QAR Data-Driven Calibration of Physics-based Aircraft Performance Models using a Machine-Learning Approach

del Pozo Domínguez, María ; López Leonés, Javier; Roling, P.C.

DOI
10.2514/6.2023-4498

Publication date
2023

Document Version
Final published version

Published in
AIAA AVIATION 2023 Forum

AIAA AVIATION Forum
12-16 June 2023, San Diego, CA and Online
AIAA AVIATION 2023 Forum

10.2514/6.2023-4498

QAR Data-Driven Calibration of Physics-based Aircraft Performance Models using a Machine-Learning Approach

María del Pozo Domínguez *
MCA Group, 28020 Madrid, Spain

Javier López Leonés †
Boeing Research & Technology - Europe, 28042 Madrid, Spain

Paul Roling ‡
Delft University of Technology, 2629 HS Delft, The Netherlands

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 research work proposes a novel approach to develop up-to-date, tail-specific performance models based on the use of Quick Access Recorder (QAR) data and machine-learning techniques. In particular, a methodology was 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.

I. Nomenclature

C_L	=	Lift coefficient
C_D	=	Drag coefficient
C_T	=	Thrust coefficient
$C_{T,idle}$	=	Thrust coefficient for idle rating
C_F	=	Fuel coefficient
$C_{F,idle}$	=	Fuel flow coefficient for idle rating
FF	=	Fuel flow
H_P	=	Pressure altitude
δ_T	=	Throttle parameter
D	=	Drag force
Th	=	Thrust force
m	=	Aircraft mass
γ	=	Path angle
T	=	Temperature
M	=	Mach number

*Data Scientist, MCA Group - Spain, C. Orense 12, 28020 Madrid, Spain.
†Associate Technical Fellow, Airspace Operational Efficiency & Autonomy, Boeing Research & Technology - Europe, Av. Sur del Aeropuerto de Barajas, 38, 28042 Madrid, Spain.
‡Lecturer, Department of Air Transport and Operations, Faculty of Aerospace Engineering, Delft University of Technology, 2629 HS Delft, The Netherlands.

II. Introduction

The discipline of aircraft performance plays a key role in aviation, being a determinant factor in aircraft design, certification and operation. 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. Additionally, aircraft performance modeling is essential in activities like post-fight analytics, trajectory prediction, flight plan generation, air traffic management, and aircraft conflict detection and resolution, among others. Therefore, it is obvious that having realistic and up-to-date aircraft performance information, and those tools and methods to derive that information is a clear necessity in aviation.

Multiple aircraft performance models exist with varying degrees of accuracy for different applications. One of the most widely accepted models is BADA [1, 2], a mass-varying, kinetic model developed and maintained by EUROCONTROL. BADA currently supports over 80% of the aircraft types in commercial aviation, and it is broadly used within the domain of ATM. BADA performance models were derived from performance data generated by the manufacturers with performance software. Reference performance data used in the development of BADA models corresponded to brand-new aircraft, and characterized an aircraft family. Thus, BADA models describe the 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 and create the need for aircraft performance monitoring, and ultimately for tailored performance modeling. The higher computing capabilities, higher quality data and more sophisticated analysis techniques that exist nowadays encourage the consideration of these factors in pursuit of enhanced aircraft performance models.

Several studies propose using different sources of data and machine-learning techniques to predict aircraft performance parameters. For example, M. Hrastovec et al. [3] presented a machine-learning model to predict aircraft performance using radar data recordings, accompanied by weather information and flight plans. Some other efforts in the field aimed to make predictions closer to real aircraft performance by predicting BADA input parameters. For example, the aircraft initial mass, which determines the mass evolution, and consequently fuel consumption and trip cost. Fang He et al. [4] 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. R. Alligier et al. [5] improved the accuracy of trajectory prediction of climbing flights by inferring the mass and thrust from recorded weather and radar data. First, the 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 the past points. Then, the thrust setting law is computed iteratively using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method, coupled to the mass estimation model. As a continuation to this work, R. Alligier et al. [6] developed a machine learning regression model to estimate aircraft mass for aircraft climb prediction. M. Hrastovec et al. [7] also used machine learning to predict BADA input parameters, specifically the k-Nearest Neighbor (k-NN) algorithm. With the model proposed predictions could be made fast enough to be used in real-time applications, but the accuracy of BADA nominal models is not improved drastically. The purpose of these studies is to substitute standard BADA input parameters, rather than to enhance BADA models.

Fuel consumption also plays an important role in aircraft performance modeling. A realistic modeling of fuel consumption and fuel flow rate is a concern for airlines, since they drive their direct operating costs. In addition, it is important to assess engine performance, as well as to estimate aircraft emissions. Existing performance models like BADA have shown insufficient capabilities in fuel flow modeling, especially when high level of accuracy is required. There are multiple studies whose goal is to model fuel flow using historical data. Remarkable are the activities carried out by the Aerospace Research Center (ARC) of the Istanbul Technical University (ITU). In [8], they presented a regression model to estimate fuel flow using deep learning and QAR data. The model consists of an Artificial Neural Network (ANN) that uses eight variables as inputs, but two out of the eight features correspond to the throttle levers, which directly represent thrust. As expected, thrust information proved to be important in the learning of the fuel flow and directly correlated with it. Throttle lever position is available in the QAR data however its evolution will depend on the flying conditions, being difficult to estimate it beforehand. Therefore a fuel consumption model based on throttle might produce large errors when used to calculate the fuel for an entire flight or used in “"what-if" simulations. In March 2019, M. Uzun et al. presented their last achievements in building a hybrid data-driven model [9] in where BADA model is used to estimate the thrust force based on the flying conditions and fuel flow is learnt from historical data. Based on this work, they developed a two-layer predictor in which the first model is in charge of estimating the thrust levers and the second model predicts the fuel flow.

Regarding the efforts focused on enhancing BADA models found in the literature, a noteworthy study was carried out by E. Casado et al. [10], who proposed a methodology to evaluate the uncertainty consequence of the use of BADA, and suggested to modify the independent coefficients of some of the models to better represent actual performance. In order 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.

Despite the many efforts that have been made to enhance BADA's estimations, as well as to build machine-learning models to predict performance parameters, no one proposes to combine BADA and machine learning to generate re-calibrated and more realistic models. The goal of this research work is to design a complete methodology that allows to calibrate BADA 4, the last family of BADA, based on historical QAR data and with the aid of 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. The resulting performance models must 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. In [11], the authors of this paper compared the performance of non-parametric machine-learning regressors trained to predict fuel flow values with BADA calibrated models, and assess the importance of various flight parameters on performance modeling by analyzing their impact on fuel flow prediction accuracy. Those data-driven models can help to predict performance parameters without modeling the underlying physics, reducing the complexity of the formulation and eventually capturing more complex relationships between input and output parameters that are commonly discarded or simplified in formulations like BADA.

The remainder of the paper is organized as follows. First, a brief overview of the structure of BADA is given in Section III. Then, the proposed methodology is described in Section IV, which includes an explanation of all the necessary steps to obtain accurate and robust performance models. Section V presents and discusses the results obtained by comparing recorded flight data with predictions calculated using BADA nominal models and BADA calibrated models. Finally, Section VI brings together the main conclusions on the work presented in this paper.

III. BADA 4: Model Overview

BADA consists of two models: Airline Procedure Model (ARPM) and Aircraft Performance Model (APM). BADA ARPM provides information on nominal aircraft operations based on information about the specific airspace procedures and operating policies of local airlines. BADA APM, focus of this work, is based on a mass-varying kinetic approach. It represents the aircraft as a point and models the underlying forces that cause aircraft motion. BADA APM is divided into four sub-models named Actions, Motion, Operations and Limitations. The Actions model computes the forces responsible for the aircraft motion, together with the fuel consumption. The Motion model contains the set of differential equations that provide the variation with time of the aircraft position, velocity and mass. The Operations model includes the instructions about the way the aircraft is typically operated. The Limitations model confines the aircraft behavior to ensure safe operation, by indicating when the aircraft would be outside the flight envelope. The structure of BADA is presented in Figure 1.

The main focus of this work is the 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. The Actions model 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.

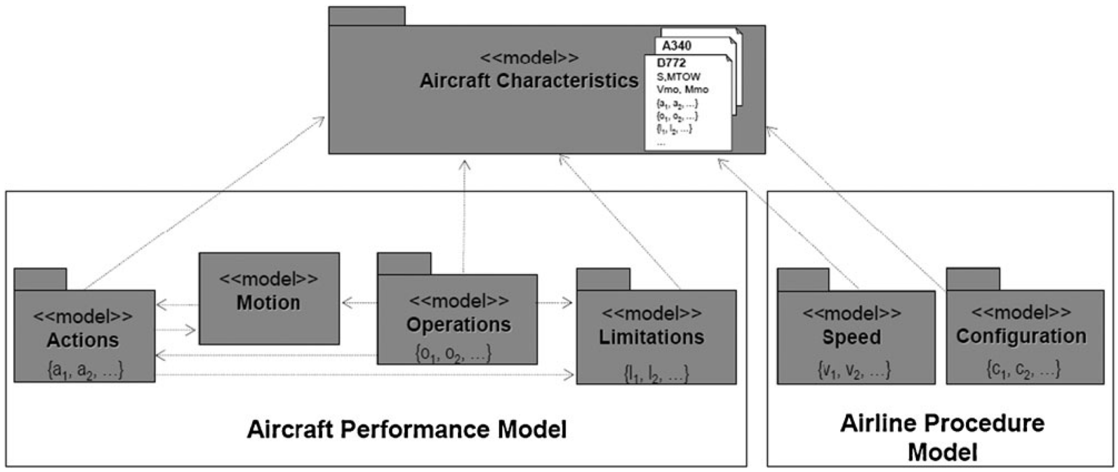

Fig. 1 Structure of BADA model [2]

IV. Methodology

This paper proposes a methodology to identify BADA 4 coefficients using QAR data as reference performance data using a set of machine-learning techniques. The methodology consists of four phases: data ingestion, data preparation, tailoring process and model evaluation. Figure 2 summarizes the main steps required to calibrate BADA nominal models, addressed one by one in this section.

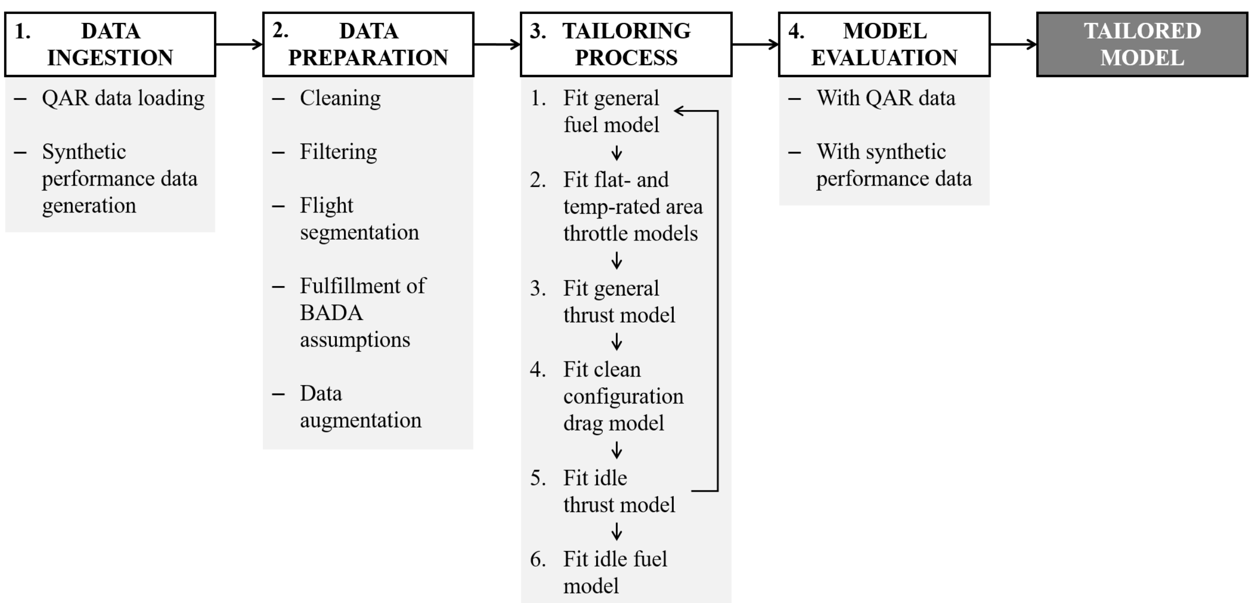

Fig. 2 Schema of the methodology developed to calibrate BADA 4 coefficients, including the main steps: data ingestion, data preparation, tailoring process and model evaluation, as well as some intermediate steps

A. Data Ingestion

Gathering data is the first step in every data-driven activity. It is needed in order to create accurate predictive models, to successfully perform real-time optimizations and to obtain relevant conclusions from post-analyses. None of these activities is 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, essential to derive performance models that are applicable over the entire flight envelope. 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 of features 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 QAR dataset, and the latter are: pressure altitude, aircraft gross weight, total air temperature, Mach number, flight path angle, and 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 work, BADA nominal models were utilized. BADA limitations model [1], 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.

B. 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. Flights with erroneous signals are discarded and outliers are eliminated. 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.

Before starting the tailoring process, it is necessary to segment flights appropriately, since different BADA models are applicable in climb, cruise and descent regimes. As any other physical model, BADA is built on multiple assumptions, so to be faithful to BADA modeling principles the preprocessing phase includes the necessary steps to guarantee that the QAR datasets used in the tailoring process fulfill as many assumptions as possible.

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 this instability of the models is to include synthetic points. 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. A set of recommendations regarding what and how many synthetic points to include was formulated. In the climb phase it is suggested to include 5 % of synthetic points with respect to the total number of QAR climb points, whereas in cruise and descent it might be sufficient to add 1 %, respectively. As mentioned above, among the total number of synthetic points added to each flight phase, the vertices of the flight envelope should be included. From the remaining points, 50 % should lie on the limits and the other 50 % should be uniformly-distributed over the unpopulated regions of the envelope. 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, Fig. 3 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 multi-variate data. Warm colors indicate higher density of flight points. Note that the color map is specific to each subplot.

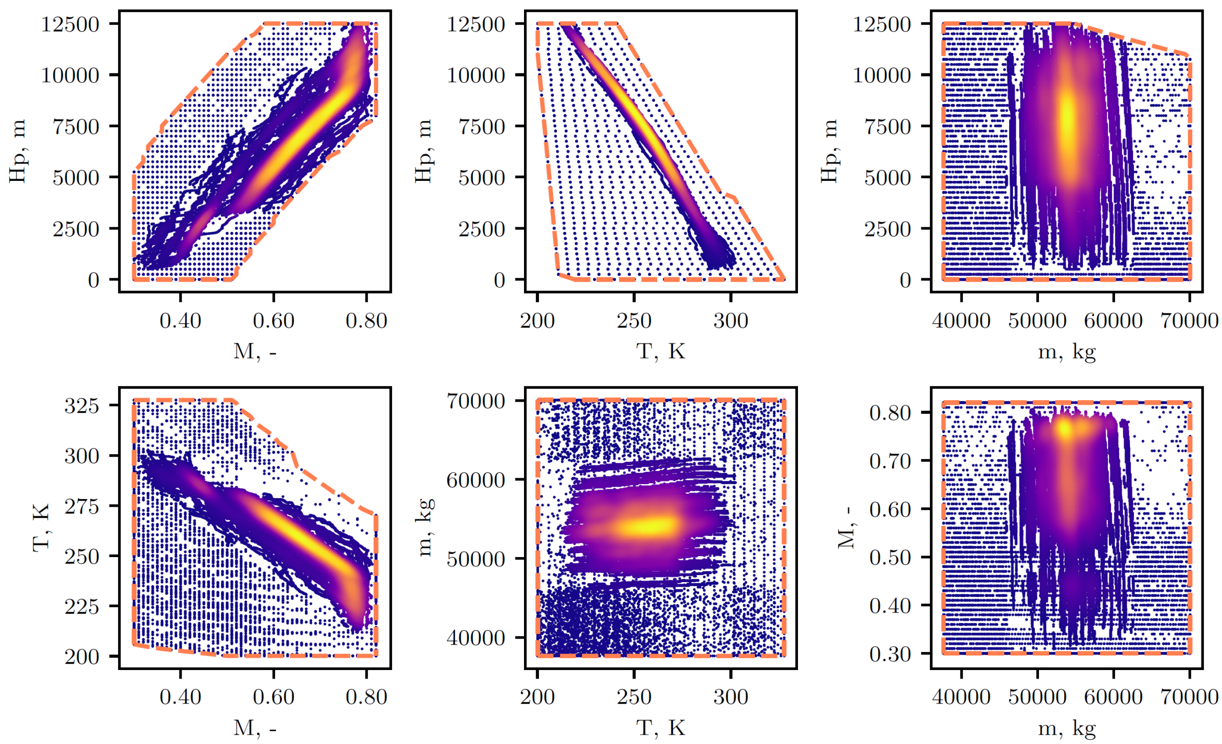

Fig. 3 Kernel-density estimate of QAR climb data augmented with synthetic performance flight data, distributed over the operational envelope of the corresponding aircraft type (– –)

C. 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 all BADA 4 coefficients based on historical flight data. In other words, the set of coefficients that best describe the tail number(s) represented in the historical data. In total, seven different sets of coefficients are contemplated and adjusted:

1) General thrust coefficients
2) Flat-rated area throttle coefficients
3) Temperature-rated area throttle coefficients
4) Clean configuration drag coefficients
5) General fuel coefficients
6) Idle thrust coefficients
7) Idle fuel coefficients

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 and motion submodels. 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 QAR-data-driven 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 when the regression coefficients are to be known. Multiple linear regressions have the form:

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

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 tailoring scheme consists of seven multivariate linear regression models corresponding to the seven different sets of coefficients. 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 modeled performance is achieved. The tailoring scheme is summarized as follows. To start with, a first approximation of the generalized fuel model is calculated. Secondly, the throttle models are computed, and afterwards the generalized thrust models are identified accordingly. The next step is to find the drag polar from the generalized fuel model (previously-obtained) and the nominal 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. 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 Mean Absolute Error (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. After the iterative process is stopped, the idle fuel coefficients are tailored. Figure 4 illustrates the intermediate steps, the interconnection between steps and the multivariate regression problems to solve in order to find the sets of coefficients that best represent the reference performance data. The nomenclature used in this figure is consistent with the one that can be found in the User Manual for the Base of Aircraft Data (BADA) Family 4 [1].

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. The developed methodology 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. The calibration of only those coefficients that are activated in BADA nominal models 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.

D. Model Evaluation

Once the QAR-data-driven coefficients have been identified, 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 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(s) characterized in the training dataset, is reserved to study the ability of calibrated models to generalize lessons learned to unseen flight data.

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. 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. Figure 5 is an example of this evaluation exercise. It represents

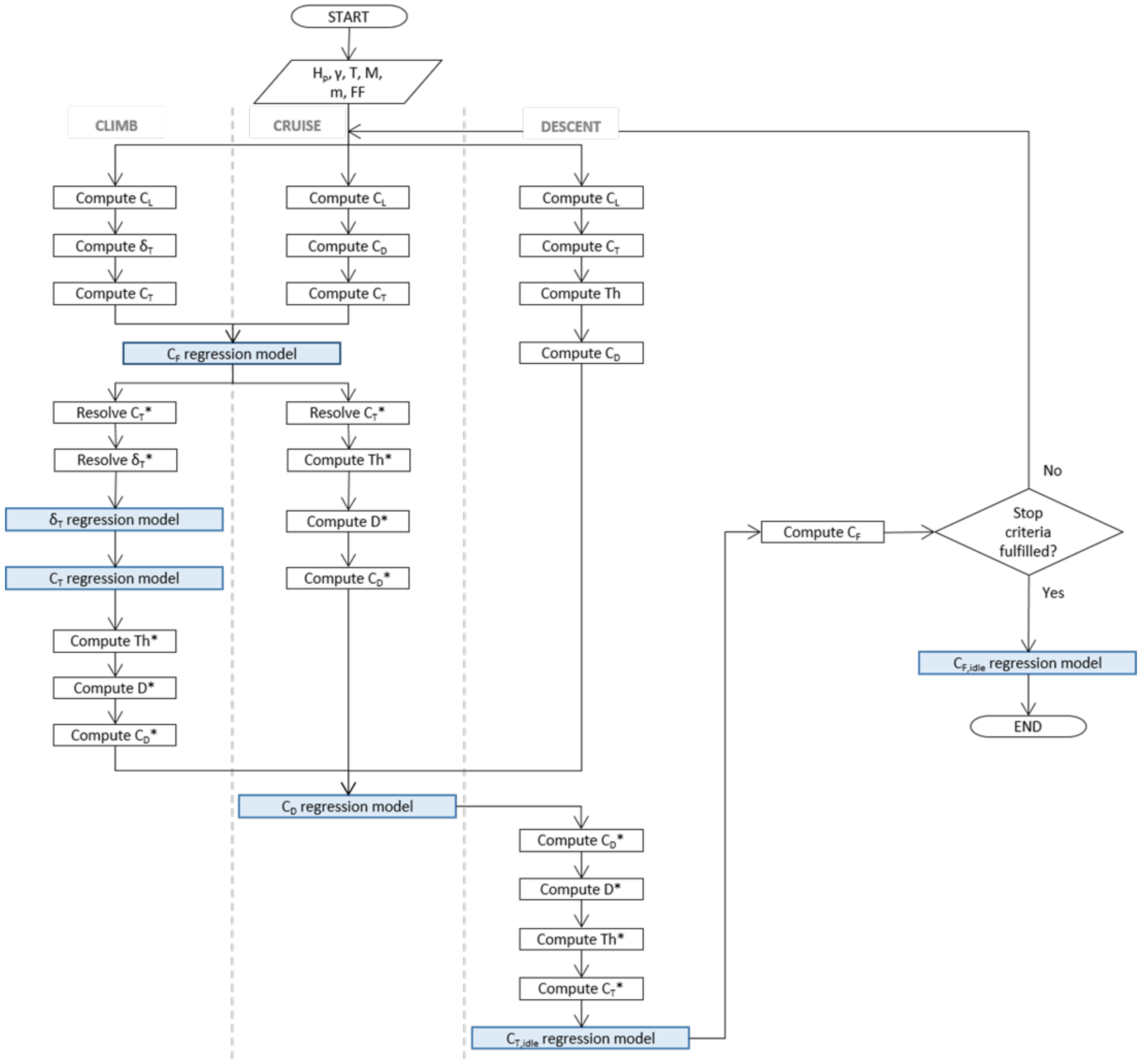

Fig. 4 Flow chart of the tailoring process including all necessary steps and the interaction between them, separated by phase of flight. '*' refers to variables with new values calculated by the regression model

the difference between BADA nominal and calibrated climb performances over the entire flight envelope of the aircraft under consideration. 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 can be use.

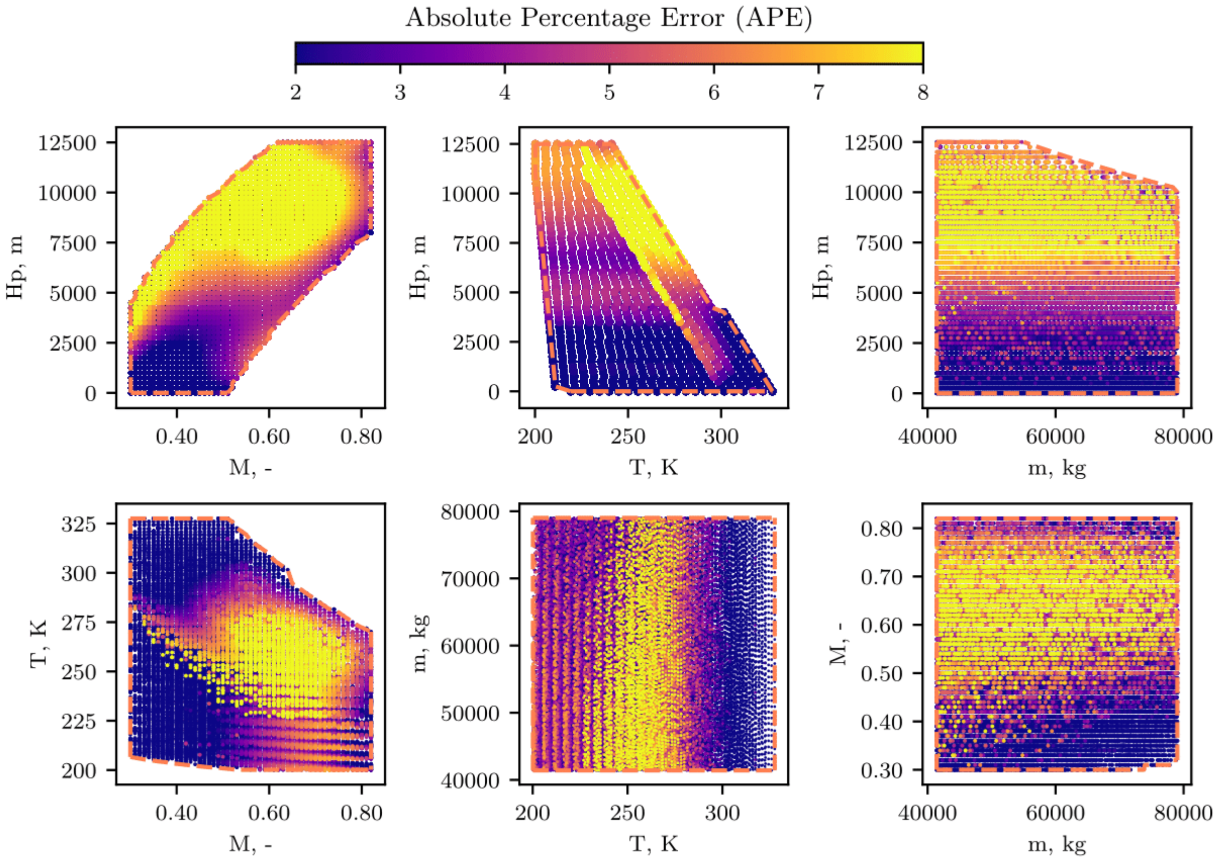

Fig. 5 APE between fuel flow modeled by BADA nominal and BADA calibrated models in each synthetic flight point generated in the data ingestion phase of the methodology

V. Results and Discussion

Following the previously-described methodology, multiple calibrations with different datasets have been successfully performed and validated. This section includes results corresponding to two calibrated models. The first one is a tail-specific calibration, meaning that it was done using QAR data from one tail number only. The second calibration was carried out with flight data from multiple tail numbers of the same aircraft type.

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

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

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{2}$$

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

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\frac{|\hat{y}_i - y_i|}{y_i} \tag{3}$$

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

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{4}$$

The 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.

A. Tail-specific calibration: narrow-body aircraft

This 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 1, 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, the calibration of BADA coefficients allows to reduce the error in fuel flow from 5.00 % to 1.50 %, considering all phases of flight. With respect to the average MAE, this is reduced from 101.27 kg/h to 30.08 kg/h, and the average RMSE is lowered from 124.31 kg/h to 43.25 kg/h.

Table 1 MAE, MAPE and RMSE of the fuel flow given by BADA nominal and BADA tail-specific calibrated models with respect to the actual fuel flow

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

Figure 6 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. The fix of this misalignment resulted in better agreement with real performance also in the cruise phase.

Even though the most noticeable enhancement is observed in fuel-flow modeling, calibrated Rate of Climb and Descent (ROCD) is realistic and consistent with actual and BADA nominal ROCDs. Fuel flow is the parameter that mainly drives this performance tailoring approach, but it is important to highlight that the calibration process adjust BADA fuel models without distorting thrust and drag models.

Last but not least, errors in modeled fuel consumption have been computed and compared 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, 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 7 consists of 4 histograms that depict the distribution of these error metrics over the testing dataset. Histograms on the left (see Figure 7(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 (7(b)), which represent MAPE in fuel consumption.

B. Tail-specific calibration: wide-body aircraft

The developed methodology was validated with QAR data from multiple different aircraft types and airlines. The results of a second tail-specific calibration are also presented in this paper. This calibration corresponds to a twin-aisle, long-range twin-jet airliner. 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.

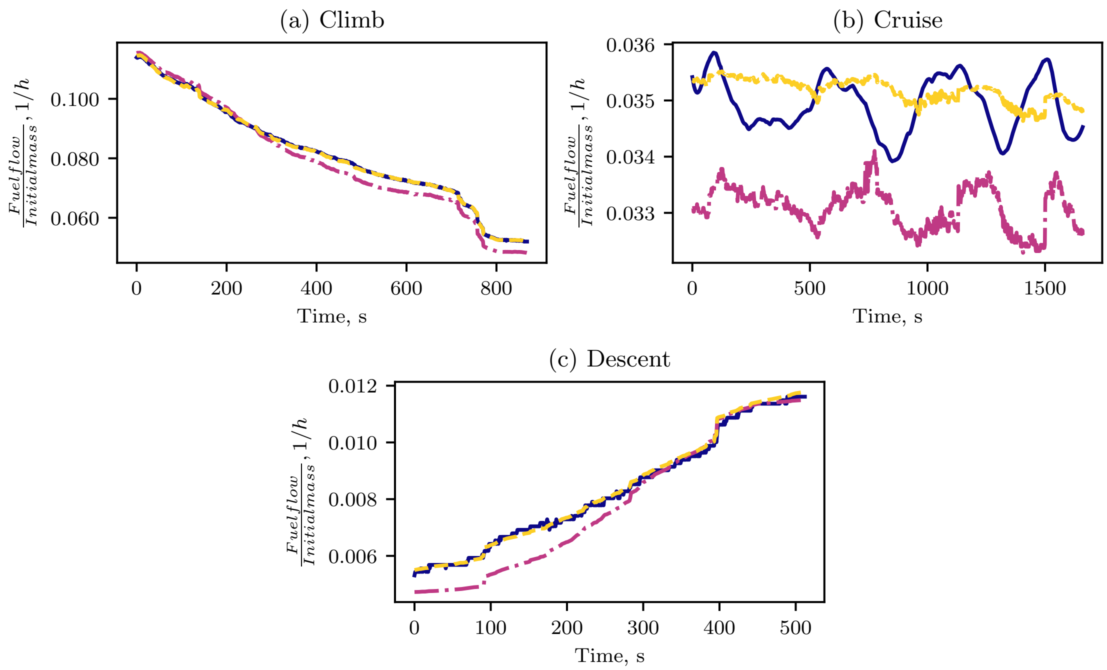

Fig. 6 Evolution of fuel flow during a testing flight according to BADA nominal model (-·-) and BADA tail-specific calibrated model (– –) of a narrow-body aircraft with respect to the actual fuel flow (—) in climb, cruise and descent

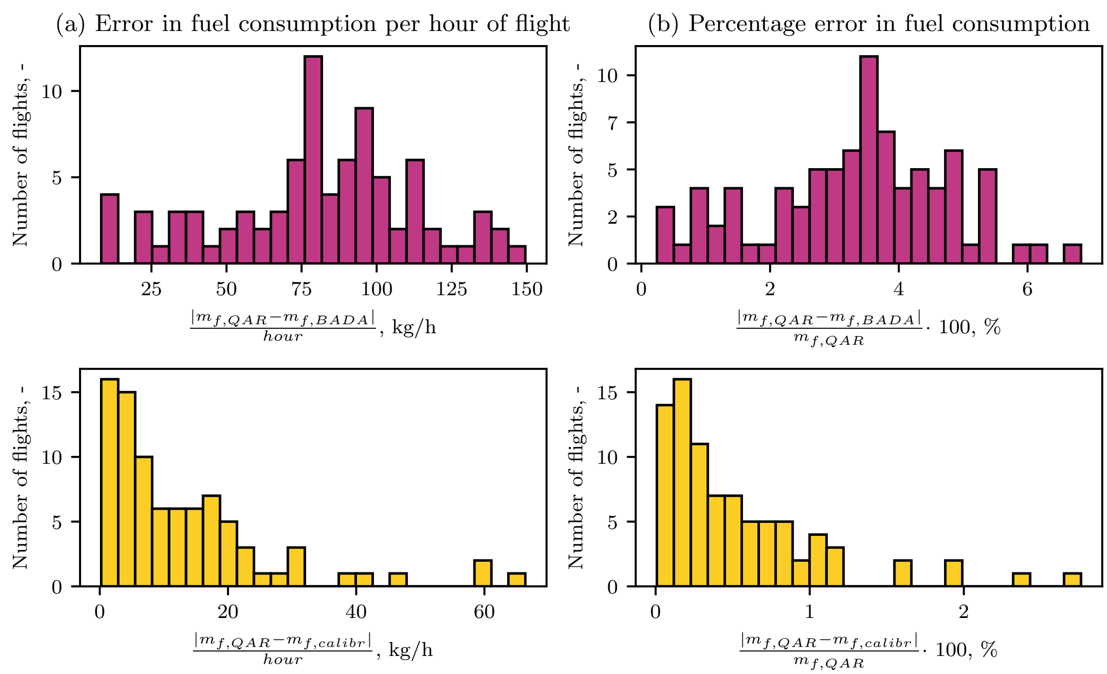

Fig. 7 Error in fuel consumption per hour of flight and percentage error between the real fuel consumption and the fuel consumption given by BADA nominal model (■) and BADA tail-specific calibrated model (■) of a narrow-body aircraft

As shown in Table 2, BADA's estimation of fuel flow in the main phases of flight also improved for wide-body aircraft. Overall, it can be said that thanks to the developed methodology the error in fuel flow is reduced from 2.33 % to 1.77 % 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 %. 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 8 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 calibration 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 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 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

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, 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 9 includes histograms that represent the distribution of these error metrics over the testing dataset. As it can be seen, the adjustment of BADA coefficients based on QAR data allows to reduce the error also in fuel consumption.

C. Generic calibration: narrow-body aircraft

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 was accomplished. This calibration uses QAR data corresponding to 85 tail numbers of the same aircraft type; a single-aisle, short- to medium-range twin-jet aircraft. 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 are presented in Table 3. In this case, the calibration reduces the average error in fuel flow from 6.82 % to 2.25 %, considering all phases of flight. In terms of MAE, the average error is decreased from 160.35 kg/h to 52.01 kg/h, and the average RMSE is lowered from 182.73 kg/h to 68.95 kg/h. Noting that the narrow-body airplane considered before and these airplanes 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.

In Figure 10, one can again notice the better match between actual and calibrated fuel flow compared to BADA

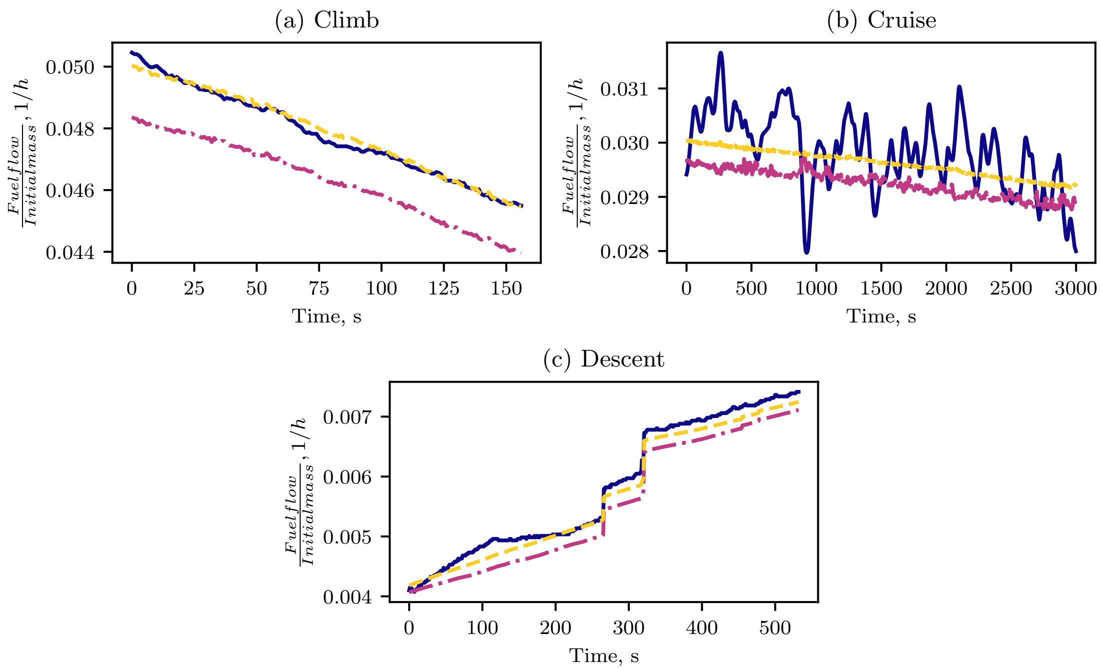

Fig. 8 Evolution of fuel flow during a testing flight according to BADA nominal model (–·–) and BADA tail-specific calibrated model (– –) of a wide-body aircraft with respect to the actual fuel flow (—) in climb, cruise and descent

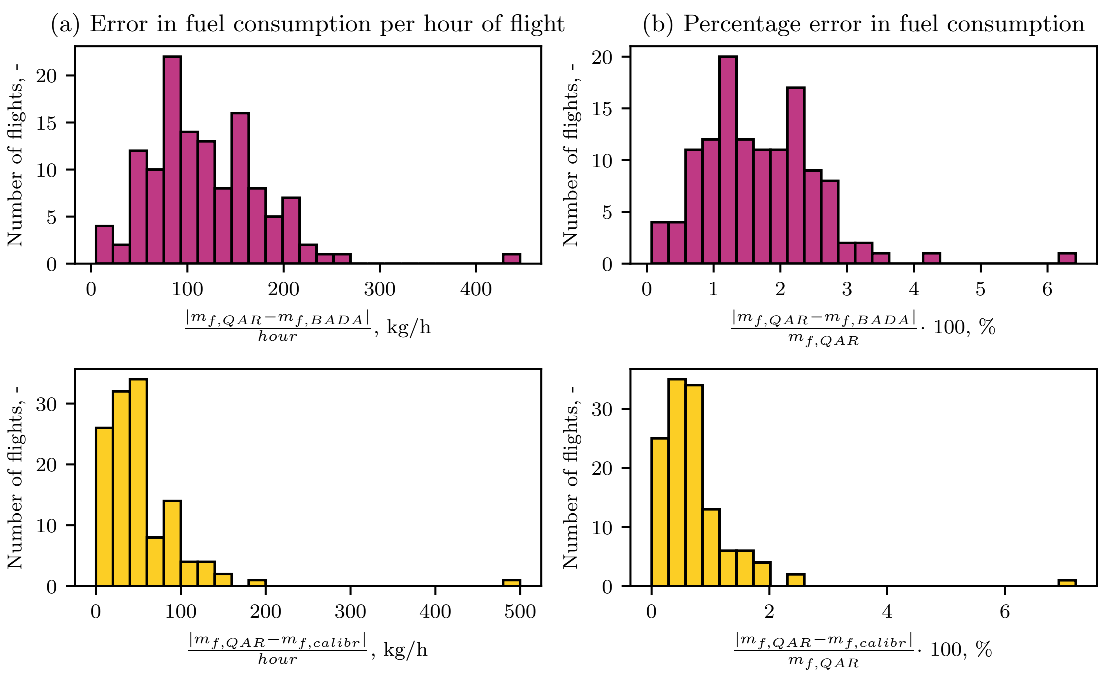

Fig. 9 Error in fuel consumption per hour of flight and percentage error between the real fuel consumption and the fuel consumption given by BADA nominal model (■) and BADA tail-specific calibrated model (■) of a wide-body aircraft

Table 3 MAE, MAPE and RMSE 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]	
	BADA	Calibration	BADA	Calibration	BADA	Calibration
Climb	230.80	54.09	4.90	1.16	245.97	66.63
Cruise	159.74	56.72	6.83	2.46	177.34	73.53
Descent	48.09	13.45	10.03	2.70	52.51	17.41

nominal fuel flow. This match is again stronger in climb and descent trajectories than in cruise, where again seems to be a de-synchronization between actual and BADA nominal curves that results in a more-accurate but more-flattened calibrated fuel flow.

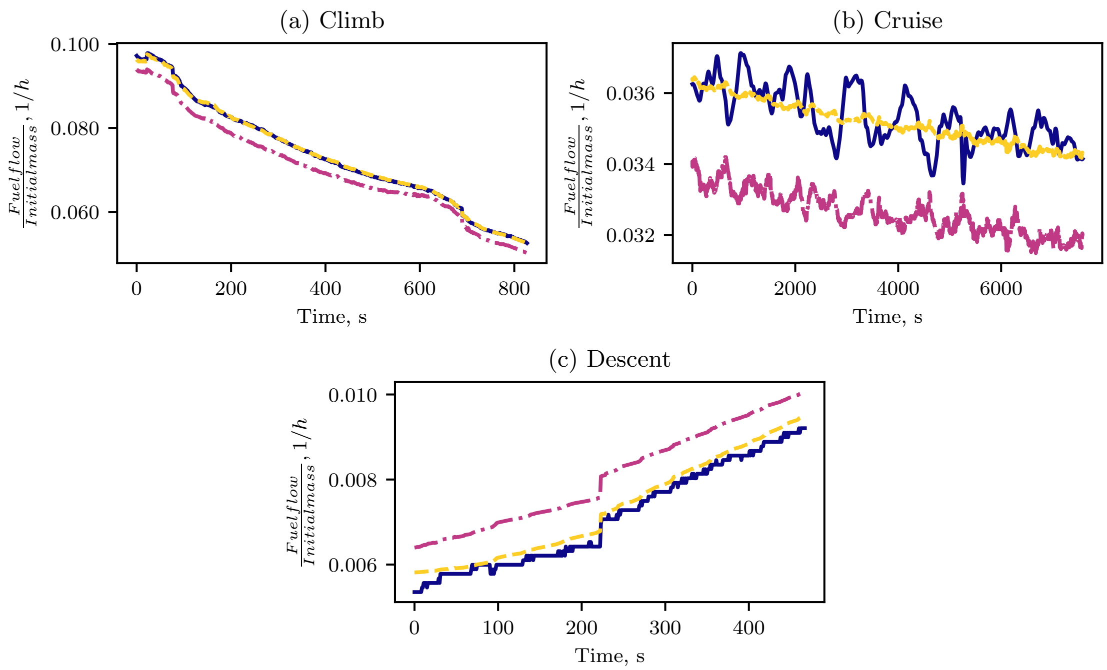

Fig. 10 Evolution of fuel flow during a testing flight according to BADA nominal models (–·–) and BADA aircraft-type-specific calibrated models (– –) with respect to the actual fuel flow (—) in climb, cruise and descent

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 underestimation of more than 130 tons of fuel in two days of operation. By using calibrated models instead, these figures would reduce to a total of 30 tons. Errors between real and modeled fuel consumption are plotted in Figure 11 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.

Based on results presented above, it can be affirmed that calibrating BADA 4 coefficients using historical QAR data and machine-learning techniques significantly improves the accuracy of BADA nominal models regardless aircraft type and operator. The results of the 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 multiple tail numbers together. As expected, generic models represent aircraft performance with less accuracy than tail-specific models. Different tail numbers can show different degradation degrees with respect to BADA, but in all cases these degrees are properly corrected with the described methodology.

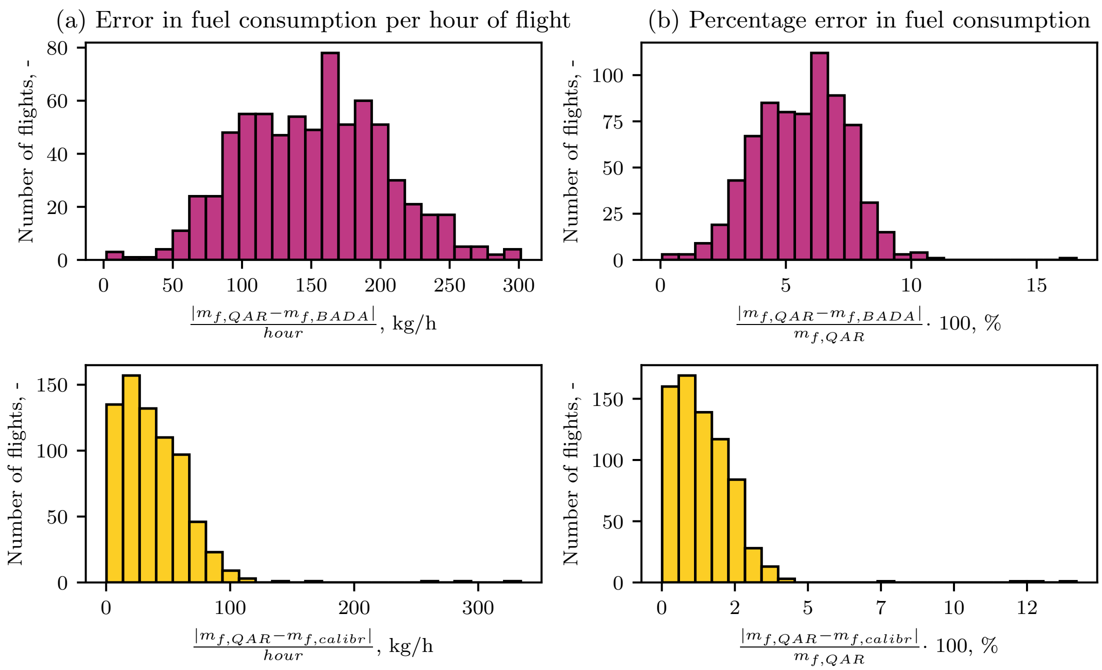

Fig. 11 Error in fuel consumption per hour of flight and percentage error between the real fuel consumption and the fuel consumption given by BADA nominal model (■) and BADA aircraft-type-specific calibrated model (■) of a narrow-body aircraft

VI. Conclusions and Future Work

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 Air Traffic Management and airspace simulation activities. Countless efforts have been devoted to model aircraft performance with enough accuracy, however the vast majority of aircraft performance models available in the literature represent average performance of particular aircraft types. These models fail to mimic performance of specific tail numbers, which means that applications based on them might lead to unreliable or misleading results.

The purpose of this work is to develop tailored 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 paper presents a methodology to calibrate BADA 4 aircraft performance models based on the use of historical QAR data and machine-learning regression algorithms. This 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 is focused on the selection of flight points that are compliant with the assumptions BADA model 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. 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 4 coefficients that best describe the reference QAR data are identified. 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).

The proposed methodology has been validated using different aircraft types and data sets. This paper includes results for two different tails corresponding to two different aircraft types and one generic calibration for an aircraft type including multiple tails. The results demonstrated that the designed methodology allows to successfully fine-tune

BADA 4 performance models based on historical QAR data, with the aid of machine-learning regression models. The developed methodology proved to be consisted and versatile, since it allows to model tail-specific performance, but also to create generic performance models applicable in all the flight envelope. The methodology is complete, since it derives throttle, thrust, drag and fuel models, which are the basic variables needed for any point-performance calculation. It is also flexible, since it can be easily adapted to derive different types of aircraft models. Regarding BADA, the system allows to choose between different polynomial orders to accommodate all memory and computing time requirements. Calibrated models demonstrated to improve the accuracy of BADA nominal models significantly, while providing the same capabilities and keeping the same level of complexity. In addition, calibrated models demonstrated to be robust over the entire flight envelope and during all phases of flight. Therefore, the use of BADA calibrated models in pre-, in- and post-flight applications like fight planning, trajectory prediction or fuel analytics, is desirable since it will ensure more realistic results. With respect to maintainability, BADA nominal models do not demand high maintenance resources. Nonetheless, aircraft performance deviation from the nominal varies with time, so for data-driven performance models to reflect these variations they need to be dynamic. Calibrated models must be updated or re-calibrated to capture performance degradation, for which an updating mechanism should be established based on degradation thresholds, maintenance actions or time. The establishment of these updating mechanisms requires the analysis of the performance variations over several months or years of QAR data.

References

[1] Eurocontrol Experimental Centre, "User Manual for the Base of Aircraft Data (BADA) Family 4," Tech. Rep. No. 12/11/22-58, March 2016.

[2] Nuic, A., Poles, D., and Mouillet, V., "BADA: An advanced aircraft performance model for present and future ATM systems," *International Journal of Adaptive Control and Signal Processing*, Vol. 24, No. 10, 2010, pp. 850–866.

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

[4] He, F., Li, L., and Zhao, W., "Aircraft Mass Estimation using Quick Access Recorder Data," *2018 IEEE/AIAA 37th Digital Avionics Systems Conference (DASC)*, IEEE, 2018, pp. 63–67.

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

[6] Alligier, R., Gianazza, D., and Durand, N., "Machine Learning and Mass Estimation Methods for Ground-based Aircraft Climb Prediction," *IEEE Transactions on Intelligent Transportation System*, Vol. 16, No. 6, 2015, pp. 3138–3149.

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

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

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

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

[11] del Pozo Domínguez, M., López Leonés, J., and Roling, P., "Modeling Tail-Specific Performance Using Historical Flight Data and Machine-Learning Techniques," Master's thesis, Delft University of Technology, The Netherlands, 2020.