UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Escola d'Enginyeria de Telecomunicació
i Aeroespacial de Castelldefels

MASTER THESIS

TFM TITLE: Estimation of aircraft performance from quick access recorder flight data

MASTER: Master in Aerospace Science and Technology (MAST)

AUTHOR: Laura Hernandez

ADVISORS: Xavier Prats
Ramon Dalmau

DATE: October 23, 2018

Title : Estimation of aircraft performance from quick access recorder flight data

Author: Laura Hernandez

Advisors: Xavier Prats
Ramon Dalmau

Date: October 23, 2018

Overview

Considering the expected air traffic growth, the innovation and development of new tools able to allow a more efficient and safer way of managing aircraft operations is necessary to achieve future expectations. In this context, it is important to be capable of accurately predict aircraft trajectories to ensure efficient aircraft operations (e.g., planning and dispatching of flights, in-flight trajectory prediction, etc.) as well as to make more robust the Air Traffic Management (ATM) system (which includes ground based systems for ATC, predicting demand in ATC sectors, etc). The way of predicting them is based on aircraft performance models (APM), i.e., sets of equations that allow to model the aircraft performance according to some specific coefficients that depend on the aircraft carrying out the flight. Therefore, the predicted trajectories accuracy will depend directly on the aircraft performance model used. If the APM does not reflect the reality, the predicted trajectories will not be accurate enough. Moreover, since these trajectories will not be longer optimal according to the real performance model, the performance in cost-efficiency and the environmental impact of aircraft operations will be decreased. Then, it is required the use of aircraft performance models as realistic as possible. The objective of this Master Thesis is the design of an algorithm capable to estimate the coefficients of the functions describing the aircraft performance model considered, which will be the Base of Aircraft Data (BADA), from Quick Access Recorder (QAR) flight data. The idea is to obtain the real coefficients values of each specific tail number aircraft to accurately model their performance for trajectory prediction purposes. To perform the coefficients' estimation, real flight data has been used, consisting in a set of trajectories carried out by more than 30 different tail number aircrafts and obtained from their respective QAR. Among these aircrafts, only one (i.e, its trajectories) will be selected to design the estimation algorithm, which consist in a nonlinear least squares optimization problem. The output will be the coefficients value that better fits to the QAR data used, i.e, that minimize the difference between the real trajectories and the generated predicted trajectories as a result of applying the estimated coefficient's value to the BADA model equations. The results show a good estimation accuracy for the four aircrafts analysed, obtaining a very low prediction error in all cases, which proves the reliability of the algorithm and the necessity of tailoring the existing performance models (in this case BADA).

The important thing is to not stop questioning.
Curiosity has its own reason for existing.
Albert Einstein

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Ramon Dalmau for getting so involved in this Thesis and for continuously supporting me during its development. Thank you for your patience, guidance and for share your knowledge, without which this project would not be what it is. Many thanks to Xavier Prats for supervise the project and for his wise tips and suggestions that have allowed to achieve the best version of this Master Thesis.

Thanks to all colleagues and friends who have invested a part of their time in helping me or giving me some advice to develop this project, especially to Marc Melgosa and Raúl Sáez for their help and support.

Many thanks to all my master colleagues, you are the best that I get from this experience and without you nothing would have been the same. Thank you for these two years of fellowship, good times and unforgettable moments.

Last but not least, thanks a lot to my family for its unconditional support in everything I decide to do. This Master Thesis is dedicated to you for being my fundamental pillar and for always being part of me in any adventure I decide to undertake.

INTRODUCTION

The future of the Air Traffic Management (ATM) system is based on the consecution of a specific set of measurable performance targets that fit to the main ATM Key Performance Areas (KPAs), particularly in the following ones: safety, efficiency, capacity and environmental impact. In this context, the ATM system performance relies directly on the prediction accuracy of future air traffic evolution and aircraft trajectories. Moreover, in order to ensure efficient aircraft operations, it is required to know as accurate as possible the current aircraft performance (e.g. drag coefficient, fuel flow, ...) and keep track of its evolution through making comparisons with respect to the theoretical one, which is present in flight manual after being manufactured and tested.

Considering that real aircraft performance usually differs from theoretical due to structure degradation and changes produced in main physical factors, it is necessary the research and development of new tools and methods capable to accurately estimate the mathematical functions and associated coefficients (i.e., the aircraft performance model) that describe the aircraft performance and, consequently, the current flight performance [1].

However, the aircraft performance model could differ for every single tail number aircraft, being very difficult to generate a generic analytical model for every single aircraft. Nevertheless, they can be estimated by observing how the aircraft behaves and how it should behave according to the performances it had just out of the factory. In this context, aircraft performance models [2] are the core of trajectory computation and prediction, playing a central role in the development and evaluation of future cockpit, flight dispatching and ATM systems. There are several existing approaches to aircraft performance modeling to support various needs for aircraft trajectory prediction, simulation and optimisation. Thus, depending on the approach and techniques used in modeling there are different forms of Aircraft Performance Models (APM). In this project, an aircraft performance model named BADA (Base of Aircraft Data) provided by Eurocontrol will be used [3].

Then, the idea of this Master Thesis would be to design the algorithm capable to estimate the coefficients of the functions describing BADA APM for a specific tail number aircraft. For this purpose, a set of real trajectories provided by QARs (Quick Access Recorder) of an A320 aircraft type fleet will be used. However, the achievement of this particular aim depends also on the consecution of other sub-objectives, which are presented below:

- **Identify the aircraft performance model coefficients to be estimated:** Identify all the coefficients of interest from BADA model equations according to the project's case study.

- **Develop the estimation algorithm:** The algorithm to be developed will be based on the least squares regression method, which makes possible the comparison of two trajectories by observing which is the squared difference between the most representative state variables of the real and predicted trajectories. These variables should represent the aircraft behaviour and they can be the horizontal or vertical position, aircraft mass, acceleration, rate of climb, etc. In this particular case, these ones will be given by the Equations Of Motion (EOM) governing the point-mass model considered. Therefore, the algorithm should provide as output the the best estimation of the coefficients' values previously identified, i.e. the values that minimizes the difference between the chosen real trajectory state variables and the corresponding predicted trajectory state variables.

- **Test the estimation algorithm:** The idea is to perform several tests in order to detect the errors that the algorithm could present and analyse the algorithm's sensitivity to the input variables and coefficients. This will allow a proper algorithm tuning to achieve accurate results.

- **Validate the results with QAR data from real flights:** What is wanted to obtain is the coefficients value that minimize the difference between the predicted trajectories and the real trajectories carried out by a specific tail number aircraft. This means that the QAR trajectories of the validation set will be compared with the result of combining BADA model equations with the coefficients already estimated by least squares. A minimum difference will indicate an accurate estimation of the coefficients.

In order to have a good understanding regarding the current situation of performance modeling and the coefficients estimation process, it is important to give a brief explanation of which are the sources that have been consulted for the development of this project.

Regarding the algorithm's design, the estimation process workflow has been inspired in the one presented in [4] in which an enhanced tailored APM based on flight recorded data is proposed. To adapt the algorithm's design to the case study selected, [5] has been reviewed since it proposes four fuel models that are generated based on recent historical high-frequency flight data to predict the fuel use of individual aircraft, which is also based on BADA and real flight data.

Although, in the end, BADA has been the aircraft performance model used in this project, [6] and [7] were consulted since they propose new approaches of aircraft performance modeling, the first one, introducing a Compromised Aircraft Performance model with Limited Accuracy (COALA) and, the second one, presenting a new approach to model aircraft performances based on machine learning. Moreover, in [8], [9] and [10] new thrust, weight and optimum climb profiles are developed to improve trajectory prediction accuracy.

Furthermore, regarding statistical prediction of aircraft trajectory, in [11] a standard point-mass model and some specific statistical regression methods, in which least squares is also included, are used to predict the altitude of climbing aircraft. Although least squares were finally selected as the regression method to be used for the coefficients' estimation, [12] and [13] were also reviewed since they examine three different approaches to aircraft parameter and dynamics estimation based on Equation-Error, Filter-Error and Output-Error methods.

Overall, after reading all the sources stated in this section, it was decided to use BADA model and least squares regression method as the best baseline on which start to build the coefficients' estimation algorithm.

CHAPTER 1. BACKGROUND

In this chapter, some theoretical background about BADA model will be provided. For a proper understanding of the coefficients estimation process, the regression method of least squares and the DYNAMO trajectory predictor will also be introduced.

1.1. Base of Aircraft Data (BADA)

The Base of Aircraft Data (BADA) [2] is a model developed and maintained by EUROCONTROL, with the active cooperation of aircraft manufacturers and operating airlines, composed by an Aircraft Model and an Atmosphere Model. As it can be observed in figure 1.1, Bada model is structured in three parts: an Aircraft Performance Model (APM), an Airline Procedure Model (ARPM) and Aircraft Characteristics.

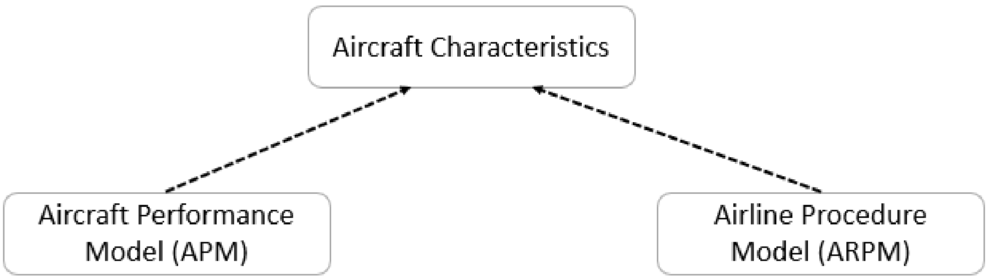

Figure 1.1: BADA Aircraft Model structure

- **Aircraft Performance Model (APM):** The APM adopted by BADA is based on a mass-varying, kinetic approach which models an aircraft as a point and requires modeling of underlying forces that cause aircraft motion [2].

- **Airline Procedure Model (ARPM):** A generic Airline Procedure Model (ARPM) is proposed in BADA, providing nominal speeds for the climb, cruise and descent phases, assuming normal aircraft operations [14].

- **Aircraft Characteristics:** Each aircraft performance sub-model in BADA is characterized with a set of coefficients, called Aircraft Characteristics, which is used by the APM and ARPM [14].

Since this project will be focused on the BADA APM, details about its sub-models will be provided in next sections. Moreover, information about Aircraft Characteristics will be also presented, as they are the coefficients describing the BADA APM functions, between which, there are the ones to be estimated in this project. BADA revision 3.12 has been the version used in this project, which can be found in source [15].

1.1.1. BADA Aircraft Performance Model

The BADA Aircraft Performance Model (APM) consists in a performance database based on the kinetic approach to aircraft performance modeling design for simulation and prediction of aircraft trajectories for purposes of ATM research and operations [2]. Aircraft performance parameters and trajectories can be calculated using information and data contained in BADA. To replicate reality in both operational and research modeling and simulation systems [16], each and every real aircraft needs a corresponding aircraft model, since it is the core of the aircraft trajectory computation system as figure 1.2 shows.

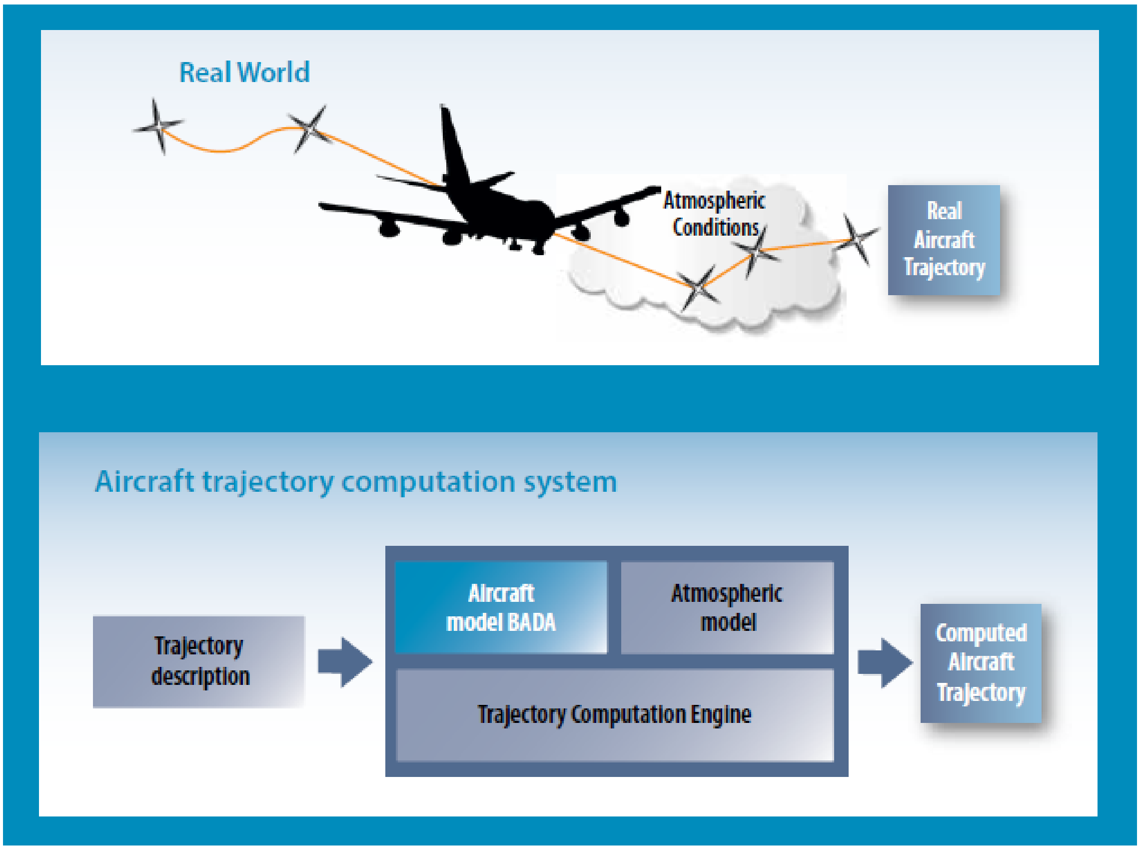

Figure 1.2: Aircraft Trajectory Computation System using BADA [16]

BADA APM is structured into four sub-models [17]: Actions, Motion, Operations and Limitations model.

- **The Actions model** is a representation of the forces acting on the aircraft. These actions are responsible for the motion of the aircraft and come from three distinct categories: aerodynamic, gravitational and propulsive. Aerodynamic actions are referring to the drag and lift forces. The gravitational action corresponds to the weight force and, finally, the representative force of the propulsive category of actions is the thrust force. Apart from thrust, the propulsive action model provides a model to compute the fuel flow, which affects the aircraft's mass. Therefore, the Actions model is composed by the following sub-models: Drag, Lift, Weight, Thrust and Fuel flow.

- **The Motion model** consists of the system of different equations that describes the aircraft motion. One of these equations is the Total-Energy Model (TEM), a physical model based on the 2nd newton's law that connects geometrical, kinematic and kinetic aspects of the aircraft motion, which is essential in order to calculate the aircraft trajectory and performance. The other equation considered is the fuel consumption model, which is also included in the Motion Model.

- **The Operations model** provides information about the way the aircraft is operated, being responsible for capturing aspects of the flight which are not directly related to the Actions or the Motion model, but are necessary to take into consideration while computing the aircraft motion. This is important since different ways of operating the aircraft may result in different trajectories.

- **The Limitations model** ensures a realistic aircraft behaviour within certain limits. These limitations are set to secure not only the realistic aircraft behavior and providing flight safety, but also to limit the degradation of the aircraft equipment.

The existing relations structure between these BADA sub-models is represented in figure 1.3:

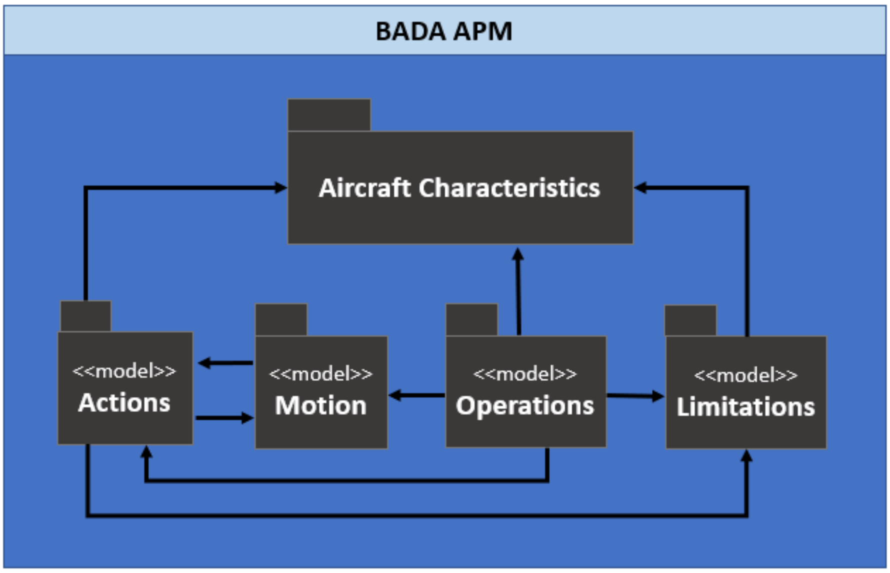

Figure 1.3: BADA Aircraft Performance Model relations structure

In this project, the Actions and Motion BADA sub-models will be considered. What is important to understand is that the Actions model equations modeling D, T and f_f are based on the TEM and fuel consumption models included in the Motion model, which constitude the baseline of BADA APM. Find in following sections the equations describing each of these models.

1.1.2. Total Energy Model and Fuel Consumption Model

The BADA APM is based on the TEM, a simplified point-mass model that equates the rate of work done by forces acting on the aircraft to the rate of increase in potential and kinetic energy [15]:

$$(T-D)\cdot v = mg_0\frac{\mathrm{d}h}{\mathrm{d}t} + mv\frac{\mathrm{d}v}{\mathrm{d}t} \tag{1.1}$$

where:

- T is the thrust force acting parallel to the aircraft velocity vector [N]
- D is the aerodynamic drag force [N]
- v is the true airspeed of the aircraft [m/s]
- m is the aircraft's mass [kg]
- g_0 is the gravitational acceleration [m/s^2]
- $\frac{\mathrm{d}}{\mathrm{d}t}$ denotes the time derivative
- h is the geometric altitude [m]

As it has been aforemetioned, the Fuel Consumption Model is also included in BADA Motion Model, which expresses the variation of mass in flight through equation 1.2:

$$m = -f_{nom} \tag{1.2}$$

The forces involved (D, T) in the TEM as well as the fuel consumption model are functions that depend on the performance model considered. These functions, which are included in the Actions model, are parametrized with some specific coefficients, or aircraft characteristics, which are the ones to be estimated and particularized for each tail number aircraft. In next sections, the Actions model equations regarding D, T and f_f modeling will be presented as well as the coefficients included in each of them.

1.1.3. Drag Model

The drag force can be calculated using the expression 1.3, where ρ is the air density (kg/m^3), S the wing reference area (m^2) and v the true airspeed (m/s).

$$D = \frac{C_D\cdot\rho\cdot v^2\cdot S}{2} \tag{1.3}$$

As it can be observed, it depends directly on the drag coefficient, which under nominal conditions is specified as a function of the lift coefficient C_L as follows [15]:

$$C_D = C_{D0} + C_{D2}\cdot\ (C_L)^2 \tag{1.4}$$

This formula is valid for all configurations. The drag coefficients are C_{D0} and C_{D2} which refer to the parasitic drag coefficient and induced drag coefficient respectively. However, depending on the configuration considered the coefficients will vary. Regarding the lift coefficient C_L, this one can be calculated assuming that the flight path angle is zero. However, a correction for a bank angle ϕ is made.

$$C_L = \frac{2 \cdot m \cdot g_0}{\rho \cdot v^2 \cdot S \cdot cos\phi} \tag{1.5}$$

Thus, the coefficients necessary to be estimated in order to model the drag force are: C_{D0} and C_{D2}.

1.1.4. Engine Thrust Model

The thrust is calculated in Newtons and includes the contributions from all engines. BADA model provides coefficients that allow the calculation of 3 different thrust levels [15], which are:

- **Maximum climb and take-off**: The maximum climb thrust at standard atmosphere conditions, $(T_{maxclimb})_{ISA}$, is calculated in Newtons as a function of the following information:

  - Engine type: either Jet, Turboprop or Piston
  - Geopotential pressure altitude, h_p [ft]
  - True airspeed, v [kt]
  - Temperature deviation from standard atmosphere, ΔT [K]

  Since all the QAR trajectories used in this project are performed by A320 aircrafts, which is a jet engine aircraft type, the equations presented below will be used to model the thrust according to this engine type. The methodology will be the same for other aircraft types, but the equations and its coefficients will vary.

  In this context, it is possible the modeling of the maximum climb and take-off thrust with the following equation:

$$(T_{maxclimb})_{ISA} = C_{T_c,1} \cdot (1 - \frac{h_p}{C_{T_c,2}} + C_{T_c,3} \cdot h_p^2) \tag{1.6}$$

  However, it is important to point out that, for all engines, the maximum climb thrust is corrected for temperature deviations from standard atmosphere, ΔT, in the following manner:

$$T_{maxclimb} = (T_{maxclimb})_{ISA} \cdot (1 - C_{T_{c,},5} \cdot \Delta T_{eff}) \tag{1.7}$$

  where:

$$\Delta T_{eff} = \Delta T - C_{T_{C},4} \tag{1.8}$$

with the limits:

$$0.0 \leq \Delta T_{eff} \cdot C_{T_{c},,5} \leq 0.04 \tag{1.9}$$

being:

$$C_{T_{c},,5} \geq 0.0 \tag{1.10}$$

This maximum climb thrust is used for both take-off and climb phases.

- **Maximum cruise**: The normal cruise thrust is by definition set equal to drag (T = D). However, the maximum amount of thrust available in cruise situation is limited. The maximum cruise thrust is calculated as a ratio of the maximum climb thrust that is [15]:

$$(T_{cruise})_{MAX} = C_{T_{cr}} \cdot T_{maxclimb} \tag{1.11}$$

In BADA, the coefficient $C_{T_{cr}}$ is currently uniformly set for all aircraft.

- **Descent**: Descent thrust is calculated as a ratio of the maximum climb thrust given by expression 1.7, with different correction factors used for high and low altitudes, and approach and landing configurations, that is [15]:
  - If $h > h_{des}$:

$$T_{des,high} = C_{T des,high} \cdot T_{maxclimb} \tag{1.12}$$

  - If $h \leq h_{des}$:
  - Cruise configuration:

$$Thr_{des,low} = C_{T des,low} \cdot T_{maxclimb} \tag{1.13}$$

  - Approach configuration:

$$T_{des,app} = C_{T des,app} \cdot T_{maxclimb} \tag{1.14}$$

  - Landing configuration:

$$T_{des,ld} = C_{T des,ld} \cdot T_{maxclimb} \tag{1.15}$$

where h_{des} is the transition altitude. Therefore, for the jet engine thrust case, and considering the modeling of an entire trajectory from the ascent phase to the descent phase, the coefficients of interest would be: $C_{T_c,1}$, $C_{T_c,2}$, $C_{T_c,3}$, $C_{T_c,4}$, $C_{T_c,5}$, $C_{T_{cr}}$, $C_{T des,high}$ and $C_{T des,low}$.

1.1.5. Fuel Flow Model

For the jet and turboprop engines, the thrust specific fuel consumption, η [kg/(min·kN)], is specified as a function of the true airspeed, v [kt]:

$$\eta = C_{f1} \cdot (1 + \frac{v}{C_{f2}}) \tag{1.16}$$

The nominal fuel flow, f_{nom} [kg/min], can then be calculated using the thrust, T:

$$f_{nom} = \eta \cdot T \tag{1.17}$$

These expressions are used in all flight phases except during idle descent and cruise, in whose case the following equations are to be used:

- **The minimum fuel flow**, f_{min} [kg/min], corresponding to idle thrust descent conditions for both jet and turboprop engines, is specified as a funtion of the geopotential pressure altitude, h_p [ft], that is:

$$f_{min} = c_{f3} \cdot (1 - \frac{h_p}{C_{f4}}) \tag{1.18}$$

- **The cruise fuel flow**, f_{cr} [kg/min], is calculated using the thrust specific fuel consumption η, the thrust T, and a cruise fuel flow factor, C_{fcr}:

$$f_{cr} = \eta \cdot T \cdot C_{fcr} \tag{1.19}$$

In this case, the coefficients extracted from the fuel consumption model are: C_{f1}, C_{f2}, C_{f3}, C_{f4} and C_{fcr}.

To sum up, the coefficients that should be estimated to model the D, T and f_f of an aircraft performing an entire flight from the ascent phase to the descent phase would be the ones listed in table 1.1.

Model	Coefficient	Meaning
Drag Model	C_{D0}	Parasitic drag coefficient (cruise)
Drag Model	C_{D2}	Induced drag coefficient (cruise)
Thrust Model	$C_{T_c,1}$	1st max. climb thrust coefficient
Thrust Model	$C_{T_c,2}$	2nd max climb thrust coefficient
Thrust Model	$C_{T_c,3}$	3rd max. climb thrust coefficient
Thrust Model	$C_{T_c,4}$	1st thrust temperature coefficient
Thrust Model	$C_{T_c,5}$	2nd thrust temperature coefficient
Thrust Model	$C_{T_{cr}}$	Cruise thrust coefficient
Thrust Model	$C_{Tdes,high}$	High altitude descent thrust coefficient
Thrust Model	$C_{Tdes,low}$	Low altitude descent thrust coefficient
Fuel Flow Model	C_{f1}	1st thrust specific fuel consumption coefficient
Fuel Flow Model	C_{f2}	2nd thrust specific fuel consumption coefficient
Fuel Flow Model	C_{f3}	1st descent fuel flow coefficient
Fuel Flow Model	C_{f4}	2nd descent fuel flow coefficient
Fuel Flow Model	C_{fcr}	Cruise fuel flow factor

Table 1.1: Drag, Thrust and Fuel flow BADA models coefficients

1.1.6. BADA Aircraft Characteristics

As it has seen in previous sections, the equations describing each aircraft model in BADA are characterized with a set of coefficients called Aircraft Characteristics, some of them included in table 1.1. In BADA, the value of these coefficients is already given, values specified per aircraft type (e.g. A320), which can be found in the aircraft corresponding Operations Performance File (OPF).

The OPF file is an ASCII file, which for a particular aircraft type specifies the value of those coefficients included in BADA actions, operations and limitations models [15]. It is constituted by three types of lines identified by the first two characters in the line. These line types with their associated two leading characters are: CC (Comment Line), CD (Data line) and FI (end-of-file line). The comment lines are provided solely for the purpose of improving the readability of the file. All coefficients are contained within the CD lines in a fixed format. The end-of-file line is included as the last line in the file in order to facilitate the reading of the file in certain computing environments. Moreover, the data is organised into a total of eight blocks, separated between them by a comment line containing the block name to make easier the identification of each section [15]. These blocks are the following listed:

- **File identification block:** The file identification block provides information on the file name, creation date and modification date.

- **Aircraft type block:** This block specifies the ICAO code, number of engines, engine type and wake category of the aircraft type considered.

- **Mass block:** It provides the value of the aircraft reference, minumum, maximum and maximum payload masses, as well as, the weight gradient on maximum altitude.

- **Flight envelope block:** The flight envelope block provides the value of BADA speed envelope parameters, such as the maximum altitude at MTOW and ISA, the temperature gradient on maximum altitude and the maximum operating speed, mach number and altitude.

- **Aerodynamics block:** The OPF aerodynamics block provides the values of the parasitic and induced drag coefficients, as well as the value in knots of the stall speed V_{stall}, which are informed for different configurations and presented in the following order:

  - CR: Cruise
  - IC: Initial Climb
  - TO: Take-Off
  - AP: Approach
  - LD: Landing

  Moreover, another parameters affecting the aircraft aerodynamics are also specified:

  - S: reference wing surface area in m^2
  - $C_{Lbo(M=0)}$: Buffet onset lift coefficient (jet only)
  - K: Buffeting gradient (jet only)
  - C_{M16}: Compressibility effects
  - $C_{D0,\Delta LDG}$: drag increment for landing gear down

- **Engine thrust block:** The Engine thrust block includes the values of the maximum climb thrust coefficients, cruise and descent thrust coefficients and reference speeds during descent.

- **Fuel consumption block:** This block specifies the values of those coefficients regarding the thrust specific fuel consumption and the descent fuel flow. The cruise fuel flow correction factor value is also informed.

- **Ground movements block:** It includes the value of the BADA coefficients in reference of ground movements, such as take-off and landing length and aircraft wingspan.

The coefficients to be taken into account are included within the aerodynamics, engine thrust and fuel consumption blocks, which are those blocks corresponding to the Actions Model. In order to visualize the structure of the file, find an example below of an OPF file for an A320 aircraft type, as it is the type of aircraft that has carried out the QAR real trajectories used in this project. The value of the coefficients have been hided since it is not public information.

```
CCCCCCCCCCCCCCCCCCCCCCCCCCC A320.OPF CCCCCCCCCCCCCCCCCCCCCCCCCCCC/
CC                                                                /
CC AIRCRAFT PERFORMANCE OPERATIONAL FILE                          /
CC                                                                /
CC                                                                /
CC File name: A320.OPF                                            /
CC                                                                /
CC Creation date: Apr 30 2002                                     /
CC                                                                /
CC Modification date: May 14 2004                                 /
CC                                                                /
CC                                                                /
CC====== Actype ==================================================/
CD   A320          2 engines   Jet                       M        /
CC   Airbus A320 212 with CFM56 5 A3 engines             wake     /
CC                                                                /
CC====== Mass (t) ================================================/
CC     reference   minimum    maximum    max payload   mass grad  /
CD     XXXXXXXX    XXXXXXXX   XXXXXXXX   XXXXXXXX      XXXXXXXX   /
CC====== Flight envelope =========================================/
CC     VMO(KCAS)     MMO      Max.Alt      Hmax        temp grad  /
CD     XXXXXXXX    XXXXXXXX   XXXXXXXX   XXXXXXXX      XXXXXXXX   /
CC====== Aerodynamics ============================================/
CC   Wing Area and Buffet coefficients (SIM)                      /
CCndrst   Surf(m2)     Clbo(M=0)     k         CM16               /
CD 5      XXXXXXXX     XXXXXXXX   XXXXXXXX   XXXXXXXX             /
CC   Configuration characteristics                                /
CC n Phase   Name   Vstall(KCAS)   CD0        CD2       unused    /
CD 1 CR      Clean  XXXXXXXX   XXXXXXXX   XXXXXXXX   XXXXXXXX     /
CD 2 IC      1      XXXXXXXX   XXXXXXXX   XXXXXXXX   XXXXXXXX     /
CD 3 TO      1+F    XXXXXXXX   XXXXXXXX   XXXXXXXX   XXXXXXXX     /
CD 4 AP      2      XXXXXXXX   XXXXXXXX   XXXXXXXX   XXXXXXXX     /
CD 5 LD      FULL   XXXXXXXX   XXXXXXXX   XXXXXXXX   XXXXXXXX     /
CC   Spoiler                                                      /
CD 1      RET                                                     /
CD 2      EXT                            XXXXXXXX   XXXXXXXX      /
CC   Gear                                                         /
CD 1      UP                                                      /
CD 2      DOWN               XXXXXXXX   XXXXXXXX   XXXXXXXX       /
CC   Brakes                                                       /
CD 1      OFF                                                     /
CD 2      ON                             XXXXXXXX   XXXXXXXX      /
CC====== Engine Thrust ===========================================/
CC      Max climb thrust coefficients (SIM)                       /
CD     XXXXXXXX   XXXXXXXX   XXXXXXXX   XXXXXXXX   XXXXXXXX       /
CC      Desc(low)   Desc(high)   Desc level   Desc(app)   Desc(ld)/
CD     XXXXXXXX   XXXXXXXX   XXXXXXXX   XXXXXXXX   XXXXXXXX       /
CC      Desc CAS    Desc Mach    unused      unused      unused   /
CD     XXXXXXXX   XXXXXXXX   XXXXXXXX   XXXXXXXX   XXXXXXXX       /
CC====== Fuel Consumption ========================================/
CC   Thrust Specific Fuel Consumption Coefficients                /
CD     XXXXXXXX   XXXXXXXX                                        /
CC   Descent Fuel Flow Coefficients                               /
CD     XXXXXXXX   XXXXXXXX                                        /
CC   Cruise Corr.   unused     unused      unused      unused     /
```

```
CD       XXXXXXXX    XXXXXXXX    XXXXXXXX     XXXXXXXX     XXXXXXXX              /
CC====== Ground ================================================================/
CC           TOL         LDL           span          length         unused       /
CD       XXXXXXXX    XXXXXXXX    XXXXXXXX     XXXXXXXX     XXXXXXXX              /
CC==============================================================================/
FI                                                                               /
```

However, all these coefficients values are given for a specific aircraft type, in this case, an A320, not for every tail number aircraft, whose estimation is the purpose of this project.

1.2. Least Squares Regression Method

Within statistical modeling field, regression analysis is a set of statistical processes designed to estimate the relationships among variables. It includes several modeling and analysis techniques that allow the proper understanding of how the value of a dependent variable (or 'criterion' variable) changes when any one of the independent variables (or 'predictors') is varied while the other independent variables keep fixed.

Regression analysis is widely used for prediction and forecasting, where its use has substantial overlap with the field of machine learning. Regression analysis is also used to understand which among the independent variables are related to the dependent variable, and to explore the forms of these relationships [18].

In all cases, a function of the independent variables called the regression function is to be estimated. In regression analysis, it is also of interest to characterize the variation of the dependent variable around the prediction of the regression function using a probability distribution. Depending on the way of defining the regression function, it is possible to differenciate two types of regression techniques:

- **Parametric:** Parametric methods are characterized by defining the regression function in terms of a finite number of unknown coefficients that are estimated from data. Linear regression is an example of this kind of technique.

- **Nonparametric:** Nonparametric regression refers to techniques that allow the regression function to lie in a specified set of functions, which may be infinite-dimensional.

Since the objective of this project is to estimate the coefficients describing the BADA model equations (regression functions), the Least Squares parametric regression method will be applied.

1.2.1. Ordinary Least Squares

Ordinary Least Squares (OLS) regression is a statistical method of analysis that attempts to model the relationship between one or more independent variables and a dependent variable by fitting a linear equation to observed data. The simplest statement of such a relationship between an independent variable, labeled x, and a dependent variable, labeled y, may be expressed as a straight line in this formula [19]:

$$y = a + bx \tag{1.20}$$

Where a is the intercept and indicates where the straight line intersects the y axis and b is the slope, which indicates the degree of the line steepness. However, since there are infinite number of lines to represent this relationship, the chosen straight line needs to be the one that minimizes the error between the predicted value of y and the actual value of $\bar{y}$. The error will be given by a and b coefficients values, since they indicates how is the line, i.e., how the relationship between y and x is.

Thus, the OLS method allows to adjust linearly the coefficients of a model function (a and b) in order to better fit a data set. The best-fitting line is calculated by minimizing the sum of the squares of the vertical deviations from each data point to the line (if a point lies on the fitted line exactly, then its vertical deviation is 0), as it can be seen in figure 1.4.

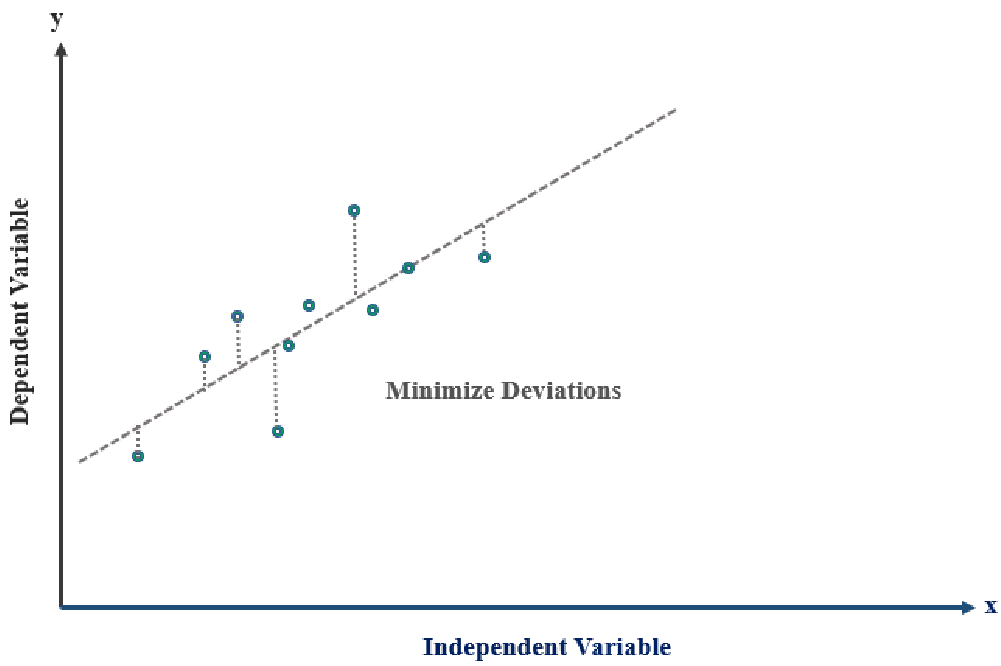

Figure 1.4: Least Squares Regression concept

Because the deviations are first squared and then summed, there are no cancellations between positive and negative values. These vertical deviations, also called as residuals, are actually, the difference between the observed data values ($\bar{y}$) and the fitted values provided by the model (y), which are obtained as results of every single equation involved.

The problem would be generally formulated in the following manner:

Given a simple data set of K points (data pairs) ($\bar{x}_i$, $\bar{y}_i$), i=1,...,K, $\bar{x}_i$ is the independent variable and $\bar{y}_i$ is the dependent variable whose values are found by observation. The model function has the form $y = f(x, \mathbf{p})$, where n adjustable coefficients are held in vector $\mathbf{p}$. The goal would be to find the value of $\mathbf{p}$ that minimizes the sum of the squared differences between the observed values of the dependent variable $\mathbf{\bar{y}_i}$ and the values predicted by the model $\mathbf{y_i}$, as expressed in next equation:

$$q = \sum_{i=1}^{K} \mathbf{W}(\mathbf{\bar{y}_i} - \mathbf{y_i})^2 \tag{1.21}$$

In this project, the observed values ($\mathbf{\bar{y}}_i$) are referring to the real trajectory states vector $\mathbf{\bar{y}_i}$ = [s_{QARi}, h_{QARi}, v_{QARi}, m_{QARi}] and the predicted values ($\mathbf{y_i}$) to the predicted trajectory states vector modelled with BADA $\mathbf{y_i}$ = [s_{BADAi}, h_{BADAi}, v_{BADAi}, m_{BADAi}], being s the distance to go, h the geometric altitude, v the True Airspeed (TAS) and m the aircraft mass. The sum of squared difference between them at each point i is what is wanted to minimize for all trajectories composing the training set. **W** is the weight vector associated to the state vector, which indicates the relative importance of each state when minimizing q.

Nevertheless, the generation of **y**, i.e., s_{BADA}, h_{BADA}, v_{BADA}, m_{BADA} depends on D, T and f_f obtained from BADA model equations, which are nonlinear functions of the coefficients (**p**) presented above. For this reason, Nonlinear Least Squares (NLS) method will be the one applied in this project.

1.2.2. Nonlinear Least Squares

Nonlinear Least-Squares method (NLS) is an extension of Linear Least-Squares regression, as it approximates the model by a linear one and then refines the coefficients by successive iterations. Thus, the unknown coefficients estimation process is conceptually the same, but allowing the use of a much larger and more general class of functions. See an example of nonlinear data fitting in figure 1.5, where the model function $f(x)$ depends on a nonlinear way to the coefficients a and b [20].

The general problem consists in fitting a set of K observations with a model that is nonlinear in n unknown coefficients ($K > n$). In the context of this project, given a set of K observed data points, $\mathbf{\bar{y}_i}$=[s_{QARi}, h_{QARi}, v_{QARi}, m_{QARi}] and a set of functions modeling $\mathbf{y_i}$=[s_{BADAi}, h_{BADAi}, v_{BADAi}, m_{BADAi}], which depend on a non-linear way of n coefficients (**p**=p_1, p_2, ... p_n) describing D, T and f_f, with $K \geq n$, the goal is to find the value of the coefficients vector **p** such that BADA model functions fit best the given data in the least squares sense, i.e. minimizing the sum of squared $\mathbf{\bar{y}_i}$-$\mathbf{y_i}$ shown in expression 1.21.

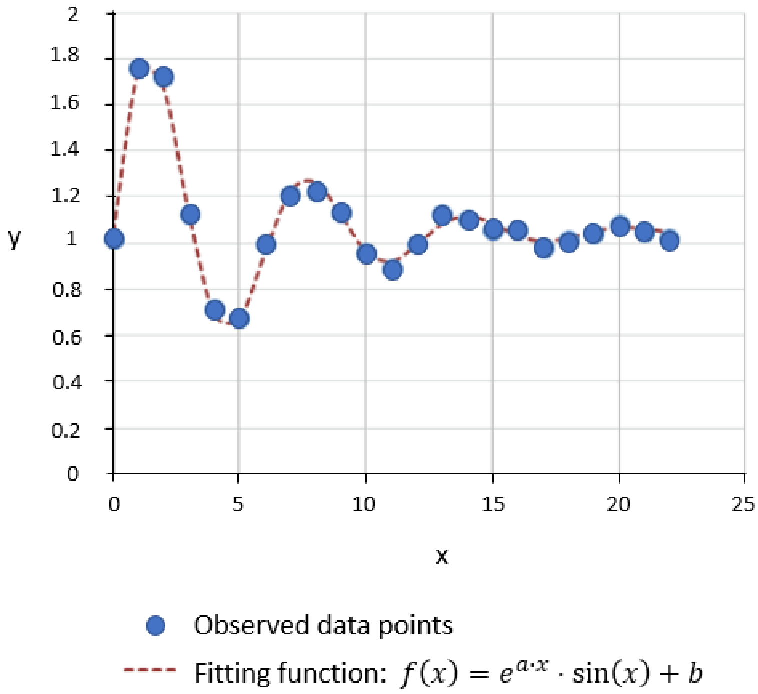

Figure 1.5: Nonlinear data fitting example [20]

However, the cost function q expressed by equation 1.21 must be adapted in the following manner to consider all the N trajectories of the training set.

$$q = \sum_{j=1}^{N} \sum_{i=1}^{K} \mathbf{W}(\bar{\mathbf{y}}_\mathbf{i} - \mathbf{y}_\mathbf{i}(\mathbf{p}))^2 \tag{1.22}$$

For this purpose, Nonlinear Least-Squares method, starting from a coefficients initial value ($\mathbf{p_0}$) equal to BADA OPF file values (for an A320), will start refining them by successive iterations until the cost function q is minimized.

1.3. DYNAMO Trajectory Predictor

DYNAMO (DYNAMic Optimiser) [21] is a trajectory prediction and optimisation framework capable to rapidly compute trajectories using realistic weather and aircraft performance data. DYNAMO generates the predicted trajectories decoupling their lateral and vertical profiles. The lateral profile prediction module is in charge of generating the sequence of waypoints from origin to destination, while the vertical profile prediction module generates the altitude and speed profiles with a fixed route. In order to understand how DYNAMO works and generates the modeled trajectories, the prediction problem must be explained.

Let us divide the aircraft trajectory into N flight phases. For each phase i, defined over the time period $[t_0^{(i)}, t_f^{(i)}]$, a state vector $\mathbf{x}^{(\mathbf{i})}(\mathbf{t})$, a control vector $\mathbf{u}^{(\mathbf{i})}(\mathbf{t})$ and parameter vector $\mathbf{p}^{(\mathbf{i})}$ are defined. In this project, the state vector $\mathbf{x}$ = $[v, h, s, m]$ is composed by the True

Airspeed (TAS), the geometric altitude, the distance to go and the mass of the aircraft and the control vector **u**=[T, γ] is composed, respectively, by the thrust and the aerodynamic flight path angle.

The dynamics of x are expressed by the following set of Ordinary Differential Equations (ODE), considering a pointmass representation of the aircraft reduced to a gamma-command model, where vertical equilibrium is assumed [22]:

$$\dot{s} = v\cos\gamma + W_s \tag{1.23}$$

$$\dot{h} = v\sin\gamma \tag{1.24}$$

$$\dot{v} = \frac{T-D}{m} - g\sin\gamma \tag{1.25}$$

$$\dot{m} = -f_f \tag{1.26}$$

where D is the aerodynamic drag, W_s the longitudinal wind speed, g is the local gravity acceleration and f_f is the total fuel flow. Find in figure 1.6 a visual representation of this point-mass model.

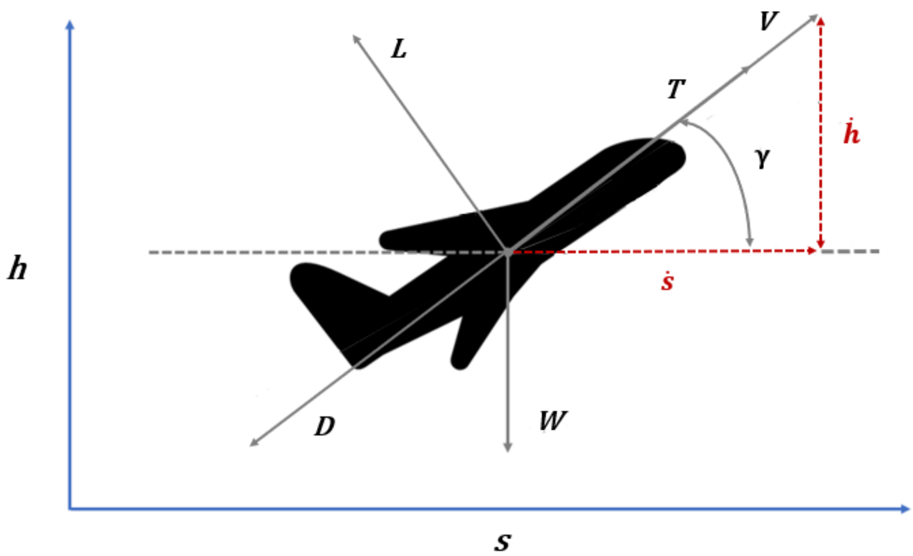

Figure 1.6: Distribution of forces of a point-mass model

From an initial aircraft state ($\mathbf{x_0}$), DYNAMO allows to compute future states, based on the flight intent, weather data and an aircraft performance model. Mathematically, the flight intent of each flight phase i could be given as a certain control vector closing the two degrees of freedom of ODE system (1.23-1.26), and an end condition. In practise, however, aircraft are not operated following specific T and γ profiles, and these controls are not known beforehand. Instead, climbs and descents are typically composed by constant

Mach (M), Callibrated Airspeed (CAS) or Energy Share Factor (k) segments performed at maximum climb and idle thrust, respectively; while in the cruise phase aircraft typically fly at constant pressure altitude (h_p) and M. In a generic formulation, two path constraints close the mathematical problem, allowing the explicit determination of the control vector **u** and consequently the obtention of an ODE system suitable for numerical integration using standard numerical procedures. The integration is performed until reaching the end condition, which triggers the switch to the following phase.

For this particular project, DYNAMO will be able to compute future states after being fed with the values of D, T and f_f obtained from BADA model equations, needed to integrate the ODE system represented by equations 1.23 - 1.26. Moreover, although DYNAMO can work with modelled weather data, in this case, it will be used real weather data extracted from QAR flights, since what is wanted to do is to generate the predicted trajectories under the same weather conditions in which real trajectories were performed to achieve a more accurate prediction.

Thus, the general role of DYNAMO in this project will be the generation of the predicted trajectories, which in the end is the generation of the cost function q to be minimized by least squares. DYNAMO will build q progressively, expressing all future states of each trajectory as a function of **p**. However, DYNAMO must be adapted to take the training set trajectories initial state vector ($\bar{\mathbf{x}}_\mathbf{0}$) as the starting point to compute the first future states of the predicted trajectories ($\mathbf{x_0}$), whose values will be used as reference to compute the following states and so on.

This is a general overview of the DYNAMO's role in this project, but more details and specifications of its adaptation according to the case study chosen, will be presented in chapter 4.

CHAPTER 2. INPUT DATA CHARACTERISATION

A Quick Access Recorder (QAR) [23] is an airborne flight recorder designed to provide quick and easy access to raw flight data, through means such as USB or cellular network connections and/or the use of standard flash memory cards. QARs are typically used by airlines to improve flight safety and operational efficiency, usually in the scope of their flight operational quality assurance plans. Like the aircraft's flight data recorder (FDR), a QAR receives its inputs from the Flight Data Acquisition Unit (FDAU), recording over 2000 flight parameters.

The QAR [24] is also able to sample data at much higher rates than the FDR and, in some cases, for longer periods of time. Unlike the FDR, the QAR usually is not required by a national Civil Aviation Authority on commercial flights and is not designed to survive an accident. Despite this, some QARs have survived to accidents and they have provided valuable information beyond what was recorded by the FDR.

The set of real flight data used in this project is composed by more than 3.000 coma separated value (csv) files, each one representing a single trajectory. These trajectories have been obtained through a set of QARs of a fleet of different tail number aircrafts of a specific airline, whose name can not be mentioned due to data confidentiality. The trajectories are discretized at one second interval represented by a file line. This is the reason why each file have different longitude in terms of number of lines, since every trajectory has different duration with respect to each others. Every line, provides the value of more than 40 different flight parameters, which thus, are informed at every second.

All flights considered are performed from a wide variety of European airports, as well as from some others from out of Europe during 3 consecutive years. Moreover, all these trajectories have been carried out using a fleet of more than 30 different tail number aircrafts, whose number of flights operated will be shown in section 2.3.

Apart from that, although the flight phase indicated in all files is 6 for all lines, plotting the altitude profile of several files it could be seen that each one represents a complete trajectory, from the ascent phase to the descent one. Thus, the flight phase field in files does not provide reliable information. However, not all trajectories have been plotted, so it is not possible to confirm that all files correspond to an entire trajectory.

Due to the confidential character of data, some information about the set of real trajectories used for this project will not be provided. However, some details about its caracterization will be presented in the following sections, as it is important to understand what kind of data we will work with.

2.1. Flight Parameters

In order to understand what kind of information it is required to work with, it has been detailed some of the parameters included in the corresponding csv files. The separation between the parameters values is not uniform, requiring a space normalization when programming with them. Since the data has been provided without any documentation or specification, the units indicated in table 2.1 are deductions done after analyse data.

Parameters	Meaning	Units
FLIGHT_NO	Unique Identifier	-
From	ICAO code Airport of Origin	-
To	ICAO code Airport of Destination	-
A/C_REG	Tail Number (A/C Registration)	-
Date	Date	-
TIME	Time	(HH:MM:SS)
groundspeed	Ground Speed	Knots
machNum	Mach Number	-
tAt	Total Air Temperature	°C
FlightPhase	Flight Phase	-
fuelFlowEngine1Fine	Fuel Flow Engine 1	kg/h
fuelFlowEngine2Fine	Fuel Flow Engine 2	kg/h
n1Eng1	Engine 1 N1	% maximum rpm
computedAirspeed	Callibrated Airspeed	Knots
radioHeight1	Radio-Altimeter Height	Ft
E1N2	Engine 1 N2	% maximum rpm
E1EGT	Exhaust Gas Temperature Engine 1	°C
PitchAttitude	Pitch Angle	Degrees(°)
RollAttitude	Roll Angle	Degrees(°)
VerticalSpd	Vertical Speed	Ft/min
gearSelDown	Gear Down	-
ApLatModes	Approach Lateral Mode	-
ApLongModes	Approach Longitudinal Mode	-
ApOpDesMode	Autopilot Open Descent Mode	-
AthrActive	Auto-Thrust Active	-
E2N1	Engine 2 N1	% maximum rpm
E2N2	Engine 2 N2	% maximum rpm
E2EGT	Exhaust Gas Temperature Engine 2	°C
SpdBrkAngle	Speed Brake Angle	Degrees(°)
autoLandWarn	Auto Landing Warning	-
Weight	Weight	kg
Altitude	Pressure Altitude	Ft
Fuel	Fuel	kg
MagHeading	Magnetic Heading	Degrees(°)
FlapLeverPosition	Flap Lever Position	-
LongitudeHiRes	Longitude	Degrees(°)
LatitudeHiRes	Latitude	Degrees(°)

Table 2.1: Flight Parameters of Real Data Set

2.2. Origin/Destination Airports

The number of different airports from which these flights have been performed is 63, being 57 European and 6 out of Europe. Conversely, there is only one airport considered as destination for all flights. As it has been aforementioned, due to data confidentiality, the name of these airports can not be provided. For this reason, the following plots show how many flights have been carried out from each airport, which are represented by a number.

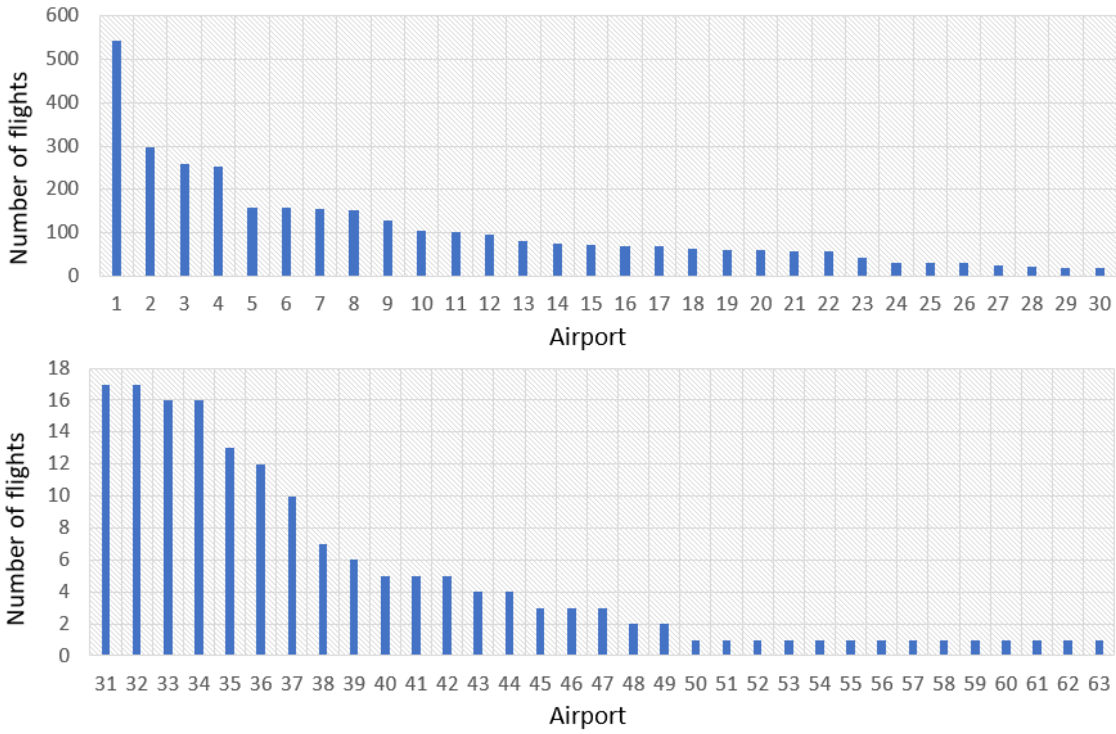

Figure 2.1: Number of flights performed per airport

2.3. Tail Number

These trajectories have been provided by the QARs of a fleet of more than 30 different tail number aircrafts, all of them corresponding to a A320-232. Considering that the purpose of this project is to estimate the degradation performance factors of a specific tail number aircraft, it is necessary to know how many trajectories have been carried out by each one. This is required in order to decide which is the aircraft chosen to fed the estimation algorithm, as the amount of trajectories performed by this aircraft must be enough for having a training data set and a validation data set. In general, since the higher the number of input flights the higher the prediction accuracy and reliability, the aircraft with a major number of flights carried out will be the one selected. Figure 2.2 shows the number of flights performed by each of the most representative 25 tail number aircrafts (in terms of amount of flights carried out), each one represented by a number. Therefore, the chosen aircraft has been the number 1, whose flights will be used to design the estimation algorithm.

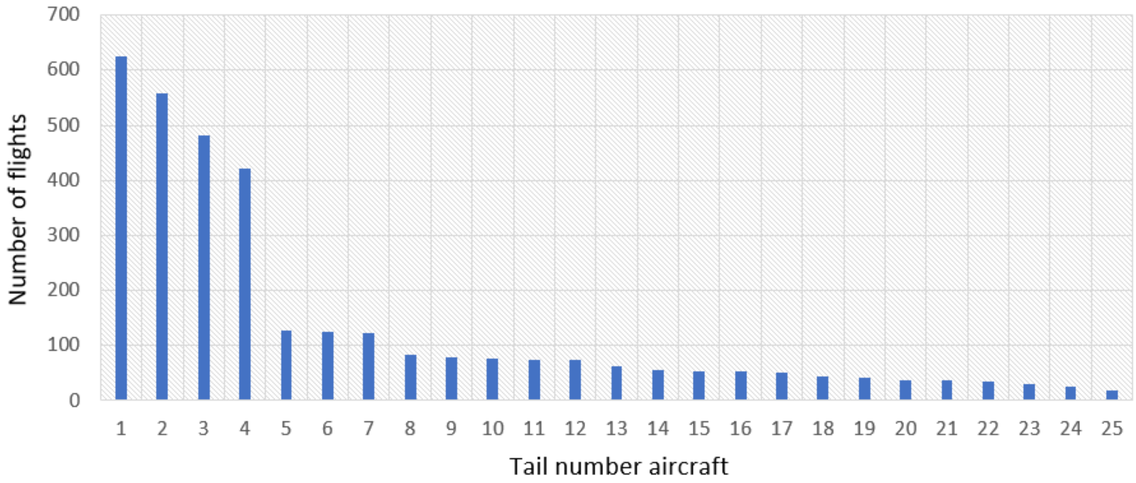

Figure 2.2: Number of flights performed by each tail number aircraft

2.4. Flights Distribution along the Time

In order to have an idea about the distribution of the flights along the time, find in the following table the percentage of flights performed per month and year:

	Year 1	Year 2	Year 3	Total
January	0%	3,2%	8,3%	**11,5%**
February	0%	2,2%	2,6%	**4,8%**
March	0%	1,6%	0%	**1,6%**
April	0%	4,0%	0%	**4,0%**
May	0%	7,0%	0%	**7,0%**
June	0,6%	7,5%	0%	**8,1%**
July	2,9%	7,4%	0%	**10,3%**
August	4,0%	8,7%	0%	**12,7%**
September	3,5%	6,0%	0%	**9,5%**
October	3,2%	5,5%	0%	**8,7%**
November	2,4%	6,6%	0%	**9.0%**
December	4,2%	8,6%	0%	**12,8%**
Total	**20,8%**	**68,3%**	**10,9%**	**100%**

Table 2.2: Distribution of fligths along months/years

Another important information is the percentage of flights distributed along the day. Although the trajectories have been carried out at different days of different years, the flights have been grouped by their initial time (indicated by the ”TIME” parameter of QAR files first line) to have an idea about how these ones are distributed along the day. In figure 2.3, it can be observed that the 55% of flights take place between 06:00 and 15:00 in the morning, followed by those ones performed during the afternoon between 15:00 and 21:00, representing a 38%. At night only a 7% is carried out between 21:00 and 06:00.

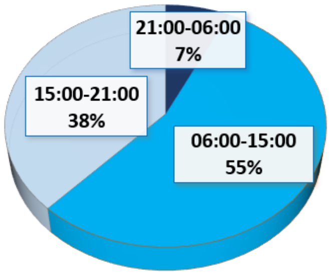

Figure 2.3: Percentage of flights vs Hour

2.5. Altitude Profile

After a previous analysis of the flight phase parameter, it is highlighted that 6 is the common value for all lines of all real trajectories. However, without information about what this value means or which flight phase it is referring to, the best option was to plot some flights and see which phases are included. All flights plotted include ascent, cruise and descent, but not all trajectories have been plotted, so it is not possible to ensure that all of them includes these three same phases. Find in figure 2.4 the altitude profile of four of the plotted trajectories.

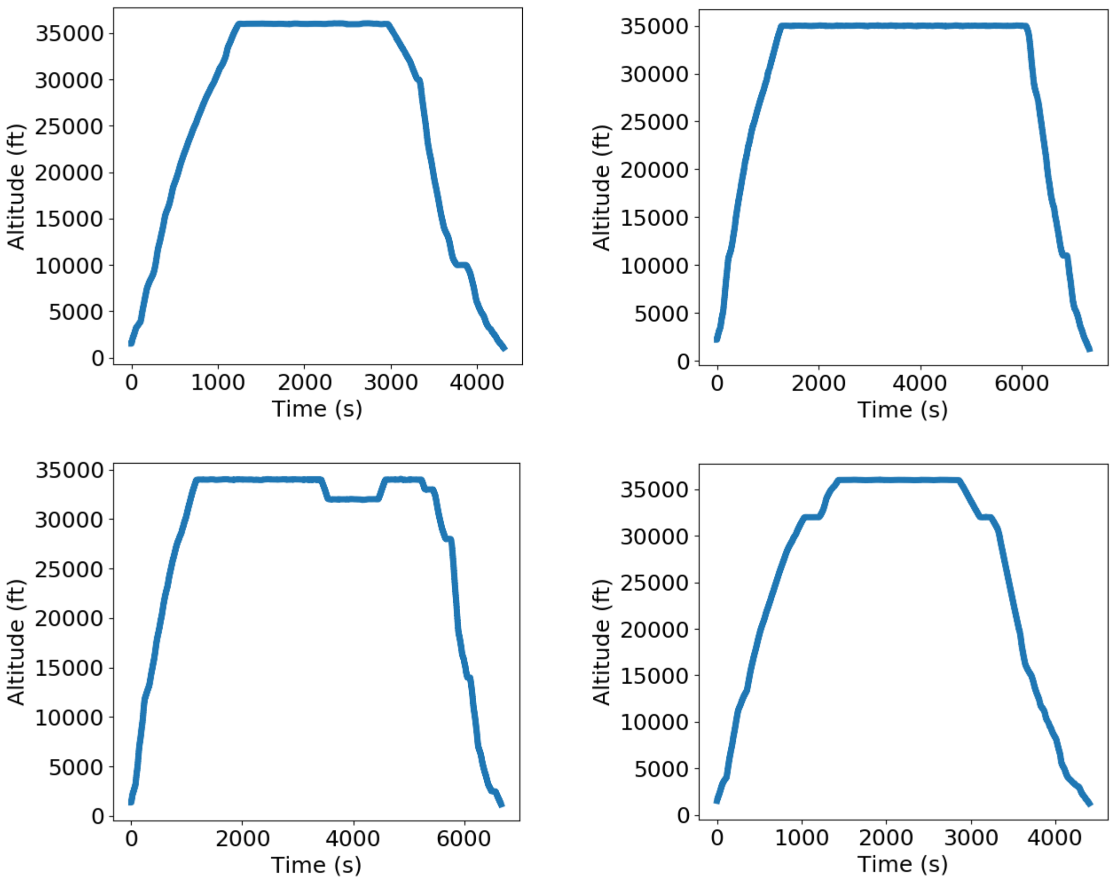

Figure 2.4: Altitude profile of 4 different flights.

CHAPTER 3. PARAMETERS ESTIMATION ALGORITHM

Once all tools and methods involved in this project have been introduced, it is important to know how they are related to and which is the process to follow for the coefficients estimation. Find in figure 3.1 the scheme of the estimation process workflow.

Before starting explaining the estimation process workflow, each component functionality will be introduced. As it is possible to observe, there are three differenciated blocks: QAR set of data, estimation architecture and validation and benchmark.

- **QAR set of data:** In chapter 2 it was shown the number of flights performed by each of the 25 most representative tail number aircrafts in terms of number of flights carried out. The aircraft represented by number 1 has been selected to design the estimation algorithm, as it is the one that performed the highest number of trajectories.

  These trajectories have been divided into two different data sets: a training set, used to train the estimation algorithm and composed by more than a half of the flights, and a validation set formed by the remaining flights that will be used to validate the results. Both sets should provide information about the initial conditions (initial trajectory states, $\mathbf{\bar{x}_0}$), trajectory states $\mathbf{\bar{x}_i}$=[s_i, h_i, v_i, m_i] and trajectory intents, which will be the inputs of the estimation architecture.

  However, the information provided by QAR files do not include any information about the intents, aspect that will constitute the main argument to justify the selection of the case study chosen.

- **Estimation Architecture:** The estimation architecture will be the result of integrating BADA models, real weather data, DYNAMO trajectory predictor and Nonlinear Least Squares (NLS) regression method.

  - **BADA model**: Actions BADA model equations, which were presented in chapter 1, will be used to model D, T and f_f as functions of the coefficients vector **p**, whose value must be estimated. These equations will be one of the DYNAMO trajectory predictor inputs, as they will be required to generate the cost function (q) that will be later minimized by least squares.
  - **Weather Forecast**: The weather forecast module pretends to provide the same weather conditions under which real trajectories were performed. For this reason, air temperature, pressure and density will be calculated from training set trajectories QAR parameters. Moreover, since longitudinal wind speed is required to integrate $\dot{s}$ as equation 1.23 expresses, it is necessary its calculation as well. The longitudinal wind speed will be treated as a new parameter of the QAR trajectories, being informed at each second like the others.

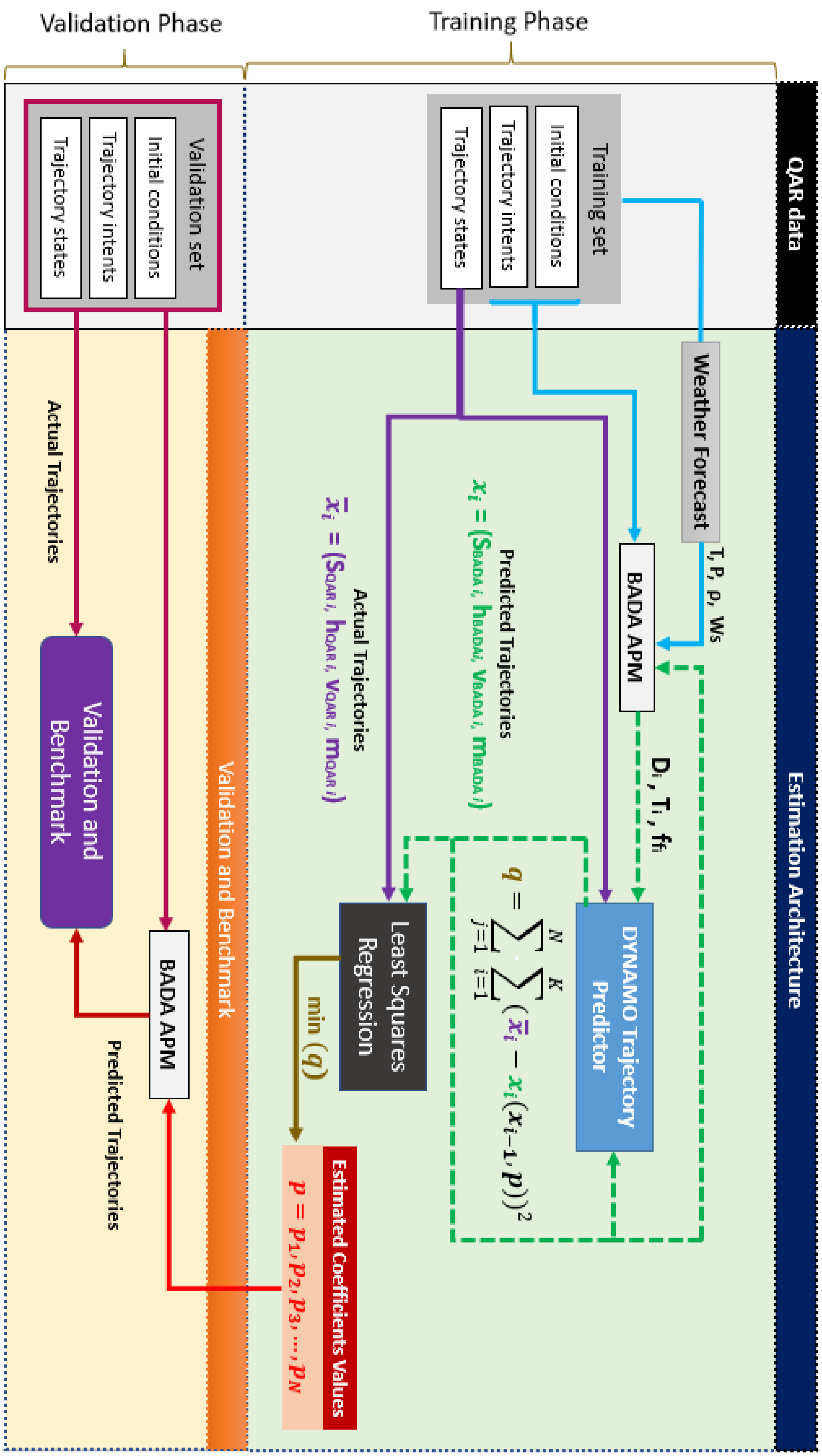

Figure 3.1: Estimation process workflow

- **DYNAMO Trajectory Predictor**: With the real weather environment and the BADA equations modeling D, T and f_f, DYNAMO will be able to integrate s, h, v and m, i.e., it will be able to generate the cost function (q) from which future states of the predicted trajectory will be computed by applying the estimated value of **p**, as result of minimize q.

- **Least Squares Regression Method**: Nonlinear least squares regression method, which will be integrated in DYNAMO framework, will refine iteratively the value of the coefficients until q is minimized, i.e., until the sum of the squared differences between the real and predicted trajectory states is minimized for all training set trajectories. The output will be the tailored coefficients vector value, that will be the main input of the validation and benchmark process.

- **Validation and Bechmark:** In order to analyse if the estimated coefficients value obtained is realistic and reliable, this one will be applied to BADA model equations for the obtention of the asociated D, T and f_f, values that will be used to integrate those predicted trajectories (i.e. the states) corresponding to the real ones forming the validation set. Actually, it is the same methodology than the followed in the estimation process but applying the coefficients previously estimated. The results must reflect a difference minimization, i.e., high similarity degree between the generated predicted trajectories and the real trajectories of the validation set.

Once the functionality and role of every component have been specified, the workflow can be detailed step by step.

- 1) The Actions BADA model equations will be the starting point of the estimation process. From the drag, thrust and fuel flow models equations, the coefficients to be estimated, represented by **p**, will be identified.

- 2) DYNAMO will be fed with the real weather environment extracted from QAR training set trajectories, composed by air temperature, pressure, density and longitudinal wind speed. Moreover, DYNAMO will use the initial states of these same trajectories ($\mathbf{\bar{x}_0}$) as the starting point of the predicted trajectories that it must generate (i.e., $\mathbf{\bar{x}_0}$ $=\mathbf{x_0}$) . Finally, the last DYNAMO's input will be the BADA model equations modeling D, T and f_f. With all this data, DYNAMO will be able to predict the future states $\mathbf{x_i}$ from which the cost function q will be generated. The cost function, at the end, will be a very long formula, expressing $\mathbf{x_i}$ as a function of the coefficients vector $\mathbf{p}$. Therefore, in order to obtain the values of $\mathbf{x_i}$=[s_i, h_i, v_i, m_i], **p** must be estimated through q minimization.

- 3) Then, Nonlinear Least Squares method will take q, which is the sum of the squared difference between the states of the predicted trajectories with respect to the reals, and it will start refining the values of $\mathbf{p}$ until q is minimized. Least squares will start the tuning process from an initial coefficients value $\mathbf{p_0}$ corresponding to the ones indicated in the BADA OPF file of the aircraft type considered (in this case, A320). The $\mathbf{p}$ output value will be the one allowing the obtention of those D, T and f_f at each point that minimize the difference between all predicted states ($\mathbf{x_i}$) with respect to the real states ($\mathbf{\bar{x}_i}$) for all training set trajectories.

- 4) Finally, in the validation and benchmark process, the estimated values of $\mathbf{p}$ will be applied to BADA equations in order to obtain the D, T and f_f values used to integrate $\mathbf{x_i}$, which will be compared with the validation set trajectories states. Again, the integration process will start with the initial states of the validation set trajectories. If the generated trajectories are "the closest" to the real trajectories, pointing out a minimization of the difference between them, the reliability of the estimation algorithm will have been proved.

This is a general overview of the estimation process. However, this one must be adapted according to the case study considered, reason why some of these elements will be explained again more in deep, together with the new changes applied.

4.1. Pre-processing of input data

The estimation process already introduced contemplates the coefficients estimation from the trajectory segment prediction associated to one or more flight phases or even the entire flight. However, it is essential to know the intents corresponding to each flight phase. In this particular case, the flight intents are unknown, as the QAR csv files do not include any parameter that indicates them. For this reason, the case study was initially focused on the cruise phase, in which aircraft typically fly at constant pressure altitude and mach number.

Nevertheless, in the end, instead of constrain the problem through the intents, it was decided to apply a certain control vector (**u**) formed by the T and γ estimated values at each point. This solution provides more flexibility as this method is more general and it can be applied to any flight phase. Although the approach changed, it was decided to continue considering only the cruise phase due to its stability and its lower complexity. However, to simplify the problem, only the values of T will be estimated while keeping fix the γ value, which will be always 0. This will allow to filter the actual cruise phase to consider only those trajectory points that keep the altitude constant. Therefore, the unknowns of this project are not only the tailored coefficients values, but also the values of T at each data point that must be applied to generate the future states of the predicted trajectories.

On one hand, it makes the dimension of the problem lower in terms of number of coefficients considered. In chapter 2 all the coefficients that could be estimated in case of consider an entire trajectory prediction were identified, but only the ones referred to the cruise phase are the coefficients of interest for this particular case study. Find in table 4.1 all the cofficients that will be estimated:

Model	Coefficient	Meaning
Total Energy Model (TEM)-Drag	C_{D0}	Parasitic drag coefficient
Total Energy Model (TEM)-Drag	C_{D2}	Induced drag coefficient
Fuel Flow Model (FF)	C_{f1}	1st thrust specific fuel consumption coefficient
Fuel Flow Model (FF)	C_{f2}	2nd thrust specific fuel consumption coefficient
Fuel Flow Model (FF)	C_{fcr}	Cruise fuel flow factor

Table 4.1: Case Study Coefficients

On the other hand, this represents an impact on the estimation architecture, since it is necessary to adapt it through including new elements and processes for the proper identification and extraction of the data corresponding to the cruise phase of every single training set trajectory. In summary, the complete trajectories will not feed the estimation process but only their most representative cruise segment. For this purpose, it is required the incorporation of a vertical speed filter and a cruise segments filter, as figure 4.1 shows.

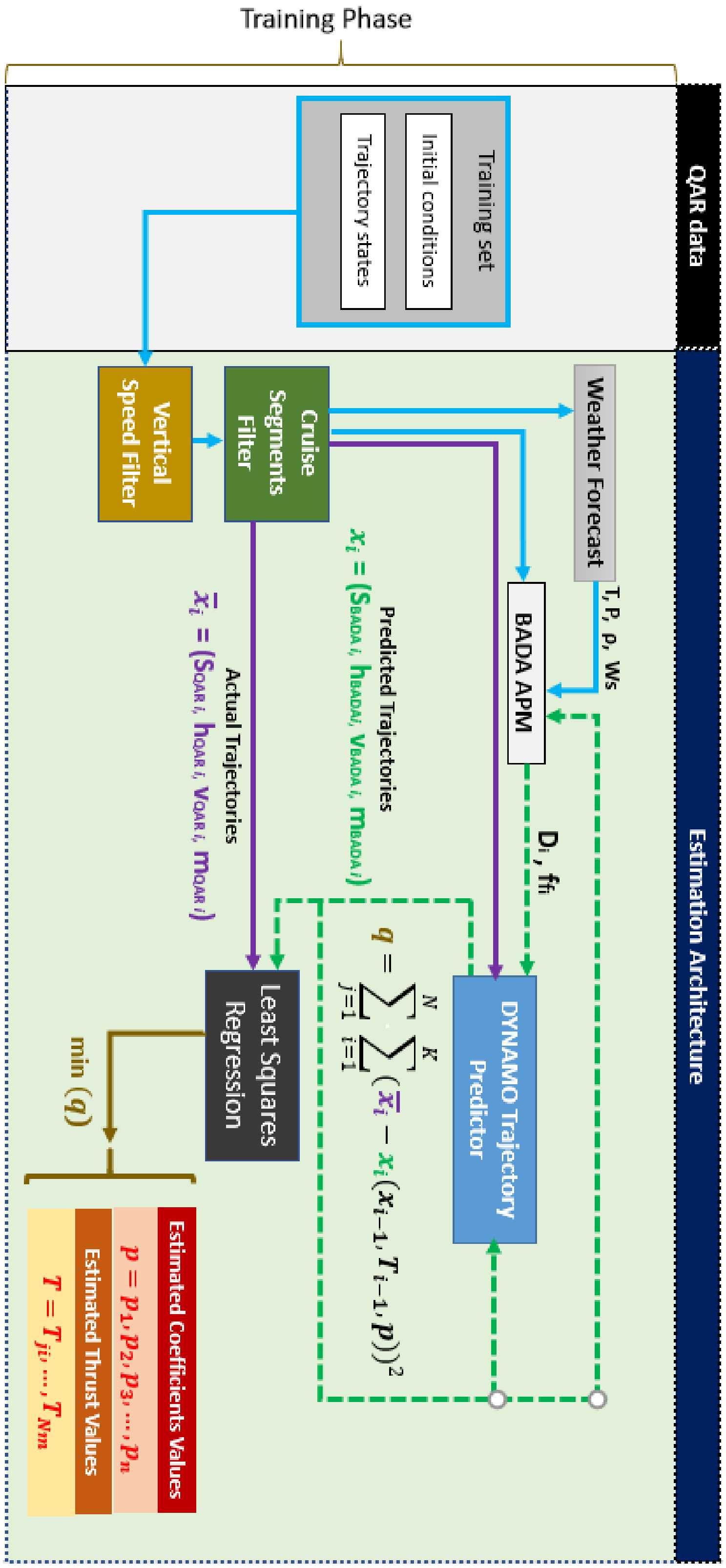

Figure 4.1: Case Study Estimation Architecture

In the following sub-sections each of the estimation algorithm input elements will be more in deep explained including their particularities associated to the case study, as well as, the vertical speed filter and segments filter, specifying how each element works and which is their functionality and role.

4.1.1. Vertical speed filter

In order to estimate the coefficients, the predicted trajectory have to be compared with the real trajectory, concretely with the most representative QAR cruise phase segment. In order to isolate it from the csv files, first of all, the complete cruise phase has to be extracted. The vertical speed filter is the element in charge of performing this extraction, process that can be illustrated in figure 4.2.

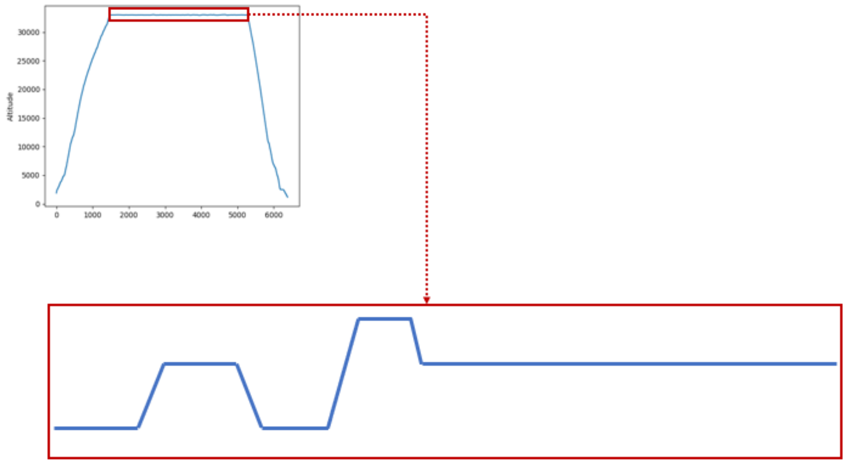

Figure 4.2: Cruise Phase Extraction

The cruise phase is the phase of flight from the top of climb (TOC) to the start of the descent toward the destination aerodrome or landing site [25]. During cruise, aircraft may fly at different cruising level/altitudes segments in which air pressure keeps constant (constant pressure altitude) as figure 4.2 shows.

In this context, vertical speed has been the QAR parameter selected to detect when the cruise phase starts and ends. Vertical speed indicates the vertical distance that aircraft is travelling per time unit, so when it is near 0 for a prolonged period of time, it can be concluded that the aircraft is neither ascending nor descending, and it is flying mantaining a constant altitude. Nevertheless, 0 is not a reliable value due to implicit data noise or small altitude fluctuations that could produced, reason why the target value of this parameter must be near 0 but enough conservative to consider these factors. Sometimes the recordings include errors in the measurements and this is something that has to be taken also into account when treating with data.

After analysing the correspondance between the altitude profile and the vertical speed profile of several trajectories, it was observed that 150 ft/min was a proper absolute value for the noise of this parameter. The vertical speed is positive in case of ascending and negative if the aircraft is descending, so by making the absolute value of this parameter, it was decided to extract the part of the trajectory with a vertical speed lower than 150, i.e., those points where the vertical speed is between 150 and -150 ft/min.

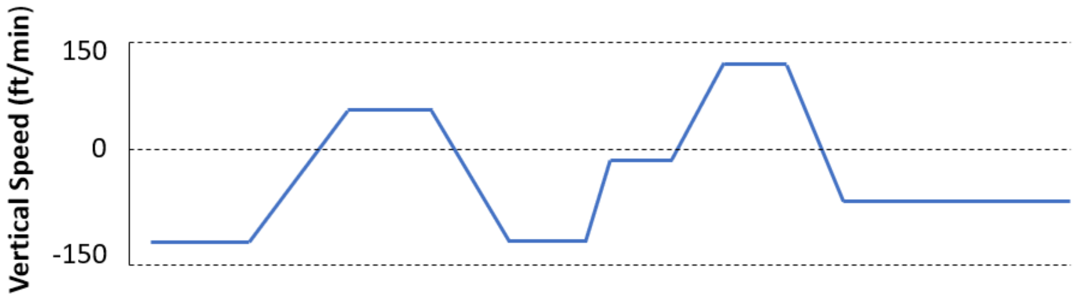

Figure 4.3: Vertical Speed bounds for cruise phase extraction

Therefore, what will be obtained will be the set of real trajectory points with an associated vertical speed lower than 150 ft/min in absolute value, points that can be consecutive or not, and even, they could belong to the climb and/or descent phases level off. This is the criterion applied to identify the phases at consant altitude of the training set trajectories. However, once the cruise phase has been isolated, the following step will be the selection of the most representative cruise segment, which will constitute the real trajectory segment to be compared with the generated by DYNAMO during the estimation process.

4.1.2. Cruise Segment Filter

Once the cruise phase has extracted, it is important to differenciate the cruising level segments included in this phase. Since there is not a ”Flight Level” parameter in the QAR csv files, firstly, it is required to know the flight level corresponding to each trajectory point. In order to do this, the ’Altitude’ QAR parameter value, which is expressed in feet, will be divided by 100 and then rounded to obtain the associated flight level. This will enable the classification of each trajectory point to a specific flight level, causing a separation of the cruise phase into flight level segments. However, the points corresponding to the same flight level can be consecutive or not.

Therefore, after all segments are identified, those with the same associated flight level are grouped. Since the idea is to select the most representative segment for the estimation process, the criterion for its identification is based on duration, reason why, the groups with a duration of 5 min or less will be discarded. Between the remaining, the flight level group selected for the experiment will be the one of major duration, which in figure 4.4 corresponds to the group of duration 2 (d2).

However, each group may be composed by the segments previously identified, and once the group has been selected, it is required to have another criterion to determine which is the most appropiate segment. The criterion will be the same used to choose the group of interest, i.e., the segment of major duration. Nevertheless, in order to determine their duration, it is needed a method capable to detect when these segments start and end.

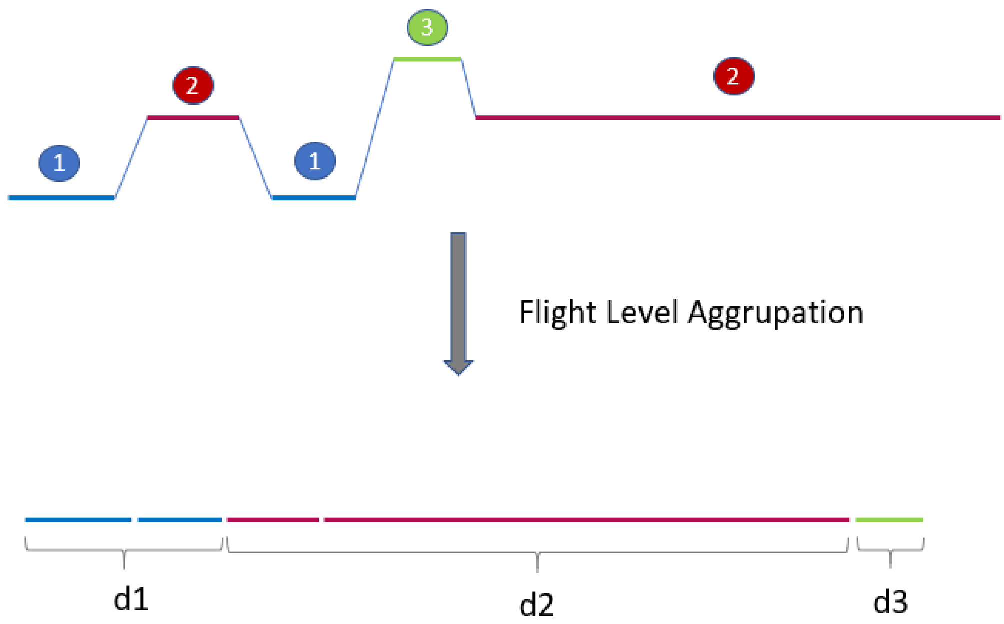

Figure 4.4: Cruise segments grouped by flight level

The method used for this purpose has been the accumulative sum. First of all, it is necessary to calculate the Δt between every trajectory point of the group considered. When the Δt is of one minute or less, it means that both points are included in the same segment, and when it is higher, it means that another segment has started. Although, the points of the real trajectories are separated by one second and, probably, it would serve as the Δt of reference, it has been considered 1 minute to be more conservative in their separation (it can exists measurement errors or imprecisions in data).

To apply the accumulative sum to this problem, the idea is to begin in the first point at 0 and then, start summing all the Δt observed in next points. The process will end when the last Δt observed is higher than 1 minute. In that moment, the counter will be restarted at 0, since it indicates a new group segment. The result will be a clear differenciation of the group segments with their respective accumulated Δt, among which, the one with the highest accumulated Δt will be the one selected. See in figure 4.5 a conceptual representation of this method.

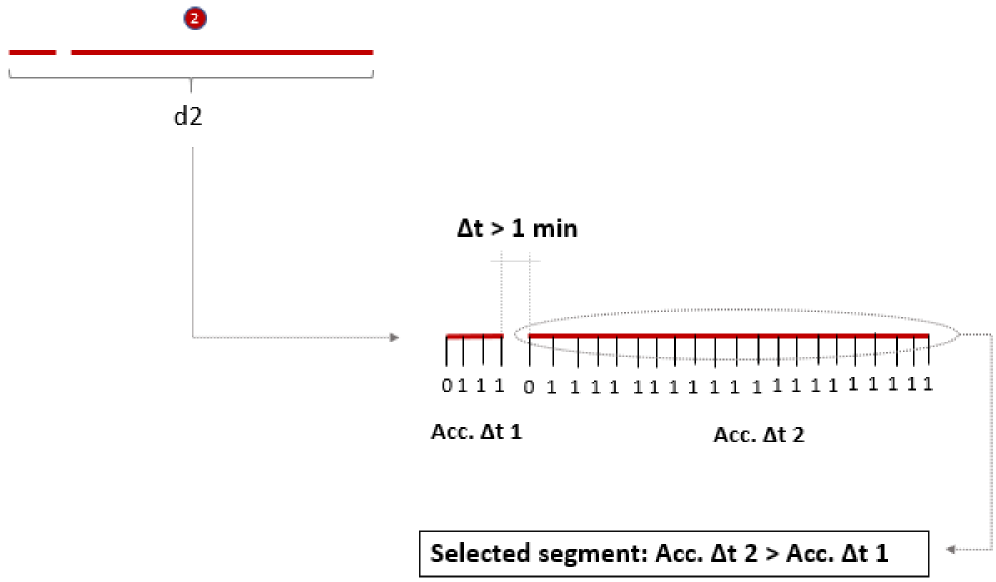

Figure 4.5: Accumulative Sum applied to cruise segment selection

4.1.3. Weather Forecast

The Weather Forecast module provides weather data to the estimation environment. The idea is to generate predicted trajectories as similar as possible to the real trajectories, reason why, the weather data will be extracted from the QAR csv files to guarantee that the weather conditions in both cases are the same.

In order to simulate these weather conditions, parameters such as temperature, pressure, density and longitudinal wind speed must be provided. However, there is no information about these parameters in the csv files, so it is required their calculation through the other QAR parameters.

The first step will be the calculation of the longitudinal wind speed, whose horizontal component is present in the point-mass model considered, and which is defined by the following expression 4.1:

$$\hat{W}_s = (\bar{S} - \hat{v}) \cdot \cos\hat{\gamma} \tag{4.1}$$

where S is the ground speed, v is the true airspeed and γ is the flight path angle. Those parameters available in QAR csv files will be represented with '¯' symbol, while the ones calculated from QAR data will be symbolized with '^'.

Contrary to what happens with the ground speed, which is one of the QAR parameters, v and γ are unknown and must be calculated too. Since v is required to obtain γ, v will be firstly determined by applying the following formula:

$$\hat{v} = \bar{M} \cdot \sqrt{\gamma_a \cdot R \cdot \hat{\tau}} \tag{4.2}$$

where M is the Mach number, γ_a the specific heat ratio of the air (1.4), R the specific gas constant for dry air (287,058 $J/kg^{-1}K^{-1}$), and finally, τ is the absolute air temperature.

τ is needed to obtain the windspeed, but at the same time is one of the required weather parameters. This one depends on M and the total air temperature (TAT) as equation 4.3 points out.

$$\hat{\tau} = \frac{\overline{TAT}}{1 + \frac{\gamma_a - 1}{2} \cdot \bar{M}^2} \tag{4.3}$$

By obtaining τ, it is already possible the determination of v. However, in order to calculate the longitudinal wind speed it is also required to know the value of γ.

In order to calculate γ, it has been assumed that the vertical speed QAR parameter is the derivative of the geometric altitude, when in reality is the derivative of the pressure altitude. Therefore, by assuming this, the vertical speed can be expressed as:

$$\bar{\dot{h}} = \hat{v} \cdot \sin \hat{\gamma} \tag{4.4}$$

Thus, γ can be isolated in the following manner:

$$\hat{\gamma} = \arcsin\left(\frac{\bar{\dot{h}}}{\hat{v}}\right) \tag{4.5}$$

The idea is to have the longitudinal wind speed associated to every single point of the real cruise segment selected. The next step will be the calculation of the air pressure and density. The air pressure can be obtained by dividing by 100 the pressure altitude QAR parameter, whose values will be used to calculate the air density together with τ:

$$\hat{\rho} = \frac{\hat{P}}{R\hat{\tau}} \tag{4.6}$$

However, the temperature, pressure and density of the air as well as the longitudinal wind speed are parameters expressed as a function of travelled distance (s), since the discretization of generated points will be defined for a set of specific "s". For this reason, the approach followed until now must change to consider s as the independent variable of the point-mass model represented by 1.23 -1.26, in which the trajectory states are expressed as a function of time. The details about all the modifications applied to the estimation algorithm in reference to this approach change are specified in section 4.2.

These are the elements in charge of pre-processing the input data. However, two more filters have been applied to correct some inconsistencies found in relation to aircraft mass. It was observed that in several consecutive points the mass were exactly the same, when it should decrease instead of remain equal, i.e., the f_f should never be equal to 0. Therefore, those points that present the same mass value than their previous points, i.e, that present a duplicated value, have been removed.

Furthermore, the other inconsistency detected has been that in some points the mass was higher than the previous point mass or lower than the following. These points have been also deleted to guarantee that the estimation algorithm is fed with realistic and consistent data. Find an example of this issue in figure 4.6.

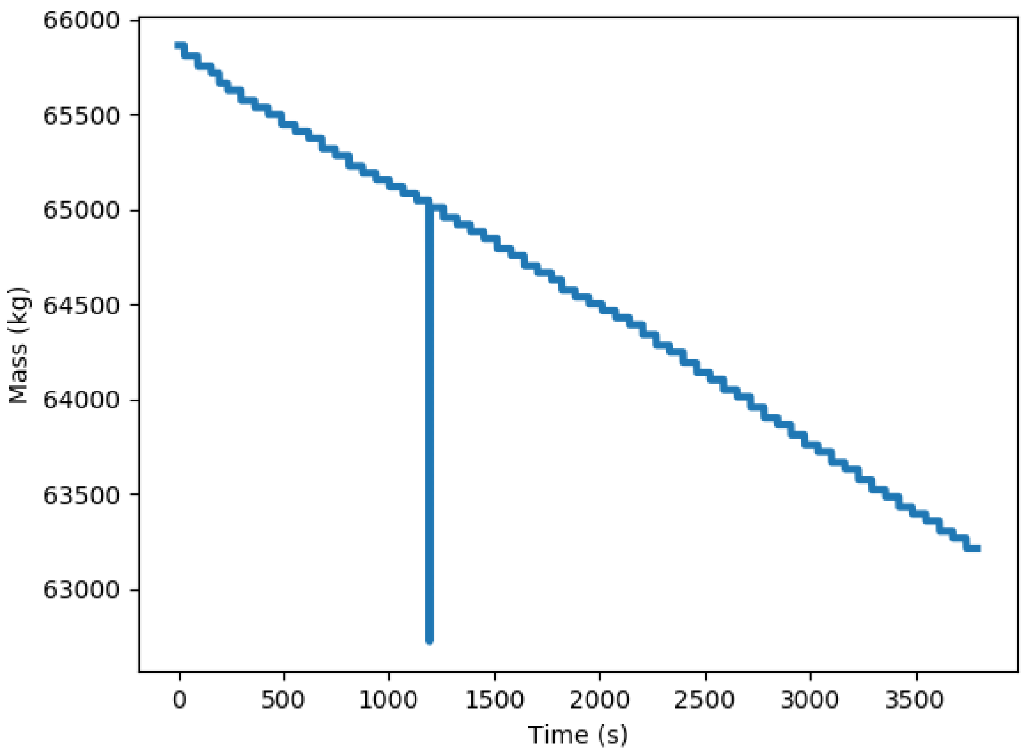

Figure 4.6: Mass values inconsistency

4.2. Case study 1 estimation algorithm

The estimation algorithm is the result of combining DYNAMO with NLS regression method, which uses the previously filtered data of the QAR training set trajectories as input to perform the coefficients estimation.

As it has been explained in previous chapters, DYNAMO computes the future predicted trajectory states, in this case, those states corresponding to the cruise segment selected ($\mathbf{x_i}$), by integrating the ODE system represented by equations 1.23 - 1.26. The states integration is performed by the fourth order Runge-Kutta method. Runge-Kutta methods [26] are one group of predictor-corrector methods that can be applied to an infinite variety of specific integration techniques. The basic idea of all Runge-Kutta methods is to move from step y_i to y_{i+1} by multiplying some estimated slope by a timestep. The difference between particular implementations involve how one estimates the slope. In the fourth-order Runge-Kutta method, which is the one used by DYNAMO, the basic idea is to combine 4 preliminary estimates to get one really good slope.

DYNAMO will integrate the first future states $\mathbf{x_1}$ starting from the initial states $\mathbf{\bar{x}_0}$ of the training set trajectories, which will feed BADA model equations to model the D and f_f associated to the following point. Therefore, with D and f_f and starting from $\mathbf{x_1}$, the second future states $\mathbf{x_2}$ will be computed, process that will be repeated until the last point is integrated. The integration process output will be the cost function (q), being the sum of the squared differences between the predicted and real states, in which the states associated to each predicted point ($\mathbf{x_i}$) will be expressed as a function of **p** and T, function that will be minimized by least squares.

First of all, since the control vector considered will be $\mathbf{u}_i$=[T_i, 0], only the thrust values applied at each point for the complete set of predicted trajectories will be estimated. Contrary, γ will be always 0, simplifying the original ODE system in the following manner:

$$\dot{s} = v + W_s \tag{4.7}$$

$$\dot{h} = 0 \tag{4.8}$$

$$\dot{v} = \frac{T - D}{m} \tag{4.9}$$

$$\dot{m} = -f_f \tag{4.10}$$

As a result, the point-mass model is reduced to a 3_{rd} order ODE system, since $\dot{h}$ is 0, which means that altitude h will be kept constant. Nevertheless, since the points to be generated will be discretized as a function of travelled distance "s", this last ODE system must be adapted to consider s as independent variable instead of time. Therefore, considering that:

$$\dot{s} = \frac{\mathrm{d}s}{\mathrm{d}t} = v + W_s \tag{4.11}$$

then:

$$t^{'} = \frac{\mathrm{d}t}{\mathrm{d}s} = \frac{1}{v + W_s} \tag{4.12}$$

$$v^{'} = \frac{\mathrm{d}v}{\mathrm{d}s} = \frac{T - D}{m}\frac{1}{v + W_s} \tag{4.13}$$

$$m^{'} = \frac{\mathrm{d}m}{\mathrm{d}s} = -f_f\frac{1}{v + W_s} \tag{4.14}$$

Therefore, $t^{'}$, $v^{'}$ and $m^{'}$ will be the new states that DYNAMO must integrate at each point.

In the end, the problem will be constrained through the application of one of the characteristic intents of the cruise phase (constant altitude) while keeping T as control variable, whose values at each point, T_i, will must be estimated as well. Therefore the unknowns of this case study will be the tailored value of the coefficients vector **p**=[$C_{D0,CR}$, $C_{D2,CR}$, C_{f1}, C_{f2}, C_{fcr}] and also T_i values.

The T_i value applied at each point must be the one that minimizes the difference between the predicted and real states at the following point $i+1$. This process will be repeated until it is reached the end of the segment. Therefore, for a cruise segment prediction, in which K points must be generated, i.e, K state vectors, $K-1$ control vectors $[T_1, 0]...[T_{K-1}, 0]$ will must be applied. Find in figure 4.7 a conceptual illustration of this process.

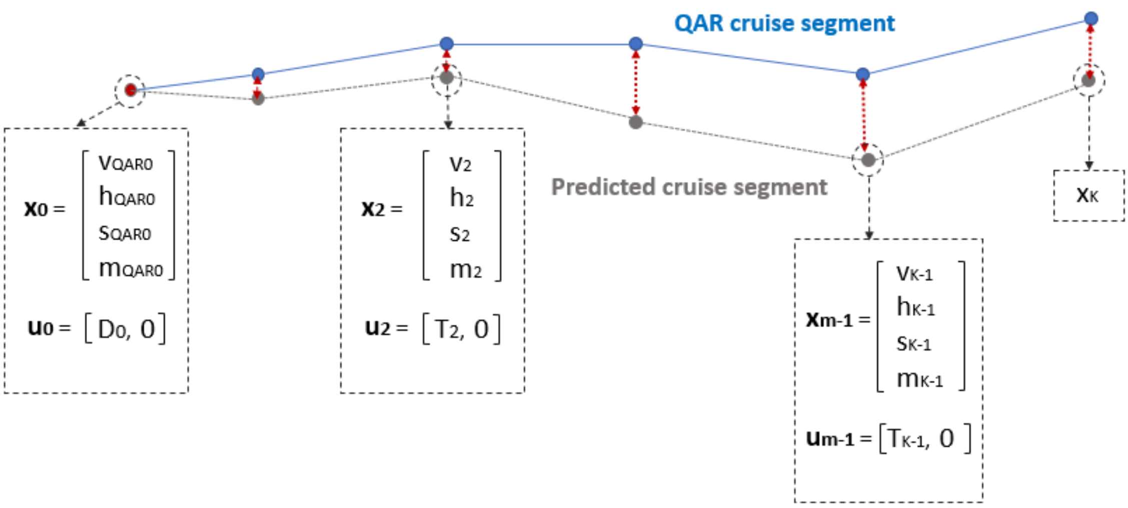

Figure 4.7: DYNAMO trajectory prediction problem resolution

As result, the cost function q generated by DYNAMO will be the sum of the squared differences between the predicted states and real states for all cruise segments of the training set, where the predicted states will be expressed as a function of **p** and the set of T_i applied:

$$q = \sum_{j=1}^{N} \sum_{i=1}^{K_j} (\bar{\mathbf{x}}_i - \mathbf{x}_i(\mathbf{x_{i-1}}, \mathbf{p}, T_{i-1}))^2 \tag{4.15}$$

where N is the number of QAR cruise segments involved in the estimation process and K_j the number of discretization points, which depends on the real trajectories being used. However, when dealing with a prediction problem, not always the prediction pretends to be optimal for all the states, in whose case, it is required to determine their relative importance by associating a specific weight to every single state. For example, if it is wanted to obtain an optimal prediction in terms of aircraft mass, the weight corresponding to m will be 1, while respect to the remaining states will be 0.

Thus, the weight vector $\mathbf{W_j}$=(W_v, W_h, W_s, W_m) must be multiplied by the squared difference between real ($\bar{\mathbf{x}}_\mathbf{i}$) and predicted states ($\mathbf{x}_i$). Thus, the final cost function q to be minimized is:

$$q = \sum_{j=1}^{N} \sum_{i=1}^{K_j} \mathbf{W_j} (\bar{\mathbf{x}}_i - \mathbf{x}_i(\mathbf{x_{i-1}}, \mathbf{p}, T_{i-1}))^2 \tag{4.16}$$

Since in this project is wanted to obtain, to the greatest extent possible, an optimal prediction for all the components of the state vector, 1 will be the value associated to each state for the sake of simplicity. Then, nonlinear least squares will minimize q by refining iteratively the values of **p** and T_i starting from an initial value of both, corresponding to the A320 BADA OPF file values as $\mathbf{p}_0$ and the initial drag D_0 (calculated by assuming BADA) as T_0.

Finally, in order to obtain reasonable results, the estimated values of **p** and T_i will be estimated within certain limits. The upper and lower bounds applied are the following:

- **Bounds of p:**
  - Upper: 10$\mathbf{p_0}$
  - Lower: 0.1$\mathbf{p_0}$
- **Bounds of $\mathbf{T_i}$:**
  - Upper: 1.2$\mathbf{D_0}$
  - Lower: 0.8$\mathbf{D_0}$

This means that the results must be values of **p** and T_i located within these limits. In case of **p**, the limits have been set according to own criteria about what would be coherent and acceptable. On the other hand, the bounds of T have been set considering this particular case study, i.e., taking into account only the cruise flight phase, in whose case it has been assumed that the values of thrust obtained will not exceed the T_{max} and T_{min} obtained from BADA. However, in future work, if it is wanted to consider also climb and descent phases, these bounds should be set according to BADA T_{max} and T_{min} values.

Actually, what the algorithm really does is to find the best combination of coefficients and thrust values, i.e., the combination that minimizes q within the limits established for **p** and T. Nevertheless, once the estimated values of **p** and T_i have been obtained and analysed, it has observed that, although they fulfill the optimality criteria (allow to minimize q), they are neither coherent nor reflecting the reality.

For this reason, it was necessary the implementation of an other approach in reference to the estimation algorithm design, which has been considered as case study 2. This second case study, will consider only the estimation of **p** by restricting the problem through the two intents characterizing the cruise phase (constant altitude and mach number). This will enable to focus the estimation to only the coefficients vector, allowing the obtention of more accurate and realistic values. However, an other filter, concretely a true airspeed filter, will be required to guarantee that the input real trajectories are only the ones performed at constant altitude and velocity.

4.3. Case study 2 estimation algorithm

In this second case study, the idea is to apply intents instead of control variables and design the estimation algorithm to only estimate the value of the coefficients for every particular tail number aircraft. Thus, the problem will be constrained through the intents corresponding to the cruise flight phase, which are:

$$\dot{h} = 0 \tag{4.17}$$

$$\dot{M} = 0 \tag{4.18}$$

When aircraft fly at constant altitude, the temperature in a relative short distance should be quite constant as well. Therefore, by observing equation 4.2 and assuming constant h, M and τ, it is possible to conclude that v (TAS) will be kept constant too, i.e, $\dot{v}$=0.

The first intent was already applied in case study 1 by considering γ as 0. Now, by applying the second intent, the ODE system will be reduced to a single differential equation, which is the one referring to aircraft mass:

$$\dot{m} = -f_f \tag{4.19}$$

Then, since:

$$\dot{v} = \frac{T-D}{m} - g\sin\gamma = 0 \tag{4.20}$$

and considering that γ is 0, thrust and drag will be equal (T=D). Therefore, the coefficients of interest will be the same that the ones estimated in case study 1. The fuel flow coefficients will be necessary to integrate $\dot{m}$, and the drag coefficients to compute D, and consequently T, which will also depend on the value of m at each point (D depends on lift coefficient (C_L), that at the same time, requires the value of m to be calculated, as equation 1.5 indicates).

The input data required is the same than in previous case study except for the longitudinal wind speed that it is not already needed. Contrary, the other weather parameters, i.e, τ, P and ρ, will have to be calculated too. Since the pre-process presented in section 4.1. allows to filter data by constant altitude, the next step will be to filter them by constant v. The idea consists in applying a true airspeed filter, identifing those cruise segments that have not been performed at constant v, which will be excluded of the estimation process. For this purpose, it is required to have a criteria to decide if v has been contant or not, criteria that, in this case, have been based on the following three indicators:

- **Standard deviation:** The standard deviation is the square root of the v variance, which is the average of the squared Δv measured from the mean.

- **Range:** The difference between the maximum and minimum value of v in a given cruise segment.

- **Mean squared deviation:** Measures the average of the squared Δv in a given cruise segment.

In the end, these indicators point out how v has varied along the trajectory segment. Therefore, the higher the values of these parameters, the less constant the true airspeed has been. However, it is required to determine these parameters thresholds from which v will not be considered constant. These thresholds have been fixed according to a previous analysis of the observed standard deviation, range and mean squared deviation of the cruise segments composing the training set. These ones must be enough restrictive to guarantee that this second intent is fulfilled, but allowing to work with the majority of cruise segments of the training set, since those ones with any of these parameters value higher than its established threshold will be excluded from the estimation process.

Once all input data have been pre-processed, the estimation process followed is the same than the one presented in previous section but considering only the estimation of **p** values. However, this time what is wanted to minimize is the difference between the predicted final mass (m_f) and the observed final mass ($\bar{m}_f$). However, the results obtained will be compared with the ones estimated when considering a mass difference minimization at each cruise segment point. The idea is to analyse if the accuracy of the estimation varies significantly or not. Therefore, the final cost functions to be minimized are:

$$q = \sum_{j=1}^{N} (\bar{m}_f - m_f(\mathbf{p}))^2 \tag{4.21}$$

which is the cost function associated to a final mass difference minimization. The q referred to a mass difference minimization at each point is the following:

$$q = \sum_{j=1}^{N} \sum_{i=1}^{K_j} (\bar{m}_i - m_i(m_{i-1}, \mathbf{p}))^2 \tag{4.22}$$

As in case study 1, nonlinear least squares method will refine the values of **p** iteratively until q is minimized, taking into account the same values of $\mathbf{p_0}$ (A320 BADA OPF file coefficients value) and the imposed upper and lower bounds to the output (i.e., $10\mathbf{p_0}$ and $0.1\mathbf{p_0}$ respectively). Find in figure 4.8 the final estimation process workflow followed for the coefficients estimation considering both mass difference minimization approaches.

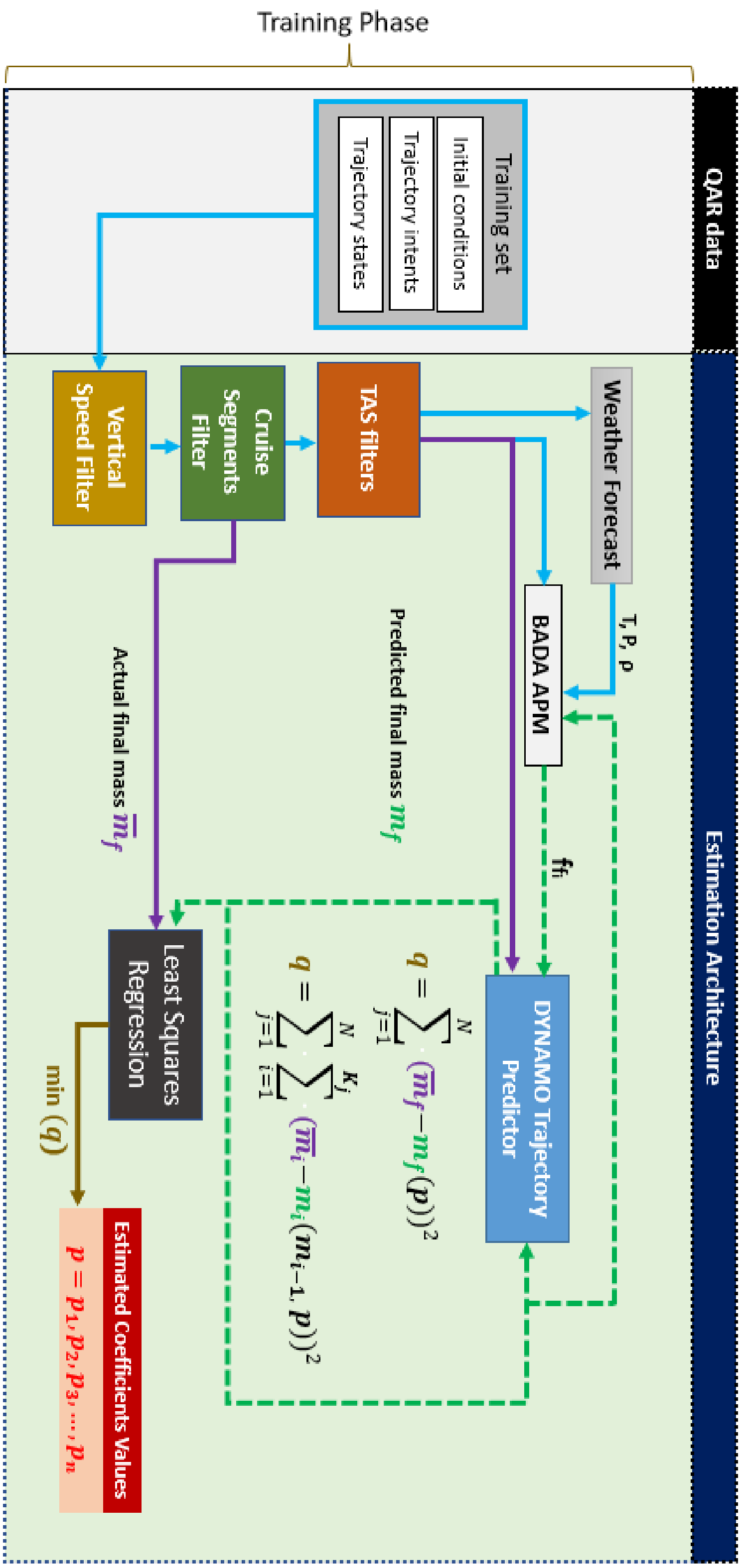

Figure 4.8: Case Study 2 Estimation Architecture

4.4. Results

In this section, an analysis of the results will be given. Since the algorithm has been designed to estimate the tailored coefficients values, the estimation of **p** has been performed for all different tail number aircrafts considered in this project.

For a proper validation of the results, the values obtained have been used to compute the final predicted mass (m_f) corresponding to the cruise segments of the validation set, with whose final observed mass ($\bar{m}_f$) have been compared. In order to analyse the reliability of the estimation algorithm, the difference between both has been displayed in a scatterplot to easily observe if the results are enough accurated. Thus, four scatterplots have been included in figure 4.9, showing the difference between m_f and $\bar{m}_f$ for all the trajectories carried out by the tail number aircrafts 1,2,3 and 4, since they are the ones that have performed more trajectories.

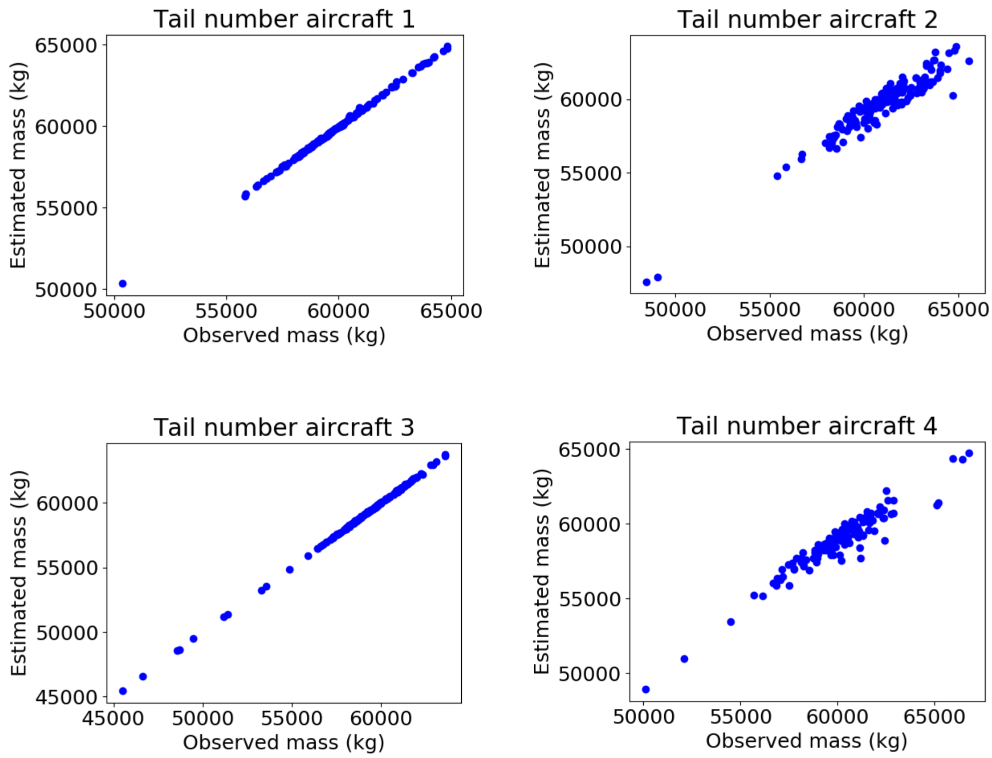

Figure 4.9: Predicted final mass values vs real final mass values per tail number aircraft

The x axis points out the $\bar{m}_f$ of every trajectory composing the validation set, while the y axis indicates the m_f computed by applying the estimated tailored value of **p** (corresponding to the tail number aircraft that is being analysed) to the BADA model equations. If the estimation was perfect, each point of the scatterplot would correspond to a value of m_f equal to the value of $\bar{m}_f$.

As it is possible to observe, the four scatterplots reflect a good estimation accuracy of **p**, since the values of $\bar{m}_f$ and m_f of each cruise segment considered are very similar, especially for tail number aircrafts 1 and 3. In case of tail number aircrafts 2 and 4, the dispersion of points ($\bar{m}_f$, m_f) is higher, aspect that denotes that the estimated value of the coefficients does not allow a prediction of m_f as accurate as in previous cases regarding tail number aircrafts 1 and 3.

Nevertheless, it is important try to identify the reason why the estimation accuracy is lower for aircrafts 2 and 4. For this purpose, instead of graphically represent how different are $\bar{m}_f$ and m_f at each point as in figure 4.9, it has been quantified this difference in kg to have an idea about which is the error's magnitude. Find in figure 4.10 the error in kg associated to the last point of each validation set cruise segment for these same 4 aircrafts.

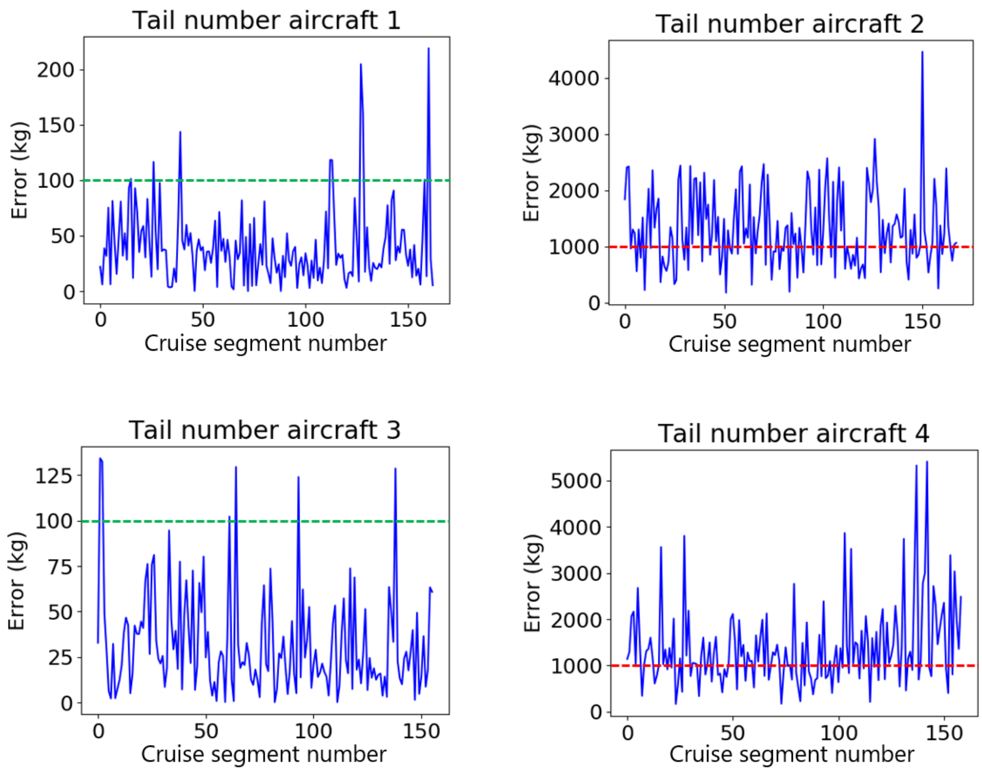

Figure 4.10: Error's magnitude considering a final mass error minimization

In plots included in figure 4.10, the x axis indicates the cruise segment number, i.e., the cruise segment identifier, whose associated error, $\bar{m}_f$ - m_f, in kg is specified in y axis. Clearly, it can be seen the correspondance of the error with the dispersion observed in figure 4.9. Overall, the error associated to tail number aircrafts 1 and 3 is quite low, presenting a difference between $\bar{m}_f$ and m_f lower than 100kg for the 94,5% and 96,1% of points respectively (all points below green line).

On the other hand, with respect to the tail number aircrafts 2 and 4, the error is of much greater proportions than in case of aircrafts 1 and 3. Firstly, there is not any point with an associated error lower than 100 kg. The minimum error found is of 180,5 kg and 160,9 kg and the maximum of 4463,2 kg and 5412,2 kg for aircrafts 2 and 4 respectively. To have a general idea, only a 36,3% and a 36,5% of the points associated to these two aircrafts present an error lower than 1000 kg (points below red line). However, although the error is significantly higher with respect to aircrafts 1 and 3, in reality it is low in terms of aircraft mass, since 1000 kg is approximately only a 1,5% of this one.

This difference in the error's magnitude could be given by the freedom that a final mass error minimization provides along all cruise segment except in last point. In order to discard this possibility, the error will also be displayed in figure 4.11 when the mass difference is minimized at each point.

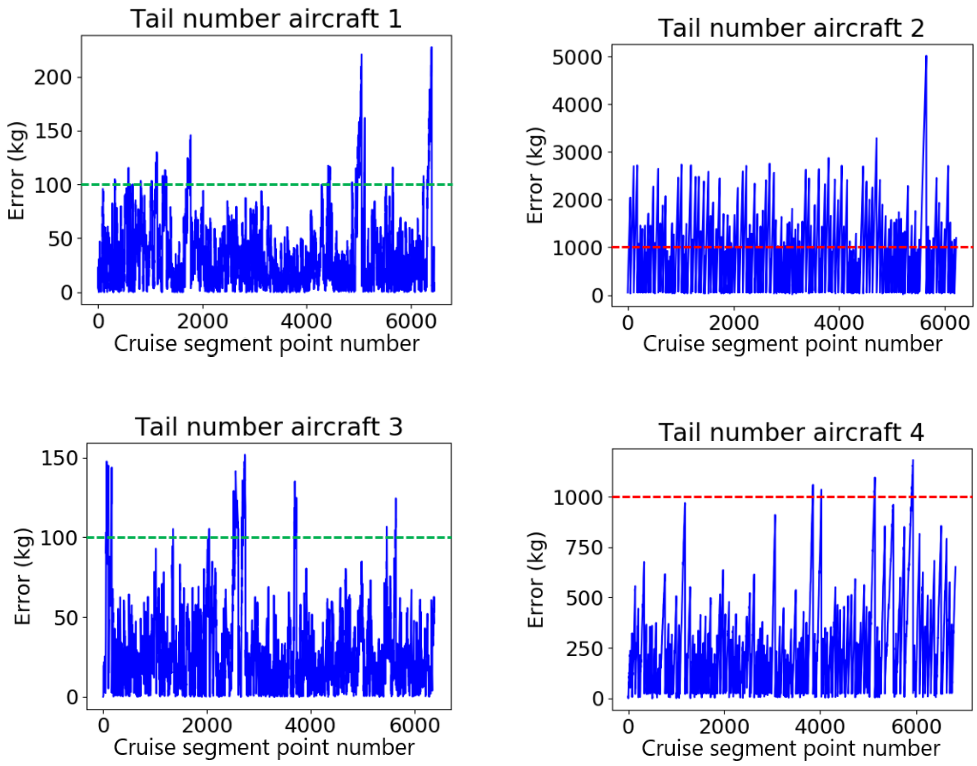

Figure 4.11: Error's magnitude considering a mass error minimization at each point

As it can be seen, the estimation accuracy slightly improves for the four aircrafts analysed. Regarding aircraft 1, the 95% of points have an associated mass difference lower than 100kg. With respect to aircraft 2, the 60 % of the points present an error lower than 1000kg from which, the 2.7 % corresponds to points with a mass difference under 100kg. The aircraft 3 is the one that reflects the most accurate estimation with a total of 96.6 % of points with a mass error lower than 100kg. Finally, aircraft 4, has been the one showing a greater difference between both approaches, whose estimation accuracy has improved a lot by applying a mass error minimization at each point. In this case a 99 % of points present an error's value below 1000kg (in front of the 36,5% considering a final mass error minimization), from which the 22.8 % are under 100kg. Find a more illustrative summary of this comparison in figure 4.12.

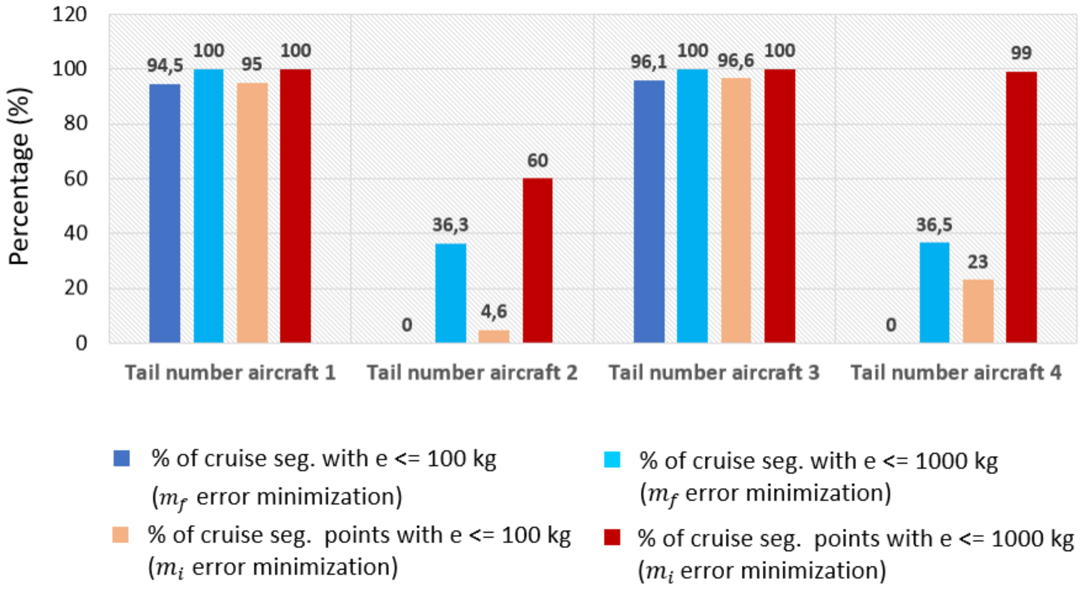

Figure 4.12: Error obtained according to the mass error minimization approach applied

In order to have more information, the range and mean value of the coefficients are displayed in figures 4.13, 4.14 and 4.15, from which it is also possible to perform an analysis about how dispersed are the estimated coefficients values when considering all set of different tail number aircrafts.

Starting from drag coefficients (figure 4.13), and applying a final mass error minimization, the 75% of the parasitic drag coefficient (C_{D0}) values are found between 0.03 and 0.08 approximately, from which the 50% are around 0.06, as mean value indicates. Furthermore, as it can be observed, the induced drag coefficient (C_{D2}) presents a higher range of values than C_{D0} since the minimum value of both are common but the maximum for C_{D2} is higher, concretely, 0.19. The distribution of the values when minimizing the mass error at each point is quite similar for these coefficients. The only aspect to be highlighted is that the range regarding C_{D0} is higher in this second case. Note that red cross in boxplots are representing $\mathbf{p_0}$, allowing to observe how different some of the tailored values of **p** are with respect to the ones provided by BADA, which are generalized by aircraft type.

Regarding the fuel flow coefficients, C_{f1} presents a more uniform distribution of the values, since the range is defined between 0.05 and 0.78 aproximately when minimizing mass error at last point, locating the 50% of values between 0.05 and 0.4 and the remaining 50% from 0.4 to 0.78. Contrary, C_{f2} concentrates the 50% of the values between 450000 to 550000, while the others are more dispersed from 210000 to 450000. Considering an error minimization at each point, the distribution sligthly varies. Concretely, C_{f1} presents a higher range of values, but the 50% is more concentrated between almost 0.4 and 0.6. Regarding C_{f2}, again, the range is higher, and consequently the dispersion, but mean value remains equal for both coefficients as it is shown in figure 4.14.

Finally, as it can be observed in figure 4.15, the cruise fuel flow factor (C_{fcr}) value is almost common for all tail number aircrafts in both mass difference minimization cases.

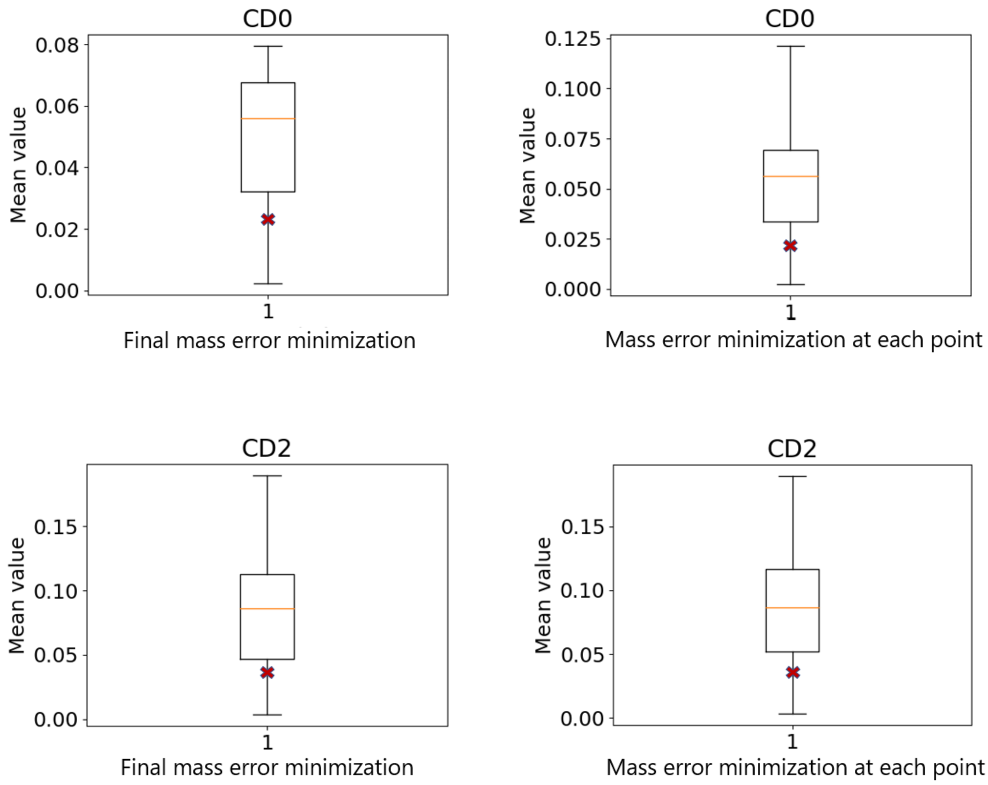

Figure 4.13: Drag coefficients range and mean value

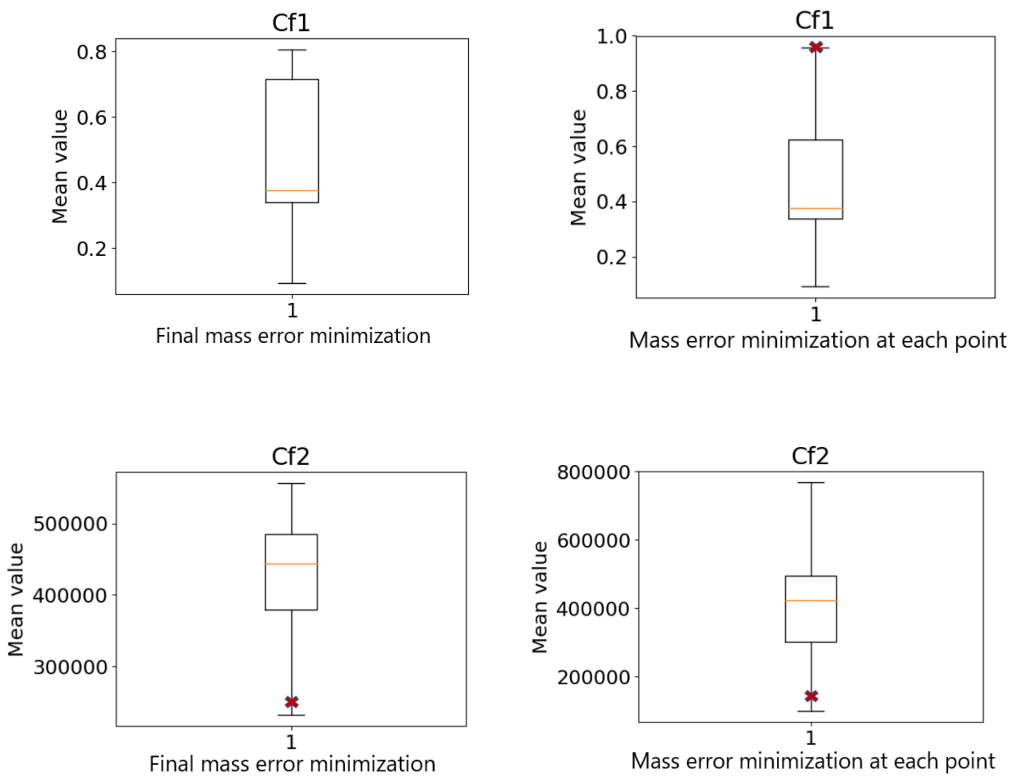

Figure 4.14: C_{f1} and C_{f2} coefficients range and mean value

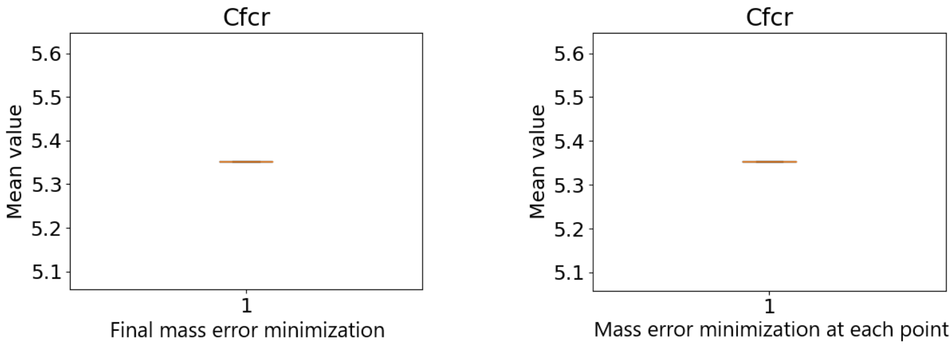

Figure 4.15: C_{fcr} coefficient range and mean value

Thus, although the mass difference minimization at each point slightly improves the estimation accuracy, there are not large differences between both approaches, obtaining results quite similar. Probably this difference on the error magnitude between tail number aircrafts 1-3 and 2-4 is due to the bounds imposed to the output that were set. Observing the estimated values of **p** for each of them, it can be seen that those corresponding to aircraft 2 and 4 present some coefficients values equal to the bounds. Concretely, for aircraft 2, the values of $CD2$, C_{f1} and C_{f2} are equal to 0.1$\mathbf{p_0}[CD2]$, 0.1$\mathbf{p_0}[C_{f1}]$ and 10$\mathbf{p_0}[C_{f2}]$ for both error minimization approaches, as well as, for aircraft 4 when minimizing final mass difference. When applying a mass error minimization at each point, the **p** values obtained for aircraft 4 includes 10$\mathbf{p_0}[CD0]$, 0.1$\mathbf{p_0}[CD2]$ and 10$\mathbf{p_0}[C_{f2}]$ as $CD0$, $CD2$ and C_{f2}.

Therefore, it can be concluded that, if the output were not bounded, the estimation accuracy obtained for tail number aircrafts 2 and 4 would be the same that the one achieved for aircrafts 1 and 3, although the values probably would not be realistic. Regarding the approach followed, the final mass error minimization would be the best option, since the cost function to be minimized is less complex allowing reduced execution times and the differences with respect to the error minimization at each point are not significant. Moreover, all coefficients except C_{fcr}, could vary a lot depending on the tail number aircraft considered. This means that, for an accurate trajectory prediction, it is not enough to have the values of these coefficients generalized by aircraft type, but they must be particularized for each different tail number aircraft, taking into account the specific performance of each individual aircraft.

CHAPTER 5. CONCLUSIONS

The main objective of this project was to design an algorithm capable to estimate the real performance coefficients value of a specific tail number aircraft, which are modelled by using BADA APM. The idea was to design a tool that made possible the building of tailored aircraft performance models.

One of the difficulties of this project was the analysis of the QAR real data to work with. This data, due to confidentiality issues, were provided without any documentation or specification and it was complicated to have a clear idea about what kind of information was being provided by each parameter, as well as, in which units they were being informed.

Moreover, an other big issue was to decide the project's case study according to the available data. During the project's development, the approach was changed several times trying to find the best case study fitting the QAR data. In the end, it was decided to focus the problem on the cruise phase, since the intents were already known. However, it would be interesting to explore other phases adapting the algorithm to be more general.

Furthermore, the estimation process has only been performed for jet aircraft. Thus, the modeling of other aircraft which are not based in jet engines would be an additional important point to take into account in future work. This offers a new criteria of tailoring aircraft performance models, in whose case, other BADA equations should be used.

The estimation algorithm developed is a code in python language responsible for the input data pre-processing (in charge of reading the QAR data files, selecting the most representative cruise segment, filtering data to solve some inconsistencies found and computing weather parameters from QAR ones) and coefficients estimation by combining the use of DYNAMO trajectory predictor and nonlinear least squares. In the future, this code could be improved in so many ways. Improvements could be applied in the model part in order to consider, as it has been commented previously, other flight phases and other aircraft types, and in the estimation part, to apply a more reliable output bounds setting criteria.

Results are promising, since they show a good estimated coefficients' fitting to the QAR data used. However, in the end, the problem was constrained through the characteristic cruise phase intents, discarding the estimation of thrust and flight path angle as controls, case that is more realistic and it could be interesting to explore. Nevertheless, unrealistic results were obtained when thrust values were estimated together with the coefficients in case study 1. For this reason, if a certain degree of freedom is considered in the system, other ways of limitting the output should be studied to achieve coherent results.

Finally, although BADA3 has been used as baseline of this project, the accuracy of the estimation algorithm could be improved if BADA4 model equations were used instead, as it provides a more accurate modeling of aircraft over the entire flight envelope and enables the modeling and simulation of future systems' advanced concepts.

Overall, it can be concluded that, although the algorithm could be improved by correcting and adapting some of its parts, this project has been a good starting point towards the obtention of tailored aircraft performance models.

BIBLIOGRAPHY

[1] Karolina Krajček, Dario Nikolić, and Anita Domitrović. Aircraft performance monitoring from flight data. *Tehnički vjesnik*, 22(5):1337–1344, 2015. 3

[2] Angela Nuic, Chantal Poinsot, Mihai-George Iagaru, Eduardo Gallo, Francisco A Navarro, and Carlos Querejeta. Advanced aircraft performance modeling for atm: Enhancements to the bada model. In *proceedings of the 24th Digital Avionics System Conference (DASC)*. AIAA/IEEE Washington, DC, 2005. 3, 5, 6

[3] Angela Nuic, Damir Poles, and Vincent 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. 3

[4] Enrique Casado. Tailoring bada 4 to individual aircraft for flight optimization. *Presented in BADA User Conference 2017*, May 2017. 4

[5] Nicholas RJ Lawrance, Salah Sukkarieh, and Bertrand Masson. Using high-frequency data for predicting fuel use of jet transport aircraft. *Journal of Aircraft*, 54(6):2115–2125, 2017. 4

[6] J Rosenow and H Fricke. *Flight performance modeling to optimize trajectories*. Deutsche Gesellschaft für Luft-und Raumfahrt-Lilienthal-Oberth eV, 2016. 4

[7] Marko Hrastovec and Franc Solina. Machine learning model for aircraft performances. *In proceedings of the 33rd Digital Avionics Systems Conference (DASC)*, 2014. 4

[8] Richard Alligier, David Gianazza, and Nicolas Durand. Energy rate prediction using an equivalent thrust setting profile. In *proceedings of the 5th International Conference on Research in Air Transportation, ICRAT 2012*, 2012. 4

[9] Richard Alligier, David Gianazza, and Nicolas 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. 4

[10] Judith Rosenow, Stanley Förster, and Hartmut Fricke. Continuous climb operations with minimum fuel burn. *proceedings of sixth SESAR Innovation days*, 2016. 4

[11] Mohammad Ghasemi Hamed, David Gianazza, Mathieu Serrurier, and Nicolas Durand. Statistical prediction of aircraft trajectory: regression methods vs point-mass model. In *proceedings of 10th USA/Europe Air Traffic Management Research and Development Seminar*. ATM Seminar, 2013. 4

[12] Eugene Morelli. Practical aspects of the equation-error method for aircraft parameter estimation. In *proceedings of the AIAA Atmospheric Flight Mechanics Conference and Exhibit*, page 6144, 2006. 4

[13] Joseph Frédéric Bonnans, Baptiste Gregorutti, Pierre Martinon, and Cédric Rommel. Aircraft dynamics identification for optimal control. In *proceedings of the 7th European Conference for Aeronautics and Aerospace Sciences (EUCASS)*, 2017. 4

[14] Damir Poles, Angela Nuic, and Vincent Mouillet. Advanced aircraft performance modeling for atm: Analysis of bada model capabilities. In *proceedings of the Digital Avionics Systems Conference (DASC), 2010 IEEE/AIAA 29th*. IEEE, 2010. 5

[15] European Organisation for the Safety of Air Navigation. *User Manual for the Base of Aircraft Data (BADA) Revision 3.12*. EUROCONTROL, August 2014. 5, 8, 9, 10, 12

[16] European Organisation for the Safety of Air Navigation. *Base of Aircraft Data (BADA) factsheet*. EUROCONTROL, 2018. 6

[17] Jakub Csanda. Using bada 4 performance model for aircraft simulation. Bachelor thesis, CZECH Technical University in Prague, May 2018. 6

[18] J. Scott Armstrong. Illusions in regression analysis. *International Journal of Forecasting*, 2012. 15

[19] International Encyclopedia of the Social Sciences. Ordinary least squares regression. https://www.encyclopedia.com/social-sciences/applied-and-social-sciences-magazines/ordinary-least-squares-regression, 2008. Accessed: 23-06-2018. 16

[20] Jan Karel Pieterse. Fitting curves to your data using least squares. https://www.jkp-ads.com/articles/leastsquares.asp, 2018. Accessed: 27-08-2018. 17

[21] Ramon Dalmau, Marc Melgosa, Santi Vilardaga, and Xavier Prats. A fast and flexible aircraft trajectory predictor and optimiser for atm research applications. *in proceedings of the International Conference for Research in Air Transportation*, 2018. 18

[22] David G Hull. *Fundamentals of airplane flight mechanics*. Springer, 2007. 19

[23] U.S. Department of Transportation. *Flight Operational Quality Assurance*. Federal Aviation Administration, 2004. 21

[24] Wikipedia. Quick access recorder. https://en.wikipedia.org/wiki/Quick_access_recorder, May 2018. Accessed: 05-03-2018. 21

[25] ECCAIRS Aviation 1.3.0.12 Data Definition Standard. *Attribute Values*. International Civil Aviation Organization (ICAO), April 2013. 33

[26] Michael Richmond. The fourth-order runge-kutta method. http://spiff.rit.edu/classes/phys559/lectures/rk/rk.html, 2018. Accessed: 19-09-2018. 39