UNIVERSIDAD POLITÉCNICA DE MADRID
ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA
AERONÁUTICA Y DEL ESPACIO

EXTRACTION OF TRAJECTORY CONTEXTUAL
FACTORS AND ITS APPLICATION IN THE
TRAJECTORY PREDICTION USING STATISTICAL
METHODS

TESIS DOCTORAL / DISSERTATION

CRISTIAN EDUARDO VERDONK GALLEGO
INGENIERO AERONÁUTICO / MSc AERONAUTICAL ENGINEER

2020

DEPARTAMENTO DE SISTEMAS
AEROESPACIALES, TRANSPORTE AÉREO Y
AEROPUERTOS
ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA
AERONÁUTICA Y DEL ESPACIO

EXTRACTION OF TRAJECTORY CONTEXTUAL
FACTORS AND ITS APPLICATION IN THE
TRAJECTORY PREDICTION USING STATISTICAL
METHODS

AUTOR

CRISTIAN EDUARDO VERDONK GALLEGU

INGENIERO AERONÁUTICO / MSc AERONAUTICAL ENGINEER

DIRECTOR DE TESIS

VÍCTOR FERNANDO GÓMEZ COMENDADOR

DOCTOR INGENIERO AERONÁUTICO / DOCTOR OF PHILOSOPHY

2020

POLITÉCNICA

Tribunal nombrado por el Sr. Rector Magfco. de la Universidad Politécnica de Madrid, el día de de 2020.

Presidente: _____

Vocal: _____

Vocal: _____

Vocal: _____

Secretario: _____

Suplente: _____

Suplente: _____

Realizado el acto de defensa y lectura de la Tesis el día de de 2020 en la E.T.S.I. / Facultad

Calificación

EL PRESIDENTE

LOS VOCALES

EL SECRETARIO

This Page is Intentionally Left Blank

AGRADECIMIENTOS / ACKNOWLEDGEMENT

El llegar a escribir una tesis doctoral es un proceso difícil. Por eso, unos agradecimientos que se quedan muy cortos a aquellas personas o entidades que me han ayudado de alguna manera en este camino.

Primero, y muy especialmente, a mi director de tesis, Fernando Gómez Comendador. Muchas gracias por toda la ayuda y guía que me ha prestado, la paciencia que ha tenido conmigo y mis “mañana te lo envío” y el apoyo incondicional.

A CRIDA A.I.E. y a José Miguel de Pablo por el apoyo que me han brindado durante todo el transcurso de la tesis. El compartir el conocimiento que atesora el personal de CRIDA es fundamental para la evolución de la investigación en ATM. Especial agradecimiento a mis compañeros Miguel García y Guillermo Orenge por su ayuda para preparar los papers que sustentan esta tesis. A mi amigo David Pérez, por abrirme a un campo completamente nuevo y ayudarme a encontrarme en el océano de datos.

To the Cranfield team, Manuel, Irfan and Dany. Also special thanks to my friends there, Marin, Tim, Iain and Mudassir. Of course, thanks to Simone, a great inspiration for me. For sure that I couldn't have reached the final results without your support and our discussions in B83 enjoying relaxed cups of tea. Al líder del equipo en Cranfield, Paco Sáez. Un mentor que ha marcado mi vida tanto profesional como personal.

Agradecerles a mis amigos, sobre todo a Carlos Ramos, por su apoyo constante durante todo este proceso, casi desde el principio. Me acuerdo de cierto viaje a Lisboa donde todos los allí presentes encontraron las palabras adecuadas para animarme.

A mi familia. Mis tíos Edu y Aurora y mis primos Juan Antonio (y Ani) y Edu. A Antonio y Toñi. A todos vosotros, por todos los momentos que hemos vivido juntos y viviremos.

Mis hermanas Myri y Ana (en orden cronológico), y mis padres, Broos y May. Bueno, hay tanto que agradecerles que me quedará corto con cualquier cosa que escriba aquí. Me hacéis sentir muy orgulloso de todos vosotros, y no hubiese llegado hasta aquí sin vosotros.

Finalmente, el más importante de todos, a Claudia, por tantos años aguantando mecha con este tema (en verdad con muchos temas), apoyándome incondicionalmente y haciendo que siguiese y siga soñando cada día.

This Page is Intentionally Left Blank

PUBLICACIONES / PUBLICATIONS

Journals

Verdonk Gallego, C.E., Gómez Comendador, V. F., Sáez Nieto, F. J., Orenga Imaz, G., & Arnaldo Valdés, R. M. (2018). **Analysis of air traffic control operational impact on aircraft vertical profiles supported by machine learning.** *Transportation Research Part C: Emerging Technologies*, vol. 95, no. October 2018, pp. 883–903, 2018. <https://doi.org/10.1016/j.trc.2018.03.017>

Verdonk Gallego, C.E., Gómez Comendador, V. F., Amaro Carmona, M. A., Arnaldo Valdés, R. M., Sáez Nieto, F. J., & García Martínez, M. (2019). **A machine learning approach to air traffic interdependency modelling and its application to trajectory prediction.**

Transportation Research Part C: Emerging Technologies, 107, no. October 2019, pp. 356–386., 2019 <https://doi.org/10.1016/j.trc.2019.08.015>

Amaro Carmona, M. A., Sáez Nieto F.J., and Verdonk Gallego, C. E., ‘**A data-driven methodology for characterization of a terminal manoeuvring area in multi-airport systems**’, *Transp. Res. Part C Emerg. Technol.*, vol. 111, pp. 185–209, Feb. 2020 <https://doi.org/10.1016/j.trc.2019.12.011>

Conferences

C. E. Verdonk Gallego and F. J. Sáez Nieto, **Discussion on Complexity and TCAS Indicators for Coherent Safety Net Transitions**, in *6th SESAR Innovation Days*, 2016, no. November, p. 8.

Verdonk Gallego, C. E., Gómez Comendador, V. F., Sáez Nieto, F. J., Martínez García, M., García Martínez, (2018). **Discussion on density-based clustering methods applied for automated identification of airspace flows.** *AIAA/IEEE Digital Avionics Systems Conference - Proceedings*, 2018-Septe, 1–10. <https://doi.org/10.1109/DASC.2018.8569219>

This Page is Intentionally Left Blank

RESUMEN

La investigación en la gestión del tránsito aéreo se ha concentrado tradicionalmente en una predicción de trayectorias centrada en la aeronave, basada en la reproducción de la dinámica de la aeronave dada una serie de restricciones e intenciones del piloto. La corriente actual para el aprovechamiento del dato en los predictores de trayectorias ha intentado precisar las entradas a los modelos, puesto que mejores entradas darán lugar a mejores salidas.

Las trayectorias voladas no dependen exclusivamente de la dinámica de las aeronaves, pero también del contexto donde se desarrollan. Tradicionalmente, el análisis del contexto se circunscribía a la situación meteorológica, mientras que otros factores de contexto eran obviados.

Esta tesis doctoral aborda la extracción de factores contextuales y su aplicación en la predicción de trayectorias, específicamente en los perfiles verticales de las mismas. Para ello, la disertación propone un modelo de datos orientado a la identificación y aislamiento de los factores contextuales operacionales. Seguidamente, la investigación realiza una serie de análisis para determinar el impacto de dichos factores en los predictores de trayectoria, usando las técnicas tradicionales de modelos de punto de masa y motores de computo basados únicamente en datos.

Finalmente, la investigación explora también las razones causales de modificaciones en la trayectoria por parte del controlador aéreo debido a tráfico circundante. Para hacerlo, la investigación propone métricas novedosas basadas en métodos estadístico para identificar interdependencias entre aeronaves. La investigación concluye evaluando el impacto de esta métrica en el perfil vertical de las aeronaves.

This Page is Intentionally Left Blank

ABSTRACT

The ATM academic community have traditionally focused on aircraft-centred trajectory prediction. Model-based trajectory prediction is based on reproducing aircraft dynamics subject to constraints and pilot's intent. Current data-driven trajectory predictors focus on precising inputs for feeding model-based trajectory predictors, as higher-fidelity inputs would result in improved predictions.

Flown trajectories do not only depend on aircraft dynamics but by their context as well. Trajectory context is usually associated to surrounding weather conditions. However, operational contextual factors have traditionally been disregarded.

This thesis addresses the extraction of trajectory contextual factors and its application in trajectory prediction, focusing on descent profiles. For doing so, the dissertation proposes a structured data model that enables the identification and isolation of the operational contextual factors. Then, the research conducts analyses of the impact of the studied factor on the trajectory predictors, introducing the most relevant as features for point-mass and data-driven trajectory predictors.

The research explores further the reasons driving modifications in the aircraft intent by the air traffic controller. Doing so, the research provides novel metric based on statistical methods to detect interdependent aircraft. The research finishes evaluating the impact of this interdependence metric in the vertical profile of the aircraft.

This Page is Intentionally Left Blank

CONTENTS

Contents	I
Figures.....	VII
Tables.....	XV
CHAPTER 1. Introduction	1
1.1. Background	1
1.2. Motivation	1
1.3. Research Scope.....	2
1.4. Research Approach	3
1.5. Thesis Structure	5
CHAPTER 2. Air Traffic Management and Trajectory Prediction.....	9
2.1. Introduction to Air Traffic Management	9
2.1.1. ATM Functions	10
2.1.2. ATM Actors	12
2.1.3. ATM Challenges and Modernisation	13
2.2. The Trajectory	15
2.2.1. Trajectory-Based Operations	15
2.2.2. Definitions and Terminology	16
2.2.3. Trajectory Planning	17
2.2.4. Flight Management.....	19
2.3. Trajectory Prediction	21
2.3.1. Trajectory Prediction Common Structure.....	22
2.3.2. Aircraft Intent Generation	25
2.3.3. Aircraft Motion and Performance Models - Classification of Trajectory Engines	25

2.4.	Framing the research.....	27
2.4.1.	Trajectory Prediction Challenges.....	27
2.4.2.	Leveraging data in ATM	30
2.4.3.	The Research Questions	32
CHAPTER 3. ATM Data Organisation and Flow Discovery		35
3.1.	Introduction.....	35
3.2.	ATM Dimensional Modelling	36
3.2.1.	The Demand Dimension	39
3.2.2.	Temporal and Spatial Dimensions	40
3.2.3.	Operational Dimensions	42
3.3.	Facts	45
3.3.1.	Trajectory Facts.....	45
3.3.2.	ATC Facts	49
3.4.	The Flow Discovery Process.....	52
3.4.1.	The Flow dimension	52
3.4.2.	A Review of the Clustering Methodology.....	53
3.5.	Research Data.....	60
3.5.1.	Flows crossing the LECBCCC sector	62
3.5.2.	Operational Vertical Procedures	65
3.6.	Summary of findings	67
CHAPTER 4. Analysis of Operational Factors.....		71
4.1.	Introduction	71
4.2.	Methodology	73
4.2.1.	Exploratory Analysis - Multilevel Linear Models	73
4.2.2.	Trajectory Computation Engines – Point-Mass Models	75
4.2.3.	Artificial Neural Networks	77
4.3.	Results.....	78

4.3.1.	Results of the MLM for initial evaluation of the ATC Operational Impact on the Trajectory Vertical Profile	78
4.3.2.	Point-Mass Models.....	85
4.3.3.	Artificial Neural Networks	88
4.4.	Discussion	90
4.5.	Summary of Findings.....	94
CHAPTER 5. Surrounding Traffic as a Contextual Factor		97
5.1.	Introduction	97
5.1.1.	Surrounding Air Traffic as a contextual factor.....	98
5.2.	Methodology	100
5.2.1.	Interdependency Modelling	100
5.2.2.	Interdependency Determination for two flights.....	110
5.2.3.	Vertical Trajectory Specification	112
5.2.4.	Contextual factors impacting on the Trajectory Vertical Profile.	114
5.3.	Results.....	115
5.3.1.	Traffic Interdependency Identification	115
5.3.1.5.	Trajectory Vertical Profile Characterisation.....	132
5.4.	Discussion	144
5.5.	Summary of Findings.....	147
CHAPTER 6. Conclusions and Further Work		151
6.1.	Overview	151
6.2.	Research conclusions.....	151
6.2.1.	ATM Data Organisation and Flow Discovery	152
6.2.2.	Analysis of Operational Factors.....	153
6.2.3.	Surrounding Air Aircraft as a factor	153
6.3.	Further work.....	154
6.3.1.	ATM Data Organisation	154

6.3.2.	Flow Discovery	154
6.3.3.	Contextual factors and point mass models.	155
6.3.4.	Surrounding Traffic	155
	References	159
	Acronyms	171
	Notation Chapter 5.....	177
	Indexes	177
	Notation (order of appearance).....	177

This Page is Intentionally Left Blank

FIGURES

Figure 1-1 Knowledge Discovery in Databases Process	4
Figure 1-2 Spiral Research Approach	5
Figure 2-1 Swiss Cheese Model for characterising a safety event[19]	11
Figure 2-2 A visual description of ATM. Objectives, functions, organisations and look-ahead times.....	12
Figure 2-3 Delay figures attributable to lack of ATC capacity in the ECAC area. Own figure made with data extracted from [27].	14
Figure 2-4 Gate-to-gate Business Trajectory process [30]	16
Figure 2-5 Notional representation of a conventional (blue) versus an optimal trajectory (black).	17
Figure 2-6 Flight management, positioning and guidance loops. The figure includes the stability loop and the control loop. The figure is based on [38] and [39].	20
Figure 2-7 Flight management, guidance and control loops, including the surveillance (ATC) control loop. To be developed.....	21
Figure 2-8 Trajectory Prediction Architecture (Bronsvoort) [35], based on [51].	23
Figure 2-9 EUROCONTROL TP Specification - Logical TP Breakdown.....	24
Figure 2-10 Notional operational stakeholder involvement in trajectory management [28], [29].....	25
Figure 2-11 Aircraft Motion Models for different TP approaches.	26
Figure 3-1 Knowledge Discovery in Databases Process. The blue rectangle describes the scope of this chapter.	36
Figure 3-2 ATM Processes.....	38
Figure 3-3 Flight dimension.....	39
Figure 3-4 Airport dimension.....	40
Figure 3-5 Temporal Dimensions.....	41
Figure 3-6 Spatial dimensions.....	42

Figure 3-7 Airspace Structure as a conjunction of operational framework and operational procedures [87].....	43
Figure 3-8 Operational definition of the Airspace Structure Dimension	44
Figure 3-9 Procedure Dimension	45
Figure 3-10 Schematic representation of the Trajectory Table Fact.....	47
Figure 3-11 Features which characterises a potential conflict [110]. The illustration reflects the relative dynamics of aircraft 2 with regard to aircraft 1. Two aircraft will be in conflict if their relative distance infringes the separation minima.	50
Figure 3-12 The nature of the flow dimension	53
Figure 3-13 Augmented Trajectory - Reference for construction.....	55
Figure 3-14 Left-hand side: Horizontal projection of a set of trajectories. Right-hand side: PCA transformation In this subplot, the components are dimensionless.	56
Figure 3-15 Considerations about the automation of the clustering process – PCA. Adaptation of J. Shlens code to a new sample.....	57
Figure 3-16 Reachability Plot for a subsample of Figure 3-14. The right-hand side column represents an equal transformed space, clustered differently based on the eps value. The right-hand side column represents the OPTICS reachability plot. Both columns are linked by colours.....	59
Figure 3-17 Automation of the density-based clustering to new samples	60
Figure 3-18 LECBCCC Sector - Barcelona Air Traffic Control Sector. [121]	61
Figure 3-19 Flow Discovery process after two iterations. Colours in left and middle columns reflect the same cluster in different projections. The right-side column illustrates the reachability plots for points displayed in the middle column.....	63
Figure 3-20 Upper airspace route structure for the LECB (Barcelona area)[121]	65
Figure 3-21 Percentage of flights which have received a specific CFL per Flow and the spatial distribution of CFL issuances. Colours differentiate flows. Associated latitude and longitude for data reflected on the left-hand side are displayed on the right-hand side.....	66
Figure 4-1 Knowledge Discovery Process - Scope of Chapter 4.....	72

Figure 4-2 Sample size for each Company broken down by Flow, RFL and Procedure. Colour shows details about Flow, maintaining same colours of Figure 3-20 and Figure 3-21. The columns are labelled by the sample size.....	79
Figure 4-3 Residuals of the LRD for a complete sample and all factors considered. Colours are common in both figures. The right-hand side of the Figure shows boxplots with whiskers with maximum 1.5 Interquartile Range (IQR).	80
Figure 4-4 Aircraft Model Frequency Distribution.....	84
Figure 4-5 Estimated mass CDF for the sample. Each colour represents an aircraft. It contains 2715 flights corresponding to B738 (blue) and 781 flights corresponding to A320 (red)...	86
Figure 4-6 Mean and Standard Deviation values of the residuals for different segmentation groups and learnt thrust laws for the training set.	86
Figure 4-7 Mean and Standard Deviation values of the residuals for different segmentation groups and learnt thrust laws for the testing set.	87
Figure 4-8 Mean and Standard Deviation values of the residuals for different segmentation groups for ANNs including them as inputs for the training set.	89
Figure 4-9 Mean and RMSE values of the residuals for different segmentation groups for ANNs including them as inputs for the testing set.	90
Figure 4-10 Mean and RMSE values of the residuals for different segmentation groups for ANNs including them as inputs, only for Flow PUMAL-LOREL	93
Figure 5-1 Knowledge Discovery Process - Scope of Chapter 5.....	98
Figure 5-2 Trajectory Predictor Logical Breakdown	100
Figure 5-3 Main Flows – LECBCCC Sector, Barcelona ATCC and recurrent patterns associated to the main flows.	102
Figure 5-4 Artificial Neural Net Predictor for estimating the time (τ) of overflying <i>γil</i> ..	104
Figure 5-5 Schematic representation of two recurrent patterns and a crossing point.....	108
Figure 5-6 Process for the generation of the Horizontal Interdependency Matrix	111
Figure 5-7 Description of the pdf of $T_{ilp} = 150$	113
Figure 5-8 Major flows and recurrent patterns. The recurrent patterns illustrated in green correspond to the PUMAL-LORES major flow, flying North-South, with destination LEPA.	

The recurrent pattern illustrated in orange corresponds to the MAMES-BCN flow, flying East-West in a cruise state.	115
Figure 5-9 Residuals for the ANN associated to the standard route associated corresponding to the flow Pumal-Lores (α_{110}) along γ_{11}	117
Figure 5-10 Residuals distribution for different values of γ_{11} for the standard route associated to the flow Pumal-Lores (α_{110}). The red series correspond to $\gamma_{11} = 30$ NM, blue series correspond to $\gamma_{11} = 60$ NM and green series correspond to $\gamma_{11} = 90$ NM	117
Figure 5-11 Residuals for the ANN TP for the standard route associated to the flow Mames-BCN (α_{210}) along γ_{21}	119
Figure 5-12 Residuals distribution for different values of γ_{21} for the standard route associated to the flow Mames-BCN (α_{210}). The red series correspond to $\gamma_{21} = 30$ NM, blue series correspond to $\gamma_{21} = 60$ NM and green series correspond to $\gamma_{21} = 90$ NM	119
Figure 5-13 QQ-plot for $\gamma_{11} = 60$ NM, corresponding to the recurrent pattern α_{110}	121
Figure 5-14 Actual Difference Time Values at the Crossing Point	122
Figure 5-15 Probability Density Function of the predicted time separation for four different cases. The purple box in each subplot is indicative of the actual time separation and the actual offset.	123
Figure 5-16 Probability of the actual time difference in terms of the residuals between actual and predicted time differences.	124
Figure 5-17 Residual distributions as a function of the actual time separation. Different colours represent different offset values.	125
Figure 5-18 Horizontal Interdependency Measure. Case of Study 1.	126
Figure 5-19 Horizontal Interdependency Measure. Case of Study 2	127
Figure 5-20 Horizontal Interdependency Measure. Case of Study 3	128
Figure 5-21 Horizontal Interdependency Measure. Case of Study 4	128
Figure 5-22 TOD Location depending on whether the encounter was classified as a conflict or not. The red colour indicates non-conflicting encounters, whereas blue columns indicate the opposite.	130

Figure 5-23 TOD Location depending on whether the encounter was classified as a conflict or not, for a sample with TOD locations larger than 30NM. The red colour indicates non-conflicting encounters, whereas green columns indicate the opposite.	131
Figure 5-24 Time to descend 50, 100 and 150FL Density Probability for non-conflicting flights with RFL370, for different TOD locations grouped in 10NM-width bins.	133
Figure 5-25 Time to descend 50, 100 and 150FL Density Probability for conflicting flights with RFL370, for different TOD locations grouped in 10NM-width bins.	134
Figure 5-26 Vertical profile Density Probability for descending 100 FL for non-conflicting aircraft with RFL370, for different locations of the TOD.	136
Figure 5-27 Vertical profile Density Probability for descending 100 FL for aircraft with RFL370, for different locations of the TOD.	137
Figure 5-28 ft1p distributions as a function of the time to descend $t_{1p} = 100$ 100FL for flights in non-conflicting scenarios. Each red point is the ft1p value corresponding to an actual time to descend value.	138
Figure 5-29 ft1p distributions as a function of the time to descend $t_{1p} = 100$ 100FL for flights in conflicting scenarios. Each red point is the ft1p corresponding to an actual time to descend value.	139
Figure 5-30 Vertical TP Residuals Distribution for non-conflicting flights (left-hand side) and conflicting flights (right-hand side). The top-row provides the residual distribution whereas the bottom row provides the values of ft1p corresponding to actual time to descend values.	140
Figure 5-31 Vertical TP Residuals Distribution for non-conflicting flights as a function of LEPA Runway Configuration.	141
Figure 5-32 Vertical TP Residuals Distribution for non-conflicting after including the airport configuration as a factor for predicting the TOD position. The top-row provides the residual distribution whereas the bottom row provides the values of ft1p corresponding to actual time to descend values. The top-row figure includes blue line corresponding to a fitted t-student distribution with parameters $\mu = -15 \text{ sec}$ and $\nu = 3.5$	142
Figure 5-33 Vertical TP Residuals Distribution for non-conflicting after including the airport configuration as a factor for predicting the TOD position for the testing dataset. The top-row provides the residual distribution whereas the bottom row provides the values of ft1p corresponding to actual time to descend values. The top-row figure includes blue line	

Figures

corresponding to a fitted t-student distribution with parameters $\mu = -18.5 \text{ sec}$ and $\nu = 2.19$	143
Figure 6-1 Typical Scenario in the LECBCCC Sector	156

This Page is Intentionally Left Blank

TABLES

Table 3-1 Surveillance Fact Table.....	46
Table 3-2 Trajectory Significant Point Fact	47
Table 3-3 ATCo Interaction Fact	49
Table 3-4 Potential Conflict Fact	51
Table 3-5 Flow Discovery Results for a sample of 10312 flights crossing the LECBCCC sector.....	63
Table 3-6 Cleared Flight Level Distribution per Flow	67
Table 4-1 Results considering different factors for the MLM and the complete sample. Training Sample 2756, Test Size 2481, New Sample: 611	79
Table 4-2 Results considering different factors for the MLM, and a sample without procedures showing a greater variance. Training Sample 2106, Test Size 1896, New Sample: 572.	81
Table 4-3 Results considering different factors for the MLM, and a sample without procedures showing a greater variance for the Flow Pumal - LEPA. Training Sample 1240, Test Size 137, New Sample: 385.....	81
Table 4-4 Results considering different factors for the MLM, and a sample without procedures showing a greater variance for the Flow Pumal - LEIB. Training Sample 507, Test Size 56, New Sample: 385.....	82
Table 4-5 Results for two level grouping depending (Procedure and RFL)	83
Table 4-6 Mean and RMSE values for residuals for B738 for the training and testing samples	91
Table 4-7 Mean and RMSE values for residuals for A320 for the training and testing samples.....	92
Table 5-1 Trajectory predictor performance as function of the mean error and the Root Mean Squared Error (RMSE) of the residuals for different values of γ_{11} for the standard route associated to the flow Pumal-Lores (α_{110}).....	118
Table 5-2 Trajectory predictor performance as function of the mean error and RMSE of the residuals for different values of γ_{21} for the standard route associated to the flow Mames-BCN (α_{210}).	120

Tables

Table 5-3 Chi-square tests for testing the independence between the position of the TOD and potential conflicts depending on the arrival runway.	147
---	-----

This Page is Intentionally Left Blank

CHAPTER 1. INTRODUCTION

1.1. Background

The Single European Sky Air Traffic Management Research (SESAR) [1] and NextGen [2] programmes aim at modernising the Air Traffic Management (ATM) system to improve cost-efficiency and capacity levels while maintaining safety.

The ICAO Global ATM Operational Concept [3] identified trajectory management at the core of the ATM modernisation process. Consequently, trajectory characterisation and prediction are cornerstones for new SESAR and NextGen solutions.

The evaluation of trajectory prediction (TP) requirements for foreseen SESAR solutions in 2014 [4] concluded that, at that moment, TP could not sustain the new Trajectory-Based Operations (TBO) concept, mainly due to a lack of understanding of the *contextual factors* driving TP uncertainties.

Simultaneously, data exploitation has increasingly become crucial to understand underlying patterns governing diverse business processes. The ATM system is gearing towards digitalisation, which will underpin operations.

Finally, traditional model-based trajectory predictors are at the core of ground ATM systems. Academic research is shifting towards enriching these models by using data [5], [6] or even, pure data-driven trajectory predictors [7].

1.2. Motivation

The ATM academic community have traditionally focused on aircraft-centred trajectory prediction. Model-based trajectory prediction is based on reproducing aircraft dynamics subject to constraints and pilot's intent, with different fidelity levels. Current data-driven trajectory predictors focus on precisising inputs for feeding model-based trajectory predictors, as higher-fidelity inputs would result in improved predictions.

However, flown trajectories do not only depend on aircraft dynamics but by their context as well. Trajectory context is usually associated to surrounding weather conditions, and as a consequence, great efforts have been conducted to improve the characterisation of the trajectory in terms of weather conditions [8], [9].

Contrary to the study of the meteorological impact, operational contextual factors have not usually been considered when characterising the trajectory. Traditionally, air traffic is only considered for the strategic and planning phases of ATM [10], often reducing the trajectory to a few significant points.

Air Traffic Control (ATC) holds the role of tactical separator as the last ATM system's resort for ensuring safe operations [11]. Thus, Air Traffic Controllers (ATCo) actions play a crucial role determining the final shape of flight trajectories, as they act upon trajectories through their decisions.

The exploitation of ATM data enables a greater understanding of the underlying causal factors driving ATC decisions. Knowledge about operational context can be achieved by modelling operational processes, and therefore, gaining understanding about underlying causal factors motivating final trajectory states.

A traditional definition of *knowledge discovery* refers to the “non-trivial extraction of implicit, previously unknown, and potentially useful information from data” [12]. The research herein described has undertaken a *knowledge discovery process* to characterise the trajectory *operational context* and its integration within the trajectory prediction process.

1.3. Research Scope

As stated by Schuster and Ochieng in [4], “*a lack of understanding of TP and Conflict Detection and Resolution driver uncertainties (especially contextual, and aircraft state, operational and performance uncertainties)*” was one the primary limitations of TP tools supporting functionalities envisaged in the initial SESAR Concept of Operations (ConOps) [13].

TP's performance is dependent on the input data. In an aircraft-centric trajectory predictor, the input data are the initial state, aircraft-specific parameters, some contextual factors such as meteorological conditions and operational constraints, and finally, the aircraft-intent information.

Aircraft do not fly aisled. The main objective of the ATM system is to enable safe operations. The ATC service is the tactical and last safety net of the ATM system, and in this regard, the ATCO does have an executive power upon aircraft trajectories.

The traditional aircraft-centric view contrasts with the resources that the ATM system deploys to prevent safety events. TP input data miss the operational contextualisation of the trajectory.

Thus, the main question that this thesis aims to answer is:

How can we characterise the impact of operational contextual factors upon a trajectory?

The answer to this question should be based on a series of questions:

1. What are the operational contextual factors impacting on an aircraft trajectory?
2. How do those operational factors impact on an aircraft trajectory?
3. How do the operational factors impact on the ATCo intention about a trajectory?

The ATM system addresses operations from strategic, planning and tactical perspectives. TP results are also characterised by the look-ahead time (LAT) to operations. In this sense, the questions enumerated before can be addressed from these different perspectives.

Traditionally, the impact of operational factors in the ATM system have been addressed from strategic and planning perspectives. Strategic and planning trajectory prediction account for air traffic interactions, i.e. an operational factor, to consider the demand the system expects at given airspace volumes during a certain period. Then, the system tries to adjust the airspace capacity to this demand strategically and in the planning phase.

This research will address these questions but from a tactical point of view. How is a vertical procedure impacting on the vertical profile of the trajectory? What is the impact of having an interaction between two aircraft in their trajectories? Can we systemise that knowledge to obtain recurrent patterns to introduce them within a TP?

1.4. Research Approach

The ATM system generates a huge amount of data daily. The exploitation of the ATM data for characterising ATM system components is at the core of the current research approach for addressing the research questions previously introduced.

Figure 1-1 Knowledge Discovery in Databases Process

The exploitation of large datasets is usually carried out through the *Knowledge Discovery in Databases* (KDD) process. Figure 1-1 presents an interpretation of the common framework for conducting KDD [14], which has been extended to include the system generating the data in the first step and the learning process to incorporate the extracted knowledge to the application domain.

The KDD process has been framed within a spiral approach for introducing the general research framework. The approach is illustrated in Figure 1-2. The research will conduct iterations aiming at answering the questions enumerated in the previous section, where the findings found on the previous step will feed forward the subsequent iteration.

The spiral approach comprehends four steps:

1. Construction of the data framework based on the ATM domain;
2. Initial data exploration to identify potential contextual factors;
3. Analysis of the impact of the identified factors on the aircraft dynamics;
4. Characterisation of the trajectories depending on the air traffic context.

Figure 1-2 Spiral Research Approach

1.5. Thesis Structure

The thesis is structured following the research approach previously introduced. The thesis is formed by seven chapters, enumerated as follows:

Chapter 1. **Introduction**

This chapter introduces the research and provides the research aim and approach.

Chapter 2. **Air Traffic Management and Trajectory Prediction**

The chapter establishes the ATM framework, the role of the trajectory prediction within ATM and its current challenges, framing it with respect to the challenges and the research questions.

Chapter 3. **ATM Data Organisation and Exploration**

The chapter describes the data organisation adopted within the research, based on the ATM domain. It provides an initial data exploration process to identify contextual factors.

Chapter 4. **Aircraft-centred Trajectory Prediction**

The chapter develops two approaches to aircraft-centred trajectory prediction,

integrating the findings of the previous chapter. These two approaches correspond to a trajectory predictor based on *point-mass models* and other corresponding to an artificial neural network (ANN) approach.

Chapter 5. **Air Traffic Contextual Factors**

The chapter addresses the modelling of flights' interdependencies and their integration within a trajectory predictor. It is built on top of the findings of the previous chapter.

Chapter 6. **Conclusions, Recommendations and Further Work**

This chapter presents the conclusions of the research, providing recommendations and recommended further work.

This Page is Intentionally Left Blank

CHAPTER 2. AIR TRAFFIC MANAGEMENT AND TRAJECTORY PREDICTION

This chapter aims at providing the research context and key definitions underpinning the document. In addition, the chapter objective is to identify key challenges for trajectory prediction, which will work as a basis for the analysis of the state-of-the-art with respect to those challenges and the research questions. Finally, the section provides the grounds for the development of a comprehensive data layer in Chapter CHAPTER 3.

The chapter is divided in four sections:

1. The first section defines *Air Traffic Management*, its objectives, functions and actors.
2. The second section defines the *trajectory* as the core of the future system, providing a definition enabling its characterisation. Finally, the section discusses which is the role of the trajectory in achieving ATM objectives.
3. The third section provides a framework for *trajectory prediction* and its role within the ATM functions.
4. The fourth section analyses the current trajectory prediction *challenges*.

2.1. Introduction to Air Traffic Management

Aeronautics is based on three pillars: air transportation; aircraft manufacturing; and finally, air traffic management and airports. Air transportation drives the business by moving passengers and cargo between airports. The air transportation industry is supplied with aircraft by manufacturers. The last pillar is composed by the airports and the *Air Traffic Management*, supporting the conduction of safe and efficient air transportation operations.

Air Traffic Management can be defined as “*the dynamic, integrated management of air traffic and airspace – safely, economically and efficiently – through the provision of facilities and seamless services in collaboration with all parties and involving airborne and ground-based functions*” [3].

This definition is complemented by the Air Traffic Management system (ATMS) definition, “*a system that provides ATM through the collaborative integration of humans, information, technology, facilities and services, supported by air and ground- and/or space-based communications, navigation and surveillance*” [15] .

The ATMS is a *safety-critical* system¹. Safety is paramount in the aviation community, due to the high societal impact of an accident, and therefore, in the air transportation business. In addition, the ATMS has to provide the necessary means to support the air traffic demand efficiently to guarantee the sustainability of air transportation.

2.1.1. ATM Functions

Following the previous definition of the ATMS as a safety-critical system, ATM main objective is to support operations of two or more aircraft sharing airspace and airport resources while being conducted safely and efficiently. ATM operations are framed in terms of the look-ahead time (LAT) to a potential safety event between two aircraft.

Two aircraft are said to be in *conflict* if their relative distance infringes horizontal and vertical separation minima criteria simultaneously. ATM is organised as a sequence of *safety nets* to mitigate the risk of the conflict happening. It has often been modelled through the *Swiss cheese model* [16] approach [17], [18], as represented in Figure 2-1.

¹ Safety-critical systems are those systems whose failure could result in loss of life, significant property damage, or damage to the environment [157] Safety-critical systems are those systems whose failure could result in loss of life, significant property damage, or damage to the environment

Figure 2-1 Swiss Cheese Model for characterising a safety event[19]

The core ATM functions cover these safety nets, representing different look-ahead times to the potential conflict, from strategic to tactical horizons. The strategic safety net has been traditionally associated to the airspace management function, which defines and manages airspace structures in the long-term. This strategic function establishes the air route structure and the long-term airspace organisation and defines strategic mechanisms to reduce the risk of an encounter between two aircraft, although the main focus is on efficiency.

The planning safety net is associated with *balancing* the demand and capacity of the system aiming at maintaining an orderly flow of aircraft (*demand*) while making an efficient use of the system resources (*capacity*). The capacity of the system is defined in terms of the number of aircraft that the tactical services supporting operations can manage at a specific time during a specific interval. The demand of the system reflects the number of flights expected to use the ATM system in that specific time and interval. The conjunction of these functions is known as *Air Traffic Flow and Capacity Management (ATFCM)*, and its objective is to ensure at pre-tactical level that a maximum number of aircraft will be supported by services operating under safe conditions.

The last safety net corresponds to the tactical layer of ATM. It is provided by means of the separation management function. Despite efforts to mitigate the risk of having two conflicting aircraft in the airspace by the previous layers, there may occur potential losses of separation.

These events are prevented by applying local modifications of the trajectories. The number of aircraft that can be safely handled during specific moments and intervals defines the capacity of the system. Finally, although the main objective of this tactical net is safety, it must pay attention to efficiency, in order to keep *orderly flows* of aircraft through the airspace, as some tactical actions upon flight trajectories might penalise the airspace user.

Figure 2-2 visually describes the objectives, functions and actors involved in the consecution of ATM. This figure was developed by Sáez Nieto and first published in [20]. It reflects the consideration of the previous paragraph, i.e. the rejection of a clear isolation between ATM layers, in this case, between *trajectory management* and *separation management*.

Figure 2-2 A visual description of ATM. Objectives, functions, organisations and look-ahead times

2.1.2. ATM Actors

The current three main ATM-related functions are *airspace management*, *air traffic flow and capacity management*, and finally, *separation management*. These functions are provided by different actors depending on the operational environment.

In typical commercial aircraft operations, the airspace management function is provided at local level by the *Air Navigation Service Provider* (ANSP), under preestablished navigation, communication and surveillance requirements for each specific operational environment, which are set by regulatory and supervisory agencies and international organisms such as ICAO.

The provision of the ATFCM function corresponds to diverse actors, typically the *Network Manager* (NM) and *Air Traffic Control Centres* (ACCs), representing capacity, and the *Flight Operator Centre*, representing demand. The ATFCM can also be split in different layers depending on the look-ahead time to the safety event, and therefore, ATFCM actions can also involve *aircraft crew* and airports.

Finally, separation management depends on the operational environment. The operational environment defines key aspects of the separation management function, such as *who* is the separator, or the minimum distances to consider that aircraft are flying safely. The operational environment considered within this thesis encompasses commercial air transport operations, and therefore, the role of *separator* is allocated to the *air traffic controller* (ATCo), through the *air traffic control* (ATC) service, whereas the pilots-in-command (PIC) execute ATC clearances.

These actors are represented in the bottom part of Figure 2-2.

2.1.3. ATM Challenges and Modernisation

The description of the ATMS objectives, functions and actors relates to a complex socio-technical multi-agent system, as introduced by Sáez Nieto in [21], formed by the conjunction of organisations, technical systems and humans [22]. The author characterises the ATM system as complex, which is a widely accepted conception (ICAO, 2005).

System complexity can be defined as the emergent behaviour that a system may show which is not explainable by a conventional analysis of the system's constituent parts [23]. As mentioned in the previous paragraph, ATM constituent elements form themselves complex systems. For example, airport operations involve management of the airport airside (ground movements, departures and arrivals), handling management and ramp operations, and finally, management of the landside (passenger operations) [24]. Any of these management functions is composed by several processes involving organisations, humans and systems, and all of them, are needed to perform at high-performance levels to contribute to the overall performance of ATM.

As it has said before, the overall performance of ATM is usually measured in terms of safety and cost-efficiency [25], [26]. In addition, other key performance areas are the impact on environment and the airspace capacity.

The provision of air traffic services shall be commensurate to the air traffic demand to contribute to the overall performance of the system. A steady increase of the ATM system demand shall be accompanied by an increase of ATM capacity meanwhile, at least, maintaining safety. The definition of the ATM capacity is a continuous subject of study, as it will be function of the capacities of the constituent components.

As it has been said, the ATM system is under pressure due to that the increment of demand is not being accompanied by the capacity. For example, the latest figure of ATFCM cumulative capacity-related delay of the ECAC area for the period January-June 2019

showed a 31% increment on with respect to January-June 2018 for an increment of 1.8% of the demand.

Figure 2-3 Delay figures attributable to lack of ATC capacity in the ECAC area. Own figure made with data extracted from [27].

These figures suggest that the ATM capacity is near its crunch value in some European areas. This challenge and the cost of the air navigation services partially drove the initiation of the SESAR and NextGen programmes in order to procure the modernisation that ATM required to handle an increasing demand.

ATM modernisation aims at shifting from an airspace-based concept of operations to a trajectory-based one by placing the trajectory at the core of information flow and introducing higher automation levels to increase the capacity of the ATM system, increase the cost-efficiency level meanwhile maintaining safety and user flexibility [3].

The provision of ATM was previously defined as the result of all the interactions of a complex socio-technical multi-agent system. TBO reflects the notion that the trajectory is the glue between the different ATM components during strategic, planning and tactical flight operations by pursuing a common view between different actors and ensuring consistency between the trajectories and the constraints that originate from ATM components [28].

2.2. The Trajectory

2.2.1. Trajectory-Based Operations

The ICAO ATM Requirements and Performance Panel (ATMRPP) introduced TBO as the process which “defines how and to what extent to include airspace users into the optimisation process, balancing flexibility for prioritization and optimising airspace users’ operations while simultaneously ensuring sufficient predictability needed to optimise the ATM network as a whole and ensure that the flight safety is never compromised” [29].

A latter definition of TBO is provided by SESAR [30]. Trajectory Based Operations “providing high predictability and accuracy of the trajectory allow a seamless process from planning to execution and a seamless process from gate-to-gate based on information sharing, with airborne and ground actors sharing consistent information throughout the Business / Mission Trajectory lifecycle”, see Figure 2-4.

These two definitions envision a user-centric ATM. A user-centric ATM means that the airspace user will fly as close as possible to its business needs. It is remarkable how flexibility is missed in the SESAR definition, while empowering *predictability*.

SESAR’s definition introduces the notion of *seamless process*. Characterising a process as *seamless* means that it shall be provided with a uniform or continuous level of quality, while combined in an inconspicuous way. A seamless *Business Trajectory lifecycle* refers then to the provision of a pro-active **trajectory management**, agreeing decisions about the trajectory in advance to limit constraints on the trajectory such as ground or airborne holdings [31].

The need of trajectory management comes from the possibility of two or more aircraft trying to use the same resource (airspace) at the same time. Even though these trajectories were users’ *preferred* trajectories, the system would need to introduce modifications to fulfil separation requirements along the whole flights.

Figure 2-4 Gate-to-gate Business Trajectory process [30]

P. Sáez Nieto sums TBO up as the need of *orchestration* of ATM systems and agents, by increasing the cohesion of the processes ensuring the provision of separation while maintaining flight and ATM efficiency [21], following the vision of G. Boy in [32].

2.2.2. Definitions and Terminology

The previous section describes a top-down approach for the provision of ATM through TBO. The nucleus element of TBO is, of course, the trajectory, the most granular element of the processes which contribute to ATM.

There are various definitions of trajectory. For example, ICAO [3] defines the trajectory as “a description of the movement of an aircraft, both in the air and on the ground, including position, time and, at least via calculation, speed and acceleration”.

EUROCONTROL has provided different definitions. For example, in conjunction with the FAA, it defines the trajectory as “the path a moving aircraft follows through the airspace and can be mathematically described by a time-order set of trajectory (state) vectors or the geometry of the flight path” [33]. This definition was simplified to “a representation of the path of an aircraft, describing the horizontal and vertical profile over time” in [34].

As J. Bronsvort points out in [35], these definitions may led to a confusion due to the use of the term *path* for defining the trajectory. While ICAO refers to the description of the

aircraft movement, EUROCONTROL refers to the *path* an aircraft follows. Path is defined as a “route between one place and another, or the direction in which something is moving”², and therefore, refers to the spatial description of an object movement. Two aircraft with different speeds but following exactly the same path will result in two different trajectories.

Given these considerations, the trajectory definition will follow the ICAO convention, also specifying that it will refer to the centre of gravity of the aircraft. Therefore, an aircraft *trajectory* is defined as “a description of the movement of the centre of gravity of an aircraft, both in the air and on the ground, including position, time and, at least via calculation, speed and acceleration”

The *path* will be defined as the spatial definition of the trajectory in 3D, whereas *track* is the projection of that same trajectory onto the Earth’s surface. The path and track definitions follow the definitions given by Bronsvoort [35].

2.2.3. Trajectory Planning

The functional purpose of the flight is to transport people and cargo safely between airports, meanwhile the operator obtains some benefits from the operation [36]. The Business Trajectory cycle generates a flight intention that covers these business objectives, while respecting constraints that ATM might impose.

Figure 2-5 Notional representation of a conventional (blue) versus an optimal trajectory (black).

² <https://dictionary.cambridge.org/dictionary/english/path>

The cost of a flight while the aircraft is in the air is driven by the fuel consumption and other time-related factors. From the point of view of fuel-related costs, the optimal trajectory would consist of a continuous climb from the origin airport followed by a continuous descent to the destination airport ([37], [38]). However, maintaining vertical separation between aircraft motivates the conventional levelled cruise segments. A notional picture between the optimal and conventional profiles is shown in Figure 2-5.

Time-related costs are dependent on balancing aircraft utilisation and other direct time-related costs such as crew wages. The temporal cost can be balanced by applying a different set of strategies, such as varying airspeeds or different climb or descent profiles [39]. Airlines balance fuel- and time-related costs by the cost index, defined as the ratio of fuel cost to all other costs³.

So, airlines conjugate the mission and the operating costs in the flight intent. The **flight intent** is defined in [33] as “the *constraints* and *preferences* that are applicable to a flight”. These constraints are given by aircraft characteristics, airport resources, airspace resources⁴ and safety requirements. There are multiple trajectories that might comply with these preferences and constraints.

Nowadays, the Airline Operations Centre (AOC), considering the business objectives and the mission, iterates the initial flight intent to obtain an optimised *flight plan* which includes the route, fuel and other factors such as alternate airports.

A *route* is defined in [34] as “the 2D trajectory of an aircraft, expressed as significant points, ATS routes or geographical points”, where significant points are defined as specified geographical locations, or waypoint, for navigation and ATS purposes. This definition is further extended by the *segment* definition as “a part of a route defined by initial and terminating points”.

The AOC will optimise altitudes and speeds for the different segments, estimated fuel consumption and additional fuel for safety requirements, subject to weather and operational constraints. For example, wind can have a significant impact on the flight costs [40]. Thus, [41] shows an example where an optimised route depending on the wind profile was a 11% longer, but 2% faster and 3% more economical in fuel terms.

³ https://www.skybrary.aero/index.php/Cost_Index

⁴ See Conditional Route (CDR) definition in [158] See Conditional Route (CDR) definition in

The result of this process is the final flight intent, which is distributed to the different actors of ATM, restricting the information which is shared among the ATM actors. For example, the AOC sends to the pilot information such as the flight plan, the specific cost index for the flight, meteorological forecast, fuel considerations, *inter alia*. However, the information which reaches the ATM system (ATFCM and ATC) is a pruned version, only including the flight plan, cruise altitude (s) and associated speed(s) (see ICAO Doc 4444 [11]).

In the future TBO paradigm this process will correspond to the Business Trajectory Cycle. The initial flight intent corresponds to the *Shared Business Trajectory* (SBT). The SBT is then continuously refined by sharing information with the other actors of ATM, in order to have a common view of the flight intention and to manage the ATM resources accordingly. The final version of the SBT becomes the *Reference Business Trajectory* (RBT), which is the **flight intent** that the operator agrees to fly and the ATM system to facilitate.

2.2.4. Flight Management

In this thesis *flight management* refers to the process by which the operation is completed according to the business objectives, respecting safety and ATM constraints.

A flight is completed by *navigating* from origin to destination. In this thesis, navigation is defined as per [42], by establishing that a *safe navigation* implies piloting an aircraft from one place to another place without getting lost, maintaining the aircraft within its flight envelope and without jeopardising the safety of cargo and passengers. Navigation is then composed by three different processes, the definition of the reference trajectory, positioning (in space and time) with respect to that reference trajectory, and finally, guidance to achieve it.

The previous section has described the process by which the flight intent is generated. However, the flight intent is an ambiguous definition, as many trajectories may fulfil the business objectives and constraints for the flight [43]. In current aircraft, the definition of the *reference trajectory* is conducted by the *Flight Management System* (FMS).

The FMS computes the 4D aircraft reference trajectory from the flight plan based on the atmosphere model and the aircraft performance data. Then, the *positioning*⁵ function, also within the FMS, provides the state of the aircraft to other functions. These functions are

⁵ The standard terminology for determining the best estimate of the current state of the aircraft is *navigation*. In this thesis, this function is defined as *positioning*.

guidance (lateral, vertical and temporal) and performance advisory, relative to the reference trajectory and the aircraft state [44].

Figure 2-6 Flight management, positioning and guidance loops. The figure includes the stability loop and the control loop. The figure is based on [38] and [39].

Figure 2-6 presents the management loop, which provides the guidance targets (guidance function). The guidance targets are provided to the *Automatic Flight Control System* (AFCS), which is able to provide control as to direct an object from its current state to the designated target state by applying high-order guidance modes [35]. Then, it is followed by the *control* loop. The control loop carries out the necessary actions to achieve a desired attitude and motion for the aircraft for navigation, maintaining the aircraft within its flight envelope and turn coordination, and it is provided by the *Flight Control System* (FCS). In this loop, the pilot can have direct control on the targets for the Control Augmentation System (CAS) through the Mode Control Panel (MCP). The stability loop is the final loop, where the aircraft automatically controls the motion to limit certain congenic instabilities such as the Dutch roll [46].

This process can be tactically amended by the ATC in order to prevent conflicts between aircraft or for efficiency purposes. ATC will issue *clearances* by which the intended trajectory is modified. This can result in *close* or *open* clearances depending on if the reference trajectory is fully defined as a result of such clearance or if an additional clearance is needed to close it. This loop is represented in Figure 2-7, where the ATC may provide clearances within the flight management loop (i.e. *fly direct* to a waypoint) or open clearances within the control loop (i.e. *follow a vector*).

To sum up, an aircraft trajectory is the result of several processes performed by actors involved by which a set of business objectives are transformed in the efficient transit of an aircraft between two points. These processes and the flight itself are profoundly impacted by the aircraft operator business objectives, weather and operational factors and the aircraft performance model.

Figure 2-7 Flight management, guidance and control loops, including the surveillance (ATC) control loop [35]

2.3. Trajectory Prediction

Trajectory management aims at reducing ATM reactive behaviour as a set of flights progress from planning to execution. This process is partially enabled by foreseeing what are the expected behaviours of the set of flights within a given look-ahead time, depending on the extent to which hazards along the future trajectory of an aircraft are considered for ATM provision⁶.

Foreseeing the expected state of an aircraft within a conflict horizon is trajectory prediction. EUROCONTROL and the FAA defined TP in a common white paper [47] as:

“The process that estimates a future trajectory of an aircraft through computation”.

This definition can be complemented by Bronsvooort’s definition of the trajectory prediction [35]:

“A process in which the future trajectory of an aircraft is computed based on a model of the aircraft’s performance, forecast meteorological conditions and a script indicating the intent of the flight”.

García Chico [48] added some additional elements for defining trajectory prediction as:

⁶ This paragraph is influenced by the definition of *Conflict Horizon*, provided in [3], as “Conflict horizon is the extent to which hazards along the future trajectory of an aircraft are considered for separation provision”.

“The process of estimating the future states of individual aircraft on the basis of the current aircraft state, estimation of pilot and controller intent, expected environmental conditions, and computer models of aircraft performance and procedures”.

Finally, [34] limits its scope to the *“calculation of the trajectory of a flight from take-off or entry to the Area of Interest of an Air Traffic Service Unit until landing or exit from the AoI, for the purpose of the support of ATM services and tools”.*

The first definition provided introduces a very generic description of a TP. Bronsvoot’s definition focuses on the trajectory prediction from the aircraft point of view, whereas Garcia Chico’s definition includes the intention of the ATCo as part of the process.

The four definitions specify TP as a process to determine the trajectory of an aircraft. The three first definitions involve the definition of the *future* trajectory of an aircraft, whereas the last one is not only defined in temporal terms, but also in spatial terms (“entry to the Area of Interest”).

Finally, the selected four references define TP as a process aimed at supporting ATM services and tools.

2.3.1. Trajectory Prediction Common Structure

The FAA / EUROCONTROL common paper presented a common structure for the TP process, divided in 4 different sub processes, i.e. *preparation*, *computation*, *update*, and finally, *export*. The preparation process “builds an initial condition” and a *behaviour model*. This behaviour model is defined as the list of manoeuvres that the aircraft plans to execute, i.e. it describes **unambiguously** how the aircraft will be operated. This terminology (“behaviour model”) evolved to ***aircraft intent*** [49], which is widely accepted by the academic community and the industry [50].

The *computation* process estimates the trajectory based on the aircraft intent. Finally, the update and export processes are related to the conformance of the computed trajectory with the trajectory constraints and the provision of the output to the TP clients respectively.

These four processes have worked as a useful framework for defining different TP architectures. Figure 2-8 and Figure 2-9 present two architectures for trajectory predictors. The first one was presented by Bronsvoot and the second one was presented by EUROCONTROL.

Figure 2-8 Trajectory Prediction Architecture (Bronsvoort) [35], based on [51].

Figure 2-8 describes the process in terms of *preparation* (Intent Generation Infrastructure) and *computation*. The intent engine depends on four different types of *adaptation data*⁷. These include:

- User Preferences Model (UPM), which indicates the airline preferences with respect to the cost index and the guidance modes,
- Environmental Model (EM), corresponding to weather predictions with respect to atmosphere and wind,
- Aircraft Performance Model (APM), which includes aircraft performance parameters and piloting models,
- Operational Context Model (OCM), which includes the definition of airspace elements such as waypoints, airways, SID, STARS and other type of airspace constraints.

It is worth noting that ICAO observed that the aircraft intent could be represented either as “the future aircraft trajectory as programmed inside the avionics” (*the reference trajectory*) or “the aircraft control parameters as managed by the automatic flight control system” [52]. Figure 2-8 refers to the second interpretation of ICAO’s aircraft intent, which may include any individual command to the FCS or the guidance modes⁸ introduced in the AFCS.

The aircraft intent is the base for the *Trajectory Computation Engine* (TCI). The TCI is composed by a *trajectory engine* which, based on a given aircraft performance model and an

⁷ Adaptation Data: Adaptation data is a set of data used to configure a given TP to a particular operational environment.

⁸ A complete explanation of the different types of guidance modes can be found on the PhD Dissertation of Javier López Leonés [65] and in the Annex C of [69].

environmental forecast, provides a **unique** description of the trajectory in its spatial and temporal components.

Figure 2-9 EUROCONTROL TP Specification - Logical TP Breakdown

The Figure 2-9 shows a similar TP architecture to Figure 2-8. The main differences between both are two. The first difference lays on the nature of both approaches. The first one is the TP from an *aircraft*-centric point of view, whereas the second one addresses it from a ground perspective. The Intent Engine in the first one includes the aircraft performance model, the contextual model and the environmental model. In the second one there is a more detailed description of the impact of ATC on the model behaviour, and then, it expresses the lack of knowledge about the User Preference Model by introducing the note of “*Conjecture Unknown Intent*”.

The second difference is that it includes an intermediate step between the *Flight Intent* and the *Aircraft Intent*, which is the generation of the Flight Script. The *flight script* is defined as the description of the potential behaviour from the ATC point of view.

As this architecture is defined from a *ground* point of view, it clearly identifies the potential sources of ATC ambiguity and their impact on the flight script. For example, it includes as relevant information that may impact on the flight script the *ATC Procedures*, *Tactical Planning and Clearances* and *Filed Flight Plans and Amendments*.

2.3.2. Aircraft Intent Generation

On the one hand, the description of the intent has been extensively described in the work from López Leonés ([49], [53]). The key problem about the intent generation is to infer the future intent of the aircraft. There are four sources of variation of the aircraft intent, which are the ground automation, controller, pilot and aircraft automation [54].

Depending on the look-ahead time for the trajectory predictor, one of these agents will be the leading source of variation in the aircraft intent. For example, Figure 2-10 shows a notional contribution of the different actors within ATM to trajectory management depending on the time-horizon for a trajectory. If a trajectory predictor were to predict the trajectory for a conflict-horizon of between 2 and 16 minutes, the leading actor impacting on the trajectory would be the ATC, with a high contribution from the aircraft side.

Figure 2-10 Notional operational stakeholder involvement in trajectory management [28], [29]

2.3.3. Aircraft Motion and Performance Models - Classification of Trajectory Engines

The aircraft intent is transformed into the predicted trajectory by means of the trajectory engine. The typical mathematical models behind the trajectory engines are point-mass models, kinematic models, and kinetic models [55]. The models depend on the level of fidelity required for the trajectory prediction client.

Assuming a rigid-body model, kinetic models provide full description of the motion of the aircraft, including translation of the centre of gravity and the three rotations. These models are also known as 6 degrees-of-freedom models.

Then, if only translations are considered and the aircraft mass is confined to the centre of gravity (CG), models are known as *point-mass* models. In these models all forces (lift, drag,

thrust) are applied in the CG, resulting in zero net-moment [56]. Using the same assumption, the Total-Energy Model equates the rate of work done by forces acting on the aircraft to the rate of increase in potential and kinetic energy [57]. These models depend on aircraft coefficients such as mass, lift coefficient or thrust coefficient to model the different forces being applied in the centre of gravity. The European Based of Aircraft Data (BADA) is the most known and used source of the required data to propagate these models [58]. This database also includes performance and operating procedures for a range of aircraft, specifying nominal cruise, climb and descent speeds among other key performance parameters.

Finally, the kinematic approach provides an aircraft motion model which is not derived from flight dynamics (forces and moments) but from other sources. For example, aircraft motion within a given airspace can be derived from pure data-driven models ([7],[59], [60]) or from performance tables [58].

A classification of different aircraft motion models is shown in Figure 2-11

Figure 2-11 Aircraft Motion Models for different TP approaches.

2.4. Framing the research

2.4.1. Trajectory Prediction Challenges

Trajectory prediction is and will be at the core of the different ATM components. However, it remains a trending area in ATM research, which can be summarised in four different challenges.

2.4.1.1. Trajectory Prediction in the current and future ATM Context

FAA / EUROCONTROL acknowledged in their common white paper that decision support tools supporting ATM services will have different performance requirements, related to accuracy, uncertainty, response times and input data requirements. [61] introduced a framework for defining the performance requirements associated to TP in the basis of a performance-based service provision⁹, which was extended in [4], and, more recently, by Gómez Comendador et al. [62], to determine the performance requirements for TP associated to the future services of a European ATM.

This applies not only for the future system but for the current system as well. The performance requirements for a TP supporting tactical operations of the ATC are completely different from those supporting planning or strategic tools. The challenge is to clearly identify which are the most relevant factors in relation to the conflict horizon of the service the TP is supporting. For example, wind intensity has a strong impact on tactical trajectory prediction, whereas the impact on planning services decisions is more limited.

In addition, the ATM environment is evolving towards the TBO paradigm, as explained before. In the transition, the intended network-vision is not fully understood by the local actors. Thus, local trajectory amendments are applied without considering the downstream impact.

As the ATM community moves forward the development of future services such as Flight-centric ATC or dynamic capacity management (DAC), those increasingly rely on greater sharing of the intentions of all actors within the ATM ecosystems [63], [64].

For this aim, it is key to guarantee interoperability within the different actors interacting within the ATM ecosystem. For example, the *Aircraft Intention Description Language* (AIDL) [65] is a language that facilitates this interoperability and that enables an unambiguous definition of the aircraft intention.

⁹ Annex K of ICAO Doc 9854 [3]

As a summary, the changing ATM environment results in a diverse and evolving set of performance requirements for TPs supporting the ATM services. It is key to understand which are the requirements from the different ATM services and which factors are more relevant depending on the time horizon of the service. Finally, the TP design shall be aligned with the new concept of operations.

2.4.1.2. The stochastic nature

Trajectory management shall be conducted considering the stochastic nature of the factors impacting on the trajectory. As F.J. Sáez defends in [42] and Henk et al. in [28], factors impacting on aircraft trajectories may be deterministic or non-deterministic. The key objective of trajectory management is to guarantee a “converging trajectory management system”, by increasing the predictability¹⁰ of the provision of ATM.

The predictability of the factors varies depending on the time scale of the TP client. For example, the wind direction at an airport is highly predictable when assessed for strategic purposes and it is not required a high granularity, whereas in tactical operations it can show a random behaviour, depending on the weather conditions of that specific day. A change of airport configuration in a complex area can have a strong impact on how the system accommodates the demand in the surrounding airspace, as it usually implies a change of its configuration as well.

If we assume current operations to have a low exchange of information between the different actors, the lack of knowledge about the intentions of the agents with regard to their function¹¹ is a source of uncertainty on the factors impacting of the trajectory. For example, aircraft parameters which are key for kinetic TP such as the thrust configuration or the navigation mode are not known for the ground trajectory predictors, and right now they can be treated as a stochastic factor [66].

So, the transition towards a more integrated trajectory management implies to have a greater knowledge about the sensitivity of it to variations in the factors, in all the flight phases, to establish mechanisms which could allow the facilitation of a seamless ATM and a proper level of orchestration between agents and trajectories.

¹⁰ Predictability: A consistent repletion of a state, course of action, behaviour, or the like, making it possible to know in advance what to expect. [28]

¹¹ ATC is mainly managing the traffic for separation purposes, whereas the pilot is *aviating, navigating and communicating* taking into account the performance objective of the airline.

This challenge is being addressed by research projects such as the COPTRA ([67], [68]), which considered the factors impacting on tactical TP as probabilistic distributions, to understand what is their impact downstream the TP process. In a related research work, Casado's work [69], [70] proposed a modelling technique to characterise the uncertainty sources affecting point-mass models based on Polynomial Chaos Expansion.

2.4.1.3. The isolated view

The analysis of uncertainty factors has traditionally focused on identifying the performance parameters which best reproduce the aircraft trajectory given an aircraft intent [71]. As it has been mentioned previously, there are factors which are exclusively within the domain of action of the airspace users (PIC or AOC). This was identified as a main area to improve TP by sharing this information.

Although it has been demonstrated that this improves the trajectory prediction in fixed waypoints along the trajectory, the prediction accuracy over other key, but movable, waypoints is not so clearly improved (see [28]).

These variable characteristic points of the trajectory are usually identified as the Top of Ascent (TOA) and the Top of Descent (TOD). Currently, these are determined by the FMS depending on the aircraft performance and the meteorological conditions that the FMS has access to. Traditionally, the research has been focused on determining the aircraft-related parameters which can impact on these variable points ([72], [73]).

However, as recent research demonstrates ([35], [74]), the impact of the separation and traffic flow functions of the ATC has a leading impact on the location of these points. A challenge for the trajectory prediction is to identify not only the most likely aircraft trajectory given an intention, but to reduce the ambiguity on the definition of the aircraft intent, which is only possible through a greater understanding of the intention of the air traffic controllers and the interdependency between the air traffic.

2.4.1.4. A complex system

So, the previous points have argued in favour of gaining understanding about the sensitivity of the trajectory predictor to the diverse factors that impact on it, as the trajectory prediction is impacted by the trajectory management in all its levels.

Although the trajectory behaviour is not deterministic due to the impact of uncertain factors such as weather, the trajectory is equally impacted by the action of ATM upon it. ATM has been previously defined as a *complex* sociotechnical multi-agent system, and it is complex due to the large number of processes which are intertwined among them and how the

variability of their outcomes impacts each other, which may be modelled following methodologies such as FRAM [75].

Some of these individual processes themselves can even be identified as complex processes, due to the magnitude of actors and the spatiotemporal nature of the decisions. For example, processes in the turnaround at the airport involve ground ATC, ground handling, passenger handling and others. Although these processes may be described individually in relation to the trajectory, small variations on their outcomes may result in deviations in the tactical operations regarding planned operations, leading to the air traffic management complexity.

The challenge for TP in this regard is to integrate the complexity at system level on the different layers of TP. At tactical level, ATC complexity is impacted by trajectory uncertainty and vice versa (as argued in the previous point) [62]. At tactical planning level, the system capacity and demand is affected by the granular capacity levels of the airspace, driven mainly by ATC complexity (*hotspots*) and by the airport processes [76]. Finally, the strategic organisation of ATM is impacted by the organisation of all the processes defining ATM and their defined capacities. If solutions are deployed at system level without considering the complexities of the compounding parts, these might not provide expected outcomes.

2.4.2. Leveraging data in ATM

2.4.2.1. Point-mass models performance parameters

Nowadays, there are well settled models for trajectory computation models for its use by ground ATM systems. The most important one, at least from an academic perspective, is the combination of the **Total Energy Model** or point-mass models with database of aircraft performance reference data, which has been already presented in Section 2.3.3.

The performance data can be either provided by manufacturers or obtained through the analysis of data. So far, the provision of these models was limited to large air navigation service provider such as EUROCONTROL, but the raise in air tracks availability through services such as FlightRadar24 or open source data platforms such as Open Sky¹² has allowed the academic community to provide these data, being the work conducted by J. Sun a cornerstone in the provision of a more “transparent, comparable and reproducible” research in ATM [77].

¹² <https://opensky-network.org/>

However, both the original work of Nuic et al. [57], [58], [78] and J. Sun work with a general sample of flights. J. Sun reflects in his PhD dissertation that the performance parameters which are obtained “may not reflect certain airspaces, countries, or airports accurately due to factors such as controller procedures or airport limitations”¹³.

Alligier et. al ([5], [6], [79]) produced a series of paper where that issue was addressed. Thus, Alligier started to estimate the mass parameter and generating thrust laws, isolating flights with common operational factors, such as the airports or the aircraft. The models that Alligier produced in his first paper were the starting point of this research work. The work of Alligier established the basis for tailored parameter estimation, focusing on the climb phase. Alligier suggested in his 2013 work that they “might try to identify classes of trajectories from the historical data, and then learn several thrust laws (one per class) instead only one” for their future work.

2.4.2.2. Pure data-driven

In addition to model-based trajectory computation engines, a new source of trajectory engines was published during the past years. Examples of this kind of engines is the work of Tastambekov et. al in [7] or the work published by Wang et al in [59]. The first one used a functional linear regression using wavelet decompositions in Sobolev space, whereas the second one used an approach mixing an automatic flow extraction and artificial neural networks.

Tastambekov mentioned in the discussion of their results that “most prediction errors can be explained by ATC controller actions that make the descending aircraft keep constant altitude for avoiding a conflicting aircraft”. That suggested that the no inclusion of that feature on their model produced errors impossible to predict, and even, the author suggested that “by essence, it is not possible to forecast such actions that are depending on the surrounding traffic”.

In his work, Wang took advantage of the automatic extraction of flows that was first proposed by Gariel [80], and then focusing on predicting the time of arrival by using neural networks. The work was further generalised in [81] but focusing on the flow extraction algorithms rather than the features that led to the predictor to accurately obtain the time of arrival.

2.4.2.3. Trajectory flows

¹³ Section 5.7.1 of [77]

The flow extraction method previously mentioned was among the first machine learning applications in air traffic management that made use of intensive air traffic data. The work conducted by Gariel et al. in [80] has been the basis of various research lines. For example, it was a key component of Conde Rocha's research [82], [83]. In addition, it was also exploited by other researchers such as [84]–[86] in a planning phase. Recently, the previously discussed work of Wang started to use it for short-term predictions.

Gariel's method will be discussed in detail in the next chapter. The work is considered as relevant for the trajectory prediction question after analysing the research of [87], where the author identified “flow-based” abstractions as a main mechanism that air traffic controllers were constantly using in order to reduce the cognitive complexity they faced when conducting their activities.

2.4.3. The Research Questions

1) Systematic ATM Data organisation enabling data-driven ATM research

Schuster postulated in [4] that one of the primary limitations of TP for being an enabler of future European ATM ConOps “was a lack of understanding of TP driver uncertainties (especially contextual, aircraft state, operational and performance uncertainties)”.

The exploitation of ATM data is key to provide figures to the previous statement. In this sense, ICAO recognised that a systematic organisation of the data [88] enables the characterisation of the performance of the system. Garcia Chico also postulated the use of structured data as a source to characterise the trajectory uncertainty [48]. To some extent, most of the time a researcher is working with ATM data is to prepare, clean and transform it¹⁴.

Thus, the first research question is:

Q1. How can ATM data be organised to enable a systematic and data-oriented analysis of the system, including Trajectory Prediction?

2) Contextual factors impacting on the descent phase

As a summary of this chapter, trajectory prediction in ATM is highly dependent on static and dynamic factors, which affect the aircraft intent generation process and the trajectory computation engines. On the one hand, previous studies established that diverse features,

¹⁴ <https://hbr.org/2018/08/what-data-scientists-really-do-according-to-35-data-scientists> or <https://www.ibm.com/cloud/blog/ibm-data-catalog-data-scientists-productivity>

such as the aircraft model or the take-off weight, have an important influence on the dispersion of trajectories. On the other hand, dynamic factors would be those related with contextual factors, such as weather conditions or operational elements. Meteorological conditions have been extensively addressed by the academic community [89]–[95], but that is not the case for ATC-related factors.

In addition, the research that was one the main motivations to start this work focused on the climb phase. In this sense, previous research also addressed the prediction of the descent phase, such as [35] and [95], but it did not address it from a data exploitation perspective.

Considering the previous remarks, the second research question is:

Q2. What are the operational contextual factors impacting on the descent profile of an aircraft trajectory?

3) Surrounding traffic factor and its impact on the aircraft intention

Air traffic as a contextual factor is usually only considered for trajectory management, with longer prediction horizons to operations. For example, ATFCM measures are based on traffic indicators with different granularities and look-ahead times, such as the sector hourly entry count or the sector occupancy [97].

Figure 2-10 identified the air traffic control as the leading actor contributing to modifications on the aircraft trajectory in the short-term. Usually, the actions of the ATC service on a trajectory are either due to trajectory management or due to interdependencies between aircraft trajectories.

As previously mentioned, it is challenging to establish causal effects driving ATC actions modifying the aircraft intent. However, it is a main source of lack of predictability of the aircraft intent in the short-term. In this sense, the last question that is posed for this thesis is:

Q3. Can the surrounding traffic be included as a factor in the definition of the aircraft intent?

CHAPTER 3. ATM DATA ORGANISATION AND FLOW DISCOVERY

3.1. Introduction

Information management is at the core of the future ATM system. As such, TBO relies on an enhanced flow of information among agents, identifying beforehand constraints impacting on airspace users' preferences.

Research in ATM is dependent on a systematic approach to how data is managed and stored. In this sense, the conjunction of air traffic services together with the CNS enablers underpin the provision of ATM. The provision of these services implies several digital fingerprints, often unstructured and from different sources.

For example, systems supporting the tactical phase of ATM are usually divided in two main groups: a Radar Data Processing System (RDPS), which processes aircraft tracks in the airspace of interest as provided by the surveillance services; and a Flight Data Processing System (FDPS), which handles the flight intentions corresponding to those aircraft, correlating them with the surveillance information. The interface of the ATCOs with these systems is the Controller Working Position (CWP).

Those systems generate data continuously such as ATCO's clearances, surveillance data or the evolution of the flight plan in the ATM system. This chapter objective is to describe the underlying rationale to evolve from raw signals to structured data. In the overall context of this thesis, this chapter covers the four initial steps of the KDD process (see Figure 3-1).

The first subsection details the business-driven structure for the data model, inspired by the data warehouse paradigm. Then, the flow discovery process is discussed. These two sections together support the last section where the datasets used in the rest of the thesis are introduced.

Figure 3-1 Knowledge Discovery in Databases Process. The blue rectangle describes the scope of this chapter.

3.2. ATM Dimensional Modelling

Data in ATM are unstructured in nature due to various reasons such as widespread of legacy systems supporting operations or the heterogeneity of the sources (airports, airport users, ANSPs, etc).

There are diverse paradigms to organise data for their exploitation to support the decision-making processes. Among them, the data warehouse paradigm is widely adopted across industries with that purpose, including ATM. For example, ICAO Manual on Global Performance on the Air Navigation System recommends the adoption of “industry-standard data warehousing concepts” for managing the storage of ATM performance data [88].

The data warehouse (DWH) paradigm [98] enables a systematic approach towards data organisation and exploitation in order to support management decisions. A DWH is a “subject-oriented, integrated, non-volatile and time-variant collection of data in support of management’s decisions” [99].

For this research, the subject-oriented attribute targets an effective airspace research activity derived from the application of the tactical layer of the ATM, the ATC. The second element is the “integrated” attribute, where, by means of Extract, Transform and Loading (ETL) processes, the data acquire a single representation independent from the analysis layer. The non-volatile attribute implies that only two processes are conducted within DWH, loading of data, and access to data, but not updating. Finally, the time-variant attribute defines the

historical perspective of the DWH and key structures always contain a time-dimension element, such as a timestamp on the records.

Following this paradigm, data are organised by means of facts and dimension tables which are representative of the business processes. Dimension tables relate to descriptors of the data, while fact tables identify business measures. These measures are often additive, to transform raw data into information by means of analysis of the aggregated data.

It has been previously mentioned that the scope of this thesis is the tactical side of ATM. However, the activity of ATC is always result of the performance of the previous ATM processes which describe univocally the context of ATC operations and the trajectories.

Figure 3-2 is a high-level representation of the main processes governing the demand of the ATM system and the capacity depending on the time horizon. The left-hand side of the figure represents the demand, which moves from a fuzzily defined demand towards the flight intents. On the right-hand side, the capacity side evolves from the long-term airspace organisation towards the daily organisation of the airspace, including resources availability or the opening schemes of airspace.

The vertical axis represents two notions, one is the temporal evolution from planning to operations, whereas the other represents the associated cost of a decision in the capacity side. The horizontal side represents the “degree of interaction” between both sides. One of the key challenges in ATM nowadays is that both sides are “blind” with respect the other.

As time moves, the planned picture becomes reality, and the system starts producing *facts* about the operational deployment. These facts can only be explained by means of their dimensional contexts, and the dimensional context is explained through the previous processes. These facts will be later the base for decisions that shape the system.

In this sense, data for trajectory predictor research are usually only characterised by their spatiotemporal dimensions (relating to the state of the aircraft) and aircraft-related dimensions (typically aircraft model and associated dimensions). How do we organise the data to characterise operationally the trajectories, in order to create the “full-picture” needed to feed the *intent* and *trajectory computation* engines?

Figure 3-2 ATM Processes

3.2.1. The Demand Dimension

From a TBO perspective, the nuclear fact¹⁵ of the ATM system operation is the trajectory and its evolution since its initial inception, as it is the element to be managed. As the aircraft intent (FPL) evolves from strategic planning to the actual flight, each new update of the aircraft intention is an *intent fact* (FPL fact), considering the final flown trajectory as a fact as well.

This statement has several implications with regard to the **demand**, as it implies that *each flight* is in substance managed separately. It represents the **core dimension** of the demand, as whenever a new flight is registered in the ATM system, there will be an operational activity with regard to it. This dimension (**flight**) is further characterised with static dimensions such as the **airports** of origin, destination and alternate, **the aircraft registration and, therefore, the aircraft model**, the **airline** and finally, **temporal** dimensions.

Figure 3-3 Flight dimension

¹⁵ Atomic fact or "grain" <https://www.kimballgroup.com/data-warehouse-business-intelligence-resources/kimball-techniques/dimensional-modeling-techniques/grain/>

These dimensions can be further organised by applying **star** or **snowflake** schemas [100], for example, the airport can be localised within a {city, country, region}, as shown in Figure 3-4. Thus, these dimensions can be organised in hierarchies for easing aggregations for the analyses [101].

Figure 3-4 Airport dimension

From a capacity point of view, ATM can be divided in an **organisational layer** (ASM) and **operational layers** (ATFCM and ATC). From a data organisation perspective, ASM defines the **operational dimensions** of the capacity (route structure and airspace volumes). On the other hand, the operational layers produce the **operational facts** with regard to the flight (changes in the aircraft intent such as regulations or tactical modifications of the trajectory) and the capacity itself (changes in the airspace configuration or momentary changes of the capacity). Depending on the needs of these processes, the demand may be summed up **spatially** and **temporally** in the decision-making processes.

3.2.2. Temporal and Spatial Dimensions

The lowest hierarchy level of any dimension is the one which is necessary to completely describe the most granular fact of the process that can be measured.

With regard to the temporal dimension, in the case of aircraft's trajectories the TMA and en-route surveillance can provide information at a *second*-level as the finest grain. The key argument supporting storing the information of the trajectory at second-level is that, in case that all ATM safety nets fail, the TCAS computes the *tau* metric at a 1Hz rate [102], which is indicative that, to completely characterise a conflict, the system shall be described at second-level.

Figure 3-5 Temporal Dimensions

As it can be seen in Figure 3-5, the temporal dimension enables the aggregation and characterisation of facts at different levels (minute, hour, etc). The following reasoning partially supports each level:

1. Minute: Minute-buckets are used to aggregate the demand in a given sector (occupancy).
2. Hour: Hourly Entry Counts are a summary of the demand which has entered an airspace volume in a given hour.
3. Day: There are different patterns in the demand depending on the day (weekday, weekend, or bank holiday).
4. Week: The week is used as reference as “design points” for capacity-related constraints.
5. Month: Traffic varies depending on the month and season of the data.
6. Year: Traffic growths are studied in a yearly basis.

Regarding the spatial dimension, the horizontal projection of each track or intent information is determined by their coordinates in a spatial projection (i.e. stereographic [103]) or by their absolute coordinates (in a projection such as WGS84). These points are real numbers, and therefore those cannot be the smallest grain of the dimension.

A spatial dimension can aggregate those by means of regions, where the smallest grain of the dimension contains spatially these points. Regions could be organised through regular and independent partitions, such as regular grids as proposed by [104] or [76] and applied in [105]–[107], or irregular and system-oriented ones, such as airspace volumes or for the operational side of the ATM.

Figure 3-6 illustrates these possibilities. The centre of the figure is dedicated for the core generator of these dimensions, which is the tuple formed by {lat, lng, FL/altitude}, in green. The airspace structure and route dimension will be defined in the next subsection, and for

now, they can be identified as spatial entities. In the case of a regular grid, the figure shows a cube of 1NMx1NMx100 ft as the finest grain of any regular dimension, just as an example.

Figure 3-6 Spatial dimensions

3.2.3. Operational Dimensions

3.2.3.1. Operational Framework

The **airspace structure** forms part of the operational framework where the aircraft will be flying, together with a set of **operational procedures** to operate them (the aircraft) in that airspace and to interface with other ones. The result of the application of these procedures and the conduction of operations themselves result in the generation of patterns which can be extracted and studied. These patterns have to be included as operational dimensions. Some can be obtained through direct extraction, transformation and load from the published documents whereas other have to be obtained from extraction of operational patterns. This framework was defined by Histon and it is shown in Figure 3-7.

Figure 3-7 Airspace Structure as a conjunction of operational framework and operational procedures [87]

Starting from the framework, the Airspace Management layer creates airspace structures (*airspace boundaries* in Figure 3-7) to be controlled. Looking it from a 2D perspective, the most elemental is the “airblock”. The union of airblocks can form a “volume”. A set of volumes can form a “sector”, which can be bandboxed to form bigger sectors, finally limited by boundaries of an ATS Unit. The basic sectors (elemental) could be arranged in different larger sectors, which is the source of different airspace configurations. All of these can be attached with flight levels in order to determine the 3D definition of the volume.

Each entity within the airspace structure dimension is also depicted by a series of features. The ASM layer works in *AIRAC Cycles*, which are valid for 28 days. The airspace structures are **slow-changing** dimensions as a result of the changes defined in each AIRAC, and therefore, they are also defined with an interval of validity [Valid from, Valid To). Finally, only some of these units are deployable in operations, and for those, capacity features are defined such as the maximum hourly entry count, the sustain peak or the monitoring value. This is illustrated in Figure 3-8.

Figure 3-8 Operational definition of the Airspace Structure Dimension

In terms of the references for the trajectories, *paths* (see Section 2.2.2) are defined by a sequence of fixes or waypoints. Fixes are defined by their 2D coordinates. If there are N waypoints, there are $\frac{(N)(N-1)}{2}$ potential combinations of them. In reality, only nearby waypoints can form segments, which are marked with different designators¹⁶.

A sequence of segments and associated FLs are defined for all flights, and it is key to signal that they are inexorably linked to the surrounding airspace structure, as the latter is a response to the interdependencies among the formers. So, each potential available sequence of segments within an airspace structure is an element of the *route dimension*, associated with the elements of the airspace structure dimension.

Finally, the most important element within the core ones as defined by Histon are *airports*, as initial and final points of any flight. Those where already defined in Figure 3-4.

3.2.3.2. Operational Procedures

ATC Procedures can be formal or informal, following Histon's taxonomy. Formal procedures may be extracted from published sources and from unsupervised learning from the data. Informal operating procedures can be only extracted from unsupervised learning.

Procedures are a way of "human automation" [108] as a way for improving safety. Procedures are typically defined for the interfaces between operational entities and also for handling the vertical profile of the aircraft trajectory, among others.

For example, the typical interface between two operational sectors is defined by where the aircraft has to be handled-off to the next controller and the trajectory conditions, i.e. speed

¹⁶ For example, the RAD - Route Availability Document is a flight-planning source, which integrates structural and ATFCM requirements. <https://ansperformance.eu/acronym/rad/>

and flight level. Procedures can also be defined as a sequence of actions to descent an aircraft, or intermediate flight levels.

Figure 3-9 Procedure Dimension

3.3. Facts

Recalling the definitions of *trajectory prediction* from Section 2.3, Eurocontrol defined TP as the calculation of the trajectory of a flight within the Area of Interest of an ATSU for the purpose of the support of ATM services and tools. So far, the **dimensions** of the ATM system have tried to depict the **Area of Interest** for an ATSU unit in a structured way.

The scope of this research was limited to tactical vertical trajectory prediction. In this sense, the following subsections only detail ATM facts with the purpose of serving this specific bit of research.

3.3.1. Trajectory Facts

Recalling again the definitions of trajectory and trajectory prediction (Sections 2.2.2 and 2.3), the trajectory is defined as the temporal evolution of the aircraft state. The TP should address the estimation of the future states given the current aircraft state, the estimation of pilot and controller intents, environmental conditions, and models of the aircraft performance and procedure in a given AoI of an ATSU. Therefore, each element of the entity **trajectory** has to be characterised by a temporal dimension, a spatial dimension, environmental dimensions (weather conditions in given cells) and operational dimensions.

The first set of facts which are described corresponds to the consideration of studying the aircraft trajectory from an isolated point of view. The first source of information forming the trajectory is the surveillance information, which is considered within the scope of this thesis to proportionate the **actual trajectory**. Each **track** as recorded by the system is a **surveillance** fact.

Table 3-1 Surveillance Fact Table

	Dimension	Description
DIMENSIONS	Flight [Flight Dimension]	It identifies the track and the flight, as registered in the system.
	Spatial Location [Spatial; Airspace Structure, Grid or Segment]	Spatial dimension containing the track.
	Timestamp [Time]	Temporal dimension registering when the track was recorded.
	Operational Sector Active [Airspace Structure]	Operational dimension identifying which ATC sector was in charge of controlling the flight.
VALUES	Latitude	Recorded latitude at the timestamp [Time].
	Longitude	Recorded longitude at the timestamp [Time].
	Flight Level	Recorded flight level.
	Ground Speed	Recorded ground speed, in knots
	Heading	Recorded track angle
	Rate of Climb or Descent	Recorded vertical speed, in ft/min
	Other	Operational information (cleared flight level, cleared heading, speed restrictions, etc.) Mode S derived information (Selected Flight Level, other)

In addition, it is capital to transform the surveillance information to a fact table where the actual flights are compared with the evolution of the aircraft intent, when available. In this table, it is key to store together how the **trajectory fact** evolves in the information flow.

Figure 3-10 Schematic representation of the Trajectory Table Fact

Figure 3-10 pictures four different views of the trajectory, representing the evolution in both the “digital” spectrum and the transformation of the digital demand to actual flown trajectories. Starting from the top of the figure, the top view represents the path of the trajectories. The middle part of the graph represents the estimated or actual time of overflow over significant points of the trajectories as a function of the arc-length over the reference trajectory (aircraft intent), including the TOD (see Section 2.4.1.3), whereas the bottom plot represents the altitude as a function of the same arc-length parameter.

The colours represent different snapshots of the trajectory (from planning to post-flight passing through in-flight). It is a schematic representation on how the trajectory evolves in the system. The following table introduces the Trajectory Significant Point Fact table (see Table 3-2).

Table 3-2 Trajectory Significant Point Fact

	Dimension	Description
	Flight [Flight Dimension]	It identifies the track and the flight, as registered in the system.

	Dimension	Description
DIMENSIONS	Spatial Location [Spatial; Airspace Structure]	Airspace structure spatial dimension containing the point the register is referring to.
	Spatial Location [Spatial; Route structure]	Route structure spatial dimension containing the point the register is referring to.
	Snapshot Timestamp [Time]	Timestamp when the snapshot was taken.
	Expected / Flown Timestamp	Expected / Flown timestamp of the trajectory point fact.
	Trajectory Id	Unique ID for all points corresponding to the same snapshot timestamp and flight dimension
	Type of record	Intent or actual
	Type of trajectory point	Waypoint, Fix, TOC, TOD, Start of levelled segment, end of levelled segment, etc.
VALUES	Arc-length	Arc-length over the segment ([spatial dimension])
	Flight Level	Flight Level
	Ground Speed	Ground speed in knots
	Heading	Recorded track angle
	Rate of Climb or Descent	Vertical speed, in ft/min
	Other	Operational information (cleared flight level, cleared heading, speed restrictions, etc.) Mode S derived information (Selected Flight Level, other) Difference of time with previous snapshot of the trajectory.

3.3.2. ATC Facts

Let's consider now the provision of air traffic control within air traffic management. The main question within the context of this chapter would be **what are the facts which describe the ATC “business” process.**

This can be seen from two perspectives: one is facilitating a safe evolution of the trajectory and the other is preventing potential proximity events to happen. For example, if an aircraft has to start descending within a given airspace volume under the control of an ATCo, the aircraft can request to start the descent, but only the ATCo can authorise it. Whenever an ATCo acts opening, restricting or closing¹⁷ a trajectory, it is a nuclear fact of the trajectory management side.

Therefore, it is key to register the activity of the air traffic controller and the impact on the aircraft operation. Table 3-3 reflects a way to measure relevant quantities with regard to the ATC activity.

Table 3-3 ATCo Interaction Fact

	Dimension	Description
DIMENSION	Flight [Flight Dimension]	An element referring to the flight dimension (which aircraft is the instruction being directed to)
	Spatial Location [Spatial; Airspace Structure, Grid or Segment]	Spatial dimension containing the location of aircraft when the ATC action was issued.
	HMI's Interaction time [Time]	Timestamp of the interaction of the ATCo with the HMI.
	Clearance Time [Time]	Timestamp corresponding to when the ATCo issued the instruction to the aircraft.
	Acknowledgement time [Time]	Timestamp corresponding to when the aircraft readback the ATC instruction
	Operational Sector Active [Airspace Structure]	Operational sector which was been controlled by the ATCo who issued the instruction.
	Type of Instruction	Assumption, transference, hand-off, Flight Level instruction, Vector instruction, speed restriction, direct to a point, resume navigation, etc.
	Procedure	Operational Procedure (if any)

¹⁷ See [34] for a definition of open and closed trajectories.

	Dimension	Description
VALUES	Position along the route	Distance with respect the initial point of the segment / route [NM]
	Overall reaction time	Interval between the initial instruction (minimum value between HMI time and clearance time) and the final of the manoeuvre of the aircraft as a response [seconds]
	Cleared Flight Level	New cleared Flight Level (FL)
	Horizontal Speed Restriction	Horizontal speed restriction, it may be Mach or Knots depending on the airspace
	Vertical Speed Restriction	Vertical speed restriction (ft/min)
	Heading	Heading restriction (degrees)
	Other	For example, point to where an aircraft is directed.

On the other hand, there are causal effects driving ATCo actions. Main operational causes for ATCo actions are **potential proximity events**. If the ATCo fails to prevent the conflict to happen, then we can have **conflict facts (losses of separation)** and even TCAS events (Traffic Advisories and Resolution Advisories [109]). Eduardo García's PhD dissertation [110] included a comprehensive review of the features that completely characterises a potential conflict. Figure 3-11 includes a notional representation of a potential conflict between two aircraft.

Figure 3-11 Features which characterise a potential conflict [110]. The illustration reflects the relative dynamics of aircraft 2 with regard to aircraft 1. Two aircraft will be in conflict if their relative distance infringes the separation minima.

The translation of the figure to a DWH framework is made by identifying the dimensions of the fact (i.e. what describes the potential conflict) and the quantities that can be **measured**.

The dimensions, at least, for a proximity event (**potential conflict**) are listed in Table 3-4.

Table 3-4 Potential Conflict Fact

	Dimension	Description
DIMENSION	Flight A [Flight Dimension]	An element referring to the flight dimension (which aircraft are involved)
	Flight B [Flight Dimension]	An element referring to the flight dimension (which aircraft are involved)
	Spatial Location [Spatial; Airspace Structure, Grid or Segment]	Where the potential conflict is about to happen.
	Predicted Conflict Time [Time]	When the conflict would evolve from potential to actual
	Predicted Minimum Separation Time [Time]	When the aircraft will be in a minimum separation.
	Operational Sector Active [Airspace Structure]	Where it was expected to happen operationally speaking .
VALUES	Overall reaction time	Interval between the detection point and a manoeuvre of the aircraft as a response [seconds]
	Time to conflict from reference point	Time to separation minima infringement [seconds]
	Time to CPA	Time to the closest point of approach [seconds]
	Other	Predicted horizontal distance, predicted vertical distance, etc.

García and Histon postulate that potential quantities to be measured are the overall reaction time, the time to conflict with regard to a reference point or the time to the closest point of approach (CPA). Other quantities might be the predicted minimum horizontal distance, predicted minimum vertical distance and so on. These are listed in Table 3-4 as values which can be measured and aggregated depending on the dimensions.

In the case of actual conflicts, the previous dimensions and quantities evolve from *potential* to *actual* quantities. Additional quantities to be measured can be added, such as the minimum 3D distance or the reaction time, considering the reference point the start of the conflict.

3.4. The Flow Discovery Process

3.4.1. The Flow Dimension

The previous section has described the data model in terms of its spatiotemporal and operational natures. From the analysis of this section, it can be seen that the *capacity* side is airspace-oriented, whereas the *demand* side is trajectory-oriented.

The operational spatial dimension was divided in two. On the one hand the airspace volumes, and on the other hand, the route structure. The former was born as a response to the latter. If the focus is placed on each specific element of the airspace structure hierarchy, it could be found the routes that lay on them. The connection between them are often referred as *flows*.

But what is a *flow* from an ATM perspective? Delahaye defends that the term *flow* “is often used but never rigorously defined” [111]. It proposes a definition based on the capacity to split a group of trajectories in “similar” trajectory classes, depending on an exact trajectory distance and a learning classifier.

The answer to the question previously stated depends on the ATM layer that is going to use the *flow* entity. For example, in Air Traffic Flow and Capacity Management it may be used to compute the aircraft rate in a given volume, i.e. the number of aircraft expected to cross a given airspace volume per time unit. The capacity of the sectors is defined as the number of aircraft which are allowed to cross a sector in one hour.

However, recalling the *flow* definition given by Delahaye, the time is not included in the definition, and only flight paths are considered. ATC usually makes use of this definition as well. For example, Histon [87] defines *standard flows* as “recurring patterns of aircraft sharing common lateral paths”, that “due to their dependence on the route structure, tend to be persistent and stable”. That research addressed the cognitive mechanisms that air traffic controllers use for carrying out their duties.

These definitions share the concept of aggregation of paths. As such, the definition that it is proposed in this thesis is that a flow in the context of ATM is an **aggregation of trajectory facts or related elements** (associated paths) depending on temporal, spatial or operational dimensions, commensurate to the application that will make use of it.

These considerations about the flow’s definition have been developed due to the need to develop elements which can work as aggregations of the **flight dimension**. Due to the

duration of the trajectory cycle management, decisions in each ATM layers consider in one way or another the aggregations of flights.

Histon postulated that *“the persistent and repeated nature of standard flows provides an important and powerful basis for simplifying abstractions. Standard flow abstractions reflect a controller’s generalized expectation of aircraft trajectories in those flows within and near a sector”*. In a way, although each flight is managed individually by the air traffic controller, when evaluating their future states, it is hypothesised that they make use of cognitive *flow* abstractions in order to evaluate faster the future state of the aircraft. Furthermore, ATCOs interact with each flight depending on the route it is flying [95], [96].

In summary, the aggregation of flight facts (actual trajectories or aircraft intents), allow the definition of an additional operational dimension, *flow*, characterised by the airspace structure element considered. This is reflected in Figure 3-12, where the basic elements to “discover” the flows within an airspace structure element are identified.

Figure 3-12 The nature of the flow dimension

This dimension is an additional feature characterising the aircraft intent or actual trajectory (flight facts) as the aircraft transits through the airspace. The most used methodology for constructing this dimension is reviewed in the following section.

3.4.2. A Review of the Clustering Methodology

The exploitation of trajectory data allows the conduction of automatic flow extraction. Unsupervised machine learning techniques have been widely applied in this field to extract

air traffic flows in a given airspace volume. The most extended methodology was first proposed by **Gariel** [80], which follows a two-step process which involves the transformation of radar tracks and then a unsupervised clustering in a reduced state space. Other methods to cluster aircraft trajectories involve the use of spectral clustering [112].

Gariel's methodology has been widely adopted by the research community, using it in different contexts. In fact, Gariel used it initially to develop an analysis on the complexity of the airspace by identifying the number of *outliers* in a planning context. Other examples are Murça-Rocha et al. making use of the methodology for clustering trajectories in a TMA environment [83], and [59] using it for short-term trajectory prediction, which was previously discussed.

In summary, Gariel's methodology is a two-step process. The first step involves the application of a Principal Component Analysis (PCA) ([111], [113]) to a previously constructed matrix derived from track data. Then, a density-based clustering algorithm (DBSCAN) [114] is applied for unsupervised clusterisation in the reduced state-space derived from the PCA.

The methodology has been widely used and described in several papers. In the following section, particular elements are introduced for its generalisation in an automated process, together with a critical discussion of potential replacements for the density-based algorithm.

3.4.2.1. Augmented Matrix

Gariel initially augmented the information provided by the trajectories by transforming them with regard to a **reference point**. First of all, the trajectories must be resampled in an equal number of points. In this resampling, the **speed** information is lost.

Figure 3-13 Augmented Trajectory - Reference for construction

Then, constructing a box centred in the reference point (illustrated in Figure 3-12), each track is augmented by incorporating not only the information about the position of each track ($P=[X \ Y \ FL]$), but also the polar coordinates of each track ($[d, \theta]$), the distances from the corners of the box to each track $[d1, d2, d3, d4]$ and the heading of the aircraft at the track point $[\phi]$. Indeed, angles were multiplexed in their sines and cosines. The next step is to create the augmented matrix, where the rows correspond to the concatenation of all the points and associated features per flight.

Once the initial matrix has been constructed, it has to be normalised per column by linearly scaling it in a given range, usually between 0 and 1. If x denotes the elements of a column of the augmented matrix, the transformed column responds to

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (1)$$

Some important considerations for the automation of this process is that it has introduced the following hyperparameters that must be kept:

- Reference Point
- Width of the rectangle
- $\min(x)$ and $\max(x)$ for each column of the augmented matrix.

Some considerations can be made with regard to the required information to construct the augmented matrix. First of all, as the posterior steps only considers the linearized data, it is equally valid to use the *latitude and longitude* coordinates of the tracks rather than the

projections, which may ease the implementation of the algorithm depending on how the track information has been recorded. Distances may be calculated by using the Haversine function rather than the Euclidean distance. In addition, other distance measures or even other significant points could be used.

In addition, it is worth to note that if expert knowledge is going to be introduced in the clustering, it has to be in this first part. For example, in case that the vertical component is not operationally relevant, that column could be excluded. In this case the algorithm will not be able to discriminate by the vertical component of the trajectory.

3.4.2.2. Dimension Reduction

The next step of the methodology is to reduce the dimensionality of matrix, by means of a Principal Component Analysis (PCA)¹⁸. In summary, PCA is a non-parametric method applied to extract the most relevant information of a given sample. Considered each column of the augmented matrix as the coordinate over a linear basis, unknown, the PCA transforms the data by projecting it in an optimal orthogonal basis formed by the eigenvectors of the covariance. This means that PCA is a **change of basis**.

This process is represented in Figure 3-14, where the figure on the left-hand side illustrates the horizontal projection of a set of trajectories and the right-hand side represents the first three principal components after applying the PCA.

Figure 3-14 Left-hand side: Horizontal projection of a set of trajectories. Right-hand side: PCA transformation. In this subplot, the components are dimensionless.

¹⁸ For a thorough review of the PCA method, it is recommended to review the tutorial developed by Jonathon Shlens [113]

With regard to the automation of the process, as the clustering is conducted in this new basis, the process must store the transformation matrix. Whenever a new sample comes, after resampling, augmenting and rescaling, the transformed matrix must be transformed as it were following the PCA process.

So, if the Singular Value Decomposition (SVD) method is used to conduct the PCA (section VI.C of [113]), the original augmented matrix of N flights and M dimensions shall be transformed first by subtracting the mean per dimension, and then transposing it and dividing by square root of the total of flights. One of the outputs of the SVD is the transformation matrix (PC), which is used to project the data over the new basis. The initial number of samples (N), the mean per dimension and the transformation matrix (PC) shall be kept for whenever a new sample of flights comes and has to be transformed.

Figure 3-15 Considerations about the automation of the clustering process – PCA. Adaptation of J. Shlens code to a new sample.

3.4.2.3. Density-based Clustering

3.4.2.4. Density-based Clustering Algorithms

The trajectories in the transformed space form arbitrarily shaped regions with varying densities. Gariel's methodology proposed DBSCAN for extracting the principal clusters within the transformed space.

DBSCAN is based on the concept of dense regions. A dense region is constituted by a minimum number of points which are sufficiently close to each other. Thus, a region is defined by two parameters, the minimum number of points (min-pts) and a distance eps. DBSCAN starts visiting a point, and then checks if the neighbouring points constitute a dense region. If so, each point within the dense region is visited to search in its surrounding for new points that may be added to the existing cluster. Once there are not new candidates to the cluster, it is closed, and the algorithm starts again with an unvisited point, until all points are labelled within a cluster or as an outlier.

Thus, DBSCAN strengths are that it does not require a prior knowledge of the number of clusters and that it is well suited to identify arbitrarily formed shapes. On the other hand, DBSCAN suffers when handling regions of different densities, because of the static nature of eps. In addition, outcomes from the algorithm are sensitive to min-pts.

Thus, Gariel's methodology can evolve considering the new developments in density-based clustering algorithms, which have addressed the cluster varying density issue. The first attempt resulted in the OPTICS algorithm [115]. OPTICS stands for Ordering Points To Identify the Cluster Structure. Main differences between DBSCAN and OPTICS reside on the nature of the solution provided by each of them. DBSCAN provides a flat partitioning of the cluster structure, whereas OPTICS enables the extraction of a hierarchical clustering [116] from a reachability plot.

OPTICS orders the sample by reachability distances of close neighbours. The result of this ordering is a reachability plot, formed by dents and valleys. A valley between two dents would represent a dense region, as points are close among them. A dent represents a region where the distances progressively increase until a new dense region is reached. A representation of the reachability plot for a subsample from the right-hand side of Figure 3-15 is presented in Figure 2.

Figure 3-16 Reachability Plot for a subsample of Figure 3-14. The right-hand side column represents an equal transformed space, clustered differently based on the eps value. The right-hand side column represents the OPTICS reachability plot. Both columns are linked by colours

Figure 3-16 presents two different clustering solutions, illustrating the effect of selecting different values of eps on the original DBSCAN algorithm. The number of clusters would be given by the number of intersections with dent left-sides. As the eps value decreases, a higher number of clusters are detected.

The last step for the OPTICS algorithm is to extract the cluster hierarchy from the reachability plot. The authors proposed the OPTICS-AutoCluster [116] algorithm, which extracts the leaf nodes of a tree built from the OPTICS reachability plot.

Other approach that could be used is HDBSCAN*, which produces a hierarchy of all possible DBSCAN* partitions, where DBSCAN* introduces a small variation from the initial version. The extraction of all possible DBSCAN* partitions is formulated as an optimisation problem over the individual qualities of the extracted clusters [117], [118].

A review of these methods was conducted in [119]. In addition, a new algorithm was proposed (RDBSCAN), also targeting Gariel's methodology.

3.4.2.5. Classifying a new sample

One of the key difficulties when deploying Gariel's or derived methodologies is how classify a new sample. When DBSCAN produces the unsupervised clustering, it assigns labels to each cluster without operational meaning. In addition, the assignation of this labels is conducted without following a specified sequence, so, depending on how the initial sample is provided to DBSCAN, it will produce different labels for the same flow.

In this work, the automation of flow discovery process has been conducted by means of complementing the initial density-based clustering algorithm with a supervised clustering method (in this case, Random Forests were used [120]). The original process produces the unsupervised flows, which are used, together with the original signals in the reduced space, in order to train a supervised classifier. This classifier is then used to train new samples, keeping the traceability with the original density-based classification. This process is illustrated in Figure 3-17.

Figure 3-17 Automation of the density-based clustering to new samples

3.5. Research Data

A couple of datasets corresponding to the sector Central (LECBCCC) within the Barcelona ATC Centre in Spain will be used in this research. The sector of interest corresponds to an en-route sector with evolution and cruise traffic, presented in Figure 3-18. The figure includes the route structure and relevant designated waypoints for the work carried out in this research. Starting from the top-left of the chart and moving clockwise, these waypoints are PUMAL, NILDU, MAMES and LORES. Relevant navigation infrastructure points are also indicated (BGR, GIR, SSL and BCN). The most relevant route structures are from PUMAL to BCN (UN859), from BCN to LORES (UZ167), from PUMAL to LORES (UN13) and from BGR to BCN (UN975). UN13 was only activated after 2015 summer, so it was not defined for the period covered for the dataset.

The data include track and ATC operational data. The training dataset corresponds to four summer months of 2014, whereas the testing dataset covers two spring months of 2015.

Figure 3-18 LECBCCC Sector - Barcelona Air Traffic Control Sector. [121]

Raw track data are reported every 4.8s, where each register is composed by latitude and longitude in WGS84 coordinates, pressure altitude in hundreds of feet as measured by the aircraft Mode C transponder, heading, ground velocity in knots and the rate of climb or descent in ft/min. These data are correlated with the operational state of the aircraft as

recorded in the ATC operational data. The datasets do not include weather or flight plan data.

Through ETL processes, those data were conformed following the previous model. A total of 47,124 flights were loaded into the DWH. These flights accounted for a total of 7,037,840 tracks in this sector and 226,039 operational ATC interactions with its HMI.

It is important to note key data that are missing. It has not been used any type information about temperature or wind. The lack of availability of those data have led to formulate some assumptions in the analysis that will be discussed in each subsection.

3.5.1. Flows crossing the LECBCCC sector

The flow discovery process is performed by applying the methodology presented in Section 3.4. Results are presented for one AIRAC cycle within the sample, which represent a total of 10,312 flights. As mentioned earlier, the flow discovery process is based on the application of Gariel's methodology, but automatically extracting the eps parameter for the DBSCAN algorithm and applying recursively that clustering technique (RDBSCAN)[119].

A visual representation of the recursive application of DBSCAN is presented in Figure 3-19. There are three different plots for each row, presenting the outcomes obtained in the first and second iteration respectively. Plots on the left column represent the 2D trajectory projection, where each colour identifies a cluster as obtained from the algorithm. The graphs in the mid column present the representations of the three main principal components for the sample after the PCA. Colours on the left- and mid-graphs represent the same cluster. Finally, the graphs on the right-hand side illustrate the reachability plot for the PCA, showing the eps parameter that has been used, and the minpts applying.

Figure 3-19 Flow Discovery process after two iterations. Colours in left and middle columns reflect the same cluster in different projections. The right-side column illustrates the reachability plots for points displayed in the middle column.

Results of the flow decomposition for this sample are presented in Table 1. The table shows obtained clusters, indicating the size of the cluster and some attributes such as the regime and the direction for entering the sector and exiting. It can be observed that main flows correspond to flights cruising from northeast to southwest and from west to east, and flights descending from north to south, mainly landing in the Balearic Islands. It is also observed a large disparity on the relative sizes of the clusters.

Table 3-5 Flow Discovery Results for a sample of 10312 flights crossing the LECBCCC sector

Flow	Regime	Size	Relative Size	Mean Flight Level Entry	Mean Flight Level Exit	Direction Entry	Direction Exit
1	CRUISE	2378	24%	360	370	NE	SW
2	CRUISE	1060	11%	370	370	W	E
3	CRUISE	581	6%	370	370	N	S
4	CRUISE	254	3%	360	350	W	E
5	DESCENT	1773	18%	370	210	N	S

Flow	Regime	Size	Relative Size	Mean Flight Level Entry	Mean Flight Level Exit	Direction Entry	Direction Exit
6	DESCENT	766	8%	370	260	N	S
7	DESCENT	624	6%	380	230	N	S
8	DESCENT	465	5%	360	330	SW	NE
9	DESCENT	346	4%	360	280	NW	SE
10	DESCENT	174	2%	340	210	NW	SE
11	DESCENT	126	1%	250	210	NE	SW
12	OUTLIERS	1182	12%	NA	NA	NA	NA
	Others	583	6%	NA	NA	NA	NA

The upper airspace structure for this area is presented in Figure 6. Flows number 5, 6 and 7 of Table 1 have been identified on the route structure. Flow number 5 corresponds to the PUMAL – LORES route, in purple in Figure 6. Flow number 6 is linked to the route ALBER – LORES, in blue, whereas flow number 7 corresponds to the route PUMAL – CORDA, in green. The first and second flows are used by flights destination Palma de Majorca (ICAO Code LEPA), whereas the third one is used by flights destination Ibiza (ICAO Code LEIB). These flows are identified from now on by the name of the entry waypoint and the destination airport.

CFLs are based on the targeted operational procedures. In addition, these procedures are different for each route, and therefore, for the different flows.

Vertical procedures can be obtained by characterising the percentage of flights that have been subject to different CFLs. Those CFLs characterised with a frequency larger than a given threshold can be considered as recurrent CFLs, and therefore, considered within a pattern. If the selected airspace structure corresponds to a elemental or collapsed sector, the lowest CFL with the highest frequency is considered as the exit flight level of the sector (XFL).

This characterisation is shown on the left-hand side of Figure 3-21. Then, vertical procedures are identified as the different sorted CFL sequences, including the XFL.

Figure 3-21 Percentage of flights which have received a specific CFL per Flow and the spatial distribution of CFL issuances. Colours differentiate flows. Associated latitude and longitude for data reflected on the left-hand side are displayed on the right-hand side

Once flights have been grouped by vertical procedure, a trajectory segmentation process is conducted to isolate descent segments, and therefore, extracting the significant point fact for each flight. The trajectory segmentation was carried out by applying the Changing Point Analysis methodology ([122], [123]).

The analysis of issued CFLs for the selected flows shows the most frequent clearances issued by ATCOs.

Table 3-6 shows the percentage and absolute number of aircraft per flow that have received a specific CFL, where the first column indicates the total number of aircraft per flow for the

whole sample. It can be observed that flows destination LEPA are frequently authorised with CFLs to FL310, FL250, and finally, FL200 as the exit flight level. On the other hand, flights destination LEIB usually receive FL310, FL250 and FL220. It could be also observed that flights destination LEPA receive in a larger percentage of intermediate CFLs, whereas LEIB flights usually are authorised for the XFL directly.

Table 3-6 Cleared Flight Level Distribution per Flow

Flow		200	220	250	310	350
PUMAL – LEIB (2,724 flights)	% of Total Flights along CFL	0.81%	93.76%	26.76%	18.28%	5.84%
	Flights along CFL	22	2,554	729	498	159
PUMAL – LEPA (7,465 flights)	% of Total Flights along CFL	78.62%	4.80%	71.01%	54.91%	7.89%
	Flights along CFL	5,869	358	5,301	4,099	589
ALBER – LEPA (3,083 flights)	% of Total Flights along CFL	81.12%	4.35%	61.37%	60.49%	7.59%
	Flights along CFL	2,501	134	1,892	1,865	234

3.6. Summary of findings

The objective of this chapter was to introduce a paradigm to structure ATM data to exploit it in the next sections of the research. In this sense, a comprehensive information management layer was built, based on the DWH paradigm.

The flow discovery process to feed flow dimension was conducted applying a modified version of Gariel's methodology, self-assessing the definition of one of the parameters and introducing a recursive application of the DBSCAN algorithm. In addition, a critical review of the algorithm was presented, together with some requirements for a complete automation in the process for its inclusion in an ETL process.

After structuring the data, a closed dataset was selected in order to conduct this research. The selected data were described. Furthermore, major flows in the airspace volume corresponding to the selected sector (LECBCCC) were extracted, together with the typical ATC operational procedures in the vertical profile for that sector.

The next sections will use these data in order to analyse the impact of operational factors on the vertical profile of the trajectory, to determine whether considered factors should be included or not in the parameter estimation process for trajectory predictors based on point-mass models or as inputs in pure data-driven approaches.

This Page is Intentionally Left Blank

CHAPTER 4.

ANALYSIS OF OPERATIONAL FACTORS

4.1. Introduction¹⁹

The previous chapter addressed the generation of a data model which could allow a comprehensive analysis of the parameters affecting trajectory prediction errors by the use of machine learning techniques.

Figure 4-1 places the scope of this chapter within the structure of this research. Considering the short-term prediction of the vertical profile of the aircraft as the general scope of the thesis, this chapter will introduce first a methodology to analyse in a broad sense the impact of a series of factors, previously identified, in the vertical profile of the trajectory. This is represented as a light green in the figure.

Thus, it was hypothesized that the integration of trajectory kinematic data and operational information could improve the current understanding about what factors are creating uncertainty, which is a current challenge for trajectory predictors.

The flight dimension provides typical operational factors which can contribute to the variations on the descent profiles, besides those caused by the weather situation. Thus, the aircraft model (weight, performance) or the airline (cost index) are candidate factors to be leading the distribution of the variance of the trajectories.

¹⁹ Methodology, results and conclusions of this chapter have been published in [147] (C. E. Verdonk Gallego, V. F. Gómez Comendador, F. J. Sáez Nieto, G. Orenga Imaz, and R. M. Arnaldo Valdés, ‘Analysis of air traffic control operational impact on aircraft vertical profiles supported by machine learning’, *Transp. Res. Part C Emerg. Technol.*, vol. 95, pp. 883–903, 2018).

Figure 4-1 Knowledge Discovery Process - Scope of Chapter 4

In addition, recalling from previous sections, the flow organisation of the airspace constitutes a main factor to understand how Air Traffic Controllers (ATCOs) handle traffic [87] considering that each route is handled differently, according to constraints established in operational procedures (see Section 3.2.3.2). Thus, ATC procedures, tactical planning and clearances are fundamental for the generation of the aircraft intent, and consequently, of the TP accuracy.

The aim of this exploratory section is to determine if the variance in the vertical profile of the trajectories may be apportioned to different sources of uncertainty, including ATC operational factors. The information management layer introduced in the previous chapter enables the aggregation of flights that have been subject to similar ATC actions, flown similar routes, operated by the same airline or with the same aircraft.

Secondly, these factors are further analysed by means of two different type of trajectory computation engines. The first one is a 3DoF model, based on point-mass models. The second one is a pure data-driven model which provides exclusively the vertical geometry and kinematic of the aircraft as a function of time. This is represented in Figure 4-1 by process coloured with a light orange shadow.

4.2. Methodology

4.2.1. Exploratory Analysis - Multilevel Linear Models

The main goal of this first part is to determine if the variance in the vertical profile of the flight trajectories may be apportioned to different sources of uncertainty, including ATC operational factors. The data model structure layer introduced in the previous section enables the aggregation of flights that have been subject to similar ATC actions, flown similar routes, operated by the same airline or with the same aircraft, in a given airspace.

The identification of a set of flights using the same route has been discussed in the previous section, whereas the airline is a descriptor of the flight dimension. Regarding the identification of flights subject to similar ATC actions, only CFLs given by the ATCOs are analysed, as this research focuses on the vertical component of the trajectory.

Recalling from Section 3.5.2, vertical procedures were obtained by characterising the percentage of flights that were subject to different CFLs. Those CFLs characterised with a frequency larger than a given threshold were considered as recurrent CFLs, and therefore, considered within an ATC pattern. The lowest CFL with the highest frequency was considered as the exit flight level of the sector (XFL). Once flights were grouped by vertical procedure, a trajectory segmentation process was conducted to isolate descent segments. Thus, trajectories in the sector of interest are reduced to the segments between the TOD and the XFL.

As introduced earlier, the main objective of exploratory work is to perform an exploratory analysis of the data, for a priori identification of factors impacting on the vertical profile by exploiting Multilevel Linear Models (MLMs).

A MLM is a set of multiple regression models which are dependent on parameters which vary at more than one level. A MLM is defined by

$$Y_{ij} = \beta_{0j} + \beta_{1j} \cdot X_{ij} + \varepsilon_{ij} , \quad (2)$$

where Y_{ij} is the response i within group j , X_{ij} refers to the independent variable for the same sample, and finally, β_{0j} and β_{1j} are the intercept and regression coefficients for the group j . If we have a representative sample of data, β_{0j} and β_{1j} could be inferred using Maximum Likelihood Estimation (MLE).

For this analysis, the response variable will be the quotient between the difference in feet between the TOD and a given flight level, and the time it has taken to reach that flight level. It is comparable to the ROCD of the aircraft if it were following a descent with constant ROCD for the whole phase. The ROCD, if a point-mass model is assumed, can be modelled as

$$ROCD = \frac{dH}{dt} = \left(\frac{T-\Delta T}{T}\right) \cdot \frac{(Thr-D)}{g \cdot m} V_{TAS} \cdot \left(1 + \left(\frac{T-\Delta T}{T}\right) \cdot \frac{V_{TAS}}{g} \cdot \frac{dV_{TAS}}{dH}\right)^{-1}, \quad (3)$$

where H is the geopotential pressure altitude; T is the air absolute temperature; ΔT is the temperature differential from the standard atmosphere at MSL (Mean Sea Level); Thr stands for Thrust; D for aerodynamic Drag; g for the gravitational acceleration; m for the aircraft mass; and V_{TAS} for the true airspeed of the aircraft. Thus, the ROCD depends on temperature, aircraft dependent coefficients (Thr , D and m), V_{TAS} , and its rate of variation with altitude.

The independent variables considered for this prior analysis are: the ground speed, V^{Gr} , which depends on the true airspeed, V^{TAS} , and the wind vector, $\vec{W}$; the distance to the destination airport from the TOD (r^{Dest}); the flight level at the TOD; the ATC vertical procedure, supposed as known; the aircraft model; and finally, the airline or company.

V^{TAS} is the magnitude of the vector difference between $\vec{V}^{Gr}$ and $\vec{W}$, which remains unknown. The analysis of the last term of (3) shows a dependence with the distance to the airport. This dependency can be observed given that, if the aircraft is close to the airport, it must reduce its speed at a higher rate to properly face the final approach. The same reasoning applies for the FL, as aircraft would need to reduce their altitude faster. Finally, the term $\frac{(Thr-D)}{g \cdot m}$ depends on the aircraft model and the airline, impacting on the speed of the aircraft, through the cost index (see Section 2.2.4). The temperature deviation at MSL (ΔT) remains unknown for this analysis. It has been assumed to be zero, making the ratio $\left(\frac{T-\Delta T}{T}\right)$ equal to one, introducing an additional source of error.

Thus, a linearized rate of descent (LRD) can be modelled as:

$$LRD_{i,j,k,l} \sim \beta_{0j,k,l} + \beta_{1j,k,l} \cdot r_{i,j,k,l}^{DEST} + \beta_{2j,k,l} \cdot V_{i,j,k,l}^{Gr} + \varepsilon_{i,j,k,l}, \quad (4)$$

where the sub-indexes stand for:

- i : is the flight i within group $\{j, k, l\}$,
- j : is the level j within the group *operational procedure*,
- k : is the level k within the group *Flight Level at the TOD*,
- l : is the level l within the group *airline*.

$\beta_{0j,k,l}$ represents the intercept term for the group formed by the tuple $\{j, k, l\}$; $\beta_{1j,k,l}$ is the slope coefficient associated to r^{Dest} for the same group; and $\beta_{2j,k,l}$ is the slope parameter for V^{Gr} for that group. The β vector is estimated using a maximum likelihood estimator for each group.

In this exploratory analysis, the set of MLMs is not intended to be used as accurate models for the ROCD, but to identify sources of uncertainty in the trajectory vertical profile. The impact of the factors on the vertical profiles will be assessed by exploring how different MLMs, obtained by introducing these factors in the models sequentially, can predict the LRD and the remaining uncertainty. Thus, 10-fold cross validations are conducted for each of these MLMs, to obtain average Root Mean Squared Errors (RMSE) for the predicted LRDs and times to reach a given flight level, and adjusted R Squared coefficients. The behaviour of the different MLMs will then be evaluated, identifying which of these factors have a stronger impact on the trajectory vertical profile.

4.2.2. Trajectory Computation Engines – Point-Mass Models

Once the exploratory analysis is conducted, the knowledge gain is exploited by incorporating it to known trajectory computation engines. The first model follows the work presented by Alligier et al. [5]. The research proposed to learn key input parameters used in point-mass models from historical data. These parameters were the mass of the aircraft and coefficients weighting a modelled thrust law. The model is based on the Total-Energy Model [58] where the variation of the work done by forces acting on the aircraft is equated to the rate of the total energy,

$$\frac{(Thr-D) \cdot V_{TAS}}{m} = V_{TAS} \frac{dV_{TAS}}{dt} + g \left(\frac{T}{T-\Delta T} \right) \frac{dH}{dt} + \frac{d\vec{W}}{dt} \cdot \vec{V}_{TAS} , \quad (5)$$

This relation can be transformed to (3), assuming no wind. The last term of that equation (3) is defined as the energy share factor (ESF), given by:

$$ESF = \left(1 + \left(\frac{T-\Delta T}{T} \right) \frac{V_{TAS}}{g} \frac{dV_{TAS}}{dH} \right)^{-1} , \quad (6)$$

This factor can be modelled as function of the Mach number [58]. So, if it is assumed that all the information is known, the vertical component of the trajectory can be computed by integrating the following expression:

$$H_t - H_0 = \int_0^t \left(\frac{(Thr - D)V_{TAS}}{m} - \frac{d\bar{W}}{dt} \cdot \vec{V}_{TAS} \right) ESF dt. \quad (7)$$

Aircraft thrust and drag are main elements participating in (5) and (7). BADA 3 models the thrust as a function of H , ΔT and V_{TAS} . Firstly, it models the thrust for maximum climb for a standard atmosphere at MSL, and then, it weights it depending on the flight phase. The drag is modelled as a polynomial function of the mass, where the coefficients depends on H , ΔT and V_{TAS} , and aircraft-related parameters such as the wing reference surface and drag coefficients. Those terms are defined as follows:

$$\begin{aligned} Thr &= c f_1(H, \Delta T, V_{TAS}) \\ D &= f_2(H, \Delta T, V_{TAS}) + f_3(H, \Delta T, V_{TAS})m^2 \end{aligned} \quad (8)$$

It can be observed that the aircraft mass and the thrust coefficient, which is denoted as c in (8), are required for obtaining the altitude evolution in (7). The approach proposed by Alligier et al. used previous tracks to estimate the mass, although alternative methods were proposed in [6] for predicting it, and then, to use historical data to learn c . The referred work was applied to climbing trajectories.

This part of the research replicates this methodology, but for flights starting the descent phase. The first step estimates the mass of the aircraft, considering previous cruise tracks. In the second step, the coefficient for the thrust law is learnt from historical data, using the previously obtained mass.

There are some modifications in the process with respect to the one presented in [5]. For the mass estimation, the c coefficient is unknown as the aircraft is in the cruise phase. In [58], it is detailed that the maximum available thrust in the cruise phase is modelled as a function of the maximum climb thrust, but it does not provide the typical value for this coefficient and phase. In order to apply the mass estimation method, the coefficient ($c_{cruise,ref}$) is obtained through the following expression:

$$c_{cruise,ref} f_1(H, 0, V_{TAS_{BADA}}(H)) - f_2(H, 0, V_{TAS_{BADA}}(H)) + f_3(H, 0, V_{TAS_{BADA}}(H))m_{des,ref}^2 = 0, \quad (9)$$

where $c_{cruise,ref}$ denotes the thrust coefficient for this phase and the reference mass, $V_{TAS_{BADA}}(H)$ is the true airspeed considering a standard atmosphere at MSL for a given pressure altitude, and $m_{des,ref}$ is the mass as declared by BADA for aircraft starting the descent phase.

Once the $c_{cruise,ref}$ constant parameter is obtained, and considering K tracks of a trajectory, the aircraft mass m_i for the trajectory i is obtained by solving the following expression:

$$m_i^4 \sum_{k=1}^K (f_{3k})^2 + m_i^3 \sum_{k=1}^K (f_{3k} \cdot \frac{Q_k}{V_{TASk}}) + m_i \sum_{k=1}^K (c_{cruise,ref} \cdot f_{1k} - f_{2k}) \cdot \frac{Q_k}{V_{TASk}} - \sum_{k=1}^K (c_{cruise,ref} \cdot f_{1k} - f_{2k})^2 = 0, \quad (10)$$

where Q_k is an observed magnitude for each track k for flight i , corresponding to the right-hand side of (5).

Once the mass is estimated for each aircraft, the second step involves learning c for the different groups used in the MLM analysis. Similarly to the process for estimating the mass, Q is an observable. The difference between the work done by the forces acting on the aircraft and this observable, for the whole set of tracks for each group to be studied, could be expressed as:

$$\varepsilon_{total} = \sum_{i=1}^N \sum_{k=1}^K \left(\frac{(c_{k,i}(T_{max,climb_{k,i}} - D_{k,i})V_{TAS_{k,i}})}{m_i} - Q_{k,i} \right)^2, \quad (11)$$

Now, c is obtained by minimising the difference between the modelled variation of the specific total-energy and the observed one, given the estimated mass. The parameter c can be modelled in different forms. In [5], a 4th-degree polynomial of the altitude is proposed. In this work, c is modelled as a constant coefficient for the sake of simplicity, obtained by minimising ε_{total} for the different groups. To solve this non-linear problem, the Levenberg-Marquardt method can be applied [124], [125].

The c coefficients for the different groups will be obtained as described previously. Then, altitudes will be predicted starting from the TOD for the different groups, discriminating by aircraft. Finally, predictions will be compared by analysing the distributions of the residuals for different times from the TOD, for the same flight samples.

4.2.3. Artificial Neural Networks

A second approach is followed by applying Artificial Neural Networks (ANN), a pure data-driven process inspired by how neurons work. An ANN is composed by a set of nodes connected by weighted edges. In each node, operations are conducted transforming inputs and passing the output through its connections to the next node.

In this work, feed-forward networks (Bishop, 1995) are considered, that may be viewed as a general framework for representing non-linear functional mappings between inputs and

outputs, by composing the outputs obtained by applying non-linear functions in each node, so called activation functions, placed in sequential layers. All neurons in the consecutive layers are fully connected. These networks are formed by an input layer (x), a set of hidden layers, and finally, an output layer. For a network with one hidden layer, this can be expressed as:

$$H(x; \theta) = \Theta(\sum_{j=1}^n \theta_j \Phi(\sum_{i=1}^p (\theta_{ij} x_i + \theta_{0j}) - \theta_0), \quad (12)$$

where Φ denotes the activation functions for the hidden layers, whereas Θ denotes the activation function of the output layer. θ_{ij} and θ_j are the weights associated to the connection between the input i and the node j in the hidden layer, and between node j and the output, respectively. Finally, θ_0 and θ_{0j} are the biases associated with the output and the neuron j on the hidden layer respectively.

The inputs of our layers would be the information that we can provide at the TOD of the trajectory for the different groups and time t . For example, if the flow of the flight is considered as a relevant variable, it will be included as an input. The output will be the altitude for time t . The activation functions for the hidden layer are sigmoid ones, whereas the output function is a linear function. The network is trained with the Levenberg-Marquardt backpropagation algorithm [126].

As in the PMM approach, different ANNs will be trained including or excluding relevant variables, and then, predictions will be compared for different times.

4.3. Results

4.3.1. Results of the MLM for initial evaluation of the ATC Operational Impact on the Trajectory Vertical Profile

The flight sample is segmented depending on the operational procedure, the flight level at the TOD (RFL), and finally, the airline or company. Only flights with carried out continuous descents are considered, for one AIRAC cycle. The aircraft model has not been included in this initial segmentation due to that introducing this additional level leads to sample sizes smaller than a desired minimum for avoiding biased estimators in the MLM modelling [127]. Nevertheless, the aircraft model is recovered later. The different partitions and their sample size used in this first exploratory analysis are represented in Figure 4-2.

Figure 4-2 Sample size for each Company broken down by Flow, RFL and Procedure. Colour shows details about Flow, maintaining same colours of Figure 3-20 and Figure 3-21. The columns are labelled by the sample size.

MLMs are trained for the sample presented above. Table 4-1 introduces the RMSE, the standard deviation on the RMSE due to the cross validation, and the RMSE for a new sample (not used for training) for two magnitudes, the estimated LRD and the estimated time to cross FL250 from the TOD. In addition, it introduces the adjusted R^2 for the model for the cross validated set and a new sample. The description of the factors included for training the different MLMs are introduced on the leftmost column, where *Complete* stands for the complete set of factors (Procedure: Proc, RFL: FL at the TOD, COM: Company).

Table 4-1 Results considering different factors for the MLM and the complete sample. Training Sample 2756, Test Size 2481, New Sample: 611

Model	Sample Size	LRD [ft/min]			Time to FL250 [s]			R ²	
		RMSE	Std RMSE (CV)	RMSE New Sample	RMSE	Std RMSE	RMSE New Sample	R2 CV	R2 NS
COMPLETE	2756	241.55	16.52	255.68	41.22	3.13	46.96	0.640	0.558
PROC + RFL	2756	252.71	11.30	268.01	43.21	2.46	46.36	0.610	0.523

Results

RFL + COM	2756	302.70	13.40	360.62	53.68	2.03	71.79	0.438	0.121
PROC	2756	294.52	16.00	289.18	50.84	2.26	52.00	0.471	0.445
Simple Regression	2756	378.89	14.59	361.07	67.35	3.02	65.20	0.242	0.134

Figure 4-3 illustrates how procedures impact on the residuals of the model. On the left-hand side of the figure, the plot illustrates the residuals distribution for the predicted LRD depending on the distance to the airport from the TOD. Additionally, each colour identifies a procedure, which relate to the right-hand side of the figure. Boxplots on the right-hand side summarise the distribution of the residuals depending on the procedure.

Figure 4-3 Residuals of the LRD for a complete sample and all factors considered. Colours are common in both figures. The right-hand side of the Figure shows boxplots with whiskers with maximum 1.5 Interquartile Range (IQR).

From Figure 4-3, it is observed that Interquartile Ranges (IQR) associated to procedures (200) and (200 – 250) are significantly larger than for other procedures. Now, these flights are removed from the sample to analyse the impact on the explained variance by the different MLMs. Results from removing these flights from the initial sample are introduced in Table 4. In addition, the sample has been segmented per flow as well, and results for this last segmentation are presented in

Table 4-3 and Table 4-4.

Table 4-2 Results considering different factors for the MLM, and a sample without procedures showing a greater variance. Training Sample 2106, Test Size 1896, New Sample: 572.

		LRD [ft/min]			Time to FL250 [s]			R ²	
Model	Sample Size	RMSE	Std RMSE (CV)	RMSE New Sample	RMSE	Std RMSE	RMSE New Sample	R ² CV	R ² NS
COMPLETE	2106	190.76	11.96	214.10	37.62	2.64	43.21	0.706	0.612
PROC + RFL	2106	204.25	13.38	220.27	39.81	3.73	41.44	0.665	0.621
RFL + COM	2106	264.44	25.05	341.10	51.13	4.72	73.16	0.430	0.024
PROC	2106	254.07	14.53	257.07	48.73	2.81	49.25	0.484	0.483
Simple Regression	2106	347.54	16.55	345.40	66.98	2.90	65.26	0.169	0.067

Table 4-3 Results considering different factors for the MLM, and a sample without procedures showing a greater variance for the Flow Pupal - LEPA. Training Sample 1240, Test Size 137, New Sample: 385.

		LRD [ft/min]			Time to FL250 [s]			R ²	
Model	Sample Size	RMSE LRD	Std RMSE (CV)	RMSE NS	RMSE	Std RMSE	RMSE New Sample	R ² CV	R ² NS
COMPLETE	1377	207.02	16.40	231.99	40.99	2.83	45.70	0.620	0.581
PROC + RFL	1377	224.14	13.27	246.05	43.96	3.13	47.00	0.560	0.536
RFL + COM	1377	207.27	12.42	231.99	40.94	3.39	45.70	0.620	0.581
PROC	1377	263.82	25.83	278.96	52.84	5.11	55.55	0.389	0.403
Simple Regression	1377	291.23	16.50	297.94	58.05	2.66	57.91	0.389	0.319

Results

Table 4-4 Results considering different factors for the MLM, and a sample without procedures showing a greater variance for the Flow Pupal - LEIB. Training Sample 507, Test Size 56, New Sample: 385.

Model	Sample Size	LRD [ft/min]			Time to FL250 [s]			R ²	
		RMSE	Std RMSE (CV)	RMSE New Sample	RMSE	Std RMSE	RMSE New Sample	R ² CV	R ² NS
COMPLETE	563	127.08	16.93	160.59	25.35	5.79	37.63	0.885	0.738
PROC + RFL	563	132.41	14.88	166.91	25.35	5.24	30.75	0.876	0.761
RFL + COM	563	133.43	24.94	159.88	27.46	11.14	33.42	0.871	0.741
PROC	563	224.10	17.99	210.53	35.89	3.68	36.33	0.647	0.619
Simple Regression	2106	333.43	40.54	317.29	55.07	5.51	52.42	0.377	0.135

From Table 4-1, it is observed that the RFL and the procedure have a greater impact on the variance explained by the model. This is reflected in the decrement of the RMSE for the LRD, the RMSE for the time to fly over FL250 and R² when these factors are removed.

Although it is detected a certain degree of explained variance, residuals show a higher variance for flights having a TOD located closer to destination. After determining the variance of the residuals by procedure, it is observed that there are some of them with more leverage to explain these larger variances. In addition, as it can be seen in Table 4-1, there is a drop on the explained variance when the procedure factor was removed from the factors to train the MLM. However, it can be observed how the best fits correspond to procedures with XFL equal to 220, which corresponds to LEIB destinations. Each destination (LEIB and LEPA) have different exit flight levels, and therefore, the procedure variable is masquerading a variable discriminating between different routes, for this sample. Table 4-2 and

Table 4-3 split the sample by flow. When the procedure variable is removed, it does not impact significantly on the explained variance.

Table 4-2 and

Table 4-3 show how the MLMs behave differently depending on the flow. Flows destination LEPA start descending later than the flow destination LEIB. For the latter, the model behaves significantly more accurately than for LEPA flows. For the 10-fold cross validation training and testing, the adjusted- R² was close to 0.90, with a RMSE for the predicted time to arrive to FL250 from the TOD of 25 seconds, with a standard deviation of 5 seconds for the RMSE. When a new sample was tested, the RMSE increased up to 37 seconds for the model including all the factors. When the airline factor was removed, the RMSE descended to 30

seconds, which indicates that including the airline factor could overfit our model from year to year, as the cost index of the airline is unknown and could vary.

For LEPA flows, explained variances drop significantly. The analysis of the residuals for procedures (200) and (200-250) shows that this larger variance in the prediction is due to the poor linear correlation between the response and the ground speed and the TOD, for the samples that are closer to the destination airport.

From a statistical point of view, the sample was not balanced. Because of this reason and that homoscedasticity conditions were not met either, ANOVA test conditions could not be fulfilled, and MLMs were chosen. In addition, the minimum sample limit of 50, together with removing those aircraft that suffered a levelled segment during the descent and those not cleared with standard flight levels, caused that almost an 80% from the initial sample was initially dismissed. Table 4-5 presents the results if the requirement of including the airline for creating the different groups is removed.

Table 4-5 Results for two level grouping depending (Procedure and RFL)

		LRD [ft/min]			Time to FL250 [s]			R ²	
Sample	Sample Size	RMSE	Std RMSE (CV)	RMSE New Sample	RMSE	Std RMSE	RMSE New Sample	R2 CV	R2 NS
ALBER - LEPA	1658	313.358	21.20	293.12	47.91	3.45	47.40	0.5271	0.5603
PUMAL - LEPA	3415	300.15	12.90	310.5	48.39	5.24	50.753	0.517	0.511
PUMAL - LEIB	1457	161.935	21.60	186.439	31.25	4.71	37.7	0.85	0.764

Results of Table 4-5 compared with the third rows of

Table 4-3 and Table 4-4 show similar values for the R², for both flows, though less variance is explained for the flow PUMAL – LEPA. It can be concluded that for this sample, the airline factor was not incorporating added value to this model.

These discussions lead to the conclusion that the main factor affecting the descent phase for this sample is the position of the TOD, that is very dependent on the flow associated to each

flight. In addition, the RFL is also impacting on the LRD. Finally, procedures may explain some variance, but to a lesser extent.

The aircraft model has not been added to this analysis, as it plays a key role in the next section. However, it has been observed that airlines operating in these routes usually fly the same aircraft model, and therefore, this variable was masquerading it. It has been observed that including the airline as a factor for training the MLMs does not have a strong impact on the explained variance.

Figure 4-4 plots the aircraft model distribution existing in the sample. It can be observed that the two most frequent aircraft were the Boeing 737-800 (ICAO code B738) and the Airbus 320 (ICAO code A320), which have been selected to conduct the rest of the analysis. Both are medium size aircraft, with similar reference masses and ranges. Finally, the number of flights operating aircraft models plotted in Figure 4-4 represents a 93% of the total sample.

Figure 4-4 Aircraft Model Frequency Distribution

After the discussion performed in the previous section, it was decided to continue the analysis by segmenting the initial sample by the following groups: RFL; RFL and flow (RFL-FLOW); RFL and company (RFL-COM); and finally, RFL, flow and operational procedure (RFL-FLOW-PROC). In order to have enough samples in the most segmented case, the initial sample was filtered in order that, in the case of RFL-FLOW-PROC, each group has a least 10 members. Then, this sample of flights has been backtracked for the other groups, to always study the same sample of flights. The total of flights considered for the analysis are 2715 flights operated with B738s and 781 operated with A320s.

Results regarding the PMM are presented first, and then results from the application of ANNs. Finally, a discussion about the results is conducted. For both analysis, comparisons are conducted for the training and testing samples, by means of the analysis of key statistics associated to the residuals derived from the prediction of the pressure altitude.

4.3.2. Point-Mass Models

PMMs are sensitive to ΔT and $\vec{W}$. Weather data were not available, and therefore, assumptions have been made regarding temperature and wind. Regarding ΔT , it has been considered zero for the whole sample. In the case of $\vec{W}$, it has been assumed that all flights are flying at V_{TAS} at the TOD. Therefore, the magnitude of the wind speed in direction of V_{TAS} for each flight is equal to the difference between the ground speed and the reference V_{TAS} for the aircraft model at a given altitude for a standard atmosphere, as provided by BADA. Then, it has been considered that the magnitude of $\vec{W}$ in direction of $\vec{V}_{TAS}$ remains constant during the entire trajectory.

Another assumption has been made regarding the ESF law. In Equation (7), the ESF law is required to conduct predictions using the total-energy model. Within this work, the ESF is considered as function of the Mach number [58]. In the following contents, and only for the PMM model case of study, it is assumed that V_{TAS} is known for the whole trajectory, considering the wind known and constant. In addition, it is assumed that the aircraft are flying following a constant Mach/CAS descent law. For both aircraft, the transition altitude between constant Mach and constant CAS is defined in BADA at FL290. The sample has not been further filtered depending on if flights have flown using a regular profile, because this additional filtering would have reduced even further the sample.

The first step of the methodology was the estimation of the mass of the aircraft. The reference mass for the B738 is 65300 kg, whereas for the A320 is 64000 kg. The cumulative distribution function (CDF) of the estimated masses, considered as a random variable, are plotted in Figure 4-5. The estimated masses are always located between the maximum and minimum weights, as provided by BADA. It is observed a larger difference between the median value and the reference mass in for A320s than for the B738s.

Figure 4-5 Estimated mass CDF for the sample. Each colour represents an aircraft. It contains 2715 flights corresponding to B738 (blue) and 781 flights corresponding to A320 (red).

After the mass estimation, the thrust law coefficients (11) are obtained for the different segmentations. Then, predictions of the pressure altitude have iteratively been conducted depending on the segmentations. The PMM predictor has been run from the TOD to 300 seconds after, for both aircraft models and for each case of sample segmentation. Mean and standard deviation values of the residuals between predicted and true pressure altitudes for the training sample are shown in Figure 4-6, which includes the prediction using BADA as a reference as well.

Figure 4-6 Mean and Standard Deviation values of the residuals for different segmentation groups and learnt thrust laws for the training set.

Figure 4-6 contains four different plots. The top row shows the results for the B738 case, whereas the bottom row shows the A320 case. The left column represents the mean of the residuals, whereas the right column shows the standard deviation. The horizontal axis represents the time for the four subplots, whereas the vertical ones consider means or standard deviations in flight levels (100ft). The colours represent different groupings, that are maintained from now on.

Biases for the BADA predictions show similar trends for both aircraft. It grows linearly with time until it stabilises around 200 seconds after the TOD. In the case of learnt thrust laws, the bias plots represent similar trends for both aircraft as well. For the first 200 seconds after the TOD, biases are almost removed, and then they start increasing in absolute values, with a more pronounced evolution for the A320 case. Bias plots show a mirrored trend in the bias evolution, with the inflection point corresponding to the timestamp where FL290 was generally crossed. This flight level corresponds to the one when the ESF law switched from constant Mach to constant CAS.

The standard deviation values increase with time for BADA and learnt thrust laws, almost following the exactly same trend, as the same dynamic model is being used. It can be observed that the grouping with a better performance corresponds to the segmentation RFL-FLOW-PROC, whereas the other three behave very similarly.

Figure 4-7 Mean and Standard Deviation values of the residuals for different segmentation groups and learnt thrust laws for the testing set.

Figure 4-7 shows the results from applying the previous model to the testing dataset. For the B738 case, the testing sample size was 605 aircraft, for a 2715 sample size for the training

one. For the A320, the testing sample size was 212 aircraft, whereas the training sample size was 781 aircraft. It can be observed that for the B738 case, the bias is slightly above zero for the initial segment between the TOD and 200 seconds, and then, it increases in absolute values. For the A320 case, the decrement on the performance starts earlier than in the training set, but its effect is increased after 200 seconds. The standard deviation curves show a high correspondence level to those presented in Figure 4-6. As for the training sample, the segmentation that has a finest performance corresponds to the RFL-FLOW-PROC case.

4.3.3. Artificial Neural Networks

A similar exercise has been conducted but using ANNs as trajectory predictor. In this case, the basic input variables that are common for training the ANNs are: time from the TOD; distance to the airport from the TOD; and finally, the observed magnitude of the ground velocity at the TOD. Then, to determine the impact of the selected features used for segmenting the sample, the corresponding variables are included as additional nodes in the input layer.

All ANNs are formed by one input layer, a 10-neurons hidden layer and an output layer. The activation function for the neurons in the hidden layer are sigmoid ones, whereas the activation function in the output layer is a linear one. Only an 85% of the sample is considered for training purposes, whereas the remaining 15% has been used as validation set, as an early stopping procedure when training the ANN for preventing overfitting. Then, another independent set is used for testing purposes.

Results are presented in Figure 4-8. The plot follows the same pattern than in Figure 4-6, with the top row dedicated to the B738 case, whereas the A320 case lies in the bottom. The bias plots can be found on the left-hand side, whereas the standard deviation values are illustrated on the right-hand side. The colours and segmenting groups are maintained from Figure 4-6.

Figure 4-8 Mean and Standard Deviation values of the residuals for different segmentation groups for ANNs including them as inputs for the training set.

The bias plots show that biases are almost removed for the different time steps. Biases for the different segmentations behave similarly, and it cannot be appreciated that including more features for training the ANNs has a relevant impact on them.

In the case of the standard deviations, the inclusion of more input variables for training the ANN results in appreciable decrements in standard deviation values along time. The results show an increased performance of this predictor compared with the BADA baseline prediction. For both aircraft, adding ‘Flow’ as a feature reduced significantly the standard deviation values, with comparison to the cases where it is not considered. For both cases and introducing RFL-FLOW-PROC features when the ANNs were trained, the standard deviation values after 300 seconds improved by more than 25% compared to the BADA ones.

Figure 4-9 Mean and RMSE values of the residuals for different segmentation groups for ANNs including them as inputs for the testing set.

Then, trained neural networks have been tested with the testing sample. Results are shown in Figure 4-9, For the B738 case, the bias plot shows that residual biases are maintained close to zero along the 300 seconds after the TOD, similarly to the results obtained for the training set. The standard deviation plot shows that groupings that include the flow information maintained an improved performance compared with the other two, similarly to the training sample.

For the A320 case, biases reproduce the B738 case, except the group which includes the company as a feature. The standard deviation plot reproduces the result for the B738 case, where it can be observed that adding ‘Procedure’ as a feature does not impact remarkably on a reduction of the standard deviation.

4.4. Discussion

Results for ANNs trajectory predictors have outperformed PMM ones, removing completely the bias in the predictions and reducing by more than 25% the standard deviation after 300 seconds. Table 4-6 and Table 4-7 summarise results plotted in Figure 4-7 and Figure 4-9.

Table 4-6 Mean and RMSE values for residuals for B738 for the training and testing samples

Aircraft Model B738		Training Set		Testing Set	
Model	Grouping	Mean of residuals (Pressure Altitude) [100ft]	RMSE of residuals (Pressure Altitude) [100ft]	Mean of residuals (Pressure Altitude) [100ft]	RMSE of residuals (Pressure Altitude) [100ft]
BADA	Complete Sample	13.935	20.921	13.608	20.443
PMM	RFL	-0.365	13.842	-0.353	13.663
PMM	RFL-FLOW	-0.333	13.812	-0.401	13.801
PMM	RFL-COM	-0.338	13.778	-0.215	13.655
PMM	RFL-FLOW-PROC	-0.407	13.347	-0.465	13.566
ANN	RFL'	-0.748	12.331	-0.192	11.596
ANN	RFL-FLOW	-0.710	10.314	-0.872	9.737
ANN	RFL-COM	-0.728	12.631	0.122	12.198
ANN	RFL-FLOW-PROC	-0.763	10.306	-0.933	9.711

B738 ANNs have performed superior than A320 ones. These results could be related to the aircraft model distribution per flow and sample size used for training. In the case of the B738, the distribution was consistent among the three studied flows (flow 3: 44.84% of flights, flow 6: 40.94% of flights, flow 7: 48.47% of flights), but four times larger. On the case of the A320, the frequency was not homogenously distributed among flows (Flow 3: 5.73% of flights, Flow 6: 13.17% of flights, Flow 7: 23.24% of flights). Indeed, the A320s were more frequent in flows that had larger variances in their behaviours, as it can be seen in Figure 4-3.

Table 4-7 Mean and RMSE values for residuals for A320 for the training and testing samples.

Aircraft Model A320		Training Set		Testing Set	
Model	Grouping	Mean of residuals (Pressure Altitude) [100ft]	RMSE of residuals (Pressure Altitude) [100ft]	Mean of residuals (Pressure Altitude) [100ft]	RMSE of residuals (Pressure Altitude) [100ft]
BADA	Complete Sample	16.215	25.230	15.197	24.810
PMM	RFL	-2.879	17.961	-3.591	18.270
PMM	RFL-FLOW	-2.859	17.952	-3.905	18.305
PMM	RFL-COM	-2.640	16.709	-4.379	18.624
PMM	RFL-FLOW-PROC	-2.679	16.701	-4.518	17.964
ANN	RFL'	-0.844	14.910	-1.0612	15.040
ANN	RFL-FLOW	-0.823	13.119	-1.819	13.274
ANN	RFL-COM	-0.835	13.049	6.798	23.171
ANN	RFL-FLOW-PROC	-0.796	12.168	-2.559	13.326

Figure 15 includes the prediction results only considering the PUMAL-LORES flow, for both aircraft. The behaviour of the predictor is consistent with the results presented in Figure 4-8, although it is observed a worse performance for both aircraft in terms of the RMSE, which has been used instead of the standard deviation for this case. Results are also consistent with findings presented in Figure 4-3 and Table 4-4, where this flow presented a larger variance on the results.

Figure 4-10 Mean and RMSE values of the residuals for different segmentation groups for ANNs including them as inputs, only for Flow PUMAL-LOREL

Results are summarised in Table 4-6 and Table 4-7. The RMSE values for the B738 for the testing sample, disregarding time, remain between 9.711 FL (RFL-FLOW) and 12.198 FL (RFL). It can be observed that the predictor performed worse for the RFL-COM case than for the RFL-Flow one, for both aircraft. This fact supports the notion that including the airline as a relevant factor for predicting the vertical profile of the aircraft is less relevant than including the flow feature, for the ANN case.

About the operational procedure, that was the main objective of this piece of work, its inclusion as a relevant feature for training the ANN does not suppose a relevant improvement in the prediction. Therefore, the analysis concludes that it is not recommendable to include this factor on the trajectory predictors, at least in the form of aggregating flights by number of received CFLs. In addition, the operational procedure has been considered as known, but it is partially unknown at the TOD, so further work would have to be carried out to predict how many CFLs a flight was about to receive, which finally is not impacting relevantly on the prediction. On the other hand, it could be considered to use as an input the difference in FL between the TOD and the first CFL, to carry out predictions for given CFLs.

In the case of the PMM models and learnt thrust laws, RMSE values show that the predictor maintained the performance from the training dataset to the testing one. Although the RMSE in the prediction remained higher than for ANNs predictors, the bias was corrected to the same extent. Results show that the data-driven process has outperformed the PMM in absolute numbers, but the PMM has demonstrated less sensitivity to the use of new samples.

It has been observed how the bias was almost removed. The changing trend after 200 seconds may be explained due to the switch from constant Mach profile to constant CAS. Two thirds of the sample used for obtaining the coefficients correspond to the Mach segment, in average, which could have driven the coefficients to reproduce more accurately this segment.

It is considered that other types of functions for the thrust coefficients may improve the prediction, together with more refined thrust and aerodynamic drag laws. Further steps in this direction should include weather data and an augmented knowledge about the thrust, drag and ESF laws.

4.5. Summary of Findings

The objective of this section was to determine the impact of operational factors on the vertical profile of the flight trajectory, to determine whether considered factors should be included or not in the parameter estimation process for trajectory predictors based on point-mass models or as inputs in pure data-driven approaches.

First, a set of Multilevel Linear Models was adopted to analyse the sensitivity of the rate of descent to the operational vertical procedures, flows, and the airline. The set of MLMs worked as multiple linear regressions, in which the coefficients associated to each independent variable were trained for different groups. The dependent variable was chosen to be a linear estimate of the rate of descent from the TOD to FL250. This response was then characterised as a function of the distance to the airport at the top of descent, and the magnitude of the ground velocity at that point.

Results have shown that there was a connection between the LRD and the position of the TOD. In cases that the TOD was farther from the airport, the response was almost linear. It was shown that the FL at the TOD impacts on the LRD as well.

The flow feature represents an important factor explaining the variance in this model. In addition, the operational procedure enabled the identification of flights with a greater variance, but once removed those, it did not represent a fundamental factor increasing the explained variance by the model, as the information that it added was related to the position of the TOD. The airline factor impacted on the amount of variance that the model could explain. However, it was demonstrated that it was not consistent when a sample from another year is tested.

Identified main factors were applied in point-mass models and neural network trajectory predictors for the pressure altitude. The predictions were made on the two most frequent

aircraft models that were present in the sample. ANNs predictors outperformed PMMs, removing the bias and maintaining the RMSE along time stable, although PMMs were more resilient against a testing set.

Conclusions from the MLM analysis were validated, identifying the flow factor and the position of the TOD as key inputs for trajectory predictors. The introduction of the company as a factor in the ANN case did not improve the performance of the predictor. On the contrary, inclusion of the procedures augmented the variance explained, but not significantly.

This conclusions led to further exploration of traffic contextual factors to predict the position of the TOD, as it constitutes a fundamental factor to predict the vertical profile of the trajectory.

CHAPTER 5.

SURROUNDING TRAFFIC AS A CONTEXTUAL FACTOR

5.1. Introduction²⁰

Chapter 3 developed a data model and detailed the process for automatic flow discovery. Chapter 4 identified that the flow factor and the position of the TOD along the route were key inputs for vertical trajectory predictors. The research question Q3 queried whether the surrounding traffic could be included as a factor in the definition of the aircraft intent.

²⁰ Methodology, results and conclusions of this chapter have been published in [159] (C. E. Verdonk Gallego, V. F. Gómez Comendador, M. A. Amaro Carmona, R. M. Arnaldo Valdés, F. J. Sáez Nieto, and M. García Martínez, “A machine learning approach to air traffic interdependency modelling and its application to trajectory prediction,” *Transp. Res. Part C Emerg. Technol.*, vol. 107, no. October 2019, pp. 356–386, Oct. 2019).

Figure 5-1 Knowledge Discovery Process - Scope of Chapter 5

Figure 5-1 establishes the framework for this chapter within the thesis. Thus, the objective that is pursued in this section is to introduce surrounding air traffic as a contextual factor for vertical trajectory predictors.

The methodology consists of trying to identify interdependencies between aircraft, trying to identify a causal factor which will lead ATC to modify the aircraft trajectory. Firstly, the chapter defines an interdependency measure between two aircraft as a probability measure depending on the performance of associated trajectory predictors. Secondly, that measure is included as a factor in a vertical trajectory predictor. The methodology approaches the problem by integrating model- and data-driven paradigms, based on ML techniques already discussed and presented in the previous section for trajectory computation and a statistic-based interdependency definition.

5.1.1. Surrounding Air Traffic as a contextual factor

The main objective of this work is to introduce air traffic as a relevant factor for the trajectory prediction. The interaction between a pair of aircraft will be considered as a feature for predictors depending on the value of an interdependency measure.

The definition of the interdependency between a pair of aircraft is a trending research topic. When this interdependency is generalised to more than two aircraft, it can be associated with air traffic complexity (see Section 2.4.1.4). Approaches vary between algorithmic ones [109], [128]–[133] and human-factors based ones [134]–[136].

Interdependency relates to potential conflicts between aircraft. Probabilistic conflict detection has been addressed in past and recent research. [137] proposed a probabilistic approach for conflict detection, which was later extended to include errors in the trajectory prediction [138]. There have been numerous research works that have extended or modified this approach, such as [139] or [140]. [95] conducted an extensive literature review about probabilistic approaches to conflict detection, including a detailed explanation of the different approaches to model the position errors. It also included an extensive review of a different approach to probabilistic conflict detection, based on the propagation of the uncertainty from a source, such as wind conditions, to the measures which are used to identify conflicts.

These approaches look for identifying when the separation minima between a pair of aircraft will be infringed. The common methodology is based on propagating the trajectory within a given look-ahead time and determining if two aircraft will lose separation by assessing the probability of their relative distance being below a given threshold for each value of the time series. More recently, [141] presented a novel approach to conflict detection and resolution in a strategic phase. It defined *conflict detection* in terms of the distance between aircraft and its comparison with the separation minima, but with a novel technique where trajectories were aligned with a 3D grid network, easing the process and enabling its scalation.

The work herein described presents two main differences with the previous approaches. Firstly, the interdependency measure is not related to distance-based measures, but to the time separation on arriving to a pre-identified location, where the separation minima between them will be surely infringed if they arrive concurrently.

This approach is based on ATCo cognitive abstractions [87], where ATCOs concentrate on critical points for controlling the air traffic. Secondly, this work addresses the probabilistic definition of the interdependency between two aircraft considering TPs' performance instead of deviations from the actual trajectories. This approach modifies [138] and others [142] by considering the uncertainty related to the temporal instead the spatial dimension. This approach is based on ATCo cognitive abstractions

Trajectory predictors will be based on ML techniques influenced by ATCo cognitive abstractions [87]. Thus, the approach followed in the previous chapter is generalised to every flow within the

airspace volume of interest. This work follows the “major flow” definition given by [143] and introduced in Section 3.4.1. Then, a probability measure of the horizontal interdependency between two aircraft is provided depending on the performance of the trajectory predictors.

The second part of the chapter exploits this horizontal interdependency measure. Firstly, vertical profiles and TOD locations are characterised depending on the interdependency measure. Secondly, this characterisation enables the introduction of this new feature into a trajectory predictor for the vertical profile of the aircraft trajectory.

Figure 5-2 introduces the logical breakdown of the trajectory predictor and how the interdependency is introduced as a relevant factor for the intent generation (TOD location) and the trajectory computation.

Figure 5-2 Trajectory Predictor Logical Breakdown

5.2. Methodology

5.2.1. Interdependency Modelling

5.2.1.1. Expected Trajectory based on Observed Traffic

A flown trajectory, η , can be described as a mapping η from a time interval $[a,b] \in \mathbb{R}$ to a state space E , with E being $\mathbb{K}^6 \times \mathbb{O}^4 \in \mathbb{R}^{10}$. The former ($\mathbb{K}^6$) corresponds to the kinematic state of the aircraft (position and velocity) as seen by the ATM ground system, whereas the latter ($\mathbb{O}^4$) reflects the operational state of the aircraft (cleared heading, cleared flight level, cleared horizontal speed, and cleared vertical speed, if any),

$$\boldsymbol{\eta}(t) = \{\boldsymbol{lat}(t), \boldsymbol{lng}(t), \boldsymbol{FL}(t), \boldsymbol{Hdg}(t), \boldsymbol{Vg}(t), \boldsymbol{Vz}(t), \boldsymbol{\varpi}(t), \boldsymbol{v}(t), \boldsymbol{\zeta}(t), \boldsymbol{q}(t)\}, \quad (13)$$

where $\boldsymbol{\varpi}, \boldsymbol{\zeta}, \boldsymbol{v}$ and $\boldsymbol{q}$ corresponds to the cleared heading, cleared flight level, horizontal speed and cleared vertical speed respectively.

Let us consider an initial sample of trajectories $\Lambda^0 = \{\eta_1(t), \dots, \eta_r(t), \dots, \eta_s(t), \dots, \eta_N(t)\}$, being N the sample size, which have flown through a given airspace. Such a set can be clustered by following the methodology presented in Section 3.4 or by any other clustering methodology. The proposed implementation allows a robust identification of the major flows within a given dataset. Thus, Λ^0 can be expressed as

$$\Lambda^0 = \bigcup_{i=1}^I \lambda_i^0, \quad (14)$$

where λ_i^0 corresponds to the major flow i , of a total of I flows. Each airspace flow λ_i^0 can be decomposed into different recurrent patterns by applying again a clustering methodology. Therefore, a major flow decomposition can be expressed as:

$$\lambda_i^0 = \bigcup_{l=1}^p \alpha_{il}^0. \quad (15)$$

where α_{il}^0 identifies a recurrent flow pattern l within a major flow i . α_{il}^0 could represent diverse pattern types, which are enumerated as follows:

1. A set of trajectories following a standard route (**Standard Flow**).
2. A set of trajectories following a recurrent pattern which differs from the standard route (**Recurrent Pattern**).
3. A set of trajectories which is not classified within categories 1 or 2.

An example of the outcomes of Gariel's Methodology and RDBSCAN* is introduced in Figure 5-3.

Figure 5-3 Main Flows – LECBCCC Sector, Barcelona ATCC and recurrent patterns associated to the main flows.

For each recurrent pattern α_{il}^0 , it can be defined a set of associated elements which characterises the normal horizontal behaviour of a flight within that pattern. Each α_{il}^0 has associated a horizontal central trajectory, $\mathbf{h}_{il}^0$. This element is formed by an ordered sequence of latitude-longitude tuples, where two consecutive tuples define a segment. Each segment length is bounded and the total horizontal central trajectory length, L_{il} , is equal to the sum of its segment lengths.

A horizontal central trajectory $\mathbf{h}_{il}^0$ is associated with a function $\mathcal{H}$ which maps the initial state of an aircraft to an offset with respect to the starting location of $\mathbf{h}_{il}^0$, assuming that the initial segment is linear. The domain of the function for the latitude and longitude is limited to the surroundings of the initial point of $\mathbf{h}_{il}^0$. Thus, $\mathcal{H}$ is defined as:

$$\begin{aligned} \mathcal{H}: \mathbb{R}^4 &\rightarrow \mathbb{R} \\ (lat, lng, Vx, Vy) &\rightarrow \mathcal{H}(lat, lng, Vx, Vy) \end{aligned} \tag{16}$$

In addition, we can define $\mathbf{h}_{il}^0$ in terms of its arc-length parameter γ_{il} [143]. The origin of $\mathbf{h}_{il}^0$ is $\gamma_{il} = 0$, and the end point corresponds to $\gamma_{il} = L_{il}$, therefore, the arc-length parameter space is Γ_{il} . This space varies for each recurrent pattern α_{il}^0 , and it ranges from 0 to L_{il} .

This function computes first the position offset with respect to $\gamma_{il} = 0$, by considering the horizontal projection of the initial state (position and velocity) of the trajectory on $\mathbf{h}_{il}^0$. Then, $\mathcal{H}$ translates the specific trajectory arc-length parameter γ to γ_{il} , considering the offset

previously computed. This is based on that the assumption that the clustering methodology would have identified group of trajectories which a high degree of similarity.

The identification of $\mathbf{h}_{il}^0$ can be based on the Principal Component Analysis (PCA) decomposition which is used in the clustering methodology (see Section 3.4.2.2). Thus, the transformed coordinates corresponding to the nearest point to the median principal components for a given α_{il}^0 is considered as the $\mathbf{h}_{il}^0$ of the recurrent pattern.

Let us assume that the clusterisation step has been able to isolate each recurrent pattern. Thus, the previous process can be applied to each trajectory η_r and then (13) is collapsed to

$$\eta_r(\gamma_{il}) \sim \{\gamma_{il}, t(\gamma_{il}), Vg(\gamma_{il}), FL(\gamma_{il}), Vz(\gamma_{il}), \varpi(\gamma_{il}), v(\gamma_{il}), \zeta(\gamma_{il}), \varrho(\gamma_{il})\}, \quad (17)$$

as there is a bijection between each pair latitude-longitude and the time t . In addition, there is a bijective projection between each pair latitude-longitude and the arc-length parameter space, Γ_{il} . Hence, the composition of these two functions provides a bijective relation between γ_{il} and t , specific for each trajectory. That relation can be generated for every trajectory in α_{il}^0 , and therefore, characterising them in terms of γ_{il} .

Once established this common parameter, the next step is to generate a TP tailored to each pattern α_{ij}^0 . In this chapter, the TP objective is to predict the time τ to overfly a location $(\gamma_{il}; \underline{X})$, where $\underline{X}$ is an input vector accounting for relevant aircraft intrinsic features. The TP's performance can be characterised as a function of γ_{il} , by observing the distribution of the residuals for the time prediction along $\mathbf{h}_{il}^0$.

Several TP approaches may be applied, such as ML-enhanced point-mass models TP [5], [6] or data-driven approaches [7], [79], [144]. In this work, it is applied ANNs as trajectory computation engines, following the work of the previous section.

The input vector for an ANN associated to α_{il}^0 is formed by γ_{il} , and a set of relevant features of the trajectory ($\underline{X}$). These features may be composed by the Requested Flight Level (RFL), the ground velocity (Vg) as measured by the ground ATM system, the wake vortex, the aircraft model, the wind speed and direction, the pressure, the temperature or any other relevant feature, which will depend on the available data. This process is introduced in Figure 5-4. The output layer corresponds to the predicted time τ .

Figure 5-4 Artificial Neural Net Predictor for estimating the time (τ) of overflying γ_{il} .

5.2.1.2. Crossing Point Interdependency Definition

When an aircraft is about to enter an airspace volume, the intention of the ATCo is fuzzily defined, and it depends on contextual factors of the trajectory, such as weather or surrounding aircraft. Approaches to tackle weather-based factors have been conducted at tactical [91] and strategic level [145], [146]. This work addresses the interdependency definition for the latter case, focusing on the definition of a probabilistic measure for the horizontal time separation of two aircraft at a crossing point.

Each flown trajectory η_r within a recurrent pattern α_{ij}^0 is associated with a flight r . For each flight, a set of potential trajectories can be generated when it enters the airspace volume of interest. These potential trajectories can be generated by means of the recurrent patterns α_{il}^0 associated to the parent flow λ_i^0 , which may fall into the Standard Flow or recurrent pattern categories.

For example, if a major flow λ_i^0 has associated three recurrent patterns, $\{\alpha_{il}^0, \alpha_{im}^0, \alpha_{in}^0\}$, it can be introduced a set of potential trajectories for each flight:

$$\widetilde{\eta}_{r,1} = \{\widetilde{\eta}_{r,il}(\gamma_{il}), \widetilde{\eta}_{r,im}(\gamma_{im}), \widetilde{\eta}_{r,in}(\gamma_{in})\}, \quad (18)$$

Each recurrent pattern, $\{\alpha_{il}^0, \alpha_{im}^0, \alpha_{in}^0\}$, has been associated with a horizontal central trajectory, $(\mathbf{h}_{il}^0, \mathbf{h}_{im}^0, \mathbf{h}_{in}^0)$, and a trajectory predictor. In addition, the performance of each trajectory predictor is characterised along each central trajectory by observing the distribution of the residuals of the overflying time predictions for a previous sample as a function of γ_{il} .

Now, the predicted time τ to arrive to a given γ_{il} as $\tau_{r,il}(\gamma_{il})$ can be defined in terms of the residual distribution of the specific TP. If now $\mathcal{T}_{r,il}(\gamma_{il})$ is considered as a random variable, defined as

$$\begin{aligned}\mathcal{T}_{r,il}(\gamma_{il}): \Omega &\rightarrow \mathbb{R} \\ \mathcal{T}_{r,il}(\gamma_{il}) &\sim ANN_{il}(\gamma_{il}; \underline{X}) + \varepsilon_{il}(\gamma_{il})\end{aligned}\tag{19}$$

where Ω is endowed with a sigma field and probability, the interdependency between two aircraft of different major flows λ_i^0 and λ_j^0 can be defined as follows.

Let us consider the definition of *Critical Point* as given by [87]. Citing this work, “a critical point is a point in the airspace with a high concentration of traffic”. Different examples that are covered by this definition are crossing points or merge points.

Let us define a crossing point as the common point between two patterns α_{il}^0 and α_{jm}^0 belonging to major flows λ_i^0 and λ_j^0 respectively. Given two discrete sequences for the arc-length parameters along $\mathbf{h}_{il}^0$ and $\mathbf{h}_{jm}^0$ respectively, it can be identified two sequences of random variables associated to $\widetilde{\eta_{r,il}}$ and $\widetilde{\eta_{s,jm}}$, given by:

$$\begin{aligned}\widetilde{\eta_{r,il}}: \{\mathcal{T}_{r,il}(\mathbf{0}), \dots, \mathcal{T}_{r,il}(\gamma_{il}), \dots, \mathcal{T}_{r,il}(L_{il})\} &= \{\mathcal{T}_{il}^0, \dots, \mathcal{T}_{il}^{\gamma_{il}}, \dots, \mathcal{T}_{il}^{L_{il}}\} \\ \widetilde{\eta_{s,jm}}: \{\mathcal{T}_{s,jm}(\mathbf{0}), \dots, \mathcal{T}_{s,jm}(\gamma_{il}), \dots, \mathcal{T}_{s,jm}(L_{il})\} &= \{\mathcal{T}_{jm}^0, \dots, \mathcal{T}_{jm}^{\gamma_{jm}}, \dots, \mathcal{T}_{jm}^{L_{jm}}\}\end{aligned}\tag{20}$$

Let us assume that every residual distribution $\varepsilon_{il}(\gamma_{il})$ and $\varepsilon_{jm}(\gamma_{jm})$ are normally distributed without bias:

$$\begin{aligned}\varepsilon_{il}(\gamma_{il}) &= \varepsilon_{il}^{\gamma_{il}} \sim N(\mathbf{0}, (\sigma_{il}^{\gamma_{il}})^2) \\ \varepsilon_{jm}(\gamma_{jm}) &= \varepsilon_{jm}^{\gamma_{jm}} \sim N(\mathbf{0}, (\sigma_{jm}^{\gamma_{jm}})^2).\end{aligned}\tag{21}$$

Then, if it is associated the value of $ANN_{il}(\gamma_{il}; \underline{X})$ to the expectation of $\mathcal{T}_{il}^{\gamma_{il}}$, it can be identified $\mathcal{T}_{il}^{\gamma_{il}}$ with a normal distribution with parameters $\mu_{il}^{\gamma_{il}}$ and $\sigma_{il}^{\gamma_{il}}$:

$$\begin{aligned}\mathcal{T}_{il}^{\gamma_{il}} &\sim N(\mu_{il}^{\gamma_{il}}, (\sigma_{il}^{\gamma_{il}})^2), \\ \mu_{il}^{\gamma_{il}} &= ANN_{il}(\gamma_{il}; \overline{X}),\end{aligned}\tag{22}$$

where $\sigma_{il}^{\gamma_{il}}$ corresponds to the standard deviation of the residuals $\varepsilon_{il}^{\gamma_{il}}$. The same logic can be applied for each $\mathcal{T}_{jm}^{\gamma_{jm}}$.

Now a random variable for a given γ_{il} and a γ_{jm} can be defined such that the time separation at γ_{il} and γ_{jm} , $\Delta\mathcal{T}_{il,jm}^{\gamma_{il},\gamma_{jm}}$, is defined as:

$$\Delta\mathcal{T}(\gamma_{il}, \gamma_{jm}) = \Delta\mathcal{T}_{il,jm}^{\gamma_{il},\gamma_{jm}} = |\mathcal{T}_{il}^{\gamma_{il}} - \mathcal{T}_{jm}^{\gamma_{jm}}|. \quad (23)$$

As $\tau_{il}^{\gamma_{il}}$ and $\tau_{jm}^{\gamma_{jm}}$ are assumed to be normally distributed, their difference will also be normally distributed. This difference can be expressed as

$$\mathcal{T}_{il}^{\gamma_{il}} - \mathcal{T}_{jm}^{\gamma_{jm}} \sim N\left(\mu_{il}^{\gamma_{il}} - \mu_{jm}^{\gamma_{jm}}, (\sigma_{il}^{\gamma_{il}})^2 + (\sigma_{jm}^{\gamma_{jm}})^2\right). \quad (24)$$

A new random variable Z can be defined so that:

$$Z = \frac{(\mathcal{T}_{il}^{\gamma_{il}} - \mathcal{T}_{jm}^{\gamma_{jm}}) - (\mu_{il}^{\gamma_{il}} - \mu_{jm}^{\gamma_{jm}})}{\sqrt{(\sigma_{il}^{\gamma_{il}})^2 + (\sigma_{jm}^{\gamma_{jm}})^2}} \quad (25)$$

Therefore,

$$\begin{aligned} \mathcal{T}_{il}^{\gamma_{il}} - \mathcal{T}_{jm}^{\gamma_{jm}} &= \sigma_Z \left(Z + \frac{\mu_Z}{\sigma_Z} \right), \\ \mu_Z &= \mu_{il}^{\gamma_{il}} - \mu_{jm}^{\gamma_{jm}}, \\ \sigma_Z^2 &= (\sigma_{il}^{\gamma_{il}})^2 + (\sigma_{jm}^{\gamma_{jm}})^2. \end{aligned} \quad (26)$$

Following the previous equation, the random variable $\Delta\mathcal{T}_{il,jm}^{\gamma_{il},\gamma_{jm}}$ may be expressed as:

$$\Delta\mathcal{T}_{il,jm}^{\gamma_{il},\gamma_{jm}} \sim |\mathcal{T}_{il}^{\gamma_{il}} - \mathcal{T}_{jm}^{\gamma_{jm}}| \sim \sqrt{(\mathcal{T}_{il}^{\gamma_{il}} - \mathcal{T}_{jm}^{\gamma_{jm}})^2} = \sigma_Z \sqrt{\left(Z + \frac{\mu_Z}{\sigma_Z}\right)^2} \quad (27)$$

Thus, $\Delta\mathcal{T}_{il,jm}^{\gamma_{il},\gamma_{jm}}$ corresponds to a scaled version of the square root of a non-central Chi-Squared distribution with one degree of freedom, being the non-centrality parameter $\lambda = \left(\frac{\mu_Z}{\sigma_Z}\right)^2$. The elementary probability of such a non-central chi-squared distribution can be expressed as:

$$f(y)dy = \frac{\sqrt{y}}{\sqrt{2\pi}} e^{\frac{1}{2}(-\lambda-y)} \cosh(\sqrt{\lambda y}) \frac{dy}{y}, y > 0 \quad (28)$$

Substituting y by z^2 , then:

$$f(y)dy = f(z^2)d(z^2) = \frac{\sqrt{z^2}}{\sqrt{2\pi}} e^{\frac{1}{2}(-\lambda - z^2)} \cosh(\sqrt{\lambda z^2}) \frac{dz^2}{z^2} \quad (29)$$

Simplifying and rescaling by σ_Z provides the probability density function (PDF) of $\Delta\mathcal{T}_{il,jm}^{\gamma_{il},\gamma_{jm}}$:

$$f_{\Delta\mathcal{T}_{il,jm}^{\gamma_{il},\gamma_{jm}}}(\Delta\tau_{il,jm}^{\gamma_{il},\gamma_{jm}}) = \frac{1}{\sigma_Z} \sqrt{\frac{2}{\pi}} \cosh\left(\frac{\Delta\tau_{il,jm}^{\gamma_{il},\gamma_{jm}} \cdot \mu_Z}{\sigma_Z^2}\right) e^{\left(-\frac{(\Delta\tau_{il,jm}^{\gamma_{il},\gamma_{jm}})^2 + \mu_Z^2}{2\sigma_Z^2}\right)} \quad (30)$$

Once this PDF has been obtained, it can be derived the interdependency relationship for a crossing point for two patterns α_{il}^0 and α_{jm}^0 . Figure 5-5 represents two recurrent patterns and the associated crossing point. α_{il}^0 is represented in green, whereas α_{jm}^0 is represented in orange. The relative distance between two aircraft if they were at γ_{il} and γ_{jm} respectively can be computed for any combination of γ_{il} and γ_{jm} . Let us declare a function D to define this distance:

$$\begin{aligned} D_{il,jm}: \mathbb{R} \times \mathbb{R} &\rightarrow \mathbb{R} \\ D_{il,jm}(\gamma_{il}, \gamma_{jm}) &= \text{Haversine}(\mathcal{H}_{il}^{-1}(\gamma_{il}), \mathcal{H}_{jm}^{-1}(\gamma_{jm}), \\ \gamma_{il} &\in [0, L_{il}], \gamma_{jm} \in [0, L_{jm}] \end{aligned} \quad (31)$$

It can be obtained γ_{il}^{CP} and γ_{jm}^{CP} such that $D_{il,jm}(\gamma_{il}^{CP}, \gamma_{jm}^{CP}) = 0$. It can also be obtained all the combinations of γ_{il} and γ_{jm} such that $D_{il,jm}(\gamma_{il}, \gamma_{jm})$ is below a given threshold.

Figure 5-5 Schematic representation of two recurrent patterns and a crossing point

Following [87]'s approach of only focusing on critical points, if we consider now γ_{il}^{CP} and γ_{jm}^{CP} , it can be observed that $\Delta\tau_{il,jm}^{\gamma_{il}^{CP}, \gamma_{jm}^{CP}}$ only approximates zero if $\gamma_{il}^{CP} \sim \gamma_{jm}^{CP}$, assuming similar aircraft dynamics for both patterns. In any other case, it is necessary to conduct a translation over one or both flows (*an offset*), to find the arc-length reference value that enables $\gamma_{il}^{CP} \sim \gamma_{jm}^{CP}$. If we introduce now $\gamma_{jm}^* = \gamma_{jm} - \gamma_{jm}^{offset}$, then it can be redefined

$$\mathcal{T}_{il}^{\gamma_{il}} - \mathcal{T}_{jm}^{\gamma_{jm}^*} \sim N\left(\mu_{il}^{\gamma_{il}} - \mu_{jm}^{\gamma_{jm}} + \mu_{jm}^{\gamma_{jm}^{offset}}, (\sigma_{il}^{\gamma_{il}})^2 + (\sigma_{jm}^{\gamma_{jm}})^2 + (\sigma_{jm}^{\gamma_{jm}^{offset}})^2\right), \quad (32)$$

, and then

$$\begin{aligned}
 f_{\Delta \mathcal{T}_{il,jm}^{\gamma_{il},\gamma_{jm}^*}}(\Delta \tau_{il,jm}^{\gamma_{il},\gamma_{jm}^*}) &= \frac{1}{\sigma_Z^*} \sqrt{\frac{2}{\pi}} \cosh\left(\frac{\Delta \tau_{il,jm}^{\gamma_{il},\gamma_{jm}^*} \cdot \mu_Z^*}{\sigma_Z^{*2}}\right) e^{-\frac{((\Delta \tau_{il,jm}^{\gamma_{il},\gamma_{jm}^*})^2 + \mu_Z^{*2})}{2\sigma_Z^{*2}}} \\
 \mu_Z^* &= \mu_{il}^{\gamma_{il}} - \mu_{jm}^{\gamma_{jm}} + \mu_{jm}^{\gamma_{jm}^{offset}} \\
 \sigma_Z^{*2} &= (\sigma_{il}^{\gamma_{il}})^2 + (\sigma_{jm}^{\gamma_{jm}})^2 + (\sigma_{jm}^{\gamma_{jm}^{offset}})^2
 \end{aligned} \tag{33}$$

Then, for each pair of potential trajectories $\widetilde{\eta_{r,il}}$ and $\widetilde{\eta_{s,jm}}$, their horizontal interdependency can be defined as a function of time and distance. Thus, it can be measured the probability of these two aircraft arriving to γ_{il} and γ_{jm}^* within a time separation of Δt_{Thres} seconds if they were following patterns α_{il}^0 and α_{jm}^0 .

$$\begin{aligned}
 I(\Delta t, \gamma_{il}, \gamma_{jm}^*) &= P(\Delta \mathcal{T}_{il,jm}^{\gamma_{il},\gamma_{jm}^*} \leq \Delta t | \gamma_{il}, \gamma_{jm}^*) = \int_{-\infty}^{\Delta t_{Thres}} (f_{\Delta \mathcal{T}_{il,jm}^{\gamma_{il},\gamma_{jm}^*}}) d(\Delta \tau_{il,jm}^{\gamma_{il},\gamma_{jm}^*}) \\
 &= \int_0^{\Delta t_{Thres}} f_{\Delta \mathcal{T}_{il,jm}^{\gamma_{il},\gamma_{jm}^*}} d(\Delta \tau_{il,jm}^{\gamma_{il},\gamma_{jm}^*}) \\
 &= \frac{1}{2} \left(\text{erf}\left(\frac{\Delta t_{Thres} - \mu_Z^*}{\sqrt{2}\sigma_Z^*}\right) + \text{erf}\left(\frac{\Delta t_{Thres} + \mu_Z^*}{\sqrt{2}\sigma_Z^*}\right) \right),
 \end{aligned} \tag{34}$$

where erf is the function error. The interdependency calculation may be generalised for the set of γ_{il} and γ_{jm}^* so that $D_{il,jm}(\gamma_{il}, \gamma_{jm}^*) \leq D_H$. So, if the intervals for both arc-length parameters that meet that conditions are $[\gamma_{il}^{min}, \gamma_{il}^{max}]$ and $[\gamma_{jm}^{min}, \gamma_{jm}^{max}]$, the generalised horizontal spatiotemporal interdependency measure for two aircraft r and s will be defined as:

$$\begin{aligned}
 \Pi_{\Delta t_{Thres}, D_H}^{il,jm}(\gamma_{il}^{min}, \gamma_{il}^{max}, \gamma_{jm}^{*,min}, \gamma_{jm}^{*,max}) &= \frac{1}{B(\gamma_{il}^{min}, \gamma_{il}^{max}, \gamma_{jm}^{*,min}, \gamma_{jm}^{*,max})} \cdot \\
 &\int_{\gamma_{jm}^{*,min}}^{\gamma_{jm}^{*,max}} \int_{\gamma_{il}^{min}}^{\gamma_{il}^{max}} bin(\gamma_{il}, \gamma_{jm}^*, D_H) \cdot I(\Delta t_{Thres}, \gamma_{il}, \gamma_{jm}^*) d\gamma_{il} d\gamma_{jm}^*
 \end{aligned} \tag{35}$$

Where $bin(\gamma_{il}, \gamma_{jm}^*, D_H)$ is a binary function which equals 1 or 0 depending on whether $D_{il,jm}(\gamma_{il}, \gamma_{jm}^*) \leq D_H$ or not, whereas $B(\gamma_{il}^{min}, \gamma_{il}^{max}, \gamma_{jm}^{*,min}, \gamma_{jm}^{*,max})$ is the sum of $bin(\gamma_{il}, \gamma_{jm}^*, D_H)$ over the intervals $[\gamma_{il}^{min}, \gamma_{il}^{max}]$ and $[\gamma_{jm}^{min}, \gamma_{jm}^{max}]$. The introduction of these two functions is for guaranteeing that Π_H is defined between 0 and 1, as there may be values of γ_{il} and γ_{jm}^* within those intervals for which $bin(\gamma_{il}, \gamma_{jm}^*, D_H)$ equal 0.

5.2.2. Interdependency Determination for two flights

Now, if there are two flights (r and s) associated to major flows λ_i^0 and λ_j^0 respectively, their expected trajectories can be generated for the different recurrent patterns. Following Eq. (18), expected trajectories for aircraft r and s would be defined by $\widetilde{\eta}_{r,1}$ and $\widetilde{\eta}_{s,j}$ respectively.

Then, it can be obtained the arc-length parameter intervals for which the distance between the pair of aircraft would be less than a given threshold for each combination of recurrent patterns $\{\alpha_{il}^0, \alpha_{jm}^0\}$. Thus, given Δt and D_H , it can be obtained a probability measure for the horizontal interdependency, $\Pi_{\Delta t, D_H}^{il, jm}$, for each combination of expected trajectories $\widetilde{\eta}_{r,u}$ and $\widetilde{\eta}_{s,jm}$.

Now, for each pair of flights, if λ_i^0 and λ_j^0 can be decomposed in p and o recurrent patterns respectively, a horizontal interdependency matrix (HIM) can be defined as

$$HIM_{r,s} = \begin{bmatrix} \Pi_{\Delta t, D_H}^{i1,j1} & \dots & \dots & \dots & \Pi_{\Delta t, D_H}^{i1,jp} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ \Pi_{\Delta t, D_H}^{il,j1} & \vdots & \Pi_{\Delta t, D_H}^{il,jm} & \vdots & \Pi_{\Delta t, D_H}^{il,jp} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \Pi_{\Delta t, D_H}^{ip,jp} & \dots & \dots & \dots & \Pi_{\Delta t, D_H}^{ip,jp} \end{bmatrix}. \quad (36)$$

Figure 5-6 introduces a schematic representation of this process. If λ_1^0 is associated with the flow represented in green and λ_2^0 with the orange one, two $\{\alpha_{11}^0, \alpha_{12}^0\}$ and three $\{\alpha_{21}^0, \alpha_{22}^0, \alpha_{23}^0\}$ associated recurrent patterns can be identified, respectively. Each of these patterns has a horizontal TP associated, with an associated residual distribution as a function of its arc-length parameter.

Figure 5-6 Process for the generation of the Horizontal Interdependency Matrix

Assuming that the total length of the recurrent patterns associated to λ_2^0 to be larger than those associated to λ_1^0 , when an aircraft r , represented in green, enters the volume of interest, a potential interdependent aircraft s represented in orange and associated to λ_2^0 , would be already at a given offset position γ_{2m}^{offset} . Then, for computing the horizontal interdependency matrix, this aircraft has to be translated backwards along its actual recurrent pattern, to identify the initial and offset positions of that aircraft for the other expected trajectories associated to λ_2^0 .

The horizontal and temporal dimensions of a potential conflict between a pair of aircraft are gathered in the horizontal interdependency matrix. When any of these values is larger than a given threshold, it implies that those trajectories are operationally restricted. For the sake of simplicity, this work will consider from now on that two aircraft are horizontally interdependent if any value of their horizontal interdependency matrix is larger than a given threshold.

$HIM_{r,s}$ will define how two aircraft r and s are related and constrained horizontally and temporally. However, a potential conflict between two aircraft is defined in three dimensions: horizontal distance; time separation; and, finally, a vertical distance. Therefore, an additional vertical distance between two aircraft has to be also defined, depending on their vertical regime at the crossing point.

5.2.3. Vertical Trajectory Specification

Once the interdependency between two aircraft has been modelled, the methodology addresses the modelling of aircraft vertical profiles. This work focuses on the modelling of the trajectory vertical profile in the airspace volume where the aircraft starts its descent. This trajectory vertical profile can be defined in terms of two segments, one depending on the position of the TOD and another one targeting the descent itself.

The aircraft flight level can be expressed similarly in terms of the arc-length parameter, following (17). The main difference with the horizontal component is that $Vz(\gamma_{il})$ can be zero. For modelling the aircraft descent profile, FL component can be transformed to:

$$\rho(\gamma) = RFL - FL(\gamma) \quad (37)$$

Now, it can be established a relation between ρ and t . $\rho(\gamma)$ equals 0 whenever the $FL(\gamma)$ corresponds to the RFL. In this piece of work, we assume that the TOD location of the trajectory, γ_{Tod} , corresponds to the last value of γ that meets such condition. Therefore, the time to descend ρ FLs since the aircraft entered the airspace volume of interest is defined as:

$$t(\rho(\gamma), \gamma_{Tod}) = \theta(\rho(\gamma)) + t(\gamma_{Tod}) \quad (38)$$

$$\gamma \geq \gamma_{Tod}; \rho(\gamma) \geq 0$$

This function is only defined for horizontal arc-length parameters larger than γ_{Tod} . θ reflects the time since the aircraft started to descend. Therefore, the vertical profile of the trajectory along the airspace volume of interest is divided in two distinct phases, the cruise one as a function of γ_{Tod} , and the descent phase, depending on ρ . Now, if it is considered a pattern α_{il}^0 , a new random variable $T_{il}^{\rho, \gamma_{Tod}}$ can be defined which characterises the time that the aircraft will take to descend ρ flight levels since the aircraft entered the airspace volume of interest, depending on the location of the TOD.

If now the dependency with the location of the TOD can be removed by marginalising this random variable over γ_{Tod} to obtain an expression for T_{il}^{ρ} . If it is identified a potential TOD location as $\gamma_{il}^{TOD} = \gamma_{il}^k$, T_{il}^{ρ, γ_k} can be defined as

$$T_{il}^{\rho, \gamma_k}(\rho, \gamma_{il}^k) \sim \mathcal{T}_{il}^{\gamma_k} + \Theta_{il}(\rho | \gamma_{il}^k) \quad (39)$$

If now the approach defined in (22) is followed, this equation is transformed to:

$$T_{il}^{\rho, \gamma^k}(\rho, \gamma_{il}^k) \sim N(\mu_{il}^{\rho, \gamma^k}, (\sigma_{il}^{\rho, \gamma^k})^2) + \Theta_{il}(\rho, \gamma_{il}^k) \quad (40)$$

If it is assumed that $\Theta_{il}(\rho, \gamma_{il}^k)$ can be treated similarly to the horizontal approach described in the previous section, we can define

$$\Theta_{il}(\rho, \gamma_{il}^k) \sim N(\mu_{\theta}^{\rho, \gamma^k}, (\sigma_{\theta}^{\rho, \gamma^k})^2) \quad (41)$$

Then, if that random variable is marginalised over γ_{il}^k and a discrete set of γ_{il}^k values is considered, of size K, then

$$T_{il}^{\rho} \sim \sum_k [P(\gamma_{il}^{TOD} = \gamma_{il}^k) \cdot T_{il}^{\rho, \gamma^k}(t; \rho, \gamma_{il}^k)] \quad (42)$$

Then, the PDF $f_{T_{il}^{\rho}}$ of T_{il}^{ρ} is defined as:

$$f_{T_{il}^{\rho}} \sim \sum_k [P(\gamma_{il}^{TOD} = \gamma_{il}^k) \cdot [f_{T_{il}^{\rho, \gamma^k}} + f_{\Theta_{il}^{\rho, \gamma^k}}]] \quad (43)$$

Eq. (43) defines $f_{T_{il}^{\rho}}$ as a mixture of gaussian distributions which depends on the distribution of the TOD location along the trajectory. An illustration describing Eq. (43) is introduced in Figure 5-7.

Figure 5-7 Description of the pdf of $T_{il}^{\rho=150}$

Figure 5-7 represents schematically different evolution profiles depending on the position of the TOD. The *blue* profile represents a descent starting at $\gamma_{il}^{TOD} = 10 \text{ NM}$, the *green* profile describes it in the case it started descending at 40 NM, whereas the orange one represents the case when it starts at 80NM. The associated PDF functions are represented over the *time* axis (in vertical) for easing the interpretation of the figure, although they should be represented in a third dimension. These PDFs are modulated depending on the probability assigned to the different TOD locations, with the *orange* case being more frequent than the other two. $f_{T_{il}^p}(\rho = 150FL)$ is represented as a dashed line generated after the composition of the three PDFs.

5.2.4. Contextual factors impacting on the Trajectory Vertical Profile.

The previous chapter demonstrated that Θ depends on the RFL, V_G , the flow λ_i^0 and the location of the TOD, among other variables. In that work, the TOD location was assumed to be known. In this work, we explore further the characterisation of the TOD location. In this analysis, we will try to determine if the interdependency measure could be considered for determining the TOD position.

The hypothesis which is being tested is that if there are two aircraft r and s associated to different major flows, the location of the TOD is independent of the interdependency between them or not,

$$P(\gamma_{il}^{TOD} | \bar{A}_r, FL_{jm}^s, HIM_{r,s}) = P(\gamma_{il}^{TOD}; \bar{A}_r) \quad (44)$$

where it is assumed that the vertical interdependency is a function of the FL of another aircraft. $\bar{A}$ is indicative of the intrinsic features of the aircraft belonging to the recurrent pattern α_{il}^0 .

A qualitative visual exploration of the location of the TOD will be conducted, together with an impact analysis on the trajectory vertical profile. In addition, a Chi-squared test will be carried out to evaluate this dependency. Once this dependency has been found or not, trajectory predictors for the vertical profile will be generated as described in (43). The vertical trajectory predictor for Θ_{il} will be defined as an Artificial Neural Network analogously to the horizontal one, as described in Eq. (21) and Eq. (22).

5.3. Results

5.3.1. Traffic Interdependency Identification

5.3.1.1. Expected Trajectories based on Observed Traffic

The first step is the extraction of the major flows λ_i^0 and recurrent patterns α_{il}^0 associated to them, as detailed in Section 3.4.2. Figure 5-8 illustrates both steps for the datasets of interest. The major flows are represented on the left-hand side laying over the sector represented in Figure 5-8. The major flows can be associated with the standard routes of the sector.

The right-hand side of Figure 5-8 corresponds to the recurrent patterns which were extracted from two major airspace flows. The first major flow, λ_1^0 from now on, corresponds to a North-South traffic flow flying from PUMAL to LORES, already studied in the previous chapter. This flow corresponds to evolution traffic towards Palma de Majorca airport, with ICAO code LEPA. The second major flow, λ_2^0 , corresponds to an East-West cruise flow. Flights within λ_2^0 usually enter the sector via MAMES and exit it via BCN.

Figure 5-8 Major flows and recurrent patterns. The recurrent patterns illustrated in green correspond to the PUMAL-LORES major flow, flying North-South, with destination LEPA. The recurrent pattern illustrated in orange corresponds to the MAMES-BCN flow, flying East-West in a cruise state.

Following Eq. (19), and based on previous results ([147]), ANN TPs were trained with interpolated data generated every 1NM along the trajectory. Additionally, the feature vector

(X) included the RFL, Vg when the aircraft entered the sector, and finally, the wake vortex of the aircraft (Light, Medium, Heavy or Jumbo). The output data was provided in seconds.

The aircraft model is a relevant feature for the trajectory prediction. However, the aircraft population of the sample was not homogenous, being largely dominated by Airbus A320 and Boeing B737-800. Other aircraft were less representative, and to prevent not having a representative population for every aircraft model, it was decided to convey the aircraft model feature into the wake vortex of the aircraft.

The ANNs were formed by one input layer, one 10-neuron hidden layer and an output layer. The activation functions of the hidden neurons are sigmoid functions. The output neuron has a linear activation function. Inputs and outputs are subject to a normalisation process which maps them between -1 and 1. The training dataset is further divided in three separate groups for training, validation (preventing overfitting) and testing. The percentages associated to each category were 70%, 15% and 15% respectively.

Equations (20), (21) and (22) assumed that the TP residuals were normally distributed. Figure 5-9 to -Figure 5-12 illustrate the behaviour of the residuals of as a function of the arc-length parameters for the sample corresponding to the training dataset. Figure 5-9 illustrates the evolution of the residuals for α_{11}^0 . This recurrent pattern corresponds to the standard route associated to the major flow λ_1^0 , with a total-length L_{11} of 119 NM. The X-axis corresponds to the arc-length parameter (γ_{11}), in NM, whereas the Y-axis corresponds to the residuals of the prediction. The boxes along the X-axis group the 50% of the residuals for the corresponding arc-length parameter. The bottom and top edges of the boxes represent the 25th and 75th percentiles, respectively. The whiskers cover 1.5 times the interquartile range previously defined.

Figure 5-10 represents the distribution of the residuals for three different arc-length values. These values were 30NM, 60NM and 90NM respectively. The horizontal axis represents the residuals in seconds while the vertical one represents their probability density. The residuals were fitted to normal distributions (continuous lines).

Figure 5-9 Residuals for the ANN associated to the standard route associated corresponding to the flow Pumal-Lores (α_{11}^0) along γ_{11} .

Figure 5-10 Residuals distribution for different values of γ_{11} for the standard route associated to the flow Pumal-Lores (α_{11}^0). The red series correspond to $\gamma_{11} = 30$ NM, blue series correspond to $\gamma_{11} = 60$ NM and green series correspond to $\gamma_{11} = 90$ NM

Table 5-1 includes the mean values and Root Mean Squared Error (RMSE) values for different values of γ_{11} . The table introduces the performance of the predictor for the training and testing datasets. It can be observed that the bias present in the residuals is concentrated around zero, meanwhile the RMSE grows with γ_{11} . The testing RMSE are similar to the training ones.

Results

Table 5-1 Trajectory predictor performance as function of the mean error and the Root Mean Squared Error (RMSE) of the residuals for different values of γ_{11} for the standard route associated to the flow Pumal-Lores (α_{11}^0).

Dataset	Metric	15NM	30NM	45NM	60NM	75NM	90NM
Training	Mean Error [s]	0.51	0.01	0.27	-0.17	-0.02	0.11
	RMSE [s]	1.59	3.23	5.11	7.31	9.67	13.92
Testing	Mean Error [s]	0.65	1.02	1.34	0.985	0.135	-1.62
	RMSE [s]	1.77	3.50	5.40	7.65	10.17	14.43

Figure 5-11 and Figure 5-12 illustrate the residual information for α_{21}^0 . This recurrent pattern corresponds to the standard route associated to the major flow λ_2^0 , with a total-length L_{21} of 105NM. Figure 5-11 illustrates the evolution of the residuals along α_{21}^0 . It can be observed that the distance between whiskers grows linearly with γ_{21} . It can be compared with Figure 5-9, where it grows linearly to start evolving quadratically after 70NM.

Figure 5-12 also shows the residual distributions for different arc-length values. It can be observed for both Figure 5-10 and Figure 5-12 that the residuals corresponding to 30NM are more concentrated around zero than the corresponding normal distribution. It is also observed that the distributions are slightly right-skewed for the recurrent pattern α_{21}^0 .

Figure 5-11 Residuals for the ANN TP for the standard route associated to the flow Mames-BCN (α_{21}^0) along γ_{21} .

Figure 5-12 Residuals distribution for different values of γ_{21} for the standard route associated to the flow Mames-BCN (α_{21}^0). The red series correspond to $\gamma_{21} = 30$ NM, blue series correspond to $\gamma_{21} = 60$ NM and green series correspond to $\gamma_{21} = 90$ NM

Table 5-2 shows the performance metric of the TP for this recurrent pattern. The considerations which have been extracted for the previous flow are repeated. We can find higher absolute values for the bias of the residuals, whereas the RMSE values are slightly larger for the training sample than for the testing sample.

Results

Table 5-2 Trajectory predictor performance as function of the mean error and RMSE of the residuals for different values of γ_{21} for the standard route associated to the flow Mames-BCN (α_{21}^0).

Dataset	Metric	15NM	30NM	45NM	60NM	75NM	90NM
Training	Mean Error [s]	0.29	-0.27	-0.135	0.09	0.04	0.18
	RMSE [s]	1.96	4.30	6.70	9.32	12.20	15.28
Testing	Mean Error [s]	0.66	1.05	1.34	0.99	0.136	-1.62
	RMSE [s]	1.77	3.50	5.40	7.65	10.17	14.43

Observing Figure 5-10 and Figure 5-12, the normality assumption may be sensed. Chi-Square goodness of fit tests have been carried out, rejecting the normality assumption. Figure 5-13 shows a QQ-plot to identify the reasons for the normality assumption being rejected. The case presented corresponds to $\alpha_{11}^0 = 1$ and $\gamma_{11} = 60$ NM. It can be observed that the behaviour corresponds to a normal distribution between the standard normal quantiles -2 and 2, and that the distribution is heavy-tailed. It means that the residuals are showing more extreme values than expected by the normal distribution. The impact of this result will be further discussed in the Section 5.4

The normality assumption is therefore partially accepted, as the bulk of residual values can be modelled as normal distributions. The next section analyses the interdependency definition results for these two recurrent patterns.

Figure 5-13 QQ-plot for $\gamma_{11} = 60$ NM, corresponding to the recurrent pattern α_{11}^0 .

5.3.1.2. Crossing Point Interdependency Definition Results

Eq. (23) introduced a probabilistic definition of the time separation with which two aircraft would have arrived at γ_{il} and γ_{jm} if they were flying following α_{il}^0 and α_{jm}^0 respectively. In addition, different recurrent patterns may have different lengths, and also aircraft do not enter the sector synchronously. Eq. (33) introduced an offset along one of the routes for considering for these two factors.

Figure 5-14 illustrates the impact of this offset on the actual time separation of two aircraft arriving to $\gamma_{11} = 64$ NM and $\gamma_{21} = 94$ NM respectively, where the offset has been assigned to the aircraft allocated to the major flow λ_2^0 . These arc-length parameters correspond to the crossing point of these two recurrent patterns. Histograms along the actual time separation dimension were generated for different offset values. The Z-axis represents the associated probability density for these histograms. It has only been represented the offset interval (0,35] NM for the sake of clarity in the figure. The training sample has been used to generate this figure.

Figure 5-14 Actual Difference Time Values at the Crossing Point

It can be observed that the actual values of $\Delta t_{11,21}^{64,94}$ are spread over the temporal dimension. It is observed how the values tend to concentrate around zero when γ_{21}^{offset} tends to 35NM. However, it can be observed that there are values of $f_{\Delta_{11,21}^{64,94}}(0)$ larger than zero for $\gamma_{21}^{offset} = 25$ NM and for $\gamma_{21}^{offset} = 30$ NM.

Figure 5-15 represents the theoretical PDFs as derived from Eq. (33). Four pairs of aircraft corresponding to the testing sample are studied for $\gamma_{11} = 64$ NM and $\gamma_{21} = 94$ NM. Theoretical $f_{\Delta_{11,21}^{64,94}-\gamma_{21}^{offset}}$ values are generated considering $\Delta\tau$ and γ_{21}^{offset} as parameters. Mean and standard deviations to be used within Eq. (33) are computed based on the residuals of the training sample.

The PDF magnitude is represented by means of a colourmap, which is shared among the four figures. The actual time separation values, Δt , corresponding to the actual offset value of the aircraft belonging to λ_2^0 , are represented as *purple* boxes in each subfigure.

Figure 5-15 Probability Density Function of the predicted time separation for four different cases. The purple box in each subplot is indicative of the actual time separation and the actual offset.

Each figure is titled depending on the predicted times τ_{11}^{64} and τ_{21}^{94} for arriving at those locations. It can be observed how the critical predicted values ($\Delta\tau = 0$ sec) vary from 32 NM to 40 NM depending on the dynamics of the involved aircraft. It can be observed how the actual values fall within high $f_{\Delta\tau_{11,21}^{64,94}-\gamma_{21}^{offset}}$ values for the four cases.

These four cases are illustrative of the results derived from Eq. (33). However, these four cases may not be representative of the whole sample. Figure 5-16 illustrates the distribution of the residuals of the actual time separation $\Delta t_{11,21}^{64,94-\gamma_{21}^{offset}}$ with respect to the expectation of the predicted time separation $\Delta\tau_{11,21}^{64,94-\gamma_{21}^{offset}}$ for those pairs of aircraft that flew following these recurrent patterns for the training sample. The horizontal axis represents the residuals in seconds, whereas the vertical axis shows the associated $f_{\Delta\tau_{11,21}^{64,94-\gamma_{21}^{offset}}}$ value for the actual value of $\Delta t_{11,21}^{64,94-\gamma_{21}^{offset}}$. Different colours represent the actual offset position of the aircraft flying the recurrent pattern α_{21}^0 , after collapsing them in 10-NM bins starting from 0.

The objective of this figure is two-fold. On the one hand, it shows the disparity of the types of encounters, as there are no gaps in the generation of the virtual bell, which resembles a normal distribution. In addition, it can be observed that the residuals are concentrated over 0, ranging from -60 to 60 seconds. The distribution of the $f_{\Delta\tau_{11,21}^{64,94-\gamma_{21}^{offset}}}$ values associated

to them shows that they are normally distributed for groups apart from the interval [30-40) NM. For that group, values are concentrated around 0 as well, but with higher associated density probability values.

On the other hand, it shows the impact of γ_{21}^{offset} and the standard deviation associated to it, in the distribution of the residuals. As the offset locations are farther from $\gamma_{21} = 0$, predictions are expected to be less accurate, as the normal distributions become wider due to the larger values of σ_Z^* .

Figure 5-16 Probability of the actual time difference in terms of the residuals between actual and predicted time differences.

Figure 5-17 plots the residuals of the predicted time separation as a function of the actual time separation. The horizontal dotted lines represent 1σ , 2σ and 3σ of the residuals. It can be observed that there is not a relation between the actual time separation and the residuals. In addition, it can be observed the values which are above 2σ . The critical cases would be those where the actual values are below the temporal threshold (120 s). These values correspond to errors between 40 and 50 seconds approximately. This prediction is carried out between 8 and 10 minutes before arriving to the crossing point. The RMSE for the testing sample was 18.65s.

Figure 5-17 Residual distributions as a function of the actual time separation. Different colours represent different offset values.

5.3.1.3. Horizontal Interdependency

Eq. (35) provided a measure of the horizontal interdependency between two aircraft r and s . The measure was function of a time separation threshold (Δt_{Thres}) and a horizontal distance threshold (D_H).

The interdependency measure answers in a quantitative manner the following question:

What is the probability that two aircraft are arriving to given locations so that their relative distance is equal or below D_H with a time separation of Δt_{Thres} or below?

For the sake of simplicity, results demonstrating Eq. (35) are illustrated by means of a series of figures. Figure 5-18 illustrates a case of two aircraft following the recurrent patterns α_{11}^0 (green) and α_{21}^0 (orange) respectively, including four charts. The top-left chart provides a representation of the horizontal projection of the actual trajectories for both aircraft.

The top-right chart illustrates the relative distance $D_{11,21}$ between them as a function of time. The relative distance is plotted in black, with reference to the left-hand vertical axis. The time reference is the timestamp corresponding to the first registered track of the aircraft which corresponds to α_{11}^0 . This chart also includes γ_{11} and γ_{21} as a function of the time. The values correspond to the right-hand side Y-axis. Finally, there are two additional reference lines, dashed and plotted in black. The horizontal reference line indicates the horizontal

separation minima distance (5NM), with respect to the left-hand side Y-axis. The vertical reference line corresponds to the timestamp when the two aircraft were at minimum distance. The bottom-right chart characterises the vertical evolution of both aircraft along time. The X-axes of both charts on the right-hand side represent the same temporal data. The left Y-axis shows the FL in hundreds of feet, whereas the right-hand side shows again γ_{11} and γ_{21} . Finally, the bottom-left chart represents $\Pi_{\Delta t_{Thres}, D_H}^{11,21}$ for diverse values of D_H (horizontal-axis) and Δt_{Thres} (vertical axis). Actual values are represented through a colourmap, which can be observed on the right-hand side on that chart.

Figure 5-18 Horizontal Interdependency Measure. Case of Study 1.

The case illustrated in the Figure 5-18 was selected based on the values of $\Pi_{\Delta t_{Thres}, D_H}^{11,21}$. The interdependency values between pairs of aircraft were computed for the training sample, when the aircraft corresponding to α_{11}^0 (green) entered the sector. Then, a set of highly interdependent pairs of aircraft was identified, without a prior analysis of actual horizontal and vertical trajectory profiles.

In this particular case, it can be observed that the metric provided a 100% certainty that the aircraft were going to be within a 2-minute temporal distance of losing horizontal separation. It can be observed that the actual distance fell below 5NM. If we observe the interdependency plot per row, the bottom one ($\Delta t_{Thres} = 30s$) reflects a low probability of being in that situation, although it grows with D_H (darker greens). If we analyse the same plot per column, from bottom to top, the interdependency values show a faster increasing rate for lower values of D_H . The reason behind this behaviour is that there are fewer discrete

values of γ_{11} and γ_{21} that meet a restrictive D_H condition. When Δt_{Thres} is relaxed and the integrand values increase, these pairs have a greater weighting than those that have relaxed conditions of D_H .

An interdependence between a pair of aircraft has three dimensions: the horizontal distance; the vertical distance; and the temporal distance. Figure 5-18 shows that the horizontal and temporal distance were expected to be infringed and were actually infringed. Therefore, the conflict should have being prevented by the ATCo by acting on the vertical layer.

The vertical evolution plot (bottom-right corner) reflects a special condition of these two recurrent patterns. The normal TOD location of the North-South flow is located around the crossing point. Although these two aircraft were not initially expected to lose vertical separation (the evolving one is flying at RFL390 and the cruise one is flying at RFL350), both aircraft were *operationally* conflicting, as the cruise flight was restricting the descent phase of the other aircraft. Thus, the evolving aircraft was first cleared to FL360, and then, once the relative distance between them started increasing again, it was further cleared to continue the descent.

Figure 5-19 Horizontal Interdependency Measure. Case of Study 2

The second case of study is illustrated in the Figure 5-19. The interdependency measure provided low probabilities of being in a tight conflict. The actual horizontal distance was larger than 5NM but below 10 NM. The analysis of the vertical regime shows that the descent starts after the flights have overflown the crossing point, but not immediately after.

Figure 5-20 Horizontal Interdependency Measure. Case of Study 3

The third case of study is illustrated in the Figure 5-20. The interdependency measure provided a high probability of losing horizontal separation. The actual horizontal distance was below 5NM. The analysis of the vertical regime shows that the evolution aircraft has an early CFL, one minute after entering the sector, to prevent the conflict to happen.

Figure 5-21 Horizontal Interdependency Measure. Case of Study 4

The fourth case of study is illustrated in the Figure 5-21. The interdependency measure provided very high probabilities of losing the horizontal separation. The actual horizontal distance was near zero. The analysis of the vertical regime shows that the evolution aircraft was authorised to descend after the flight crossed the critical point. In addition, the RFL was

390, indicating that the descent should have started earlier, confirming the interdependence between these two aircraft.

The observation of these four cases outlined that these aircraft were in operational vertical conflicts, as the cruise flow restricted the descent of the evolution one. Consequently, the vertical interdependence between these two major flows was defined by establishing that, if the RFL of the evolving flight matched or was larger than the RFL of the cruise flight, both aircraft were vertically interdependent.

5.3.1.4. Interdependence Effect on TOD Locations

Summing up the previous sections, two aircraft r and s were expected to be in a horizontal conflict if their $HIM_{r,s}$, following Eq. (36), had any of its components above a given threshold. The $HIM_{r,s}$ values were calculated when the α_{11}^0 flight entered the sector.

For this study, the parameters Δt_{Thres} and D_H were set to 120s and 5NM respectively. These two choices were based upon two operational premises. The Short-Term Conflict Alert (STCA) look-ahead time (LAT) varies from 90 seconds to 120 seconds depending on the operational environment. In this Spanish en-route airspace is set to 120 seconds. The horizontal threshold corresponds to the horizontal separation minima for en-route airspaces. After the discussion of the previous section, it was decided to establish the HIM threshold at a 95% level. Finally, the vertical interdependence condition included *operational* conflicts, depending on whether there were restrictions on the evolution flights or not.

Eq. (43) modelled the trajectory vertical profile as function of the TOD location. It was observed in Figure 5-18 and Figure 5-19-Figure 5-21 that conflicting scenarios impacted on the TOD locations for these specific cases. Figure 5-22 shows the generalisation of this analysis for the training sample. The figure presents the TOD locations for all aircraft flying α_{11}^0 with RFL370 depending on whether they were expected to be in a potential conflict when they entered the sector (blue series) or not (red series).

Figure 5-22 discretises the horizontal axis (TOD location) in 1NM-width bins. The analysis of the red series (non-conflicting scenarios) shows that TODs are distributed from 30NM to 90NM, with the more frequent values ranging between 55NM and 65NM. In the case of the conflicting flights, the TOD locations are spread over the γ_{11} range. It can be observed the remarkable number of flights with an early CFL (γ_{11} between 0NM and 35NM), which was negligible for non-conflicting flights.

It can be observed how the number of flights with a TOD location between 60NM and 65NM drops with respect to the non-conflicting aircraft, to rise again after $\gamma_{11} = 65\text{NM}$. Figure 5-23 illustrates this specific case, by censoring the sample for flights with TOD locations below 30NM (green series). The analysis demonstrates that the TOD locations are more frequent after 65NM for the conflicting aircraft.

Figure 5-22 TOD Location depending on whether the encounter was classified as a conflict or not. The red colour indicates non-conflicting encounters, whereas blue columns indicate the opposite.

Figure 5-23 TOD Location depending on whether the encounter was classified as a conflict or not, for a sample with TOD locations larger than 30NM. The red colour indicates non-conflicting encounters, whereas green columns indicate the opposite.

An example of the first case (early TOD) is shown in Figure 5-20, whereas two examples of the latter (TOD after 65NM) are represented in Figure 5-19 and Figure 5-21. Figure 5-18 illustrated a third case, where the flight was cleared to descend to an intermediate FL, until the potential conflict disappeared.

Figure 5-22 and Figure 5-23 show that the hypothesis formulated in Eq. (44) may be rejected. The application of the chi-square test for testing the independence of two random variables for the training dataset, $\chi^2(8, N = 2425) = 113.04$, $p < 1e-9$ resulted in a strong rejection of the null hypothesis as formulated in Eq. (44), and therefore, the independence of the TOD position and the categorisation by the conflict measure can be ruled out. The result is repeated again for the testing sample, where the $\chi^2(8, N = 613) = 83.08$, $p < 1e-9$ yielded a strong rejection of the null hypothesis as well.

The next step was to analyse the impact of the TOD locations on the vertical profile of the actual trajectories, depending on whether the flights were in an interdependent scenario or not.

5.3.1.5. Trajectory Vertical Profile Characterisation

Figure 5-24 illustrates the PDFs associated to T_{1l}^ρ for different values of ρ for non-conflicting flights. T_{1l}^ρ indicates the time to descend ρ FLs (including levelled segments) since an aircraft enters the sector. As the total lengths of α_{11}^0 and α_{12}^0 were similar, both samples were merged for conducting the current analysis. That resulted in removing the index l from T_{1l}^ρ , for transforming it to T_1^ρ . Figure 5-27 represents similar information, but for conflicting flights.

The first row corresponds to $\rho = 50\text{FL}$, the second row to $\rho = 100\text{FL}$ and the third row to $\rho = 150\text{FL}$. TOD locations have been grouped in 10NM-width bins to obtain visual meaningful information. In addition, times are also binned in 60-seconds categories for the sake of clarity.

Figure 5-24 Time to descend 50, 100 and 150FL Density Probability for non-conflicting flights with RFL370, for different TOD locations grouped in 10NM-width bins.

Figure 5-25 Time to descend 50, 100 and 150FL Density Probability for conflicting flights with RFL370, for different TOD locations grouped in 10NM-width bins.

Starting with the non-conflicting flights, the actual values of the studied variable are very spread. For example, values range from 500s to 1100s for $\rho = 100\text{FL}$, being the median value around 720s. A similar ranges of values (700 s) can be seen for the other two rows.

For the conflicting cases, it can be observed how distributions do not have clear medians for the first and second rows, showing a different pattern than for the mirrored cases for non-conflicting flights. On the other hand, the third row shows a closer behaviour to the non-conflicting flights, but with a heavy right tail.

The explanation for these behaviours can be extracted from observing Figure 5-26. It splits the second row of Figure 5-24 by TOD category, maintaining the colour characterisation. If we observe closely the early CFL cases (TODs from 10 to 30NM), the values of the studied time are spread more widely, whereas for the other cases, they have a more normal-like performance. The higher prevalence of these cases for the conflicting flights provokes that the mixtures of gaussian does not have a clearly defined median (second row of the Figure 5-27).

The last step is to include this information in a trajectory predictor for the vertical profile of the trajectory. For doing so, we used an ANN TP to model Θ_{il} (see Eq. (41) and [81], [147]). The neural networks had the same structure than the horizontal ones, and the inputs for them were ρ, γ_{1l}^{TOD} , RFL, the wake vortex and V_g at $\gamma_{1l} = 0$. and

Finally, the TOD locations were modelled depending on the observed frequency for the sample $[P(\gamma_{il}^{TOD} = \gamma_{il}^k)]$, depending on whether the flights were conflicting with other aircraft or not.

Figure 5-26 Vertical profile Density Probability for descending 100 FL for non-conflicting aircraft with RFL370, for different locations of the TOD.

Figure 5-27 Vertical profile Density Probability for descending 100 FL for aircraft with RFL370, for different locations of the TOD.

5.3.1.6. Vertical Profile Trajectory Predictor Results

Following Eq. (43), each flight had a specific PDF $f_{T_1^\rho}$ depending on the predicted values for arriving to the TOD locations and the descent profiles as a function of the TOD. Figure 5-28 and Figure 5-29 illustrate these curves for the training sample. The former illustrates non-conflicting cases whereas the latter does the opposite. The left-hand side of the figures represents different $f_{T_1^\rho}$ as a function of Vg and the right-hand side represents $f_{T_1^\rho}$ as a function of t_1^ρ (time to descend ρ FLs). Each red point corresponds to the density value corresponding to an actual value. Both cases represent RFL370, wake vortex category “M” (*medium*) and $\rho = 100$ FLs. These parameters have been selected due to their representativeness of the whole sample.

Figure 5-28 $f_{t_1^\rho}$ distributions as a function of the time to descend $t_1^{\rho=100}$ 100FL for flights in non-conflicting scenarios. Each red point is the $f_{t_1^\rho}$ value corresponding to an actual time to descend value.

Figure 5-28 illustrates that the artificial PDFs have a clear median value, showing symmetry as well. The left-hand side of the figure illustrates that, as Vg increases, the PDFs are wider, and the median values decrease. This effect can be identified on the right hand-side as well. For example, for the extreme values around 1000 seconds, which correspond to low Vg speeds, the curves have larger peaks than for higher values of Vg .

The distribution of the actual values shows that they are more spread than in Figure 5-15. There are values corresponding to the median of their associated PDFs, but there are a high proportion of points on the tails of the distributions, with low density values associated. Figure 5-29 illustrates the case for the conflicting cases. It can be observed that it shows a similar pattern than for non-conflicting flights.

Figure 5-29 $f_{t_1^p}$ distributions as a function of the time to descend $t_1^{p=100}$ 100FL for flights in conflicting scenarios. Each red point is the $f_{t_1^p}$ corresponding to an actual time to descend value.

Figure 5-30 specifies the difference between the actual value of $t_1^{p=100}$ and the median value of the theoretical probability density function $f_{t_1^p} = 100$. Columns discriminate the residuals depending on whether the aircraft were conflicting with other traffic (right-hand side) or not (left-hand side). The top row introduces a histogram of the distribution of the residuals, based on 10-second bins. The bottom row provides the actual value of the theoretical probability density function $f_{t_1^p} = 100$ as a function of the residuals.

It can be observed that the range of these residuals varies from -150 to 250 seconds. There are also two identifiable mode values of the density distribution of the residuals, located around -30 and 130 seconds respectively. These patterns can be detected for both conflicting and non-conflicting cases.

These patterns are indicative of a hidden binary variable. This restriction removes potential variables which vary in a broader range from being the source of this binary effect, such as wind direction or temperature.

The previous chapter identified the distance of the TOD to the destination airport as a relevant factor impacting on the vertical profile of the trajectories. The sample that has been used within this work is characterised to be flying towards LEPA. The generalisation of the previous finding to this sample, in addition to the binary behaviour of the residual distribution, leads to a unique factor that has a binary behaviour and can reflect the distance from the TOD to the airport, which is **the airport runway configuration**.

Figure 5-30 Vertical TP Residuals Distribution for non-conflicting flights (left-hand side) and conflicting flights (right-hand side). The top-row provides the residual distribution whereas the bottom row provides the values of $f_{t_1^p}$ corresponding to actual time to descend values.

LEPA airport has two parallel runways, which can be operated in westerlies (24R/L) or easterlies (06R/L) configurations. Figure 5-31 classifies the residuals as a function of the runway configuration. It can be observed that the binary behaviour is fully characterised by this factor, which indicates that it should be introduced for the generation of the vertical probability density functions $f_{t_1^p}$.

Figure 5-31 Vertical TP Residuals Distribution for non-conflicting flights as a function of LEPA Runway Configuration.

Eq. (43) introduced $f_{T_{it}^{\rho}}$ as a function of the probability mass function of the TOD location along the horizontal central trajectory. If now a model for the TOD location is trained as a function of different parameters, such as Vg , RFL, the wake vortex and the airport configuration, k probability mass distributions can be generated for each one of the possible TOD categories. These mass distributions are generated by characterising the actual values of the TOD when that specific category was predicted.

Figure 5-32 Vertical TP Residuals Distribution for non-conflicting after including the airport configuration as a factor for predicting the TOD position. The top-row provides the residual distribution whereas the bottom row provides the values of $f_{t_1^{\rho}}$ corresponding to actual time to descend values. The top-row figure includes blue line corresponding to a fitted t-student distribution with parameters $\mu = -15 \text{ sec}$ and $\nu = 3.5$.

A simple ensemble predictor was used for the prediction of the TOD location for non-conflicting aircraft with RFL370. Results of the integration of the runway configuration by using this model are illustrated in Figure 5-32. The figure is organised as Figure 5-30. It shows a healthier residual distribution, where the residuals were found to be following a t-student distribution, with a μ parameter of -15 seconds and $\nu = 3.5$ (degrees of freedom). The Root Mean Square Error (RMSE) of the prediction was 108 seconds.

Figure 5-33 Vertical TP Residuals Distribution for non-conflicting after including the airport configuration as a factor for predicting the TOD position for the testing dataset. The top-row provides the residual distribution whereas the bottom row provides the values of $f_{t_1^{\rho=100}}$ corresponding to actual time to descend values. The top-row figure includes blue line corresponding to a fitted t-student distribution with parameters $\mu = -18.5 \text{ sec}$ and $\nu = 2.19$.

Figure 5-33 represents the results for the testing dataset. It shows a similar residual distribution that for the training dataset. The residuals were found to be following a t-student distribution, with a μ parameter of -18.5 seconds and $\nu = 2.19$ (degrees of freedom), showing heavier tails than for the training dataset. The Root Mean Square Error (RMSE) of the prediction was 109.92 seconds

5.4. Discussion

The results have been organised in seven sections. The present section discusses key findings from each of them. Section 3.5 detailed the datasets that were used for this study. The datasets corresponded to an en-route sector with evolution flows which may conflict with cruise flows.

Section 5.3.1.1 demonstrated that the clustering approach which has been used within this study was able to discriminate between the major airspace flows in the sector of interest. In addition, the application of the same methodology to each major airspace flow provided their recurrent patterns. It can be observed on the right-hand side of Figure 5-8 that there are some trajectories which do not follow exactly the horizontal central trajectory of the Mames-BCN flow (orange). The difference is subtle and only affects a small portion of the whole trajectory, but it might have a relevant operational impact. Further work could consider developing a clustering process to improve the discrimination between those minor differences, which may improve the results of the application of this methodology.

Section 5.3.1.1 also discussed the normality assumption for the TP's residual distributions. Figure 5-9-Figure 5-12 characterised the dependency of the residuals with the arc-length parameter along the horizontal central trajectory for each studied recurrent pattern. The normality assumption was rejected due to the behaviours of the residuals in the tails of the distribution. It was showed that rare cases were more frequent than expected, which may be propagated to the interdependency measure. In any case, the extreme combined error is below 60s (see Figure 5-17).

Section 5.3.1.2 presented results about the horizontal interdependency measure between flights which belong to two recurrent patterns. The actual time separation at the crossing point varied depending on aircraft dynamics and the initial offset applied to one of the recurrent patterns. It was demonstrated that each aircraft pair had a specific probability density distribution associated to the expected time separation at the crossing point as a function of the offset. It was also demonstrated that the actual time separation values corresponded to high probability density values of the modelled approach. This was further elaborated by analysing the density value associated to each residual. The total RMSE for the actual time separation and the predicted one was below 20s for the testing dataset, without bias.

The definition of a probability measure of the horizontal interdependence between two aircraft r and s was presented in 5.3.1.3. It was shown how the measure varied depending on

relevant parameters. It was demonstrated that the measure provided a clear indication about when two aircraft were expected to be horizontally and temporally interdependent by exploring four different cases. In addition, it was demonstrated that when the metric raised high-probability values, there was an action upon the vertical profile of the aircraft by the ATCo to guarantee the separation minima. The section introduced a *conflict* definition involving also *operational* vertical conflicts.

The novel contribution was that it shifted from a distance-based to a time-based definition. The distance-based definition has been used in different research such as [95], [137], [138], [140], [148]. In addition, the modelling of the time separation as a folded normal distribution was similar to the one presented by [139], but for time-based separations. Another relevant contribution is that the interdependency measure considers not only the standard routes, but also other recurrent patterns (typical ATC actions). This enables the identification of potential constraints on the trajectories and the ATC actions that may be issued. This approach was also introduced in [149], but with other operational purpose.

Based on the insights gained from Section 5.3.1.4, the validity of the metric was demonstrated by analysing the TOD location distribution depending on whether the evolution aircraft was expected to lose separation with other aircraft or not. It was shown that a relevant proportion of the aircraft in conflicting scenarios were given early CFLs along the route, compared with the normal behaviour. In addition, it was shown that there was a drop on the number of aircraft starting to descend around the crossing point position for conflicting aircraft. The dependency between the position of the TOD and the predicted conflicts was validated through Chi-Square tests.

Figures showed that there were cases with early CFLs for non-conflicting aircraft. The studied recurrent pattern often interacts with the analysed cruise flow, but it can also interact with other major airspace flows that are less frequent, or even with itself. Future work should address the definition of the interdependency probability measures for other type of ATC conflicts, such as two aircraft flying in the same flow or the aggregation of different interdependencies. An example in this direction is the introduction of *flow abstraction* factors, as defined in [87], in [149], where the arrival times are predicted based on the situation of a precedent aircraft within the same terminal flow.

Section 5.3.1.5 characterised the mixtures of gaussian associated to the descent profiles of these aircraft depending on whether they were expected to be in conflicting or non-conflicting situations. It was shown that the descent times for conflicting traffic was more

spread than for non-conflicting flights. It was also shown that the impact of the potential conflict identification was reduced to the initial phase of the descent, which showed a greater variability.

Finally, Section 5.3.1.6 exploited the results of the two previous sections in order to design a trajectory predictor for the vertical profile of the trajectories. The analysis of the residuals demonstrated that there was a binary hidden variable. This insight, together with conclusions of the previous chapter led to the identification of the airport runway configuration as a key hidden variable.

Once this variable was introduced in a predictor for the TOD location, that binary pattern disappeared. It was shown that they could be fitted to a known distribution, easing the handling of the uncertainty surrounding the vertical profile for this major airspace flow in the future. The predictor was tested with the testing dataset, yielding similar results.

The prediction RMSEs for both datasets were still large, although the bias on the prediction was almost removed. The prediction was carried out when the evolution aircraft entered the sector. In addition, the vertical and horizontal ANN predictors did not include relevant features for aircraft dynamics such as the wind direction and intensity, temperature, pressure, the aircraft model or intrinsic intention features such as the expected guidance mode [150] or the aircraft inferred mass [79]. The integration of richer datasets may reduce errors on the prediction, with longer look-ahead times. The prediction RMSEs for both datasets were still large, although the bias on the prediction was almost removed.

Trajectory predictors based on state-dependent hybrid mode transitions, such as [148], [150], [151] could benefit from this work to characterise the probability vector governing the transitions between states. Trajectory predictors based on state-dependent hybrid mode transitions, such as [148], [150], [151] could benefit from this work to characterise the probability vector governing the transitions between states.

The conclusion of this chapter should be aimed at unveiling if the traffic information shall be included in the trajectory predictors. Table 5-3 shows the results of conducting Chi-square tests for testing whether the position of the TOD is independent from the potential conflicts, as determined by the interdependency measure. The results have been discriminated depending on the arrival runway, including all the recurrent patterns of the evolution flow. It can be observed that the independency hypothesis is strongly rejected for all cases. The impact of this dependency affects the trajectories in a short-term period after the TOD, as it was observed in Figure 5-24 and Figure 5-25. Therefore, the conclusion is that the inclusion

of this factor would improve short-term trajectory predictions, although its impact when longer look-ahead times are considered shall be evaluated.

Table 5-3 Chi-square tests for testing the independence between the position of the TOD and potential conflicts depending on the arrival runway.

Chi-square tests	Runway 24	Runway 06
Training	$\chi^2(9, N = 1796) = 269.29$, $p < 1e-9$	$\chi^2(10, N = 629) = 113.04$, $p < 1e-9$
Testing	$\chi^2(9, N = 422) = 58.32$, $p < 1e-6$	$\chi^2(9, N = 191) = 31.36$, $p < 1e-3$

5.5. Summary of Findings

The objective of this chapter was to consider surrounding air traffic as a contextual factor for trajectory prediction. To that end, it was proposed an interdependency measure at crossing points, combining data- and model-driven approaches. The interdependency measure was based on a probabilistic approach to air traffic interdependence supported by well-known machine learning algorithms.

The interdependency measure was used to characterise vertical profiles. This characterisation was defined in terms of intrinsic aircraft features in addition to the mentioned air traffic interdependency. The vertical profile was modelled based on the location of the Top of Descent and different profiles for the descent phase, depending on those features.

The work used the previously described clustering techniques to extract the major airspace flows of the airspace volume of interest. The work extended the classification of the trajectories by extracting recurrent patterns within each major airspace flow. Then, trajectory computation engines based on Artificial Neural Networks were developed for each recurrent pattern.

After the application of the horizontal interdependency measure to the datasets, it was demonstrated that it enabled the identification of pairs of aircraft that were expected to lose horizontal separation. In addition, the vertical profiles of aircraft that were identified as horizontally interdependent were shown to diverge from a normal behaviour. As a

conclusion, the proposed air traffic interdependency measure was shown to be a valid precursor of air traffic control actions on involved flights.

The characterisation of the vertical profiles was reinforced by using surrounding air traffic as a relevant feature. In addition, the results indicated a dependency of the descent profiles on other features, such as the requested flight level or the ground velocity. In addition, the analysis of the residuals of the vertical predictors resulted in the detection of a hidden binary factor.

The airport runway configuration demonstrated to be a relevant feature for defining the location of the TOD. The inclusion of this factor in a vertical trajectory predictor resulted on the residuals following a known statistical distribution, which can lead to an enhanced handling of uncertainty in the vertical profile of the trajectory.

The final conclusion of this paper is that the surrounding traffic information can impact on the position of the TOD. This dependency affects the trajectories in the short-term after the TOD, which is relevant in high-density scenarios where an ATC action is constrained by surrounding traffic and in scenarios where there are extended AMAN functionalities, where the position of the TOD may be crucial for handling the uncertainty of the time of arrival.

This Page is Intentionally Left Blank

CHAPTER 6. CONCLUSIONS AND FURTHER WORK

6.1. Overview

This chapter concludes the research and presents potential future lines of research to continue the work in terms of the TP framework presented in the second chapter of the thesis.

Then, Chapter 3 introduced a data framework for the exploitation of ATM data oriented to research and to measure the performance of the system. In addition, the chapter reviewed the most relevant process to cluster trajectories and how the structured data enabled the automatic extraction of ATC vertical procedures.

Chapter 4 and chapter 5 addressed the main research questions. Chapter 4 presented an analysis of the impact of operational factors in the vertical profile of the aircraft trajectories. The outcomes of that analysis were further validated by their inclusion in two type of trajectory computation engines.

Chapter 5 addressed the impact of the surrounding aircraft in the vertical profile of the aircraft. To this end, a methodology to identify potential horizontal interdependencies as a function of the time separation was proposed. It was demonstrated that the metric had explanatory power over the position of the TOD for the selected flow of aircraft. In addition, the runway configuration of the destination airport was identified as relevant factor for the position of the TOD for the studied flow.

6.2. Research conclusions

The main driver of this dissertation, as stated in the introduction was that “*a lack of understanding of TP and Conflict Detection and Resolution driver uncertainties (especially contextual, and aircraft state, operational and performance uncertainties)*” was one the primary limitations of TP tools supporting functionalities envisaged in the initial SESAR ConOps. Thus, the thesis main goal was to answer the question:

How can we characterise the impact of operational contextual factors upon a trajectory?

The methodology which was introduced in Chapter 1 established a series of steps, which were further framed in the current state-of-the-art of trajectory prediction of Chapter 2 in a series of research questions. The questions herein presented limited the scope to the **descent** phase of the trajectory.

6.2.1. ATM Data Organisation and Flow Discovery

The first question was:

Q1. How can ATM data be organised to enable a systematic and data-oriented analysis of the system, including Trajectory Prediction?

Chapter 3 addressed this question by introducing a data warehouse paradigm to structure ATM data for its exploitation for research and performance measurement purposes, aligned with ICAO's recommendations.

In addition, due to the operational relevance and its presence in the state-of-the-art of TP, a review of the most used methodology to extract flows from radar tracks was conducted.

The results of the chapter were presented as the data set which was then used in the rest of the dissertation. The structure of the DWH enabled the extraction of the ATC vertical procedures, relating also ATC interactions with the flights and trajectories, enabling going a step further in the analysis of the impact of ATC on the aircraft intent.

Conclusion 1

The data warehouse paradigm applied to ATM can be extended to operationally characterise the trajectories in a large scale. In particular, the DWH paradigm is useful to establish the key elements to measure the performance of the ATC service.

Conclusion 2

Gariel's methodology is useful to conduct an exploratory analysis of the flows within a given airspace volume. However, its escalation requires to store outcomes of the intermediate steps for being able to produce a stable flow classification.

Conclusion 3

Conforming the flow dimension and the ATC operational activities enables an automatic identification of the ATC vertical procedures over a given flow of aircraft.

6.2.2. Analysis of Operational Factors

The second question presented for this dissertation was:

Q2. What are the operational contextual factors impacting on the descent profile of an aircraft trajectory?

Chapter 4 introduced a methodology based on Multilevel models allowing a broad identification of which factors were having a stronger impact on the vertical profile of an aircraft trajectory. The outcomes of that analysis were further validated by tailoring two different trajectory calculation engines depending on these factors.

These two engines were a point-mass model and a data-driven predictor based on artificial neural networks. The main conclusions after the analysis are:

Conclusion 4

The average rate of descent was dependent on the position of the TOD. The flow dimension represents an important feature explaining the variance on the rate of descents and the vertical operational procedure was not a factor with a relevant explanatory power. The airline operator had intra-year explanatory power, but it decays in an inter-year scope.

Conclusion 5

Point mass models benefit from the introduction of operational factors to train the parameters governing the thrust-law. In addition, given the appropriate isolation of the operational features, pure data-driven kinematic trajectory engines outperform those basic point-mass models in terms of variance of the residuals of the prediction.

6.2.3. Surrounding Air Aircraft as a factor

The third and last question addressed was:

Q3. Can the surrounding traffic be included as a factor in the definition of the aircraft intent?

The Chapter 5 analysed the interdependency between two aircraft by inferring potential conflicts. The interdependency measure was based on a probabilistic approach to air traffic interdependence supported by well-known machine learning algorithms.

After the application of the horizontal interdependency measure to the datasets, it was demonstrated that it enabled the identification of pairs of aircraft that were expected to lose horizontal separation. In addition, the vertical profiles of aircraft that were identified as horizontally interdependent were shown to diverge from a normal behaviour.

Conclusion 6

The proposed time-based interdependency measure identified correctly horizontal interdependencies between aircraft. The measure was validated as a precursor of ATC actions preventing conflicts to happen.

Conclusion 7

The kinematic trajectory computation engine based on artificial neural networks and flow abstractions provides accurate results for short-term trajectory prediction. The residuals of the horizontal predictors follow a normal distribution, which eases its use in probabilistic environments.

Conclusion 8

The runway configuration is an important contextual factor for the prediction of top of descent for the dataset studied. The inclusion of this factor for a kinematic vertical trajectory predictor led to residuals following known probability distributions, which can lead to an enhanced handling of uncertainty in the vertical profile of the trajectory

6.3. Further work

6.3.1. ATM Data Organisation

Trajectory management in ATM relies on air traffic metrics such as the hourly entry count or the sector occupancy. These metrics may be insufficient to characterise the complexity associated to the traffic flying through a given airspace volume. The generalisation of the proposed interdependency measure in Chapter 5 may help to develop new trajectory management metrics and procedures to manage complex operational environments.

In terms of the ATM data organisation, future work could generalise the use of this metric as a method to identify the ATC facts. A systematic analysis of this fact table, using methodologies such as ([110], [152], [153]) can help to measure the safety performance of the provision of the ATC services, not only by studying the actual conflicts between aircraft, but by studying the safety performance in normal operations.

6.3.2. Flow Discovery

The methodology which was reviewed in this thesis suffers for the hyperparameter settings. The use of the data model as explained leads to an enriched featuring of the flights. The usage of semantic analysis to characterise the sequence of segments of the aircraft intent may help to provide more accurate results in the flow classification.

The sequence of points suffers from variability in the presence of all the waypoints, and therefore, it is recommended its use in combination with unsupervised techniques, as Gariel's methodology.

In addition, new concept of operations for the route structure (free routing) will redefine the flows within an airspace. The use of these techniques can help to depict accurately the new airspace flows.

6.3.3. Contextual factors and point mass models.

Future research could address similarities between flows in different sectors, to identify if the behaviour shown is sector-dependent or it could be reproduced in other sectors as well. The research could be further extended by introducing weather-related data for improving the point-mass predictors and learning the ESF law from the data. In addition, other data-driven approaches could be explored, such as Gradient-Boosting-Machines [154] .

6.3.4. Surrounding Traffic

Future work could address the generalisation of the interdependency metric for other crossing points and other airspaces. Moreover, other conflicts of different nature can be modelled, such as flow-based interdependencies. In addition, the aggregation of the interdependency measure for more than two major airspace flows could help to gain a greater understanding of the underlying patterns governing horizontal ATC actions on flights.

The interdependency measure is the initial step for a more comprehensive metric aggregating not only pairwise encounters but also other types such as flow-based conflict or sequencing / merging conflicts [87].

The methodology is scalable. Nonetheless, from an operational point of view, the flow-based trajectory prediction supporting the horizontal interdependency metric shall be complemented by:

1. A complete characterisation and modelling of other ATC abstractions and types of conflicts, such as flow-based conflicts and other types of interactions in critical points, such as sequencing, merging and metering.
2. The extension of the matrix presented in Equation 36 to represent the complete state of the airspace volume of interest.

When the leading aircraft reaches or is about to reach the cleared flight level, the trailing one is cleared to descend to that level.

- $\{r,v\}, \{s,u\}$: A merge-based interaction (MI), a type of critical point. The LORES waypoint (at the bottom of the sector) is the coordination waypoint with the TMA of Palma de Majorca ACC. The ATCO sequences the aircraft, merges them (horizontally and vertically) and finally, meters the horizontal distance between them.
- $\{r,t\}, \{v,t\}, \{u,t\}$: The interdependency case modelled (HI) and studied in the research.

The research line should continue modelling flow-based and merge-based interactions. Next steps should define the $FIM_{r,s}$ and the $MIM_{r,v}$ metrics, similarly to the $HIM_{r,t}$ defined in this paper. The research is intended to continue by defining the *Complexity matrix* (see Equation (46), which is a realisation of the complexity metric as defined in [87]

$$Complexity \sim \begin{bmatrix} AIC_{r,r} & FIM_{s,r} & HIM_{r,t} & MIM_{r,u} & MIM_{r,v} \\ FIM_{s,r} & AIC_{s,s} & HIM_{s,t} & MIM_{s,u} & MIM_{s,v} \\ HIM_{t,r} & HIM_{t,s} & AIC_{t,t} & HIM_{t,u} & HIM_{t,v} \\ MIM_{u,r} & MIM_{u,s} & HIM_{u,t} & AIC_{u,u} & FIM_{u,v} \\ MIM_{v,r} & MIM_{v,s} & HIM_{v,t} & FIM_{v,u} & AIC_{v,v} \end{bmatrix} \quad (46)$$

The following points apply to the matrix defined in the previous equation:

- Each element will be a construct as per Eq. (45). When this matrix is populated, it would be possible to characterise the airspace volume of interest at a given moment, and thus, identifying all the interdependencies within the sector and how the sector is constrained. ATCos or ground automation looking for resolution trajectories free of induced conflicts can look at this matrix to prone the candidate resolution trajectories.
- Each aircraft will have its own element in the matrix representing the preferred trajectory from a performance point of view (Aircraft Performance Constraint, $APC_{r,r}$).

The case of multiple conflicts will be a subset of the matrix previously defined. For future research addressing conflict resolution of multiple aircraft, assembling these matrixes can help to identify the constraints to compute the resolution trajectories following works as [155], [156] or [141] which prevent induced conflicts and preventing other safety nets to be triggered.

REFERENCES

- [1] SESAR JU, ‘European ATM Master Plan 2015’, 2015.
- [2] Federal Aviation Administration, ‘NextGen Implementation Plan’, 2015.
- [3] International Civil Aviation Organization, ‘Doc 9854 Global Air Traffic Management Operational Concept’, *ICAO TECHNICAL PUBLICATIONS The*, no. First Edition. 2005.
- [4] W. Schuster and W. Ochieng, ‘Performance requirements of future Trajectory Prediction and Conflict Detection and Resolution tools within SESAR and NextGen: Framework for the derivation and discussion’, *J. Air Transp. Manag.*, vol. 35, no. 0, pp. 92–101, 2014.
- [5] R. Alligier, D. Gianazza, and N. Durand, ‘Learning the aircraft mass and thrust to improve the ground-based trajectory prediction of climbing flights’, *Transp. Res. Part C Emerg. Technol.*, vol. 36, no. 0, pp. 45–60, 2013.
- [6] R. Alligier, D. Gianazza, and N. Durand, ‘Machine Learning and Mass Estimation Methods for Ground-Based Aircraft Climb Prediction’, *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 6, pp. 3138–3149, 2015.
- [7] K. Tastambekov, S. Puechmorel, D. Delahaye, and C. Rabut, ‘Aircraft trajectory forecasting using local functional regression in Sobolev space’, *Transp. Res. Part C Emerg. Technol.*, vol. 39, no. 0, pp. 1–22, 2013.
- [8] D. González-Arribas, M. Soler, and M. Sanjurjo-Rivo, ‘Robust Aircraft Trajectory Planning Under Wind Uncertainty Using Optimal Control’, *J. Guid. Control. Dyn.*, vol. 41, no. 3, pp. 673–688, Mar. 2017.
- [9] A. Franco, D. Rivas, and A. Valenzuela, ‘Probabilistic aircraft trajectory prediction in cruise flight considering ensemble wind forecasts’, *Aerosp. Sci. Technol.*, vol. 82–83, pp. 350–362, Nov. 2018.
- [10] EUROCONTROL, ‘Atfcm users manual’, 2018.
- [11] I. C. A. Organization, ‘Procedures for Air Navigation Services - Air Traffic Management - PANS-ATM Doc 4444’, no. June, p. 432, 2009.
- [12] C. J. Frawley, William J, Piatetsky-Shapiro, Gregory, Matheus, ‘Knowledge Discovery in Databases: An Overview’, *AI Magazine*, pp. 57–70, 1992.
- [13] SESAR JU, ‘SESAR Concept of Operations Step 1’, 2012.

References

- [14] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, 'The KDD process for extracting useful knowledge from volumes of data', *Commun. ACM*, vol. 39, no. 11, pp. 27–34, 2002.
- [15] ICAO, *ICAO Doc 4444: Procedures for Air navigation Services*, vol. 4444. 2016.
- [16] J. Reason, 'Human error: Models and management', *West. J. Med.*, vol. 172, no. 6, pp. 393–396, 2000.
- [17] EUROCONTROL, 'Safety Nets - Ensuring Effectiveness Guide', 2009.
- [18] EUROCONTROL, 'European Organisation for the Safety of Air Navigation Eurocontrol Experimental Centre Revisiting the « Swiss Cheese » Model of Accidents', Bretigni, 2006.
- [19] B. Kirwan, 'Incident reduction and risk migration', *Saf. Sci.*, vol. 49, no. 1, pp. 11–20, 2011.
- [20] AGENT Consortium, 'D1.1.Report on AGENT functional and non-functional requirements', 2016.
- [21] F. J. Sáez Nieto, 'The long journey toward a higher level of automation in ATM as safety critical, sociotechnical and multi-Agent system', *Proc. Inst. Mech. Eng. Part G J. Aerosp. Eng.*, vol. 230, no. 9, pp. 1533–1547, 2016.
- [22] F. Netjasov and M. Janic, 'A review of research on risk and safety modelling in civil aviation', *J. Air Transp. Manag.*, vol. 14, no. 4, pp. 213–220, 2008.
- [23] J. L. Casti, 'Complexity', *Encyclopædia Britannica*. 2017.
- [24] P. García Fernández, *Estado del arte de las nuevas tecnologías TIC aplicadas al entorno aeroportuario*, 1st Editio. Madrid: ISDEFE, 2012.
- [25] European Commission, 'SES Performance and Charging Schemes', 2019. [Online]. Available: https://ec.europa.eu/transport/modes/air/single-european-sky/ses-performance-and-charging/performance-and-charging-schemes_en.
- [26] European Commission, *Regulation (EU) No 390/2013 - Laying down a performance scheme for air navigation services and network functions*, no. 3. 2013, pp. 1–7.
- [27] E.- Performance Review Unit, 'EUROCONTROL Area Key Performance Figures', 2019. [Online]. Available: https://ansperformance.eu/download/xls/En-Route_ATFM_Delay_AUA.xlsx. [Accessed: 18-Jul-2019].
- [28] R. Hof, Henk; Nunez, Olivia; Pugh, Richard; Flohr, 'ATMRPP Proposal for the development of TBO concept. Managing unpredictable evolution in trajectory based operations'. Toulouse, 2014.
- [29] ICAO, 'Tbo Concept By the Icao Atmrpp', pp. 1–50, 2015.
- [30] SESAR Joint Undertaking, 'SESAR 2020 Concept Of Operations', no. 2017, 2017.
- [31] Y. Xu, 'Cost-based Linear Holding Practice and Collaborative Air Traffic Flow Management

- under Trajectory Based Operations’, no. September, 2018.
- [32] G. A. Boy, ‘The orchestra: A conceptual model for function allocation and scenario-based engineering in multi-agent safety-critical systems’, *VTT Symp. (Valtion Tek. Tutkimuskeskus)*, no. 258, pp. 187–193, 2009.
 - [33] E. / FAA, ‘ACTION PLAN 16 Common Trajectory Prediction Capability’, 2010. [Online]. Available: <https://sites.google.com/site/trajectorymgt/ap16---white-paper#TOC-White-Paper-:-Common-Trajectory-Prediction-Related-Terminology>.
 - [34] EUROCONTROL, ‘EUROCONTROL Specification for Trajectory Prediction’, Brussels, 2017.
 - [35] J. Bronsvort, ‘Contributions to trajectory prediction theory and its application to arrival management for air traffic control’, 2014.
 - [36] N. A. Stanton, D. Harris, and A. Starr, ‘The future flight deck: Modelling dual, single and distributed crewing options’, *Appl. Ergon.*, vol. 53, pp. 331–342, 2016.
 - [37] M. Soler, A. Olivares, E. Staffetti, and D. Zapata, ‘Framework for Aircraft Trajectory Planning Toward an Efficient Air Traffic Management’, *J. Aircr.*, vol. 49, no. 1, pp. 341–348, 2012.
 - [38] X. Dalmau, Ramon; Prats, ‘How much fuel and time can be saved in a perfect flight trajectory ? Continuous cruise climbs vs . conventional operations’, in *“Proceedings of the 6th International Congress on Research in Air Transportation (ICRAT)*, 2014.
 - [39] B. Roberson, ‘Fuel Conservation Strategies’, no. Ci, p. 19, 2007.
 - [40] A. Franco and D. Rivas, ‘Minimum-Cost Cruise at Constant Altitude of Commercial Aircraft Including Wind Effects’, *J. Guid. Control. Dyn.*, vol. 34, no. 4, pp. 1253–1260, 2012.
 - [41] S. Altus, ‘Effective Flight Plans Can Help Airlines Economize’, *AERO QUATERLY*, no. 35, pp. 27–30, 2009.
 - [42] F. J. Saez Nieto, *Navegación Aérea. Posicionamiento Guiado y Gestión del Tráfico Aéreo*. Garceta, 2012.
 - [43] S. Swierstra, C. Garcia-avello, J. Bronsvort, I. Bayraktutar, and G. McDonald, “‘ The Dawn of the Trajectory ’”, *Ninth USA/Europe Air Traffic Manag. Res. Dev. Semin.*, 2011.
 - [44] R. Walter, ‘Flight management systems’, *Digit. Avion. Handbook, Third Ed.*, Jan. 2017.
 - [45] J. Bronsvort, G. McDonald, M. Vilaplana, and J. A. Besada, ‘Two-Stage Approach to Integrated Air/Ground Trajectory Prediction’, *J. Guid. Control. Dyn.*, vol. 37, no. 6, pp. 2035–2039, 2014.
 - [46] M. V. Cook, *Flight Dynamics Principles*, 2nd ed. Cranfield: Elsevier, 2007.

References

- [47] C. Trajectory and P. Capability, ‘White Paper Common Trajectory Predictor Structure and Terminology’, no. January, pp. 1–25, 2010.
- [48] J. Garcia-Chico, R. Vivona, and K. Cate, ‘Characterizing Intent Maneuvers from Operational Data: Step Towards Trajectory Prediction Uncertainty Estimation’, no. August, 2012.
- [49] J. López-Leonés, M. A. Vilaplana, E. Gallo, F. A. Navarro, and C. Querejeta, ‘The aircraft intent description language: A key enabler for air-ground Synchronization in trajectory-based operations’, *AIAA/IEEE Digit. Avion. Syst. Conf. - Proc.*, pp. 1–12, 2007.
- [50] EUROCONTROL, *Trajectory Management Exchange Formats Definition (SESAR P10.02.02)*. Brussels, pp. 1–19.
- [51] M. A. Vilaplana, ‘REACT Project Overview’. Seville, 2008.
- [52] ICAO, *Manual on Air Traffic Management System Requirements*. 2008.
- [53] J. A. Besada, G. Frontera, J. Crespo, E. Casado, and J. Lopez-Leones, ‘Automated aircraft trajectory prediction based on formal intent-related language processing’, *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 3, pp. 1067–1082, 2013.
- [54] S. Swierstra, ‘Flight Object A Recommendation for Flight Script and Trajectory Description / Action Plan 16 Common Trajectory Prediction Capabilities’, 2006.
- [55] B. Musialek, C. Munafo, H. Ryan, and M. Paglione, ‘Literature Survey of Trajectory Predictor Technology, Technical Report DOT/FAA/TCTN11/1’, 2010.
- [56] R. Slattery, M. Field, and Y. Zhao, ‘Trajectory Synthesis for Air Traffic Automation’, vol. 20, no. 2, 1997.
- [57] A. Nuic, C. Poinso, M. Iagaru, E. Gallo, F. A. Navarro, and C. Querejeta, ‘Advanced Aircraft Performance Modeling for ATM: Enhancements to the Bada Model’, pp. 2.B.4-1-2.B.4–1, 2006.
- [58] A. Nuic, ‘User Manual for the Base of Aircraft Data (BADA) Revision 3.12 EEC Technical/Scientific Report No. 14/04/24-44’, 2014.
- [59] Z. Wang and D. Delahaye, ‘Short-term 4D Trajectory Prediction Using Machine Learning Methods’, *Sesar Innov. Days 2017*, no. November, pp. 1–9, 2017.
- [60] M. G. Hamed, D. Gianazza, M. Serrurier, and N. Durand, ‘Statistical prediction of aircraft trajectory: regression methods vs point-mass model’, in *10th USA/Europe ATM R&D Seminar*, 2013, p. pp.
- [61] S. Mondoloni, S. Swierstra, and M. Paglione, ‘Assessing trajectory prediction performance - Metrics definition’, in *AIAA/IEEE Digital Avionics Systems Conference - Proceedings*, 2005, vol. 1, pp. 3.C.1-31–13 Vol. 1.
- [62] R. M. Gómez Comendador, Victor Fernando; Arnaldo Valdés, A. Vidosavljevic, M. Sanchez

- Cidoncha, and S. Zheng, ‘Impact of Trajectories ’ Uncertainty in Existing ATC Complexity Methodologies and Metrics for DAC and FCA SESAR Concepts’, 2019.
- [63] Airbus; EUROCONTROL; INDRA; NATS; SKYGUIDE; THALES, ‘PEGASE - Demonstration Report (LSD.01.04)’, Brussels, 2016.
- [64] SESAR JU, ‘P05.05.02 D01 - Concept for use of AOC data to improve Trajectory Prediction’, Brussels, 2010.
- [65] J. Lopez Leones, ‘Definition of an aircraft intent description language for air traffic management applications’, *PQDT - UK Irel.*, no. February, 2008.
- [66] H. Huang, K. Roy, and C. Tomlin, ‘Probabilistic Estimation of State-Dependent Hybrid Mode Transitions for Aircraft Arrival Time Prediction’, *Proc. 제날 프로시딩즈/ 기술 ...*, no. August, 2012.
- [67] EUROCONTROL, ‘COPTRA - COMBINING PROBABLE TRAJECTORIES’, 2019. [Online]. Available: <https://www.sesarju.eu/projects/coptra>.
- [68] D. Z. Crida, ‘COPTRA-D2.1-Techniques-to-determine-trajectory-uncertainty-and-modelling-V1.1’, no. February, pp. 1–43, 2017.
- [69] E. J. Casado Magaña, ‘Trajectory Prediction Uncertainty Modelling for Air Traffic Management’, 2016.
- [70] E. Casado, C. Goodchild, and M. Vilaplana, ‘Identification and Initial Characterization of Sources of Uncertainty Affecting the Performance of Future Trajectory Management Automation Systems’, *Proc. 2nd Int. Conf. Appl. Theory Autom. Command Control Syst.*, pp. 170–175, 2012.
- [71] J. Sun, J. Ellerbroek, and J. M. Hoekstra, ‘WRAP: An open-source kinematic aircraft performance model’, *Transp. Res. Part C Emerg. Technol.*, vol. 98, no. November 2018, pp. 118–138, 2019.
- [72] C. M. Johnson, ‘Analysis of Top of Descent (TOD) uncertainty’, *AIAA/IEEE Digit. Avion. Syst. Conf. - Proc.*, pp. 2E3-1-2E3-10, 2011.
- [73] L. Stell, ‘Prediction of Top of Descent Location for Idle-thrust Descents’, in *Ninth USA/Europe Air Traffic Management Research and Development Seminar*, 2011, pp. 1–12.
- [74] G. Enea, J. Bronsvort, S. L. Mondoloni, and R. M. Sgorcea, ‘Advanced Trajectory Modeling with Air-Ground Data Exchange’, pp. 1–14, 2018.
- [75] E. Hollnagel, *FRAM: The Functional Resonance Analysis Method*. 2018.
- [76] N. Schefers, M. A. Piera, J. J. Ramos, and J. Nosedal, ‘Causal Analysis of Airline Trajectory Preferences to Improve Airspace Capacity’, *Procedia Comput. Sci.*, vol. 104, no. December 2016, pp. 321–328, 2016.

- [77] J. Sun, *Open Aircraft Performance Modeling Based on an Analysis of Aircraft Surveillance Data*. 2019.
- [78] A. Nuic, C. Poinot, M. Iagaru, E. Gallo, F. A. Navarro, and C. Querejeta, ‘ADVANCED AIRCRAFT PERFORMANCE MODELING FOR ATM: ENHANCEMENTS TO THE BADA MODEL BADA Model overview’, 2005.
- [79] R. Alligier and D. Gianazza, ‘Learning aircraft operational factors to improve aircraft climb prediction: A large scale multi-airport study’, *Transp. Res. Part C Emerg. Technol.*, vol. 96, no. January, pp. 72–95, 2018.
- [80] M. Gariel, A. N. Srivastava, and E. Feron, ‘Trajectory clustering and an application to airspace monitoring’, *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 4, pp. 1511–1524, 2011.
- [81] Z. Wang, M. Liang, and D. Delahaye, ‘A hybrid machine learning model for short-term estimated time of arrival prediction in terminal manoeuvring area’, *Transp. Res. Part C Emerg. Technol.*, vol. 95, pp. 280–294, Oct. 2018.
- [82] R. J. Hansman, H. Balakrishnan, R. DeLaura, M. Conde Rocha Murca, T. Reynolds, and R. Jordan, ‘Trajectory Clustering and Classification for Characterization of Air Traffic Flows’, *16th AIAA Aviat. Technol. Integr. Oper. Conf.*, no. June, pp. 1–15, 2016.
- [83] M. C. R. Murça, R. J. Hansman, L. Li, and P. Ren, ‘Flight trajectory data analytics for characterization of air traffic flows: A comparative analysis of terminal area operations between New York, Hong Kong and Sao Paulo’, *Transp. Res. Part C Emerg. Technol.*, vol. 97, no. February, pp. 324–347, 2018.
- [84] E. Salaün, M. Gariel, A. Vela, and E. Feron, ‘Aircraft Proximity Maps Based on Data-Driven Flow Modeling’, *J. Guid. Control. Dyn.*, vol. 35, no. 2, pp. 563–577, 2011.
- [85] J. Clarke, E. Feron, E. Salaun, M. Gariel, A. Vela, and W. Singhose, ‘Determining bounds on controller workload rates at an intersection’, in *American Control Conference (ACC), 2010*, 2014, pp. 3063–3068.
- [86] A. Marzuoli, M. Gariel, A. Vela, and E. Feron, ‘Data-Based Modeling and Optimization of En Route Traffic’, *J. Guid. Control. Dyn.*, vol. 37, no. 6, pp. 1930–1945, 2014.
- [87] J. M. Histon and R. J. Hansman, ‘Mitigating Complexity in Air Traffic Control: The Role of Structure-Based Abstractions’, *MIT Int. Cent. Air Transp. Dep. Aeronaut. Astronaut. Massachusetts Inst. Technol.*, no. August, 2008.
- [88] ICAO, *Manual on Global Performance of the Air Navigation System*. 2009.
- [89] M. P. Scheele and P. C. Siegmund, ‘Estimating Errors in Trajectory Forecasts Using Ensemble Predictions’, *J. Appl. Meteorol.*, vol. 40, no. 7, pp. 1223–1232, 2002.
- [90] P. Weitz, ‘Determination and visualization of uncertainties in 4D-trajectory prediction’,

-
- Integr. Commun. Navig. Surveill. Conf. ICNS*, pp. 1–9, 2013.
- [91] E. Hernandez, A. Valenzuela, and D. Rivas, ‘Probabilistic Aircraft Conflict Detection Considering Ensemble Weather Forecast’, *6th SESAR Innov. Days*, no. November, 2016.
 - [92] J. Sun, H. Vu, J. Ellerbroek, and J. M. Hoekstra, ‘Weather field reconstruction using aircraft surveillance data and a novel meteo-particle model’, *PLoS One*, vol. 13, no. 10, p. e0205029, Oct. 2018.
 - [93] D. Rivas, A. Franco, and A. Valenzuela, ‘Analysis of aircraft trajectory uncertainty using Ensemble Weather Forecasts’, *Eur. Conf. Aeronaut. Sp. Sci.*, 2017.
 - [94] A. Lee, S. Weygandt, B. Schwartz, and J. Murphy, ‘Performance of Trajectory Models with Wind Uncertainty’, *AIAA Model. Simul. Technol. Conf.*, pp. 1–16, 2012.
 - [95] E. Hernández-Romero, A. Valenzuela, and D. Rivas, ‘A probabilistic approach to measure aircraft conflict severity considering wind forecast uncertainty’, *Aerosp. Sci. Technol.*, vol. 86, pp. 401–414, 2019.
 - [96] C. M. Johnson, T. M. Corporation, A. Aviation, and S. Development, ‘ANALYSIS OF TOP OF DESCENT (TOD) UNCERTAINTY Experimental Design’, *2011 IEEE/AIAA 30th Digit. Avion. Syst. Conf.*, pp. 2E3-1-2E3-10, 2011.
 - [97] J. Niarchakou, S., Simon Selva, ‘EUROCONTROL ATFCM Operations Manual’, Brussels, 2017.
 - [98] R. Kimball, *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling (Second Edition)*. 1996.
 - [99] W. H. Inmon, *Building the data warehouse*. 2005.
 - [100] M. Kimball, Ralph; Ross, *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling (Second Edition)*. 2002.
 - [101] L. Leonardi *et al.*, ‘A general framework for trajectory data warehousing and visual OLAP’, *Geoinformatica*, vol. 18, no. 2, pp. 273–312, 2014.
 - [102] F. A. Administration, ‘Introduction to TCAS II’, 2011.
 - [103] European Organisation for the Safety of Air Navigation (EUROCONTROL), ‘Application of ASTERIX to ARTAS’, 2009.
 - [104] S. Ruiz, M. A. Piera, J. Nosedal, and A. Ranieri, ‘Strategic de-confliction in the presence of a large number of 4D trajectories using a causal modeling approach’, *Transp. Res. Part C Emerg. Technol.*, vol. 39, pp. 129–147, 2014.
 - [105] A. Jardines, M. Soler, and J. Garc, ‘Data-Driven Occupancy Prediction in Adverse Weather Conditions using Thunderstorm and Traffic Observations’, 2019, no. December.

- [106] Y. Huo, D. Delahaye, and M. Sbihi, ‘Integrated Traffic Flow Based Optimization of Airport and Terminal Area Civil Aviation University of China’, no. December, 2019.
- [107] E. Hern and D. Delahaye, ‘Metaheuristic approach to probabilistic aircraft conflict detection and resolution considering Ensemble Prediction Systems’, no. Cd, 2019.
- [108] G. A. Boy, *Orchestrating Human-Centered Design*, no. 1. Springer, 2013.
- [109] C. E. Verdonk Gallego, F. Javier, and S. Nieto, ‘Discussion on Complexity and TCAS Indicators for Coherent Safety Net Transitions’, in *6th SESAR Innovation Days*, 2016, no. November.
- [110] E. J. García González, ‘Development of a 3-Dimensional mathematical collision risk model based on recorded aircraft trajectories to estimate the safety level in high density en-route airspaces dissertation’, 2013.
- [111] D. Delahaye, S. Puechmorel, P. Tsiotras, and E. Feron, ‘Mathematical models for aircraft trajectory design: A survey’, *Lect. Notes Electr. Eng.*, vol. 290 LNEE, pp. 205–247, 2014.
- [112] M. Enriquez, ‘Identifying Temporally Persistent Flows in the Terminal Airspace via Spectral Clustering’, in *Tenth USA/Europe Air Traffic Management Research and Development Seminar (ATM2013) Airport*, 2013.
- [113] J. Shlens, ‘A Tutorial on Principal Component Analysis’, 2014.
- [114] X. Ester, Martin; Kriegel, Hans-Peter; Sander, J.; Xiaoewei, ‘A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise’, in *Second International Conference on Knowledge Discovery and Data Mining*, 1996, vol. 2, pp. 226–231.
- [115] M. Ankerst, M. M. Breunig, H. Kriegel, and J. Sander, ‘OPTICS: Ordering Points To Identify the Clustering Structure’, 1999, pp. 49–60.
- [116] N. Niu, J. Sander, X. Qin, A. Kovarsky, and Z. Lu, ‘Automatic Extraction of Clusters from Hierarchical Clustering Representations’, pp. 75–87, 2007.
- [117] A. Zimek, R. J. G. B. Campello, J. Sander, D. Moulavi, and P. A. Jaskowiak, ‘Density-Based Clustering Validation’, *Proc. 2014 SIAM Int. Conf. Data Min.*, no. i, pp. 839–847, 2014.
- [118] R. J. G. B. Campello, D. Moulavi, and J. Sander, ‘Density-based clustering based on hierarchical density estimates’, in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2013, vol. 7819 LNAI, no. PART 2, pp. 160–172.
- [119] C. E. Verdonk Gallego, V. F. Gómez Comendador, F. J. Sáez Nieto, and M. Martínez García, ‘Discussion on density-based clustering methods applied for automated identification of airspace flows’, in *AIAA/IEEE Digital Avionics Systems Conference - Proceedings*, 2018, vol. 2018-Septe, pp. 1–10.

-
- [120] L. Breiman, ‘ST4_Method_Random_Forest’, *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
 - [121] A. ENAIRE, ‘Aeronautical Information Publication’, *Insignia*, 2016. [Online]. Available: https://ais.enaire.es/insignia/index_en.aspx. [Accessed: 29-Oct-2018].
 - [122] W. A. Taylor, ‘Change-Point Analysis: A Powerful New Tool For Detecting Changes’, 2000.
 - [123] A. Eckstein, ‘Automated flight track taxonomy for measuring benefits from performance based navigation’, *Proc. 2009 Integr. Commun. Navig. Surveill. Conf. ICNS 2009*, pp. 1–12, 2009.
 - [124] K. Levenberg, ‘A method for the solution of certain non-linear problems in least squares’, *Q. Appl. Math.*, vol. 2, no. 2, pp. 164–168, 2016.
 - [125] D. W. Marquardt, ‘An Algorithm for Least-Squares Estimation of Nonlinear Parameters’, *J. Soc. Ind. Appl. Math.*, vol. 11, no. 2, pp. 431–441, 2005.
 - [126] M. T. Hagan and M. B. Menhaj, ‘Training Feedforward Networks with the Marquardt Algorithm’, *IEEE Trans. Neural Networks*, vol. 5, no. 6, pp. 989–993, 1994.
 - [127] C. J. M. Maas and J. J. Hox, ‘Sufficient sample sizes for multilevel modeling’, *Methodology*, vol. 1, no. 3, pp. 86–92, Jan. 2005.
 - [128] D. Delahaye and S. Puechmorel, ‘Air traffic complexity: towards intrinsic metrics’, in *3rd USA/Europe Air Traffic Management R&D Seminar*, 2000, no. June, pp. 1–11.
 - [129] D. Delahaye, S. Puechmorel, R. J. Hansman, and J. M. Histon, ‘Air Traffic Complexity Map Based on Non Linear Dynamical Systems’, in *5th USA/Europe ATM R&D Seminar*, 2003.
 - [130] D. Delahaye and S. Puechmorel, ‘Air traffic complexity based on dynamical systems’, *Proc. IEEE Conf. Decis. Control*, pp. 2069–2074, 2010.
 - [131] J. Djokic, B. Lorenz, and H. Fricke, ‘Air traffic control complexity as workload driver’, *Transp. Res. Part C Emerg. Technol.*, vol. 18, no. 6, pp. 930–936, 2010.
 - [132] M. Prandini, L. Piroddi, S. Puechmorel, and S. L. Brázdilová, ‘New Generation Air Traffic Management Systems’, *Transportation (Amst.)*, vol. 12, no. 3, pp. 809–818, 2011.
 - [133] M. Radanovic, M. A. Piera Eroles, T. Koca, and J. J. Ramos Gonzalez, ‘Surrounding traffic complexity analysis for efficient and stable conflict resolution’, *Transp. Res. Part C Emerg. Technol.*, vol. 95, pp. 105–124, Oct. 2018.
 - [134] C. D. Wickens, ‘Multiple Resources and Mental Workload’, *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 50, no. 3, pp. 449–455, 2008.
 - [135] G. M. Flynn, ‘European Organisation for the Safety of Air Navigation Eurocontrol Experimental Centre Cognitive Complexity in Air Traffic Control a Literature Review’, *EEC note*, no. 04, 2004.

- [136] N. Suárez, P. López, E. Puntero, and S. Rodriguez, ‘Quantifying Air Traffic Controller Mental Workload’, *Fourth SESAR Innov. Days*, no. November, pp. 1–6, 2014.
- [137] R. A. Paielli and H. Erzberger, ‘Conflict Probability for Free Flight’, *J. Guid. Control. Dyn.*, vol. 20, no. 3, pp. 588–596, 2008.
- [138] H. Erzberger, R. A. Paielli, D. R. Isaacson, and M. M. Eshow, ‘Conflict detection and resolution in the presence of prediction error’, in *1st USA/Europe Air Traffic Management R&D Seminar, Saclay, France*, 1997, pp. 17–20.
- [139] R. Irvine, ‘A geometrical approach to conflict probability estimation’, *Encounter*, vol. 10, no. 2, pp. 1–15, 2001.
- [140] L. Piroddi and M. Prandini, ‘A geometric approach to air traffic complexity evaluation for strategic trajectory management’, *Proc. IEEE Conf. Decis. Control*, pp. 2075–2080, 2010.
- [141] S. Ayhan, P. Costas, and H. Samet, ‘Prescriptive analytics system for long-range aircraft conflict detection and resolution’, in *Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2018, pp. 239–248.
- [142] G. Chaloulos and J. Lygeros, ‘Effect of Wind Correlation on Aircraft Conflict Probability’, *J. Guid. Control. Dyn.*, vol. 30, no. 6, pp. 1742–1752, 2007.
- [143] D. Delahaye *et al.*, ‘Mathematical Models for Aircraft Trajectory Design : A Survey’, *Lect. notes Electr. Eng.*, vol. 290, pp. 205–247, 2014.
- [144] A. de Leege, M. van Paassen, and M. Mulder, ‘A Machine Learning Approach to Trajectory Prediction’, in *AIAA Guidance, Navigation, and Control (GNC) Conference*, 2013.
- [145] S. Ayhan and H. Samet, ‘Aircraft Trajectory Prediction Made Easy with Predictive Analytics’, in *KDD ’16 Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 21–30.
- [146] Y. Pang, H. Yao, J. Hu, and Y. Liu, ‘A Recurrent Neural Network Approach for Aircraft Trajectory Prediction with Weather Features From Sherlock’, *AIAA Aviat. 2019 Forum*, no. June, pp. 1–14, 2019.
- [147] C. E. Verdonk Gallego, V. F. Gómez Comendador, F. J. Sáez Nieto, G. Orenga Imaz, and R. M. Arnaldo Valdés, ‘Analysis of air traffic control operational impact on aircraft vertical profiles supported by machine learning’, *Transp. Res. Part C Emerg. Technol.*, vol. 95, no. July 2017, pp. 883–903, 2018.
- [148] W. Liu and I. Hwang, ‘Probabilistic Trajectory Prediction and Conflict Detection for Air Traffic Control’, *J. Guid. Control. Dyn.*, vol. 34, no. 6, pp. 1183–1188, 2011.
- [149] S. Hong and K. Lee, ‘Trajectory Prediction for Vectored Area Navigation Arrivals’, *J. Aerosp. Inf. Syst.*, vol. 12, no. 7, pp. 490–502, 2015.

-
- [150] X. Dalmau Codina, Ramon; Pérez-Batlle, Marc; Prats Menéndez, ‘Real-time Identification of Guidance Modes in Aircraft Descents Using Surveillance Data’, in *37th IEEE/AIAA Digital Avionics Systems Conference (DASC’18)*, 2018, pp. 1295–1304.
- [151] H. Huang, K. Roy, and C. Tomlin, ‘Probabilistic Estimation of State-Dependent Hybrid Mode Transitions for Aircraft Arrival Time Prediction’, in *AIAA Guidance, Navigation and Control Conference and Exhibit*, 2007, no. August.
- [152] A. del R. Barragán Montes, ‘FINDING A CAUSAL METHOD FOR THE PREDICTION AND ASSESSMENT OF SAFETY PERFORMANCE IN AN AIR TRAFFIC MANAGEMENT SERVICE’, Universidad Politécnica de Madrid, 2017.
- [153] R. B. Montes, A. D. Rojas, V. F. G. Comendador, R. A. Valdés, and L. P. Sanz, ‘A novel predictability performance metric and its forecast using machine learning techniques’, *AIAA/IEEE Digit. Avion. Syst. Conf. - Proc.*, vol. 2018-Sept, 2018.
- [154] R. Alligier, D. Gianazza, and N. Durand, ‘Predicting Aircraft Descent Length with Machine Learning’, *7th Int. Conf. Res. Air Transp. (ICRAT), Philadelphia, United States*, pp. 1–8, 2016.
- [155] M. Radanovic, M. A. Piera Eroles, K. Thimjo, and F. J. Saez Nieto, ‘Self-Reorganized Supporting Tools for Conflict Resolution in High-Density Airspace Volumes’, in *Twelfth USA/Europe Air Traffic Management Research and Development Seminar*, 2017, p. 10.
- [156] S. Bae and C. Lee, ‘A New Multiple Commercial Aircraft Routing and Scheduling Algorithm in Terminal Manoeuvring Area’, *2018 IEEE/AIAA 37th Digit. Avion. Syst. Conf.*, pp. 1–9, 2018.
- [157] J. Knight C, ‘Safety Critical Systems : Challenges and Directions 3 . WHAT GOES WRONG WITH SAFETY-’, pp. 547–550, 2002.
- [158] EUROCONTROL, ‘EUROCONTROL Specification for Airspace Management (ASM) Support System Requirements supporting the ASM processes at local and FAB level Part I - Baseline Requirements EUROCONTROL-SPEC-166’. pp. 0–55, 2017.
- [159] C. E. Verdonk Gallego, V. F. Gómez Comendador, M. A. Amaro Carmona, R. M. Arnaldo Valdés, F. J. Sáez Nieto, and M. García Martínez, ‘A machine learning approach to air traffic interdependency modelling and its application to trajectory prediction’, *Transp. Res. Part C Emerg. Technol.*, vol. 107, no. November 2018, pp. 356–386, Oct. 2019.

ACRONYMS

ACC	Air Traffic Control Centre
AFCS	Automatic Flight Control System
AIDL	Aircraft Intention Description Language
AIRAC	Aeronautical Information Regulation And Control
AMAN	Arrival Manager
ANN	Artificila Neural Network
ANOVA	ANalysis Of VAriance
ANSP	Air Navigation Service Provider
AOC	Aircraft Operator Centre
APM	Aircraft Performance Model
ASM	Airspace Management
ATC	Air Traffic Control Centre
ATCO	Air Traffic Controller
ATFCM	Air Traffic Flow and Capacity Management
ATFM	Air Traffic Flow Management
ATM	Air Traffic Management
ATMRPP	ICAO ATM Requirements and Performance Panel
ATMS	Air Traffic Management System
ATS	Air Traffic Services
ATSU	Air Traffic Services Unit

Acronyms

BADA	Base of Aircraft Data
-------------	-----------------------

CAS	Calibrated Airspeed
------------	---------------------

CDF	Cumulative Distribution Function
------------	----------------------------------

CFL	Cleared Flight Level
------------	----------------------

CG	Centre of Gravity
-----------	-------------------

CNS	Communication, Navigation and Surveillance
------------	--

CPA	Closest Ppoint of Approach
------------	----------------------------

CWP	Controller Working Position
------------	-----------------------------

DAC	Dynamic Airspace Configuration
------------	--------------------------------

DBSCAN	Density-based clustering algorithm
---------------	------------------------------------

DWH	Data warehouse
------------	----------------

ECAC	European Civil Aviation Conference
-------------	------------------------------------

ESF	Energy Sharing Factor
------------	-----------------------

ETL	Extract, Transform and Load
------------	-----------------------------

FAA	Federal Aviation Authority (USA)
------------	----------------------------------

FCS	Flight Control System
------------	-----------------------

FDPS	Flight Data Processing System
-------------	-------------------------------

FL	Flight Level
-----------	--------------

FMS	Flight Management System
------------	--------------------------

FPL	Flight Plan
------------	-------------

FRAM	The Functional Resonance Analysis
-------------	-----------------------------------

HIM	Horizontal Interdependency Matrix
HMI	Human-Machine Interface
ICAO	International Civil Aviation Organisation
IQR	Interquartile Range
KDD	Knowledge Discovery Process
LAT	Look-ahead Time
LECB	Barcelona ACC
LEIB	Ibiza Airport
LEPA	Palma de Majorca Airport
LRD	Linear Rate of Descent
MCP	Mode Control Panel
ML	Machine Learning
MLE	Maximum Likelihood Estimator
MLM	Multilevel Linear Models
MSL	Mean Sea Level
NM	Network Manager
OCM	Operational Context Model
OPTICS	Ordering Points To Identify the Clustering Structure
PC	Principal Component
PCA	Principal Components Analysis
PDF	Probability Density Function

Acronyms

PIC	Pilot-in-Command
PMM	Point-mass Model
RBT	Reference Business Trajectory
RDPS	Radar Data Processing System
RFL	Requested Flight Level
RMSE	Root mean squared error
ROCD	Rate of Climb/Descent
SBT	Shared Business Trajectory
SES	Single European Sky
SESAR	Single European Sky ATM Research
SID	Standard Instrumental Departure
STAR	Standard Terminal Arrival Route
STCA	Short-Term Conflict Alert
SVD	Singular-Value Decomposition
TBO	Trajectory-Based Operations
TCAS	Traffic Collision Avoidance System
TCI	Trajectory Computation Infrastructure
TMA	Terminal Maneuvering Area
TOA	Top of Ascent
TOC	Top of Climb
TOD	Top of Descent

TP	Trajectory Prediction / Trajectory Predictor
-----------	--

UPM	User Preferences Model
------------	------------------------

XFL	Exit Flight Level
------------	-------------------

This Page is Intentionally Left Blank

NOTATION CHAPTER 5

Indexes

i, j, k	Major airspace flows
l, m, n	Recurrent patterns within a major airspace flow
p, o, q	Total number of recurrent patterns within a major airspace flow
r, s	Flight r or Flight s within a traffic sample.

Notation (order of appearance)

η	Flown Trajectory
Lat	Latitude
Lng	Longitude
t	Time
Hdg	Heading
Vg	Ground Velocity
Vz	Vertical speed
ϖ	Cleared Heading
ζ	Cleared Flight Level
v	Cleared Horizontal Speed
ϱ	Cleared Vertical Speed
Λ^0	Sample of trajectories
λ_i^0	Set of trajectories forming the major airspace flow i
α_{il}^0	Set of trajectories within λ_i^0 which constitute the recurrent pattern l within λ_i^0
$\mathcal{H}_{il}^0$	Horizontal Central Trajectory of the recurrent pattern α_{il}^0
γ_{il}	Arc-length parameter associated to $\mathcal{H}_{il}^0$
Γ_{il}	Arc-length space associated to $\mathcal{H}_{il}^0$
L_{il}	Total length of $\mathcal{H}_{il}^0$
$\underline{X}$	Feature vector for the horizontal trajectory predictor
$\widetilde{\eta}_{r,i}$	Set of predicted trajectories for the aircraft r for a major airspace flow i
$\widetilde{\eta}_{r,il}$ flow i	Predicted trajectory for the aircraft r for a recurrent pattern l of a major airspace flow i
$\tau_{il}^{\gamma_{il}}$	Predicted time τ to arrive to a given location γ_{il} along $\mathcal{H}_{il}^0$
$\mathcal{T}_{il}^{\gamma_{il}}$	Random Variable associated to $\tau_{il}^{\gamma_{il}}$
$\varepsilon_{il}^{\gamma_{il}}$	Residuals associated to the TP for the recurrent pattern α_{il}^0 at the location γ_{il}
$\sigma_{il}^{\gamma_{il}}$	Standard deviation of the $\varepsilon_{il}^{\gamma_{il}}$ assuming a normal distribution.

$\mu_{il}^{\gamma_{il}}$	Expected value of $\mathcal{T}_{il}^{\gamma_{il}}$
$\Delta\mathcal{T}_{il,jm}^{\gamma_{il},\gamma_{jm}}$	Random variable associated to the time difference with which two aircraft are arriving to two locations γ_{il} and γ_{jm} respectively.
Z	Auxiliar random variable to transform $\mathcal{T}_{il}^{\gamma_{il}} - \mathcal{T}_{jm}^{\gamma_{jm}}$
$f_{\Delta\mathcal{T}_{il,jm}^{\gamma_{il},\gamma_{jm}}}$	Probability Distribution Function associated to $\Delta\mathcal{T}_{il,jm}^{\gamma_{il},\gamma_{jm}}$
$D_{il,jm}$	Function defining the distance between two aircraft if they were at γ_{il} and γ_{jm} respectively.
Δt_{Thres}	Temporal threshold
D_H	Horizontal Distance threshold
$\Pi_{\Delta t, D_H}^{il,jm}$	Interdependency Probability Measure associated to two aircraft flying recurrent patterns α_{il}^0 and α_{jm}^0 , given temporal and distance thresholds.
$HIM_{r,s}$	Horizontal Interdependency Matrix for two aircraft r and s
ρ	Number of FL that an aircraft has descended from the RFL.
γ_{Tod}	Location along $\mathcal{H}_{il}^0$ when an aircraft starts descending
θ	Time since the aircraft started descending
t_{il}^{ρ,γ_k}	Time for descending ρ FLs if an aircraft starts its descent phase at γ_k when flying along $\mathcal{H}_{il}^0$
T_{il}^{ρ,γ_k}	Random variable associated to t_{il}^{ρ,γ_k}
$\Theta_{il}^{\rho,\gamma_{il}^k}$	Random variable associated to the time θ required to descend ρ FLs if an aircraft starts its descent phase at γ_k when flying along $\mathcal{H}_{il}^0$