Feature	%
Pressure altitude	93.95	Pressure altitude	37.09	Pressure altitude	41.86
Mach number	5.11	Mach number	21.83	Mach number	25.29
Total air temperature	0.56	True airspeed	13.39	True airspeed	17.54
Gross weight	0.38	N1	11.95	Total air temperature	9.00
		Total air temperature	9.66	Ground speed	6.07
		Ground speed	4.88	Others	< 1
		N2	1.11		
		Others	< 1		

the methodology implemented, non-physical models for a different tail number were created.

5.3.1.2. Wide-body Aircraft

To derive purely data-driven performance models for a wide-body aircraft, QAR data from the same airplane considered in Section 4.4.1.2 are used. In this case, Dataset 1 is formed by 427 flights that took place between the end of December 2012 and June 2013. Dataset 2 includes these flights, plus 243 additional flights. In total, it is composed by 670 flights that happened between the end of December 2012 and October 2013. Finally, Dataset 3 consists of 126 flights from November and December 2013. Once again, this third dataset has been used to obtain the results presented below.

Regarding *climb-fuel-flow modeling* of the wide-body aircraft, and in contrast to what happened for the narrow-body aircraft, hybrid models are able to capture actual fuel flow better than non-physical models. With respect to the purely data-driven models, ML 1 is the most accurate model with a MAPE

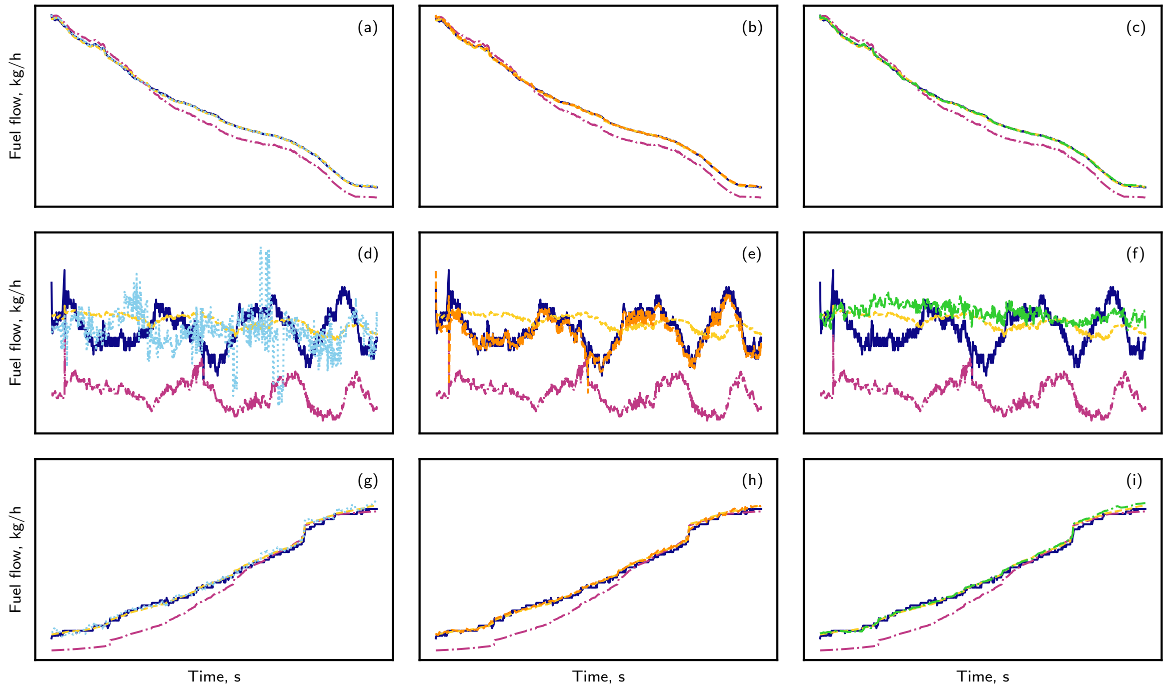


Figure 5.4: Evolution of fuel flow during a testing flight of the narrow-body aircraft, according to BADA nominal (–·–), BADA calibrated (– –), ML 1 (·····), ML 2 (– –) and ML 3 (–·–) models, with respect to actual fuel flow (——) in climb (a-c), cruise (d-f) and descent (g-i)

of 0.58 %, followed by ML 2 with a MAPE of 0.62 %, and ML 3 with a MAPE of 0.73 %. In terms of MAE and RMSE, this trend maintains. These results suggest that multivariate linear regressors mimic actual climb fuel flow more accurately than tree-based ensembles, unlike the case of the narrow-body airplane. In addition, they indicate that the consideration of additional variables introduces noise to the models, and therefore, is detrimental. Table 5.8 provides the ranking of feature importance in ML 1, ML 2 and ML 3 models. As in the case of the ML 1 model developed for the narrow-body aircraft, the most important input variables are pressure altitude, Mach number and total air temperature, while gross weight is not determinant. The evolution of actual fuel flow in a climb trajectory, together with the fuel flow modeled by BADA nominal, BADA calibrated and purely data-driven models, is illustrated in panels (a-c) of Figure 5.5. These results support the hypothesis that MCMB fuel flow can be accurately modeled using BADA input variables only. BADA calibrated and ML 1 models are not only the most accurate models, but also the only suitable for pre-flight applications.

Table 5.7: Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE) and Pearson correlation (R) of climb fuel flow given by BADA nominal, BADA calibrated and purely data-driven models with respect to actual fuel flow, for the wide-body airplane

	BADA	BADA Calibrated	ML 1	ML 2	ML 3
MAE [kg/h]	578.79	85.26	102.45	108.81	129.17
MAPE [%]	3.29	0.49	0.58	0.62	0.73
RMSE [kg/h]	623.14	119.53	144.82	149.15	179.57
R [-]	0.999	0.999	0.999	0.999	0.998

As discussed in Section 4.4.1.2, BADA nominal models capture cruise fuel flow more accurately in the case of the wide-body aircraft than in the case of the narrow-body aircraft. Still, *cruise-fuel-flow modeling* was enhanced thanks to the hybrid tailoring approach described in Section 4. As shown in Table 5.9, purely data-driven models also describe aircraft performance more realistically than BADA nominal models. Cruise-fuel-flow modeling is highly sensitive to the features utilized to build the non-physical models. Once again, ML 2 is the most accurate model in terms of mean error, and it is also the model that best reproduce the tendency of actual fuel flow. The MAPE of ML 3 is 1.53 % higher

Table 5.8: Ranking of feature importance in the prediction of climb fuel flow of the purely data-driven models, for the wide-body airplane

ML 1		ML 2		ML 3	
Feature	%	Feature	%	Feature	%
Pressure altitude	67.23	Pressure altitude	49.96	Pressure altitude	47.65
Mach number	24.40	Mach number	16.54	Mach number	17.33
Total air temperature	7.14	True airspeed	12.87	True airspeed	14.90
Gross weight	1.23	Ground speed	7.92	Ground speed	11.51
		N1	5.30	Total air temperature	3.61
		Total air temperature	2.33	Flight path angle	1.52
		Angle of attack	1.03	Angle of attack	1.03
		Others	< 1	Others	< 1

than the MAPE of ML 2, and 0.05 % lower than the MAPE of ML 1. If this last model is compared with BADA calibrated models, the latest exhibits lower mean errors. However, BADA calibrated models cannot capture fuel flow fluctuations. Despite the differences in mean error, and unlike BADA nominal and BADA calibrated models, purely data-driven models can reproduce the tendency of actual fuel flow, regardless of the used input variables. What model to use to predict cruise fuel flow would firstly depend on the input variables available. If the model is needed for pre-flight applications, and such applications demand instantaneous fuel flow predictions or engine performance modeling, ML 1 would be the best choice. On the other hand, BADA calibrated models would provide more realistic predictions of accumulated fuel consumption, for example. If performance parameters other than fuel flow or fuel consumption need to be estimated, BADA calibrated models are the best option. Panels (d-e) in Figure 5.5 depict the evolution of fuel flow during a cruise segment according to all the performance models being compared.

Table 5.9: Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE) and Pearson correlation (R) of cruise fuel flow given by BADA nominal, BADA calibrated and purely data-driven models with respect to actual fuel flow, for the wide-body airplane

	BADA	BADA Calibrated	ML 1	ML 2	ML 3
MAE [kg/h]	188.67	163.80	174.78	46.60	161.54
MAPE [%]	2.52	2.21	2.34	0.63	2.16
RMSE [kg/h]	247.73	225.47	245.10	63.22	225.74
R [-]	0.940	0.941	0.928	0.997	0.940

The ranking of features in cruise-fuel-flow modeling is presented in Table 5.10, which shows that gross weight is the most important variable in all purely data-driven models, like in the models of the narrow-body aircraft. Based on the performance of ML 2 model, it is confirmed that engine information is key to achieve a high level of accuracy in cruise-fuel-flow modeling. The total importance of engine variables in this model is 31.6 %. Like in the case described above, the consideration of additional flight parameters enhance cruise performance modeling. Although angle of attack and stabilizer position are not as influencing in this case, they both affect aerodynamic forces, and hence, thrust and fuel flow. The importance of longitudinal acceleration is higher in the case of the wide-body airplane (2.69 %) than of the narrow-body airplane (1.67 %). Either way, this demonstrates than even though it is ideally assumed that aircraft cruise at constant speed, in reality this is not always the case, and this impacts fuel flow.

With respect to *descent-fuel-flow modeling*, mention that in contrast to the case of the narrow-body aircraft, the precision of sensors is not a concern in this case. As it can be seen in Table 5.11, ML 2 and ML 3 models increase the accuracy of BADA calibrated models in 0.4 % and 0.1 %, respectively. In panels (h-i) of Figure 5.5 it can be observed to what extent these models mimic real performance. They both capture the unexpected tendency at the beginning of the descent trajectory, which is not reflected by BADA nominal, nor BADA calibrated models, as discussed in Section 4.4.1.2. ML 1 is the least accurate tail-specific model in terms of MAE, MAPE and RMSE. It is also the least correlated with actual fuel flow.

Table 5.12 includes the ranking of input variables of ML 1, ML 2 and ML 3 based on their importance

Table 5.10: Ranking of feature importance in the prediction of cruise fuel flow of the purely data-driven models, for the wide-body airplane

ML 1		ML 2		ML 3	
Feature	%	Feature	%	Feature	%
Gross weight	87.01	Gross weight	43.64	Gross weight	56.43
Total air temperature	5.57	Pressure altitude	20.34	Pressure altitude	20.93
Mach number	4.51	Thrust levers	13.16	Total air temperature	5.95
Pressure altitude	2.91	N1	9.59	Longitudinal acceleration	2.69
		N2	8.85	Angle of attack	2.66
		Total air temperature	2.71	Mach number	2.39
		Others	< 1	True airspeed	2.11
				Stabilizer position	1.26
				Others	< 1

Table 5.11: Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE) and Pearson correlation (R) of descent fuel flow given by BADA nominal, BADA calibrated and purely data-driven models with respect to actual fuel flow, for the wide-body airplane

	BADA	BADA Calibrated	ML 1	ML 2	ML 3
MAE [kg/h]	84.35	37.61	42.36	31.63	37.13
MAPE [%]	5.28	2.41	2.70	2.01	2.31
RMSE [kg/h]	94.14	47.17	56.12	37.33	45.02
R [-]	0.990	0.995	0.991	0.998	0.997

in the building of decision trees. Once more time, pressure altitude is essential and gross weight barely impacts fuel flow when a fixed engine rating governs the flight. N1 and ground speed are among the most important variables of ML 2. While N1 is clearly correlated with fuel flow, it is believed that the importance of ground speed is mainly statistical, because this parameter is taken into account to build the decision trees of model ML 2 but not of model ML 3. Regarding ML 3, mention that its relevant input variables are also considered in ML 1, but its performance is better. The information provided by the additional parameters is not meaningful, but results in a model (algorithm and hyperparameters) that better describe the physical relationships between the input variables and with the target, as reflected in the results.

Table 5.12: Ranking of feature importance in the prediction of descent fuel flow of the purely data-driven models, for the wide-body airplane

ML 1		ML 2		ML 3	
Feature	%	Feature	%	Feature	%
Pressure altitude	95.44	Pressure altitude	50.07	Pressure altitude	58.66
Mach number	2.22	Mach number	17.97	Mach number	23.88
Gross weight	1.38	N1	13.37	Total air temperature	16.09
Total air temperature	0.96	Ground speed	9.30	Others	< 1
		Total air temperature	8.16		
		Others	< 1		

In this section the implemented methodology has been validated using a different tail number and aircraft type. Tree-based ensembles have demonstrated to be effective in modeling climb, cruise and descent fuel flow, independently of the aircraft type. The results obtained corroborate that non-physical models can notably improve the accuracy of BADA nominal models. These models, fed with the adequate input variables, can perform even better than BADA calibrated models.

5.4. Conclusion

This chapter deals with the purely data-driven approach designed to develop tail-specific models based on QAR data. This approach allowed to compare the performance of non-parametric machine-learning regressors with BADA nominal and calibrated models, which are physical parametric models. It also

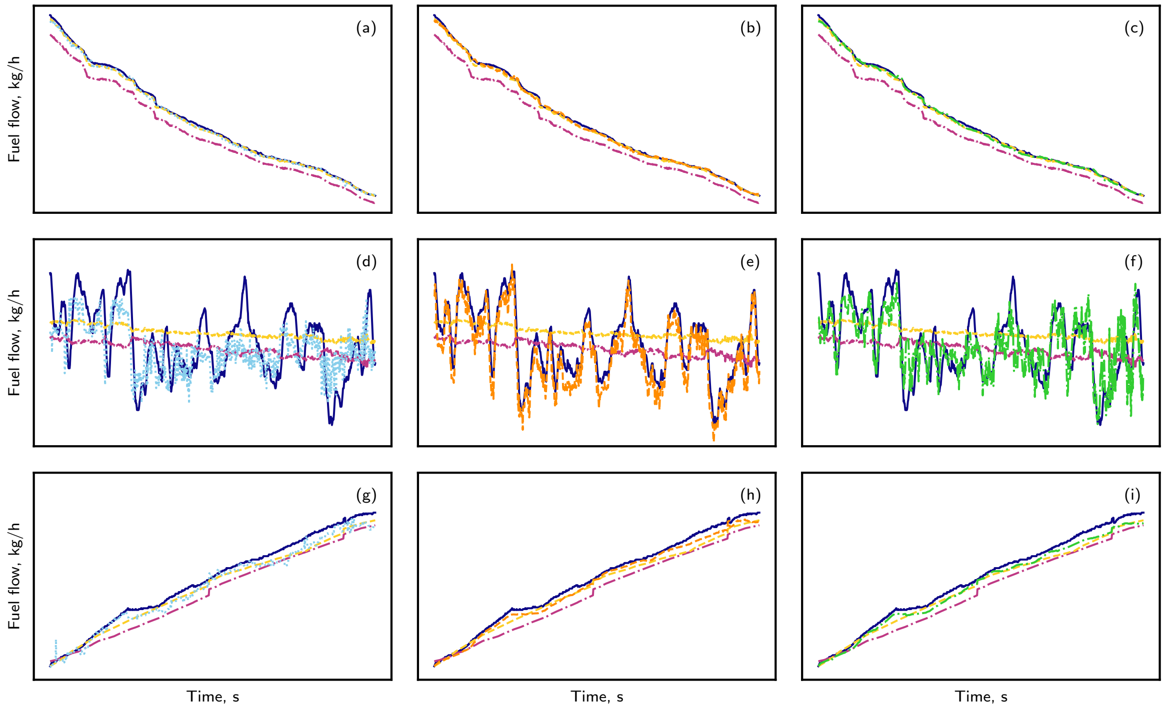


Figure 5.5: Evolution of fuel flow during a testing flight of the wide-body aircraft, according to BADA nominal (–·–·), BADA calibrated (– –), ML 1 (······), ML 2 (– –) and ML 3 (–·–·) models, with respect to actual fuel flow (——) in climb (a-c), cruise (d-f) and descent (g-i)

enabled the assessment of the impact of distinct parameters on aircraft performance modeling. The chapter starts with a brief overview of the machine-learning techniques that are applicable to this research work. To assess the motivations of this part of the research work, it was decided to focus on tree-based ensemble models. The second section of the chapter describes the methodology that was followed to create the non-physical models, which includes six different phases: data ingestion, data preparation, feature selection, model selection, model training and model evaluation, described in Sections 5.2.1 - 5.2.5. Lastly, this chapter discussed the results of applying the previously described methodology using two sets of QAR data, corresponding to a narrow-body airplane and a wide-body airplane. Results in this chapter include error metrics and rankings of input variables based on their importance in the construction of the models.

Regarding climb performance, it was concluded that three flight parameters are sufficient to model fuel flow when the aircraft is climbing at MCMB and clean configuration. These parameters are: pressure altitude, total air temperature and Mach number. In the case of the narrow-body aircraft, tree-based ensembles described climb performance more accurately than multivariate linear regressors. On the other hand, the mean errors of multivariate linear regressors were lower in the case of the wide-body aircraft. However, differences between physical and non-physical models were of 0.09 % in both cases.

With respect to cruise-fuel-flow modeling, it was confirmed that the accuracy of purely data-driven models is very sensitive to the features used to build the decision trees of the ensembles. If MAPEs lower than 2 % are required it is necessary to include engine parameters, with which MAPEs of 0.69 % and 0.62 % are achieved for the narrow- and wide-body airplanes, respectively. However, it was demonstrated that the consideration of parameters like angle of attack, stabilizer position and longitudinal acceleration results in more accurate performance models. Even though multivariate linear regressors exhibit lower mean errors, they fail at capturing the tendency of actual fuel flow. This is more evident when dealing with noisy data. On the other hand, ensembles performed worst in average but were able to simulate the real tendency of fuel flow.

In terms of descent-fuel-flow modeling of the narrow-body aircraft, the level of accuracy achieved was in the order of the precision of the sensors that record fuel flow, which makes the interpretation of results difficult. Unlike climb-fuel-flow modeling, error metrics proved that feeding LIDL fuel flow models with engine and additional flight parameters culminates in more accurate models. When BADA input variables are the only features, multivariate linear regressors are more convenient than tree-based ensembles. However, it is desirable to validate these conclusions with a different tail number. The purely data-driven models generated for the narrow-body aircraft provide valuable insights but do not allow to support these conclusions.

6

Conclusions and Recommendations

In this research, it was investigated how can tail-specific performance models be developed from historical flight data, as well as how can machine-learning techniques help to achieve this goal. Two different approaches were undertaken: a hybrid approach that consists on calibrating BADA performance models, and a purely data-driven approach that uses tree-based ensembles to predict aircraft performance. This work is the first to provide a comparison between hybrid and non-physical tail-specific performance models.

6.1. Conclusions

Aircraft performance plays a significant role in aircraft design, certification and operation, which is the scope of this research project. Aircraft performance dictates aircraft operation strategies. It is monitored by airlines to optimize the utilization of their resources in order to maximize profit, and it is latent in ATM and ATC activities. Countless efforts have been devoted to model aircraft performance. The vast majority of aircraft performance models available in the literature represent average performance of particular aircraft types. However, they fail to mimic performance of specific tail numbers, which means that applications based on them might lead to unreliable or misleading results. This is the case of BADA performance models, the most widely used nowadays.

The purpose of this research project is to develop tail-specific performance models based on historical flight data and using machine-learning techniques, in order to enrich the applications for which average performance is not accurate enough. By means of using historical QAR data as reference performance data, tailored models can account for performance degradation due to aging, performance variations due to user-configurable specifications or maintenance actions, etc. This research goal is pursued through two different approaches: by fine-tuning physical models like BADA and by creating purely data-driven, non-physical models.

This thesis project addresses some important research gaps in the field of aircraft performance tailoring. First, the lack of methods to calibrate BADA performance models using machine-learning techniques, and to develop tail-specific non-physical models applicable on-ground. Second, the absence of a detailed procedure describing data preprocessing, including: treatment of data noise and inconsistencies, feature selection, data scaling and normalization when needed, etc. Third, the lack of efforts to complement historical QAR data with synthetic performance data, in order to guarantee that tailored performance models are robust and consistent over the entire flight envelope. In order to meet the objective of the project and fill the gaps noted above, the work was divided in two efforts: hybrid aircraft performance tailoring and purely data-driven performance tailoring.

For the first part of the thesis, a methodology to calibrate BADA aircraft performance models based on the use of historical QAR data and machine-learning regression algorithms was developed. The designed methodology includes four phases: data ingestion, data preparation, tailoring process and model evaluation. Data ingestion consists on loading QAR data and generating synthetic performance data. The preparation phase, an important phase of the methodology that should not be underestimated, is focused on the selection of flight points that are compliant with the assumptions BADA is built on, and on the cleaning and smoothing of flight signals. To be able to use calibrated models to simulate any possible flight trajectory, it is necessary to add synthetic data to the training dataset before the tailoring process. A set of recommendations on what and how many synthetic points to include has been defined. Data augmentation has a negligible impact on the accuracy of calibrated models. However, it significantly helps their stability and robustness. Once the training dataset is prepared, the tailoring process can be launched. This process consists of a fitting scheme through which the sets of BADA coefficients that best describe the reference QAR data are identified. Before starting the tailoring process, it might be beneficial to deactivate certain coefficients. The disregard of high-degree terms results in calibrated models that are more stable and robust over the entire flight envelope. Thus, reducing the amount of synthetic points required and the associated bias of the training dataset. The deactivation of coefficients proved to be useful when time computing requirements are a priority. Every tailoring process must be followed by an extensive validation exercise, that should include test QAR data (to study to what extent the model is able to generalize to new, previously unseen data), and synthetic data (to guarantee that the model is stable over the entire operational flight envelope).

Three calibrations were carried out: two tail-specific calibrations and one generic calibration for an aircraft type. Results demonstrated that the designed methodology allows to successfully fine-tune BADA performance models based on historical QAR data, with the aid of machine-learning regression models. The developed methodology proved to be consisted, complete and versatile, since it allows to model tail-specific performance, but also to create generic performance models. Calibrated models demonstrated to improve the accuracy of BADA nominal models significantly, while being robust over the entire flight envelope. Therefore, the use of BADA calibrated models in pre-, in- and post-flight applications is desirable since it will ensure more realistic results.

For the second part of the thesis, a methodology to create purely data-driven, non-physical models based on historical QAR data and machine-learning regression algorithms was developed. The designed methodology includes six phases: data ingestion, data preparation, feature selection, model selection, model training and model evaluation. In this approach, the data ingestion phase only involved QAR data. The data preparation certainly has similarities with the preprocessing of the hybrid approach: it includes cleaning and filtering of signals, and segmentation of flights in climb, cruise and descent. Furthermore, it incorporates certain simplifications to guarantee faithfulness to BADA modeling principles. This limits the scope of the study and allows to relate its outcomes to BADA models. After data are prepared, the optimum set of features, hyperparameters and algorithms are selected for each scenario. Later, selected models are trained and evaluated on different sets of flights.

The two main purposes behind this part of the research project were: to compare the performance of non-parametric machine-learning regressors with BADA calibrated models, and to assess the importance of various flight parameters on performance modeling by analyzing their impact on prediction accuracy. Several non-physical models were created for two tail numbers of different aircraft types. Based on the obtained results, it can be concluded that climb, cruise and descent fuel flow can be modeled using QAR data and tree-based ensembles, regardless the aircraft type. Results demonstrated that purely data-driven models can significantly improve the accuracy of BADA nominal models. Certain non-physical models, fed with the adequate input variables, can also further improve the accuracy of tail-specific hybrid models. Regarding climb-performance modeling, three flight parameters are sufficient to model fuel flow when the aircraft is climbing at MCMB and clean configuration. In terms of cruise-performance modeling, the accuracy of non-physical models is very sensitive to the input variables used to build the models. With regard to descent-performance modeling, three flight parameters are responsible for the prediction of fuel flow. However, the consideration of engine and additional flight parameters seems to be beneficial for the construction of the tree-based ensembles.

Despite the differences in accuracy between hybrid and non-physical models they all significantly improve BADA nominal models. However, each tailoring approach has its own advantages and disadvantages. On one hand, the hybrid approach relies on a consolidated and worldwide-used aircraft performance model that describes the underlying physics governing aircraft motion. Therefore, hybrid models are more interpretable. In addition, the hybrid approach culminates in complete performance models, including gravitational, aerodynamic, thrust and fuel consumption models. Thus, hybrid tailored performance models are applicable to a wider range of applications. However, since these models are based on physical models, they inherit their assumptions and simplifications. On the other hand, non-physical models based on QAR data can only predict those parameters recorded by aircraft sensors, like fuel flow or ROCD. This limits the range of applications they can be used in. Nonetheless, the purely data-driven approach provides more flexibility in what parameters to consider. It allows to include non-physical parameters that might impact aircraft operation, and consequently, aircraft performance. In this case, model input variables are not restricted by modeling simplifications. In short, the purely data-driven approach is not worse, or better, than the hybrid. What approach and model to used is determined first by the application, and secondly by the demanded level of accuracy.

6.2. Recommendations for Future Work

Even though it is believed that the designed methodologies are complete, consistent and versatile. They can be further improved. Based on the work discussed in this and the preceding chapters, several recommendations for future work have been formulated. These are divided based on the tailoring approach they refer to. The suggestions applicable to both tailoring approaches are:

- To improve the data preprocessing routine, which should include the appropriate filtering of all parameters that proved to be key in the tailoring process and in the construction of decision trees. For example, descent fuel flow. As mentioned in Section 5.3, in the case of the narrow body aircraft the generation of non-physical, descent fuel flow models was influenced by the precision of the sensors that record fuel flow, which can be prevented by applying the right filtering techniques.
- To replicate the work with additional tail numbers and aircraft types. So far, the designed methodologies have only been validated with two tail numbers of two different aircraft types. It is desirable to validate the conclusions with more tail numbers of the same and different aircraft types, of the same and different aircraft manufacturer.
- To assess and quantify the impact of using tail-specific performance models in real-life applications, like fuel analytics and flight plan generation.

The recommendations for future work specific to the hybrid approach are:

- To include other engine ratings in the methodology, like maximum cruise (MCRZ) and maximum take-off (MTKF).
- To also calibrate non-clean drag coefficients.

Regarding the purely data-driven, non-physical approach, the recommendations for future work are:

- To consider non-physical parameters in the development of purely data-driven models, which would allow to determine their impact on aircraft performance modeling. Examples of these non-physical parameters are: departure and destination airports, flight time, etc.
- To validate the models outside the region of the flight envelope in which QAR data are agglomerated. This would allow to compare the capabilities of generalization of hybrid and purely data-driven model outside the typically-operated envelope. It is expected that in order to generate purely data-driven models robust over the entire flight envelope, QAR data will need to be augmented with synthetic performance data. A sensitivity analysis of what and how many points to include in the training of the models would be required.

As discussed in Chapter 3, aircraft performance is dynamic as it varies with time. For QAR-data-driven models to reflect performance deterioration, consequences of maintenance actions, etc., they must also be dynamic. One of the initial research sub-goals was to establish an updating mechanism to account for deviations of aircraft performance with time. During the development of the project, it was decided to leave this sub-goal out of this Master Thesis, but it is definitely something to consider for future research projects.

Currently, tail-specific performance models use machine learning to estimate performance at the moment at which predictions are made. This is sufficient to simulate ”what-if” scenarios and calculate the corresponding performance. However, some other ATM applications may demand the prediction of performance in the future, rather than in the present. Currently, there is no research work that explores the use of time series forecasting to model aircraft performance degradation. Time series forecasting is used to make predictions about the future, and it is an important machine-learning technique to consider when the problem to solve involves a time component. This would be a technique to bear in mind if performance degradation proves to be noticeable in the short-medium term.

A

BADA 4: Model Overview

Base of Aircraft Data (BADA), introduced in Section 2.1.2, is an aircraft performance model developed and maintained by EUROCONTROL. BADA is a mass-varying, kinetic approach applicable for aircraft trajectory simulation and prediction within the domain of ATM.

As shown in Figure 2.1, BADA consists of two major models: an Aircraft Performance Model (APM) and an Airline Procedure Model (ARPM). At the same time, the APM is further divided into four sub-models named actions, motion, operations and limitations. The actions model, detailed in Section A.2, computes the forces responsible for the aircraft motion, together with fuel consumption. The motion model contains the differential equations that provide the variation of aircraft position, velocity and mass with time. The operations model includes the instructions about the way the aircraft is operated, and the limitations model confines the aircraft behavior to ensure safe operation.

The actions model, which is the main focus of this work, is expressed in the form of polynomial expressions with their corresponding sets of coefficients. Whereas polynomials are unique and common to every aircraft family, coefficients are unique to each aircraft type. These coefficients were computed to achieve the best fit between calculated and reference performance data. In other words, they aimed to be the set of coefficients that best describe the performance of the aircraft family under consideration. Reference performance data used in the development of BADA models corresponded to brand-new aircraft. It was generated with performance software, and it was provided to EUROCONTROL by the manufacturers.

BADA models describe average performance of an aircraft type, independent of airline configuration and customizations. Also over time aircraft suffer structure, engine and aerodynamic deterioration, as well as maintenance actions. These factors, which vary with tail number, make aircraft performance deviate from the theoretical. Therefore, BADA nominal models cannot accurately represent the real performance of a specific aircraft. To do so, BADA models must be identified using reference data corresponding to that particular aircraft, and must be updated taking into account performance degradation. The former is addressed in this part of the project, in which BADA 4 actions model is adjusted to specific tail numbers using QAR data as the reference performance data, together with some machine-learning techniques.

A.1. Model Assumptions

In order to properly address the tailoring of BADA performance models, it is important to analyze and understand the causes of their inaccuracies. These can be due to discrepancies between the performance of the aircraft under consideration and the reference brand-new aircraft, or due to the simplifications inherent in the models. Every model is a simplification of the real world, so it may ignore some variables that affect the modeled phenomena and influence results. BADA models were built on several assumptions that deteriorate its accuracy, but that allow the models to stay within reasonable complexity, maintenance and computing requirements.

From the Concept Document for the Base of Aircraft Data (BADA) Family 4 [61], the main hypotheses that affect this research work have been extracted:

1. The aircraft is represented as a point, with all its mass concentrated in its center of gravity.

2. The position of the center of gravity with respect to any fixed point of the airframe depends on the fuel load and the payload distribution, so it varies very slowly with time. Therefore, it can be assumed known and stationary.
3. The air humidity does not affect the value of any other atmospheric property.
4. The variation of wind speed and direction with time is negligible compared to the aircraft speed, so wind is considered stationary.
5. The inertial reactive force caused by the aircraft loss of mass due to fuel burn is already taken into account in the propulsive force.
6. The sideslip angle is always zero, since the flight is assumed to be symmetric with respect to the aircraft plane of symmetry.
7. The aircraft is considered symmetric; geometrically, and in terms of mass distribution.
8. The thrust force produced by the engines is contained within the aircraft plane of symmetry.
9. Aeroelastic effects are neglected, since the aircraft is considered a rigid solid.
10. The variation of Mach number and altitude during cruise is assumed quasi-steady.
11. The external shape of the wing is not included in the aerodynamic models.
12. Wing and tail drags are not considered in the computation of the aerodynamic pitch moment, since they are one order of magnitude lower than their lifts, and they act much closer to the center of gravity.
13. The influence of the amount of air bleed extracted from the engines is eliminated, by considering that it is known and constant.
14. Even though the flight path angle is considered in the formulation of the fundamental equations of climb and descent trajectories, the aerodynamic model computes the lift coefficient assuming that the flight path angle is zero [60].
15. The position of the throttle lever in the flight deck is modeled assuming that it is controlled by means of an automatic control law, or engine rating, and not manually.
16. Four different engine ratings are modeled: low idle thrust (LIDL), maximum climb thrust (MCMB), maximum cruise thrust (MCRZ) and maximum take-off thrust (MTKF). However, the possibility of selecting reduced thrust modes in take-off and climb are not considered. There are two means of selecting reduced thrust: the derate thrust setting, and the assumed temperature method.
17. The bank angle is considered in the modeling. However, if the value of the angle is not known this is assumed to be zero.
18. Even though the aerodynamic configuration includes high lift devices, landing gear and speed brakes, it ignores the influence of the position of the control surfaces and horizontal stabilizer.
19. In cruise, the variation of true airspeed with time can be neglected since it is generally very small. In addition, the flight path angle can be considered zero, as altitude changes very slowly.
20. A standard constant gravity field is assumed, which results in identical geodetic and geopotential altitudes.

A.2. Actions Model

The main focus of the hybrid performance tailoring effort is BADA 4 actions model, responsible for the computation of the forces acting on the aircraft, and hence, responsible for its motion. Actions are classified in three main categories: gravitational, aerodynamical and propulsive. They include four forces acting on the aircraft: weight, lift, drag and thrust, together with fuel consumption, the derivative of weight over time [60].

A.2.1. Gravitational Model

The gravitational model includes the weight force, computed as:

$$W = m \cdot g_0 \tag{A.1}$$

Where:
m is the aircraft mass [kg]
g_0 is the gravitational acceleration [m/s^2]

A.2.2. Aerodynamic Model

The aerodynamic model describes the lift coefficient, lift force, drag coefficient and drag force. These aerodynamic forces depend on the atmospheric properties, Mach number and aerodynamic configuration. For this project, only clean configuration was considered.

Lift coefficient

Assuming that the flight path angle is zero, BADA 4 calculates the lift coefficient as follows:

$$C_L = \frac{2mg_0}{\delta p_0 \kappa S M^2 \cos\phi} \tag{A.2}$$

Where:
m is the aircraft mass [kg]
g_0 is the gravitational acceleration [m/s^2]
δ is the pressure ratio [-]

$$\delta = \frac{p}{p_0} \tag{A.3}$$

p_0 is the standard atmospheric pressure at MSL (Mean Sea Level) [Pa]
κ is the adiabatic index of air [-]
S is the wing reference area [m^2]
M is the Mach number [-]
ϕ is the bank angle [rad]

Lift force

Once the lift computed is known, the lift force is determined in the standard manner:

$$L = \frac{1}{2}\delta p_0 \kappa S M^2 C_L \tag{A.4}$$

Where:
δ is the pressure ratio [-]
p_0 is the standard atmospheric pressure at MSL [Pa]
κ is the adiabatic index of air [-]
S is the wing reference area [m^2]
M is the Mach number [-]

Drag coefficient

The drag coefficient is modeled as a function of the lift coefficient and the Mach number. BADA 4 proposes to use equations (3.2-3), (3.2-4), (3.2-5) and (3.2-6) in [60] to compute the drag coefficient for clean configuration, which consider the compressibility of air at high Mach numbers.

Drag force

To finish with the aircraft aerodynamics, the drag force is computed from the drag coefficient in the standard manner:

$$D = \frac{1}{2}\delta p_0 \kappa S M^2 C_D \tag{A.5}$$

Where:
δ is the pressure ratio [-]
p_0 is the standard atmospheric pressure at MSL [Pa]
κ is the adiabatic index of air [-]
S is the wing reference area [m^2]
M is the Mach number [-]

A.2.3. Engine Thrust Model

BADA 4 provides three different thrust models depending on the engine type: turbofan, turboprop or piston engines. This work focuses on the turbofan model, since the used flight data was gathered by aircraft with turbofan engines. Turbofan engines are operated either by direct control of the throttle, or through the use of engine ratings. BADA 4 distinguishes between four ratings: low idle thrust (LIDL), maximum climb thrust (MCMB), maximum cruise thrust (MCRZ) and maximum take-off thrust (MTKF).

Thrust coefficient

If the flight is governed by the idle rating, BADA 4 suggests that the thrust coefficient can directly be computed as a function of the Mach number and the atmospheric conditions, using equation (3.3-2) in [60].

Otherwise, the generalized thrust coefficient depends on the Mach number and the throttle parameter, so first it is necessary to model the throttle parameter. Turbofan engines behave differently if they are operated in the flat-rated area or in the temperature-rated area. Turbofan engines operate in the flat-rated area if the atmospheric conditions result in a temperature deviation lower than the kink point, and in the temperature-rated area otherwise:

$$\delta_T = \begin{cases} \delta_{T,flat} & \text{when } \Delta T \leq \Delta T_{kink} \\ \delta_{T,temp} & \text{when } \Delta T > \Delta T_{kink} \end{cases} \tag{A.6}$$

Where:
- δ_T is the throttle parameter [-]
- $\delta_{T,flat}$ is the throttle parameter in the flat-rated area [-]
- $\delta_{T,temp}$ is the throttle parameter in the temperature-rated area [-]
- ΔT is the temperature differential at MSL [K]
- ΔT_{kink} is the kink point [K]

BADA 4 estimates the throttle parameter of a turbofan aircraft operated in the flat-rated area as a function of the Mach number and the pressure ratio, as it can be seen in equation (3.3-5) in [60]. In the case of a turbofan aircraft operated in the temperature-rated area the throttle parameter is modeled as in equation (3.3-6) in [60]. This throttle parameter also depends on the total temperature ratio (θ_t):

$$\theta_t = \theta \cdot \left(1 + \frac{M^2 \cdot (\kappa - 1)}{2}\right) \tag{A.7}$$

Where:
- θ is the temperature ratio [-]

$$\theta = \frac{T}{T_0} \tag{A.8}$$

- T_0 is the standard atmospheric temperature at MSL [K]
- M is the Mach number [-]
- κ is the adiabatic index of air [-]

Once the throttle parameter has been modeled, the thrust coefficient is calculated using equation (3.3-3) in [60]. This formula is applicable when no engine rating is used (in other words, when the aircraft is operated through direct throttle parameter input), and when non-idle ratings (MCMB, MCRZ, MTKF) are utilized.

Thrust force

Finally, the thrust force is calculated from the thrust coefficient:

$$Th = \delta W_{m_{ref}} C_T \tag{A.9}$$

Where:
- δ is the pressure ratio [-]
- m_{ref} is the reference mass [kg]
- $W_{m_{ref}}$ is the weight force at m_{ref} [N]
- C_T is the thrust coefficient [-]

A.2.4. Fuel Consumption Model

Even though fuel consumption is not a force (or action), it affects aircraft mass and therefore, its weight. Thus, it is included in BADA 4 actions model. As in the case of the engine thrust model, BADA 4 differentiates between three separate fuel consumption models depending on the type of engine. Once again, the turbofan model is the model of interest. The fuel consumption model also varies depending on the engine rating.

Fuel coefficient

If the aircraft is operated through the idle rating, BADA 4 computes the fuel coefficient as a function of the Mach number and the atmospheric conditions, as indicated in equation (3.4-3) in [60].

Otherwise, the fuel coefficient is modeled as a function of the general thrust coefficient and Mach number, using equation (3.4-4) in [60]. Note that this expression is valid when the engine rating is different than idle (e.g. MCMB, MCRZ or MTKF), and when the aircraft is flown through a direct throttle parameter input.

Fuel flow

To finish with BADA 4 actions model, fuel flow is computed as:

$$F = \delta\theta^{\frac{1}{2}} W_{m_{ref}} a_0 L_{HV}^{-1} C_F \tag{A.10}$$

As mentioned earlier, the coefficients in BADA models were computed to achieve the best fit between calculated and reference performance data. Accuracy of current BADA models rely on the quality of the reference data with which the coefficients were identified. This explains why in many occasions some of the coefficients are equal to zero, and the polynomials result in simpler expressions.

A.3. Motion Model

The motion model is also a fundamental pillar in the development of a hybrid methodology to model tail-specific performance. The motion model describes the variation of those variables that describe the aircraft trajectory, such as the center of gravity position and velocity or the aircraft mass, with time. BADA 4 models the aircraft motion through the Total-Energy Model (TEM), based on the forces described in Section A.2.

The TEM equates the rate of work done by the forces acting on the aircraft to the rate of increase in potential and kinetic energy, as follows [60]:

$$(Th - D) \cdot v_{TAS} = mg_0 \frac{dh}{dt} + mv_{TAS}\frac{dv_{TAS}}{dt} \tag{A.11}$$

Where:

Th is the thrust acting parallel to the aircraft velocity vector [N]
D is the aerodynamic drag force [N]
m is the aircraft mass [kg]
h is the geodetic altitude [m]
g_0 is the gravitational acceleration [m/s^2]
v_{TAS} is the true airspeed [m/s]
$\frac{d}{dt}$ is the time derivative [s^{-1}]

Equation A.11 is commonly used to compute the corresponding rate of climb or descent (ROCD). Rearranging the equation and isolating the vertical speed in the left hand side gives:

$$\frac{dh}{dt} = \frac{(Th - D) \cdot v_{TAS}}{mg_0}\left[1 + \left(\frac{v_{TAS}}{g_0}\right) \cdot \left(\frac{dv_{TAS}}{dh}\right)\right]^{-1} \tag{A.12}$$

The assumption of standard constant gravity results in identical geodetic and geopotential altitude. Thus:

$$\frac{dH_p}{dt} = \frac{T - \Delta T}{T} \cdot \frac{(Th - D) \cdot v_{TAS}}{mg_0} \cdot \left[1 + \left(\frac{v_{TAS}}{g_0}\right) \cdot \left(\frac{dv_{TAS}}{dh}\right)\right]^{-1} \tag{A.13}$$

Where:

T is the atmosphere temperature [K]

ΔT is the temperature differential [K]

Once the variation of the aircraft geopotential pressure altitude with time is known, the rate of climb (ROC) and rate of descent (ROD) can be computed as follows:

$$ROC = \frac{dH_p}{dt} \tag{A.14}$$

$$ROD = -\frac{dH_p}{dt} \tag{A.15}$$

The last term of equation A.13 corresponds to the energy share factor (ESF), which represents how much of the available power is allocated to climb and not to acceleration.

B

Purely Data-driven Models: Feature and Model Selection

B.1. Tuned Hyperparameters

Table B.1: Hyperparameters of adaptive boosting regressors tuned in the developed model-selection routine

Hyperparameter	Definition [67]
learning_rate	Rate that shrinks the contribution of each tree
n_estimators	Maximum number of estimators at which boosting is terminated
loss	Loss function to use when updating the weights for each estimator after each boosting iteration

Table B.2: Hyperparameters of gradient boosting regressors tuned in the developed model-selection routine

Hyperparameter	Definition [67]
learning_rate	Rate that shrinks the contribution of each tree
max_depth	Maximum depth of the individual regression estimators
max_leaf_nodes	Total number of terminal nodes in a tree
min_samples_leaf	Minimum number of samples required to be at a terminal node
min_samples_split	Minimum number of samples required to split an internal node
n_estimators	Number of boosting stages to perform

Table B.3: Hyperparameters of extreme gradient boosting regressors tuned in the developed model-selection routine

Hyperparameter	Definition [68]
alpha	L1 (Lasso regression) regularization term on weights
gamma	Minimum loss reduction required to make further partitions on leaf nodes
lambda	L2 (Ridge regression) regularization term on weights
learning_rate	Rate that shrinks the contribution of each tree
max_depth	Maximum depth of a tree
min_child_weight	Minimum sum of instance weight (hessian) needed in a child
n_estimators	Number of boosting stages to perform

Table B.4: Hyperparameters of random forest regressors tuned in the developed model-selection routine

Hyperparameter	Definition [67]
max_depth	Maximum depth of the tree
max_leaf_nodes	Total number of terminal nodes in a tree
min_samples_leaf	Minimum number of samples required to be at a terminal node
min_samples_split	Minimum number of samples required to split an internal node
n_estimators	Number of trees in the forest

Table B.5: Hyperparameters of extreme random forest regressors tuned in the developed model-selection routine

Hyperparameter	Definition [67]
max_depth	Maximum depth of the tree
max_leaf_nodes	Total number of terminal nodes in a tree
min_samples_leaf	Minimum number of samples required to be at a terminal node
min_samples_split	Minimum number of samples required to split an internal node
n_estimators	Number of trees in the forest

B.2. Selected Features

Climb

Table B.6: Initial and selected features to model the climb fuel flow of the narrow- and wide-body aircraft

	Narrow-body aircraft				Wide-body aircraft			
	ML 2		ML 3		ML 2		ML 3	
	Initially	Selected	Initially	Selected	Initially	Selected	Initially	Selected
Gross weight	X	X	X	X	X	X	X	X
Mach number	X	X	X	X	X	X	X	X
Pressure altitude	X	X	X	X	X	X	X	X
Total air temperature	X	X	X	X	X	X	X	X
True airspeed	X	X	X	X	X	X	X	X
Vertical speed	X	X	X	X	X	X	X	
Flight path angle	X	X	X	X	X	X	X	X
Angle of attack	X	X	X	X	X	X	X	X
Course angle	X	X	X	X				
Drift angle	X	X	X	X	X	X	X	
Track angle	X	X	X	X	X	X	X	X
Longitudinal acceleration	X	X	X	X	X		X	
Vertical acceleration	X		X		X		X	
Aileron position	X		X		X	X	X	
Elevator position	X	X	X	X	X		X	
Rudder position	X		X		X		X	
Stabilizer position	X	X	X	X	X	X	X	X
Ground speed	X	X	X	X	X	X	X	X
Heading angle	X	X	X	X	X	X	X	
Wind speed	X	X	X	X	X	X	X	X
Wind direction	X	X	X	X	X	X	X	X
Thrust levers	X	X			X			
N1	X	X			X	X		
N2	X	X			X	X		

Cruise

Table B.7: Initial and selected features to model the cruise fuel flow of the narrow- and wide-body aircraft

	Narrow-body aircraft				Wide-body aircraft			
	ML 2		ML 3		ML 2		ML 3	
	Initially	Selected	Initially	Selected	Initially	Selected	Initially	Selected
Gross weight	X	X	X	X	X	X	X	X
Mach number	X	X	X	X	X	X	X	X
Pressure altitude	X	X	X	X	X	X	X	X
Total air temperature	X	X	X	X	X	X	X	X
True airspeed	X	X	X	X	X	X	X	X
Vertical speed	X		X		X		X	
Flight path angle	X		X		X		X	
Angle of attack	X	X	X	X	X	X	X	X
Course angle	X	X	X	X				
Drift angle	X	X	X	X	X		X	X
Track angle	X	X	X	X	X		X	X
Longitudinal acceleration	X	X	X	X	X		X	X
Vertical acceleration	X		X		X		X	
Aileron position	X		X		X		X	
Elevator position	X	X	X		X		X	X
Rudder position	X		X		X		X	
Stabilizer position	X	X	X	X	X		X	X
Ground speed	X	X	X	X	X		X	X
Heading angle	X	X	X	X	X		X	X
Wind speed	X	X	X	X	X		X	X
Wind direction	X	X	X	X	X		X	X
Thrust levers	X	X			X	X		
N1	X	X			X	X		
N2	X	X			X	X		

Descent

Table B.8: Initial and selected features to model the descent fuel flow of the narrow- and wide-body aircraft

	Narrow-body aircraft				Wide-body aircraft			
	ML 2		ML 3		ML 2		ML 3	
	Initially	Selected	Initially	Selected	Initially	Selected	Initially	Selected
Gross weight	X		X	X	X	X	X	X
Mach number	X	X	X	X	X	X	X	X
Pressure altitude	X	X	X	X	X	X	X	X
Total air temperature	X	X	X	X	X	X	X	X
True airspeed	X	X	X	X	X		X	
Vertical speed	X		X		X		X	
Flight path angle	X		X	X	X		X	
Angle of attack	X		X		X		X	
Course angle	X		X	X				
Drift angle	X		X		X		X	
Track angle	X	X	X	X	X	X	X	X
Longitudinal acceleration	X		X		X		X	
Vertical acceleration	X		X		X		X	
Aileron position	X		X		X		X	
Elevator position	X		X		X		X	
Rudder position	X		X		X		X	
Stabilizer position	X		X		X	X	X	X
Ground speed	X	X	X	X	X	X	X	
Heading angle	X	X	X	X	X	X	X	X
Wind speed	X		X		X		X	
Wind direction	X		X		X	X	X	
Thrust levers	X				X	X		
N1	X	X			X	X		
N2	X	X			X	X		

B.3. Selected Models

Narrow-body Aircraft

Table B.9: Ensemble methods and hyperparameters selected to model the fuel flow of the narrow-body aircraft in climb, cruise and descent

	Climb	Cruise	Descent
ML 1	Extreme random forest • max_depth = 25 • max_leaf_nodes = None • min_samples_leaf = 1 • min_samples_split = 2 • n_estimators = 500	Extreme random forest • max_depth = 40 • max_leaf_nodes = None • min_samples_leaf = 1 • min_samples_split = 6 • n_estimators = 300	Extreme gradient boosting • alpha = 0 • gamma = 0 • lambda = 1 • learning_rate = 0.9 • max_depth = 40 • min_child_weight = 1 • n_estimators = 700
ML 2	Extreme random forest • max_depth = 25 • max_leaf_nodes = None • min_samples_leaf = 1 • min_samples_split = 2 • n_estimators = 400	Extreme random forest • max_depth = 25 • max_leaf_nodes = None • min_samples_leaf = 2 • min_samples_split = 8 • n_estimators = 500	Extreme random forest • max_depth = 30 • max_leaf_nodes = None • min_samples_leaf = 1 • min_samples_split = 2 • n_estimators = 500
ML 3	Extreme random forest • max_depth = 30 • max_leaf_nodes = None • min_samples_leaf = 1 • min_samples_split = 2 • n_estimators = 300	Extreme random forest • max_depth = 30 • max_leaf_nodes = None • min_samples_leaf = 1 • min_samples_split = 2 • n_estimators = 500	Extreme random forest • max_depth = 25 • max_leaf_nodes = None • min_samples_leaf = 1 • min_samples_split = 2 • n_estimators = 500

Wide-body aircraft

Table B.10: Ensemble methods and hyperparameters selected to model the fuel flow of the wide-body aircraft in climb, cruise and descent

	Climb	Cruise	Descent
ML 1	Extreme random forest • max_depth = 25 • max_leaf_nodes = None • min_samples_leaf = 1 • min_samples_split = 2 • n_estimators = 500	Random forest • max_depth = 25 • max_leaf_nodes = 15000 • min_samples_leaf = 2 • min_samples_split = 12 • n_estimators = 600	Extreme gradient boosting • alpha = 0 • gamma = 0 • lambda = 1 • learning_rate = 0.8 • max_depth = 20 • min_child_weight = 2 • n_estimators = 1000
ML 2	Extreme random forest • max_depth = 25 • max_leaf_nodes = None • min_samples_leaf = 1 • min_samples_split = 2 • n_estimators = 500	Extreme random forest • max_depth = 40 • max_leaf_nodes = None • min_samples_leaf = 1 • min_samples_split = 4 • n_estimators = 500	Extreme random forest • max_depth = 25 • max_leaf_nodes = None • min_samples_leaf = 1 • min_samples_split = 2 • n_estimators = 400
ML 3	Extreme random forest • max_depth = 25 • max_leaf_nodes = None • min_samples_leaf = 1 • min_samples_split = 2 • n_estimators = 500	Extreme random forest • max_depth = 50 • max_leaf_nodes = None • min_samples_leaf = 1 • min_samples_split = 2 • n_estimators = 200	Extreme random forest • max_depth = 40 • max_leaf_nodes = None • min_samples_leaf = 1 • min_samples_split = 2 • n_estimators = 400

Bibliography

[1] S. K. Ojha. *Flight Performance of Aircraft*, chapter 1. Aircraft, Basic Forces, and Performance. American Institute of Aeronautics and Astronautics, 1995.

[2] *FAA-H-8083-25B. Pilot's Handbook of Aeronautical Knowledge*, chapter 11. Aircraft Performance. United States Department of Transportation. Federal Aviation Administration, 2016.

[3] A. Nuic, D. Poles, and V. Mouillet. BADA: An advanced aircraft performance model for present and future ATM systems. *International Journal of Adaptive Control and Signal Processing*, 24(10): 850–866, 2010.

[4] International Air Transport Association. IATA forecast predicts 8.2 billion air travelers in 2037. Press Release No.: 62. Available at: https://www.iata.org/pressroom/pr/Pages/2018-10-24-02.aspx, 24 October 2018. Accessed on: 15/02/19.

[5] K. Krajček, D. Nikolić, and A. Domitrović. Aircraft performance monitoring from flight data. *Tehnicki vjesnik/Technical Gazette*, 22(5), 2015.

[6] *Getting trips with aircraft performance monitoring*. AIRBUS. Flight Operations Support and Line Assistance, December 2002.

[7] E. Chu, D. Gorinevsky, and S. Boyd. Detecting aircraft performance anomalies from cruise flight data. In *AIAA Infotech@ Aerospace 2010*, page 3307. 2010.

[8] A. Nuic, C. Poinsot, M. Iagaru, E. Gallo, F. A. Navarro, and C. Querejeta. Advanced aircraft performance modeling for ATM: Enhancements to the BADA model. In *24th Digital Avionics System Conference*, pages 1–14. AIAA/IEEE Washington, DC, 2005.

[9] A. Nuic. User Manual for the Base of Aircraft Data (BADA) Revision 3.12. Technical Report 14/04/24-44, EUROCONTROL, 2014.

[10] E. Gallo, F. A. Navarro, A. Nuic, and M. Iagaru. Advanced aircraft performance modeling for ATM: BADA 4.0 results. In *25th Digital Avionics System Conference*, pages 1–12. AIAA/IEEE Portland, Oregon, 2006.

[11] D. Poles, A. Nuic, and V. Mouillet. Advanced aircraft performance modeling for ATM: Analysis of BADA model capabilities. In *29th Digital Avionics System Conference*. AIAA/IEEE Salt Lake City, Utah, 2010.

[12] BADA: aircraft performance model. [Online]. Available at: https://simulations.eurocontrol.int/solutions/bada-aircraft-performance-model/. Accessed on: 25/03/19.

[13] A. Harada, Y. Miyamoto, Y. Miyazawa, and K. Kunabiki. Accuracy Evaluation of an Aircraft Performance Model with Airliner Flight Data. *Transactions of the Japan Society for Aeronautical and Space Sciences, Aerospace Technology Japan*, 11:79–85, 2013.

[14] V. Mouillet, A. Nuic, E. Casado, and J. López. Evaluation of the Applicability of a Modern Aircraft Performance Model to Trajectory Optimization. In *2018 IEEE/AIAA 37th Digital Avionics Systems Conference (DASC)*, pages 257–265. IEEE, 2018.

[15] A. Suchkov, S. Swierstra, and A. Nuic. Aircraft performance modeling for air traffic management applications. In *5th USA/Europe Air Traffic Management Research and Development Seminar*, pages 23–27, 2003.

[16] H. Erzberger. CTAS: Computer Intelligence for Air Traffic Control in the Terminal Area. *Technical Memorandum. National Aeronautics and Space Administration*, July 1992.

[17] R. Vivona and K. T. Cate. Center-TRACON Automation System (CTAS) En-route Trajectory Predictor. Requirements and Capabilities. *National Aeronautics and Space Administration*, 2013.

[18] M. Abramson and K. Ali. Integrating the Base of Aircraft Data (BADA) in CTAS Trajectory Synthesizer. *National Aeronautics and Space Administration*, September 2012.

[19] I. Metz, J. Hoekstra, J. Ellerbroek, and D. Kügler. Aircraft performance for open air traffic simulations. In *AIAA Modeling and Simulation Technologies Conference*, page 3522, 2016.

[20] T. W. Gloudemans. Aircraft Performance Parameter Estimation using Global ADS-B and Open Data. Master Thesis, Delft University of Technology, 2016.

[21] J. Sun, J. Ellerbroek, and J. M. Hoekstra. Modeling aircraft performance parameters with open ADS-B data. In *Proceedings of the 12th USA/Europe Air Traffic Management Research and Development Seminar*. FAA/EUROCONTROL, 2017.

[22] J. Sun, J. Ellerbroek, and J. M. Hoekstra. WRAP: An open-source kinematic aircraft performance model. *Transportation Research Part C: Emerging Technologies*, 98:118–138, 2019.

[23] J. Sun, J. Ellerbroek, and J. M. Hoekstra. Aircraft initial mass estimation using Bayesian inference method. *Transportation Research Part C: Emerging Technologies*, 90:59–73, 2018.

[24] M. Kaiser, M. Schultz, and H. Fricke. Enhanced Jet Performance Model for High Precision 4D Flight Path Prediction. In *Proceedings of the 1st International Conference on Application and Theory of Automation in Command and Control Systems*, pages 33–40. IRIT Press, 2011.

[25] J. Rosenow and H. Fricke. *Flight performance modeling to optimize trajectories*. Deutsche Gesellschaft für Luft-und Raumfahrt-Lilienthal-Oberth eV, 2016.

[26] M. D. Dupuy and M. Porretta. Preliminary results for a robust trajectory prediction method using advanced flight data. In *2007 IEEE/AIAA 26th Digital Avionics Systems Conference*, pages 1–C. IEEE, 2007.

[27] W. Schuster, W. Ochieng, and M. Porretta. High-performance trajectory prediction for civil aircraft. In *29th Digital Avionics Systems Conference (DASC)*, pages 1–C. IEEE, 2010.

[28] M. Porretta, M. D. Dupuy, W. Schuster, A. Majumdar, and W. Ochieng. Performance Evaluation of a Novel 4D Trajectory Prediction Model for Civil Aircraft. *The Journal of Navigation*, 61(3): 393–420, 2008.

[29] S. Gillet, A. Nuic, and V. Mouillet. Enhancement in realism of ATC simulation by improving aircraft behaviour models. In *29th Digital Avionics Systems Conference (DASC)*, pages 2–D. IEEE, 2010.

[30] Fang H., Lishuai L., and Weizun Z. Aircraft Mass Estimation using Quick Access Recorder Data. In *2018 IEEE/AIAA 37th Digital Avionics Systems Conference (DASC)*, pages 63–67. IEEE, 2018.

[31] R. Alligier, D. Gianazza, and N. Durand. Learning the aircraft mass and thrust to improve the ground-based trajectory prediction of climbing flights. *Transportation Research Part C: Emerging Technologies*, 36:45–60, 2013.

[32] M. Hrastovec and F. Solina. Machine Learning Model for Aircraft Performances. In *2014 IEEE/AIAA 33rd Digital Avionics Systems Conference (DASC)*, pages 8C4–1. IEEE, 2014.

[33] M. Hrastovec and F. Solina. Prediction of aircraft performances based on data collected by air traffic control centers. *Transportation Research Part C: Emerging Technologies*, 73:167–182, December 2016.

[34] R. Alligier, D. Gianazza, and N. Durand. Machine Learning and Mass Estimation Methods for Ground-based Aircraft Climb Prediction. *IEEE Transactions on Intelligent Transportation System*, 16(6):3138–3149, 2015.

[35] R. Alligier, D. Gianazza, and N. Durand. Machine Learning Applied to Airspeed Prediction During Climb. In *ATM seminar 2015, 11th USA/Europe Air Traffic Management Research and Development Seminar*, 2015.

[36] Y. S. Chati and H. Balakrishnan. Statistical Modeling of Aircraft Engine Fuel Flow Rate. In *30th Congress of the International Council of the Aeronautical Sciences*, 2016.

[37] Y. S. Chati and H. Balakrishnan. A Gaussian Process Regression Approach to Model Aircraft Engine Fuel Flow Rate. In *Proceedings of the 8th International Conference on Cyber-Physical Systems*, pages 131–140. ACM, 2017.

[38] Zhen P., Chenczhi ., and Jingkai Z. A Model of Fuel Consumption Estimation and Abnormality Detection based on Airplane Flight Data Analysis. In *2018 IEEE/AIAA 37th Digital Avionics Systems Conference (DASC)*, pages 251–256. IEEE, 2018.

[39] J. Sun, J. Ellerbroek, and J. M. Hoekstra. Large-Scale Flight Phase Identification from ADS-B Data Using Machine Learning Methods. In *7th International Conference on Research in Air Transportation*, 2016.

[40] J. Sun, J. Ellerbroek, and J. M. Hoekstra. Flight Extraction and Phase Identification for Large Automatic Dependent Surveillance-Broadcast Datasets. *Journal of Aerospace Information Systems*, 14(10):566–572, 2017.

[41] S. Barker. Aircraft Performance Monitoring. In *Airbus, Regional Seminar*, 2013.

[42] M. Foueri. Aircraft Performance Degradation. In *16th Performance and Operations Conference*, 2009.

[43] M. Linder, J. Rosenow, and H. Fricke. Potential of Integrated Aircraft Rotation and Flight Scheduling by using Individual Tail Sign Performance. In *Deutscher Kongress für Luft- und Raumfahrt*. Braunschweig, 2016.

[44] AIAA. Guide for the Verification and Validation of Computational Fluid Dynamics Simulations. *American Institute of Aeronautics and Astronautics, Reston, VA*, (AIAA G-077-1998), 1998.

[45] T. G. Puranik, E. D. Harrinson, S. Min, I. Chakraborty, and D. N. Mavris. A Framework for General Aviation Aircraft Performance Model Calibration and Validation. In *2018 Aviation Technology, Integration, and Operations Conference*, page 3191, 2018.

[46] E. Casado, C. Goodchild, and M. Vilaplana. Sensitivity of Trajectory Prediction Accuracy to Aircraft Performance Uncertainty. In *Conference, Guidance, Navigation, and Control and Co-located Conferences*, number AIAA 2013-5045. AIAA, September 2013.

[47] E. Casado. Method for Modeling Aircraft Performance Through Adaptive Aircraft Performance Models, July, 30 2015. US Patent. No. 20150210405A1.

[48] A. Alizadeh, N. Humaira, and E. Koyuncu. Learning-based Aircraft Trajectory Planning Enhancement. In *AIAA Scitech 2019 Forum*, page 0138, 2019.

[49] M. Uzun, M. U. Demirezen, E. Koyuncu, and G. Inalhan. Design of a Hybrid Digital-twin Flight Performance Model through Machine Learning. In *2019 IEEE Aerospace Conference*, 2019.

[50] S. Raschka and V. Mirjalili. *Python Machine Learning*. Packt Publishing Ltd, 2 edition, 2017. ISBN: 978-1-78712-593-3.

[51] W. Koehrsen. Random forest in python. *Medium*, December 2017.

[52] Y. S. Abu-Mostafa, M. Magdon-Ismail, and L. Hsuan-Tien. *Learning From Data: A short Course*. AMLbook.com, 2012. ISBN: 978-1-60049-006-4.

[53] A. Géron. *Hands-On Machine Learning with Scikit-Learn & TensorFlow: concepts, tools, and techniques to build intelligent systems*. O'Reilly Media, Inc., 2017. ISBN: 978-1-491-96229-9.

[54] G. James annd D. Witten, T. Hastie, and R. Tibshirani. *An Introduction to Statistical Learning with Applications in R*. Springer, 2015. ISBN: 978-1-4614-7137-0.

[55] A. M. P. de Leege, M. M. van Passen, and M. Mulder. A Machine Learning Approach to Trajectory Prediction. In *AIAA Guidance, Navigation, and Control (GNC) Conference*, page 4782, 2013.

[56] Xiangmin G., Renli L., Liang S., and Yang L. A study of 4D trajectory prediction based on machine deep learning. In *2016 12th World Congress on Intelligent Control and Automation (WCICA)*, pages 24–27. IEEE, 2016.

[57] C. S. Kern, I. P. de Medeiros, and T. Yoneyama. Data-Driven Aircraft Estimated Time of Arrival Prediction. In *2015 Annual IEEE Systems Conference (SysCon) Proceedings*, pages 727–733. IEEE, 2015.

[58] Zhengyi W., Man L., and D. Delahaye. A hybrid machine learning model for short-term estimated time of arrival prediction in terminal manoeuvring area. *Transportation Research Part C: Emerging Technologies*, 95:280–294, 2018.

[59] Kerosene prices. https://www.globalpetrolprices.com/kerosene_prices/. Accessed: 19-11-2019.

[60] A. Nuic and V. Mouillet. User Manual for the Base of Aircraft Data (BADA) Family 4. Technical Report 12/11/22-58, EUROCONTROL, March 2016.

[61] Concept Document for the Base of Aircraft Data (BADA) Family 4. Technical Report 12/11/22-57, EUROCONTROL, November 2012.

[62] C. Querejeta E. Gallo and M. A. Vilaplana. AMEBA Improvement Report. Technical report, Boeing Research & Technology Europe, May 2005.

[63] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer science & business media, 2013.

[64] C. Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5):206–215, 2019.

[65] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik. Gene selection for cancer classification using support vector machines. *Machine learning*, 46(1-3):389–422, 2002.

[66] J. H. Torrie. *Principles and Procedures of Statistics: with Special Reference to the Biological Sciences*. McGraw-Hill, 1960.

[67] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[68] Xgboost documentation. https://xgboost.readthedocs.io/en/latest/index.html. Accessed: 16-02-2020.