Transportation Research Part C: Emerging Technologies, Volume 107, October 2019, pp. 356-386
DOI: 10.1016/j.trc.2019.08.015

A Machine Learning approach to Air Traffic Interdependency modelling and its application to Trajectory Prediction

Christian Eduardo Verdonk Gallego*,[1] ,

Dr Víctor Fernando Gómez Comendador[1],

Manuel Angel Amaro Carmona[2],

Dr Rosa María Arnaldo Valdés[1]

Dr Francisco Javier Sáez Nieto[2],

Miguel García Martínez[3],

[1] *Aeronautical Systems, Air Transport and Airports Department, Universidad Politécnica de Madrid, Madrid, Spain*
[2] *Centre for Aeronautics, School of Aerospace, Transport and Manufacturing. Cranfield University, Cranfield, United Kingdom*
[3] *CRIDA A.I.E., Madrid, Spain*

Abstract

Air Traffic Management is evolving towards a Trajectory-Based Operations paradigm. Trajectory prediction will hold a key role supporting its deployment, but it is limited by a lack of understanding of air traffic associated uncertainties, specifically contextual factors.

Trajectory predictors are usually based on modelling aircraft dynamics based on intrinsic aircraft features. These aircraft operate within a known air route structure and under given meteorological conditions. However, actual aircraft trajectories are modified by the air traffic control depending on potential conflicts with other traffics. This paper introduces surrounding air traffic as a feature for ground-based trajectory prediction.

The introduction of air traffic as a contextual factor is addressed by identifying aircraft which are likely to lose the horizontal separation. For doing so, this paper develops a probabilistic horizontal interdependency measure between aircraft supported by machine learning algorithms, addressing time separations at crossing points.

Then, vertical profiles of flight trajectories are characterised depending on this factor and other intrinsic features. The paper has focused on the descent phase of the trajectories, using datasets corresponding to an en-route Spanish airspace volume.

The proposed interdependency measure demonstrates to identify in advance conflicting situations between pairs of aircraft for this use case. This is validated by identifying associated air traffic control actions upon them and their impact on the vertical profile of the trajectories.

Finally, a trajectory predictor for the vertical profile of the trajectory is developed, considering the interdependency measure and other operational factors. The paper concludes that the air traffic can be included as a factor for the trajectory prediction, impacting on the location of the top of descent for the specific case which has been studied.

Keywords

Machine Learning; Trajectory Prediction; Air Traffic Management; Conflict; Vertical profile; Air Traffic Control; Flows; Artificial Neural Networks

Published by Elsevier. This is the Author Accepted Manuscript issued with: Creative Commons Attribution Non-Commercial No Derivatives License (CC:BY:NC:ND 4.0).
The final published version (version of record) is available online at DOI:10.1016/j.trc.2019.08.015. Please refer to any applicable publisher terms of use

1. Introduction

Air Traffic Management (ATM) is evolving towards a Trajectory-Based Operations (TBO) paradigm. Trajectory prediction (TP) is at the core of the different components of the ATM system (International Civil Aviation Organization, 2005). The performance of the trajectory prediction affects the overall performance of the ATM system (Sáez Nieto, 2016), but it is limited by a lack of understanding of TP uncertainties, specifically contextual factors (Schuster and Ochieng, 2014).

Contextual factors are usually associated with meteorological conditions and system constraints. The prediction of weather factors, mainly wind, has been addressed by academia and industry (Andrienko et al., 2014; Sun et al., 2018), and also the weather impact in different contexts (Franco et al., 2018; González-Arribas et al., 2017).

Air traffic as a contextual factor is usually considered for trajectory management, with longer prediction horizons to operations. For example, Air Traffic Flow and Capacity Management (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 (Niarchakou, S., Simon Selva, 2017). An airspace sector is a volume of interest controlled by an air traffic controller (ATCo).

In this sense, a planned trajectory represents the most likely behaviour of a flight through a volume of interest (EUROCONTROL, 2017). It is derived from the integration of data sources such as flight plans, ATC procedures, or the aircraft historical behaviour. The outcome of this process is the intent of the aircraft (Besada et al., 2013).

Aircraft intents are used to generate aircraft expected trajectories that usually deviate with respect to the actual ones due to the stochastic nature of the predictors (Casado Magaña, 2016). A trajectory prediction can be computed by systems or be based on cognitive abstractions of the air traffic controllers (Histon and Hansman, 2008). The ATC service, fulfilling the separation management function of ATM, removes potential conflicts between aircraft (García González, 2013) by issuing tactical instructions. Those instructions affect the trajectories downstream impacting on the ATM trajectory management function. For example, an early tactical descent instruction may impact on the Arrival Manager (AMAN) function due to the variation of the aircraft performance.

The objective of this paper is to introduce surrounding air traffic as a contextual factor for trajectory predictors. Firstly, the paper 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 well-known machine learning (ML) techniques for trajectory computation and a statistic-based interdependency definition.

The proposed methodology is presented in Section 2. Section 3 and Section 4 present the dataset, the results and a discussion about them. Finally, Section 5 presents the conclusions of this paper and proposes further work.

2. Materials and Methods

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. Approaches vary between algorithmic ones (Delahaye et al., 2003; Delahaye and Puechmorel, 2010, 2000; Djokic et al., 2010; Prandini et al., 2011; Radanovic et al., 2018; Verdonk Gallego et al., 2016) and human-factors based ones (Flynn, 2004; Suárez et al., 2014; Wickens, 2008).

Interdependency relates to potential conflicts between aircraft. Probabilistic conflict detection has been addressed in past and recent research. (Paielli and Erzberger, 2008) proposed a probabilistic approach for conflict detection, which was later extended to include errors in the trajectory prediction (Erzberger et al., 1997). There have been numerous research works that have extended or modified this approach, such as (Irvine, 2001) or (Piroddi and Prandini, 2010). (Hernández-Romero et al., 2019) 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 are 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, (Ayhan et al., 2018) 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.

This paper 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 (Histon and Hansman, 2008), where ATCOs concentrate on critical points for controlling the air traffic. Secondly, this paper addresses the probabilistic definition of the interdependency between two aircraft considering TPs' performance instead of deviations from the actual trajectories. This approach modifies (Erzberger et al., 1997) and others (Chaloulos and Lygeros, 2007) by considering the uncertainty related to the temporal instead the spatial dimension.

Trajectory predictors will be based on ML techniques influenced by ATCo cognitive abstractions (Histon and Hansman, 2008). Thus, the approach followed in (Verdonk Gallego et al., 2018b) is generalised to every flow within the airspace volume of interest. This work follows the "major flow" definition given by

(Delahaye et al., 2014). 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 paper exploits this horizontal interdependency measure. Firstly, vertical profiles and Top of Descent (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 1 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. This TP logical breakdown is based on (EUROCONTROL, 2017).

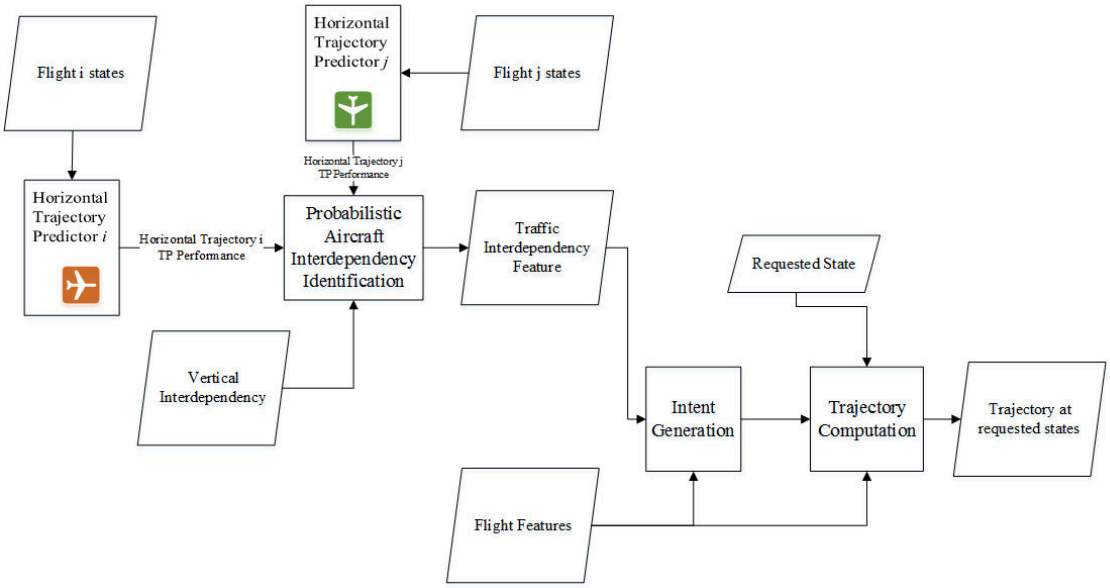


Figure 1 Trajectory Predictor Logical Breakdown

The following subsections detail the methods used in each step of the methodology. Section 2.1 covers the determination of the aircraft expected trajectories. Then, the derivation of the interdependency measure is conducted in Sections 2.2 and 2.3. Finally, the trajectory vertical profile is modelled based on the TOD location and the descent profile (Section 2.4), where the interdependency measure may affect both (Section 2.5).

2.1. 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 $\mathbb{K}^6$ x $\mathbb{O}^4 \in \mathbb{R}^{10}$. 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),

$$\eta(t) = \{lat(t), lng(t), FL(t), Hdg(t), Vg(t), Vz(t), \varpi(t), \upsilon(t), \zeta(t), \varrho(t)\}, \quad (1)$$

where ϖ, ζ, υ and ϱ corresponds to the cleared heading, cleared flight level, horizontal speed and cleared vertical speed respectively.

Let us consider an initial sample of trajectories $\Lambda^0 = \{\eta_1(t), \cdots, \eta_r(t), \cdots \eta_s(t), \cdots \eta_N(t)\}$, being N the sample size, which have flown through a given airspace. Such a set can be clustered by following Gariel's methodology (Gariel et al., 2011), modified in its last step to include an implementation of RDBSCAN* (Verdonk Gallego et al., 2018a), 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

$$\Lambda^0 = \bigcup_{i=1}^{I} \lambda_i^0, \tag{2}$$

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:

$$\lambda_i^0 = \bigcup_{l=1}^{p} \alpha_{il}^0. \tag{3}$$

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 2.

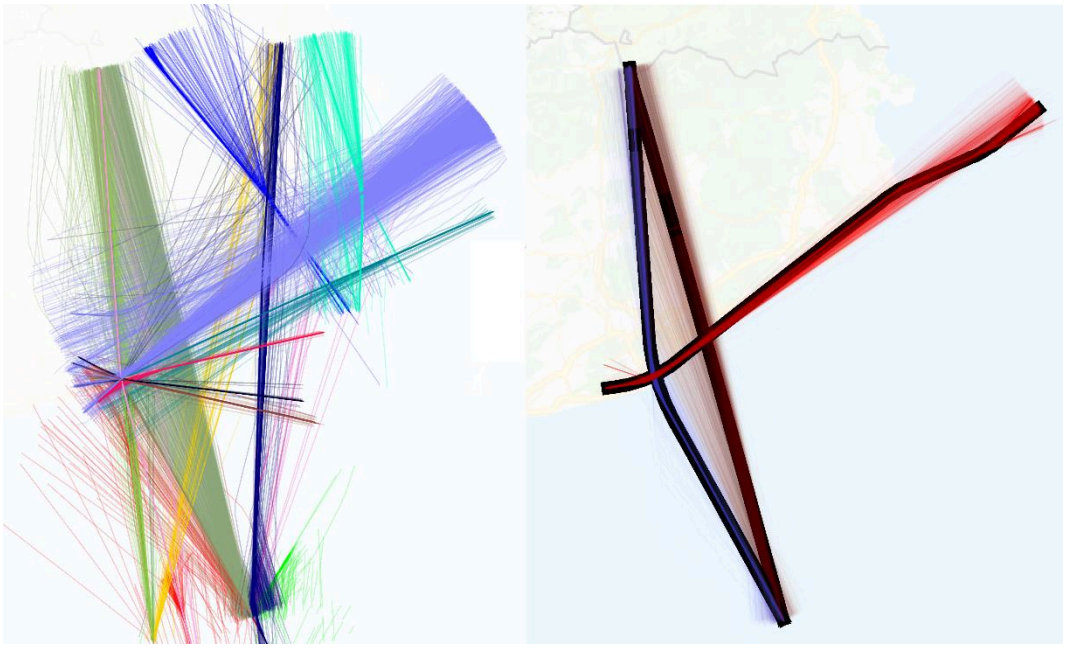


Figure 2 Main Flows – LECBCCC Sector, Barcelona ATCC and recurrent patterns associated to the main flows.

For each recurrent pattern α_{il}^0 we can define 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, $\boldsymbol{\hbar_{il}^0}$. 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 $\boldsymbol{\hbar_{il}^0}$ 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 $\boldsymbol{\hbar_{il}^0}$, 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 $\boldsymbol{\hbar_{il}^0}$. Thus, $\mathcal{H}$ is defined as:

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

In addition, we can define $\boldsymbol{\hbar_{il}^0}$ in terms of its arc-length parameter γ_{il} (Delahaye et al., 2014). The origin of $\boldsymbol{\hbar_{il}^0}$ is γ_{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 γ_{il} =0, by considering the horizontal projection of the initial state (position and velocity) of the trajectory on $\boldsymbol{\hbar_{il}^0}$. 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 $\boldsymbol{\hbar_{il}^0}$ can be based on the Principal Component Analysis (PCA) decomposition which is used in the clustering methodology (see (Gariel et al., 2011)). Thus, the transformed coordinates corresponding to the nearest point to the median principal components for a given α_{il}^0 is considered as the $\boldsymbol{\hbar_{il}^0}$ 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 (1) is collapsed to

$$\eta_r(\gamma_{il}) \sim \{\gamma_{il}, t(\gamma_{il}), Vg(\gamma_{il}), FL(\gamma_{il}), Vz(\gamma_{il}), \varpi(\gamma_{il}), \upsilon(\gamma_{il}), \zeta(\gamma_{il}), \varrho(\gamma_{il})\}, \tag{5}$$

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 paper, 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 $\boldsymbol{\hbar}_{il}^{0}$.

Several TP approaches may be applied, such as ML-enhanced point-mass models TP (Alligier et al., 2015, 2013) or data-driven approaches (Alligier and Gianazza, 2018; de Leege et al., 2013; Tastambekov et al., 2013). In this work, we apply an Artificial Neural Net (ANN) TP, following previous works in the literature (Verdonk Gallego et al., 2018b; Wang et al., 2018).

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 3. The output layer corresponds to the predicted time τ.

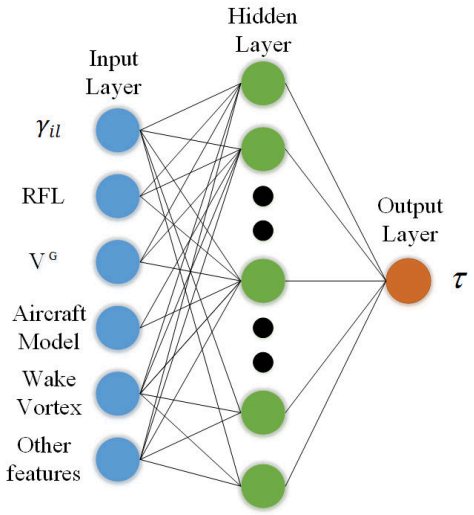


Figure 3 Artificial Neural Net Predictor for estimating the time (τ) of overflying γ_{il}.

2.2. 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 (Hernandez et al., 2016) and strategic level (Ayhan and Samet, 2016; Pang et al., 2019). This paper 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, $\{\alpha_{il}^{0}, \alpha_{im}^{0}, \alpha_{in}^{0}\}$ we can introduce a set of potential trajectories for each flight:

$$\widetilde{\eta_{r,\imath}} = \{\widetilde{\eta_{r,il}}(\gamma_{il}), \widetilde{\eta_{r,im}}(\gamma_{im}), \widetilde{\eta_{r,in}}(\gamma_{in})\}, \quad (6)$$

Each recurrent pattern, $\{\alpha_{il}^0, \alpha_{im}^0, \alpha_{in}^0\}$, has been associated with a horizontal central trajectory, $(\boldsymbol{\hbar_{il}^0}, \boldsymbol{\hbar_{im}^0}, \boldsymbol{\hbar_{in}^0})$, 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, we can define the predicted time τ to arrive to a given γ_{il} as $\tau_{r,il}(\gamma_{il})$ in terms of the residual distribution of the specific TP. If we consider now $\mathcal{T}_{r,il}(\gamma_{il})$ as a random variable defined as

$$\begin{aligned} &\mathcal{T}_{r,il}(\gamma_{il}): \Omega \to \mathbb{R} \\ &\mathcal{T}_{r,il}(\gamma_{il}) \sim ANN_{il}(\gamma_{il}; \underline{X}) + \varepsilon_{il}(\gamma_{il}) \end{aligned} \quad (7)$$

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

Let us consider the definition of Critical Point as given by (Histon and Hansman, 2008). 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. This interdependency definition focuses on crossing 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 $\boldsymbol{\hbar_{il}^0}$ and $\boldsymbol{\hbar_{jm}^0}$ respectively, we can identify two sequences of random variables associated to $\widetilde{\eta_{r,il}}$ and $\widetilde{\eta_{s,jm}}$, given by:

$$\begin{aligned} &\widetilde{\eta_{r,il}}: \{\mathcal{T}_{r,il}(0), \dots, \mathcal{T}_{r,il}(\gamma_{il}), \dots, \mathcal{T}_{r,il}(L_{il})\} = \{\mathcal{T}_{il}^0, \dots, \mathcal{T}_{il}^{\gamma_{il}}, \dots \mathcal{T}_{il}^{L_{il}}\} \\ &\widetilde{\eta_{s,jm}}: \{\mathcal{T}_{s,jm}(0), \dots, \mathcal{T}_{s,jm}(\gamma_{il}), \dots, \mathcal{T}_{s,jm}(L_{il})\} = \{\mathcal{T}_{jm}^0, \dots, \mathcal{T}_{jm}^{\gamma_{jm}}, \dots \mathcal{T}_{jm}^{L_{jm}}\} \end{aligned} \quad (8)$$

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

$$\begin{aligned} \varepsilon_{il}(\gamma_{il}) = \varepsilon_{il}^{\gamma_{il}} \sim N(0, (\sigma_{il}^{\gamma_{il}})^2) \\ \varepsilon_{jm}(\gamma_{jm}) = \varepsilon_{jm}^{\gamma_{jm}} \sim N\left(0, (\sigma_{jm}^{\gamma_{jm}})^2\right). \end{aligned} \quad (9)$$

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

$$\begin{aligned} \mathcal{T}_{il}^{\gamma_{il}} \sim N(\mu_{il}^{\gamma_{il}}, (\sigma_{il}^{\gamma_{il}})^2), \\ \mu_{il}^{\gamma_{il}} = ANN_{il}(\gamma_{il}; \vec{X}), \end{aligned} \quad (10)$$

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

$$\Delta\mathcal{T}(\gamma_{il},\gamma_{jm}) = \Delta\mathcal{T}_{il,jm}^{\gamma_{il},\gamma_{jm}} = \left|\mathcal{T}_{il}^{\gamma_{il}} - \mathcal{T}_{jm}^{\gamma_{jm}}\right|. \tag{11}$$

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

$$\mathcal{T}_{il}^{\gamma_{il}} - \mathcal{T}_{jm}^{\gamma_{jm}} \sim N\left(\mu_{il}^{\gamma_{il}} - \mu_{jm}^{\gamma_{jm}}, \left(\sigma_{il}^{\gamma_{il}}\right)^2 + \left(\sigma_{jm}^{\gamma_{jm}}\right)^2\right). \tag{12}$$

A new random variable *Z* can be defined so that:

$$Z = \frac{\left(\mathcal{T}_{il}^{\gamma_{il}} - \mathcal{T}_{jm}^{\gamma_{jm}}\right) - \left(\mu_{il}^{\gamma_{il}} - \mu_{il}^{\gamma_{jm}}\right)}{\sqrt{\left(\sigma_{il}^{\gamma_{il}}\right)^2 + \left(\sigma_{jm}^{\gamma_{jm}}\right)^2}} \tag{13}$$

Therefore,

$$\begin{aligned} &\mathcal{T}_{il}^{\gamma_{il}} - \mathcal{T}_{jm}^{\gamma_{jm}} = \sigma_Z\left(Z + \frac{\mu_Z}{\sigma_Z}\right), \\ &\mu_Z = \mu_{il}^{\gamma_{il}} - \mu_{jm}^{\gamma_{il}}, \\ &{\sigma_Z}^2 = (\sigma_{il}^{\gamma_{il}})^2 + (\sigma_{jm}^{\gamma_{jm}})^2. \end{aligned} \tag{14}$$

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

$$\Delta\mathcal{T}_{il,jm}^{\gamma_{il},\gamma_{jm}} \sim \left|\mathcal{T}_{il}^{\gamma_{il}} - \mathcal{T}_{jm}^{\gamma_{jm}}\right| \sim \sqrt{(\mathcal{T}_{il}^{\gamma_{il}} - \mathcal{T}_{jm}^{\gamma_{jm}})^2} = \sigma_Z\sqrt{\left(Z + \frac{\mu_Z}{\sigma_Z}\right)^2} \tag{15}$$

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:

$$f(y)dy = \frac{\sqrt{y}}{\sqrt{2\pi}}e^{\frac{1}{2}(-\lambda-y)}\cosh\left(\sqrt{\lambda y}\right)\frac{dy}{y}, y > 0 \tag{16}$$

Substituting y by z^2, then:

$$f(y)dy = f(z^2)d(z^2) = \frac{\sqrt{z^2}}{\sqrt{2\pi}}e^{\frac{1}{2}(-\lambda-z^2)}\cosh\left(\sqrt{\lambda z^2}\right)\frac{dz^2}{z^2} \tag{17}$$

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

$$f_{\Delta\mathcal{T}_{il,jm}^{\gamma_{il},\gamma_{jm}}}\left(\Delta\tau_{il,jm}^{\gamma_{il},\gamma_{jm}}\right) = \frac{1}{\sigma_Z}\sqrt{\frac{2}{\pi}}\cosh\left(\frac{\Delta\tau_{il,jm}^{\gamma_{il},\gamma_{jm}}\cdot\mu_Z}{{\sigma_Z}^2}\right)e^{\left(-\frac{(\Delta\tau_{il,jm}^{\gamma_{il},\gamma_{jm}})^2+{\mu_Z}^2}{2{\sigma_Z}^2}\right)} \tag{18}$$

Once this PDF has been obtained, we can derive the interdependency relationship for a crossing point for two patterns α_{il}^0 and α_{jm}^0. Figure 4 represents two recurrent patterns and the associated crossing point. α_{il}^0 is represented in green, whereas α_{jm}^0 is represented in orange. We can compute the relative distance between two aircraft if they were at γ_{il} and γ_{jm} respectively, for any combination of γ_{il} and γ_{jm}. Let us declare a function *D* to define this distance:

$$D_{il,jm}: \mathbb{R}\ x\ \mathbb{R} \rightarrow \mathbb{R}$$
$$D_{il,jm}(\gamma_{il}, \gamma_{jm}) = Haversine({\mathcal{H}_{il}}^{-1}(\gamma_{il}), {\mathcal{H}_{jm}}^{-1}(\gamma_{jm}), \quad (19)$$
$$\gamma_{il} \epsilon [0, L_{il}], \gamma_{jm} \epsilon [0, L_{jm}]$$

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

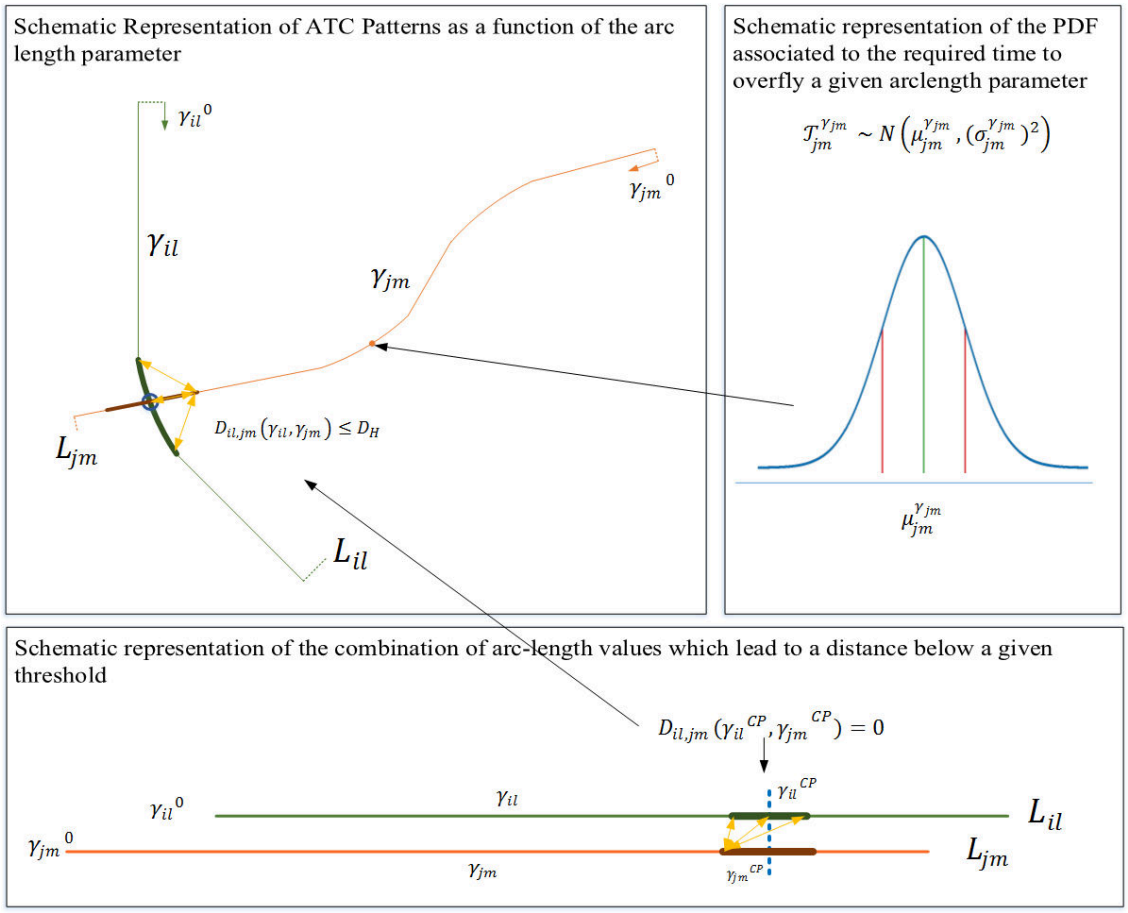


Figure 4 Schematic representation of two recurrent patterns and a crossing point

Following (Histon and Hansman, 2008)'s approach of only focusing on critical points, if we consider now ${\gamma_{il}}^{CP}$ and ${\gamma_{jm}}^{CP}$, we can observe 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 we can redefine

$$\mathcal{T}_{il}^{\gamma_{il}} - \mathcal{T}_{jm}^{{\gamma_{jm}}^{*}} \sim N\left(\mu_{il}^{\gamma_{il}} - \mu_{jm}^{\gamma_{jm}} + \mu_{{\gamma_{jm}}^{offset}}, (\sigma_{il}^{\gamma_{il}})^2 + (\sigma_{jm}^{\gamma_{jm}})^2 + (\sigma_{{\gamma_{jm}}^{offset}})^2\right), \quad (20)$$

, and then

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

$$\mu_Z{}^*=\mu_{il}^{\gamma_{il}}-\mu_{jm}^{\gamma_{jm}}+\mu_{jm}{}^{\gamma_{jm}{}^{offset}} \tag{21}$$

$$\sigma_Z{}^{*2}=\left(\sigma_{il}^{\gamma_{il}}\right)^2+\left(\sigma_{jm}^{\gamma_{jm}}\right)^2+\left(\sigma_{jm}^{\gamma_{jm}{}^{offset}}\right)^2$$

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

$$\begin{aligned}I\left(\Delta t,\gamma_{il},\gamma_{jm}{}^*\right)&=P\left(\Delta\mathcal{T}_{il,jm}^{\gamma_{il},\gamma_{jm}{}^*}\le\Delta t\middle|\gamma_{il},\gamma_{jm}{}^*\right)=\int_{-\infty}^{\Delta t_{Thres}}\left(f_{\Delta\mathcal{T}_{il,jm}^{\gamma_{il},\gamma_{jm}{}^*}}\right)d\left(\Delta\tau_{il,jm}^{\gamma_{il},\gamma_{jm}{}^*}\right)\\&=\int_0^{\Delta t_{Thres}}f_{\Delta\mathcal{T}_{il,jm}^{\gamma_{il},\gamma_{jm}{}^*}}d\left(\Delta\tau_{il,jm}^{\gamma_{il},\gamma_{jm}{}^*}\right)\\&=\frac{1}{2}\left(\mathrm{erf}\left(\frac{\Delta t_{Thres}-\mu_Z{}^*}{\sqrt{2}\sigma_Z{}^*}\right)+\mathrm{erf}\left(\frac{\Delta t_{Thres}+\mu_Z{}^*}{\sqrt{2}\sigma_Z{}^*}\right)\right),\end{aligned} \tag{22}$$

where erf is the function error. The interdependency calculation may be generalised for the set of γ_{il} and $\gamma_{jm}{}^*$ so that $D_{il,jm}\left(\gamma_{il},\gamma_{jm}{}^*\right)\le D_H$. 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:

$$\begin{aligned}&\Pi_{\Delta t_{Thres},D_H}^{il,jm}\left(\gamma_{il}{}^{min},\gamma_{il}{}^{max},\gamma_{jm}{}^{*,min},\gamma_{jm}{}^{*,max}\right)=\frac{1}{B\left(\gamma_{il}{}^{min},\gamma_{il}{}^{max},\gamma_{jm}{}^{*,min},\gamma_{jm}{}^{*,max}\right)}\cdot\\&\int_{\gamma_{jm}{}^{*,min}}^{\gamma_{jm}{}^{*,max}}\int_{\gamma_{il}{}^{min}}^{\gamma_{il}{}^{max}}bin\left(\gamma_{il},\gamma_{jm}{}^*,D_H\right)\cdot I\left(\Delta t_{Thres},\gamma_{il},\gamma_{jm}{}^*\right)d\gamma_{il}\,d\,\gamma_{jm}{}^*\end{aligned} \tag{23}$$

Where $bin\left(\gamma_{il},\gamma_{jm}{}^*,D_H\right)$ is a binary function which equals 1 or 0 depending on whether $D_{il,jm}\left(\gamma_{il},\gamma_{jm}{}^*\right)\le D_H$ or not, whereas $B\left(\gamma_{il}{}^{min},\gamma_{il}{}^{max},\gamma_{jm}{}^{*,min},\gamma_{jm}{}^{*,max}\right)$ is the sum of $bin\left(\gamma_{il},\gamma_{jm}{}^*,D_H\right)$ 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 $\gamma_{jm}{}^*$ within those intervals for which $bin\left(\gamma_{il},\gamma_{jm}{}^*,D_H\right)$ equal 0.

2.3. Interdependency Determination for two flights

Now, if we have two flights (*r* and *s*) associated to major flows λ_i^0 and λ_j^0 respectively, we can generate their expected trajectories for the recurrent patterns. Following Eq. (6), expected trajectories for aircraft *r* and *s* would be defined by $\widetilde{\eta_{r,1}}$ and $\widetilde{\eta_{s,1}}$ respectively.

Then, we can obtain 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, $\Pi^{il,jm}_{\Delta t,D_H}$, for each combination of expected trajectories $\widetilde{\eta_{r,il}}$ 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, we can define a horizontal interdependency matrix (HIM),

$$HIM_{r,s} = \begin{bmatrix} \Pi^{i1,j1}_{\Delta t,D_H} & \cdots & \cdots & \cdots & \Pi^{i1,jo}_{\Delta t,D_H} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ \Pi^{il,j1}_{\Delta t,D_H} & \vdots & \Pi^{il,jm}_{\Delta t,D_H} & \vdots & \Pi^{il,jo}_{\Delta t,D_H} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \Pi^{ip,j1}_{\Delta t,D_H} & \cdots & \cdots & \cdots & \Pi^{ip,jo}_{\Delta t,D_H} \end{bmatrix}. \quad (24)$$

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

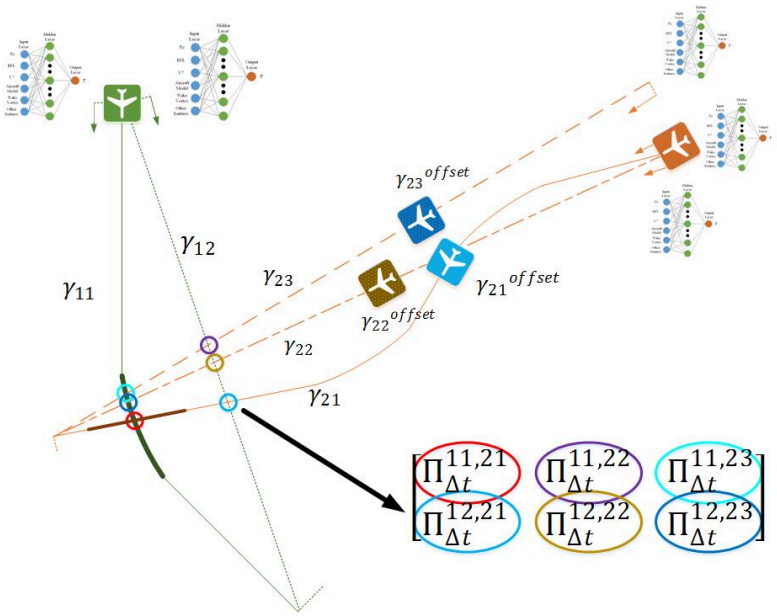


Figure 5 Process for the generation of the Horizontal Interdependency Matrix

Assuming 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 ${\gamma_{2m}}^{offset}$. 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 are gathered in the horizontal interdependency matrix. When any of those values is larger than a given threshold, it implies that those trajectories are operationally constrained. For the sake of simplicity, this paper 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 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 regimes at the crossing point.

2.4. Vertical Trajectory Specification

Once the interdependency between two aircraft has been modelled, the methodology addresses trajectory vertical profiles. This paper focuses on the modelling of the trajectory vertical profile in the airspace volume where the aircraft starts its descent. This 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 similarly expressed in terms of the arc-length parameter, following (5). The main difference with the horizontal component is that $Vz(\gamma_{il})$ can be zero. For modelling the descent profile, we can transform the FL component to:

$$\rho(\gamma) = RFL - FL(\gamma) \tag{25}$$

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

$$\begin{gathered} t(\rho(\gamma), \gamma_{Tod}) = \theta\big(\rho(\gamma)\big) + t(\gamma_{Tod}) \\ \gamma \geq \gamma_{Tod};\ \rho(\gamma) \geq 0 \end{gathered} \tag{26}$$

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, we divide the vertical profile of the trajectory along the airspace volume of interest in two phases, the cruise one as a function of γ_{Tod} , and the descent phase, depending on ρ. Now, if we consider a pattern α_{il}^0, we can define a new random variable $T_{il}^{\rho,\gamma_{Tod}}$ 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 we want to remove the dependency with the location of the TOD, we need to marginalise this random variable over γ_{Tod} to obtain an expression for T_{il}^{ρ}. If we identify a potential TOD location as $\gamma_{il}^{TOD} = \gamma_{il}^{k}$, we can define T_{il}^{ρ,γ_k} as

$$T_{il}^{\rho,\gamma_k}\big(\rho, \gamma_{il}^{k}\big) \sim \mathcal{T}_{il}^{\gamma_k} + \Theta_{il}(\rho|\gamma_{il}^{k}) \tag{27}$$

If we now follow the approach defined in (10), this equation is transformed to:

$$T_{il}^{\rho,\gamma_k}\big(\rho, \gamma_{il}^{k}\big) \sim N\left(\mu_{il}^{\gamma_{il}^{k}}, (\sigma_{il}^{\gamma_{il}^{k}})^2\right) + \Theta_{il}(\rho, \gamma_{il}^{k}) \tag{28}$$

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

$$\Theta_{il}(\rho, \gamma_{il}^k) \sim N(\mu_\theta^{\rho,\gamma^k}, \left(\sigma_\theta^{\rho,\gamma^k}\right)^2) \tag{29}$$

Then, if we marginalise over γ_{il}^k and we consider a discrete set of γ_{il}^k values, of size K, then

$$T_{il}^\rho \sim \sum_k [P(\gamma_{il}^{TOD} = \gamma_{il}^k) \cdot T_{il}^{\rho,\gamma_k}(t;\ \rho, \gamma_{il}^k)] \tag{30}$$

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

$$f_{T_{il}^\rho} \sim \sum_k [P(\gamma_{il}^{TOD} = \gamma_{il}^k) \cdot \left[f_{\mathcal{T}_{il}^{\gamma_{il}^k}} + f_{\Theta_{il}^{\rho,\gamma_{il}^k}}\right]] \tag{31}$$

Eq. (31) 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. (31) is introduced in Figure 6.

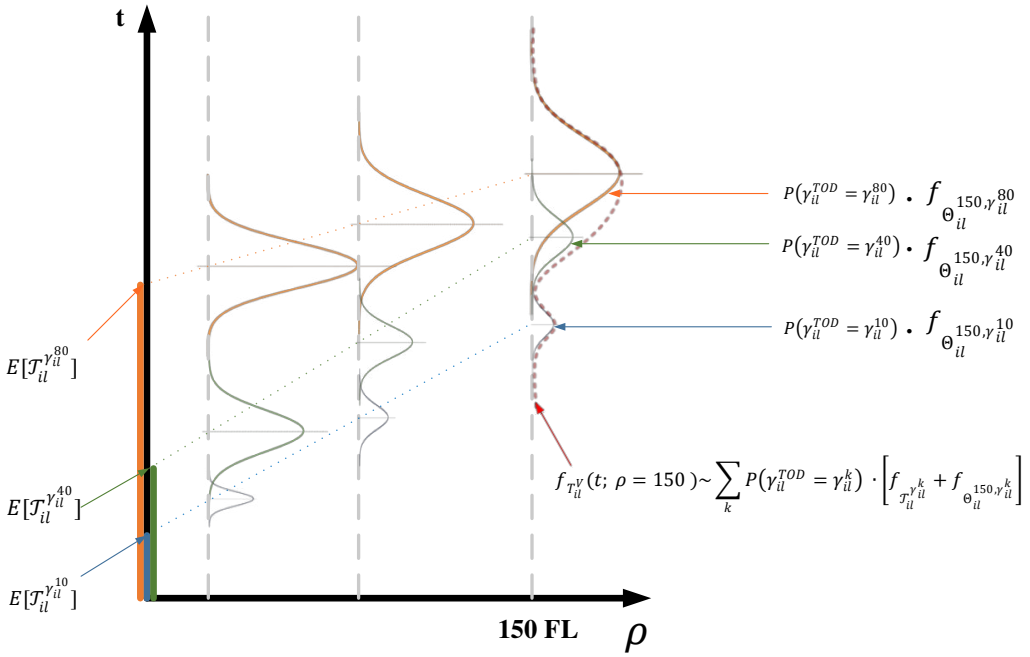


Figure 6 Description of the pdf of $T_{il}^{\rho=150}$

Figure 6 represents schematically different evolution profiles depending on the position of the TOD. The *blue* profile represents a descent starting at $\gamma_{il}^{TOD} = 10$ 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. $f_{T_{il}^\rho}(\rho = 150FL)$ is represented as a dashed line generated after the composition of the three PDFs.

2.5. Contextual factors impacting on the Trajectory Vertical Profile.

Previous works (Verdonk Gallego et al., 2018b) 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 (see Section 2.3) could be considered for determining the TOD position.

The hypothesis we are testing is that if we have two aircraft *r* and *s* associated to different major flows, the location of the TOD is independent of the interdependency between them or not,

$$P\left(\gamma_{il}^{TOD} \mid \bar{A}_r, \mathrm{FL}_{jm}^s, HIM_{r,s}\right) = P(\gamma_{il}^{TOD}; \bar{A}_r) \tag{32}$$

where we assume 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 (31). The vertical trajectory predictor for Θ_{il} will be defined as an Artificial Neural Network analogously to the horizontal one, as described in Eq. (9) and Eq. (10).

3. Results

3.1. Dataset

A couple of datasets corresponding to the sector *Central* (LECBCCC) within the Barcelona ATC Centre in Spain are used in this study. The sector of interest corresponds to an en-route sector with evolution and cruise traffic, presented in Figure 7. The figure includes the route structure and relevant designated waypoints for the work carried out in this paper. 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. This dataset has been already used in (Verdonk Gallego et al., 2018b). The training dataset corresponds to four summer months of 2014, whereas the testing dataset covers two spring months of 2015.

Flown trajectories are expressed following Eq. (3). 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 of 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.

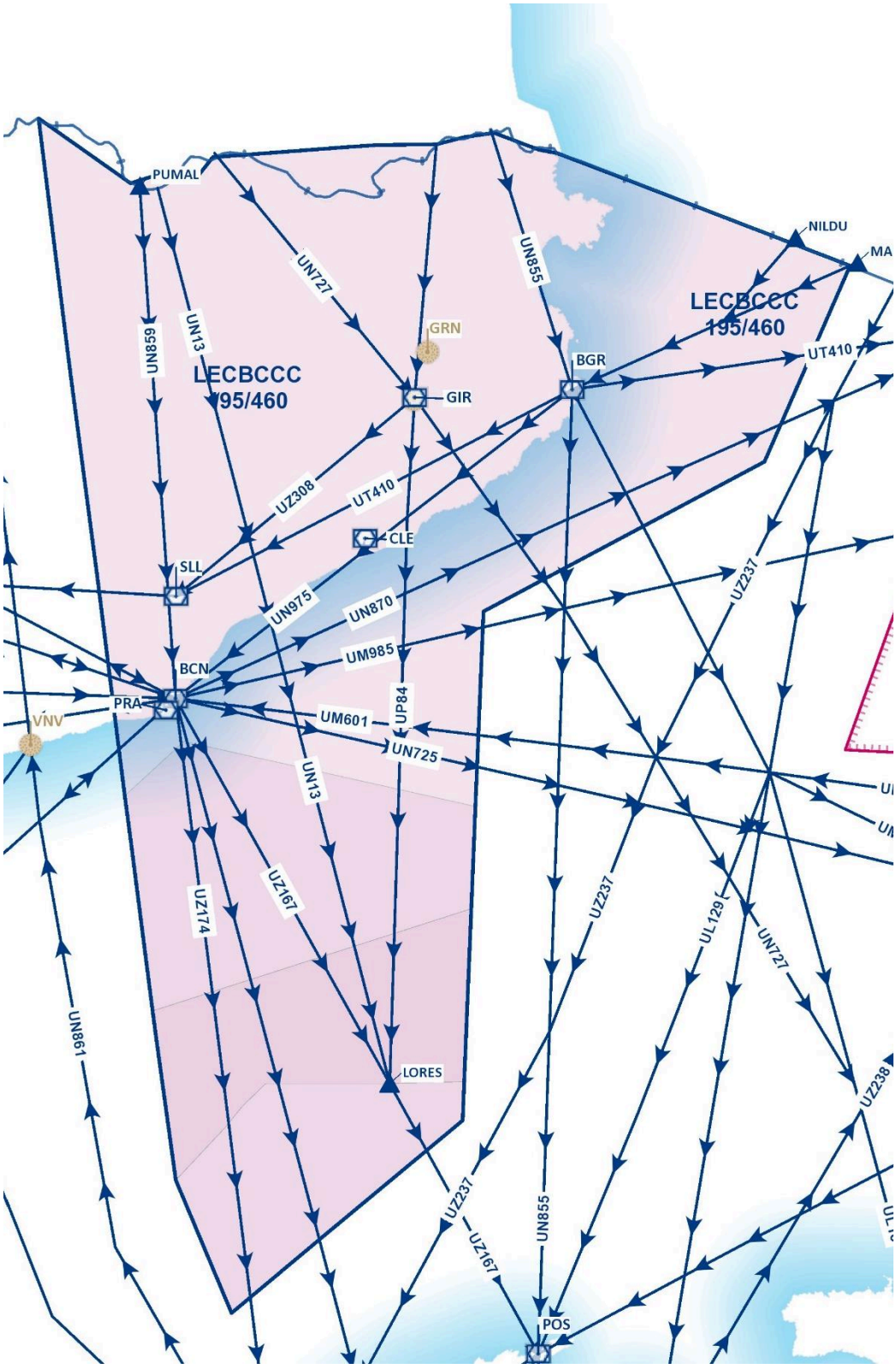


Figure 7 LECBCCC Sector - Barcelona Air Traffic Control Sector. (ENAIRE, 2016)

3.2. 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. Figure 8 illustrates both steps for the datasets of interest. The major flows are represented on the left-hand overlaying the sector represented in Figure 7. The major flows can be associated with the standard routes of the sector.

The right-hand side of Figure 8 corresponds to the recurrent patterns 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. 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.

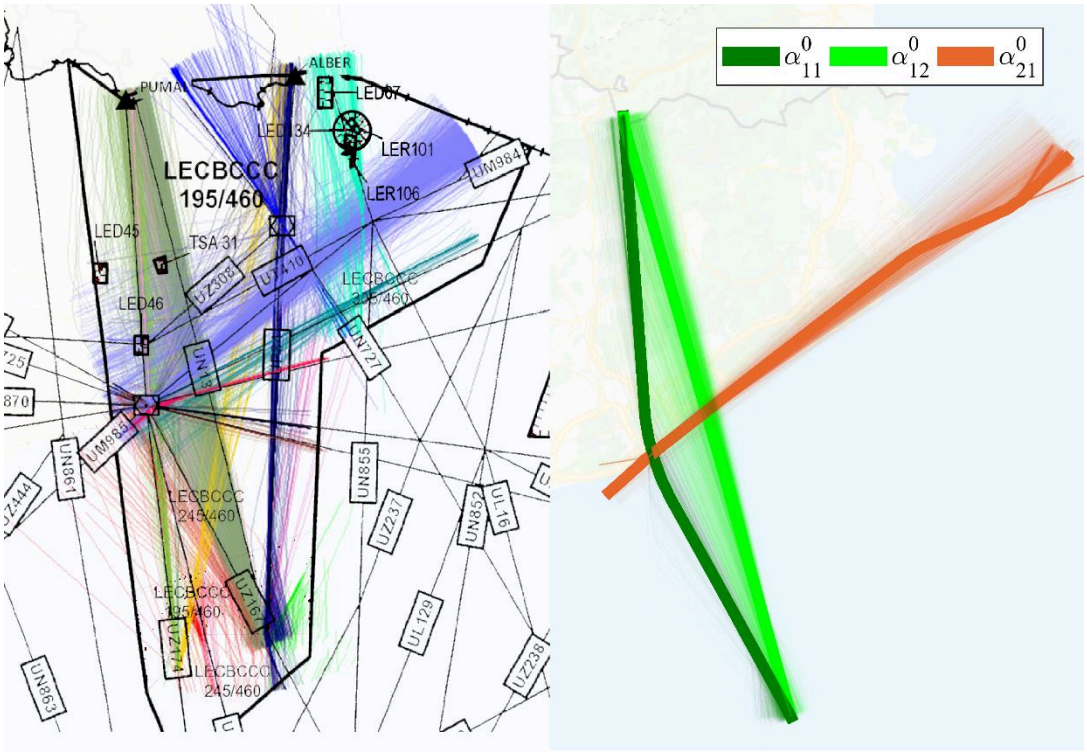


Figure 8 Major flows and recurrent patterns. The recurrent patterns illustrated in green correspond to the PUMAL-LORES major flow, flying North-South, with destination LEPA. The recurrent pattern illustrated in orange corresponds to the MAMES-BCN flow, flying East-West in a cruise state.

Following Eq. (7), and based on previous results (Verdonk Gallego et al., 2018b), ANN TPs were trained with interpolated data generated every 1NM along the trajectory. Additionally, the feature vector ($\underline{X}$) 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 (to prevent overfitting) and testing. The percentages associated to each category were 70%, 15% and 15% respectively.

Equations (8), (9) and (10) assumed that the TP residuals were normally distributed. Figure 9 illustrates the behaviour of the residuals (see Eq. (9)) as a function of the arc-length parameters for the sample corresponding to the training dataset. Figure 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 in seconds. 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 10 represents the distribution of the residuals for three different arc-length values (30NM, 60NM and 90NM). The horizontal axis represents the residuals in seconds while the vertical one represents the probability density. The residuals were fitted to normal distributions (continuous lines).

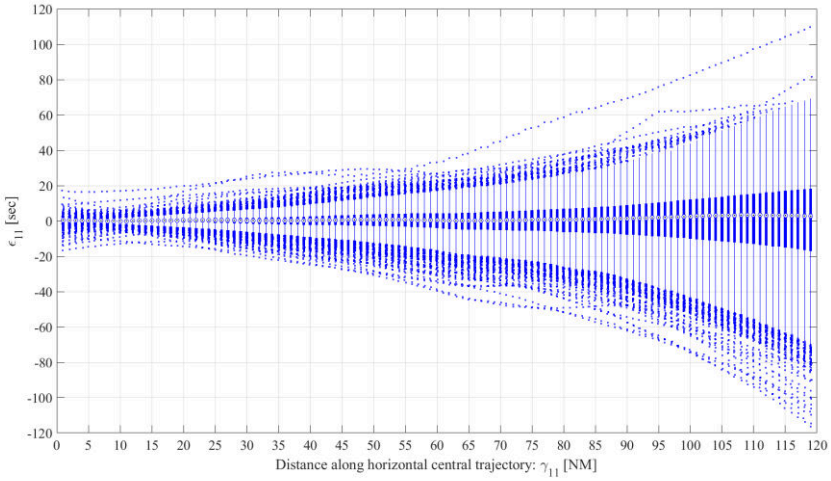


Figure 9 Residuals for the ANN associated to the standard route associated corresponding to the flow Pumal-Lores (α_{11}^{0}) along γ_{11}.

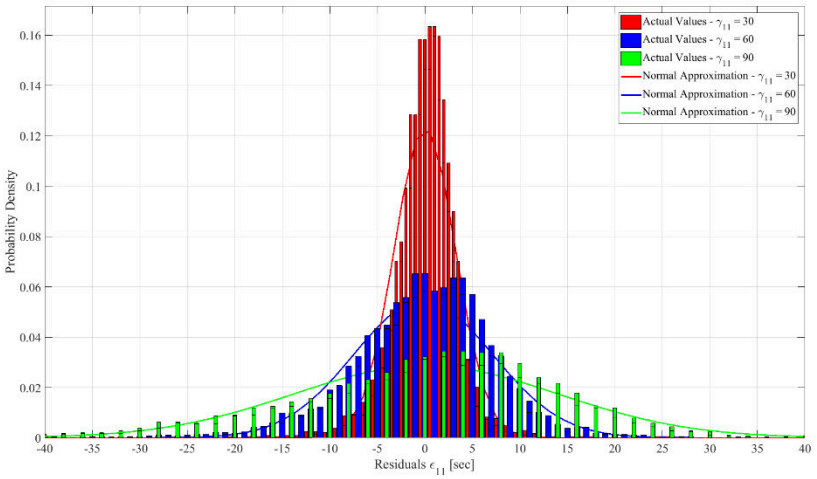


Figure 10 Residuals distribution for different values of γ_{11} for the standard route associated to the flow Pumal-Lores (α_{11}^0). The red series correspond to $\gamma_{11} = 30$ NM, blue series correspond to $\gamma_{11} = 60$ NM and green series correspond to $\gamma_{11} = 90$ NM

Table I includes the mean and Root Mean Squared Error (RMSE) values for selected 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 values are similar to the training ones.

Table I Trajectory predictor performance as a 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

The normality assumption may be sensed observing Figure 10. Chi-Square goodness of fit tests have been carried out, rejecting the normality assumption. Figure 11 shows a QQ-plot to identify reasons for the normality assumption rejection. The case presented corresponds to α_{11}^0. 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 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.

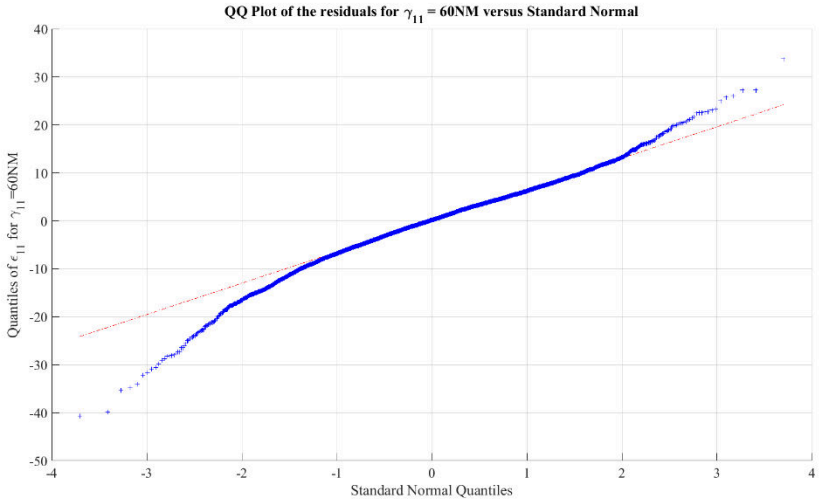


Figure 11 QQ-plot for $\gamma_{11} = 60$ NM, corresponding to the recurrent pattern α_{11}^0.

3.3. Crossing Point Interdependency Definition Results

Eq. (11) 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. (21) introduced an offset along one of the routes for considering for these two factors.

Figure 12 illustrates the impact of this offset on the actual time separation of two aircraft arriving at $\gamma_{11} =$ 64 NM and $\gamma_{21} = 94$ NM respectively, where the offset is assigned 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. Only the offset interval (0,35] NM is presented for the sake of clarity in the figure. The training sample has been used to generate this figure.

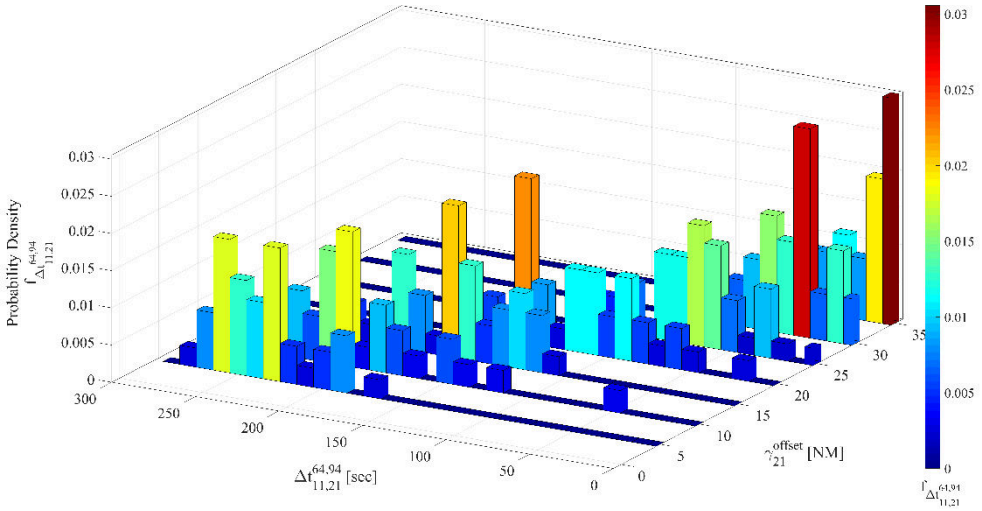


Figure 12 Actual Difference Time Values at the Crossing Point

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

Figure 13 represents the theoretical PDFs as derived from Eq. (21). Four pairs of aircraft corresponding to the testing sample are studied for $\gamma_{11} = 64$ NM and $\gamma_{21} = 94$ NM. Theoretical $f_{\Delta\mathcal{T}_{11,21}^{64,94-\gamma_{21}offset}}$ values are generated considering $\Delta\tau$ and ${\gamma_{21}}^{offset}$ as parameters. Mean and standard deviations to be used in Eq. (21) 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 purple boxes in each figure represent the actual time separation values, Δt, which correspond to the actual offset value of the aircraft belonging to λ_2^0.

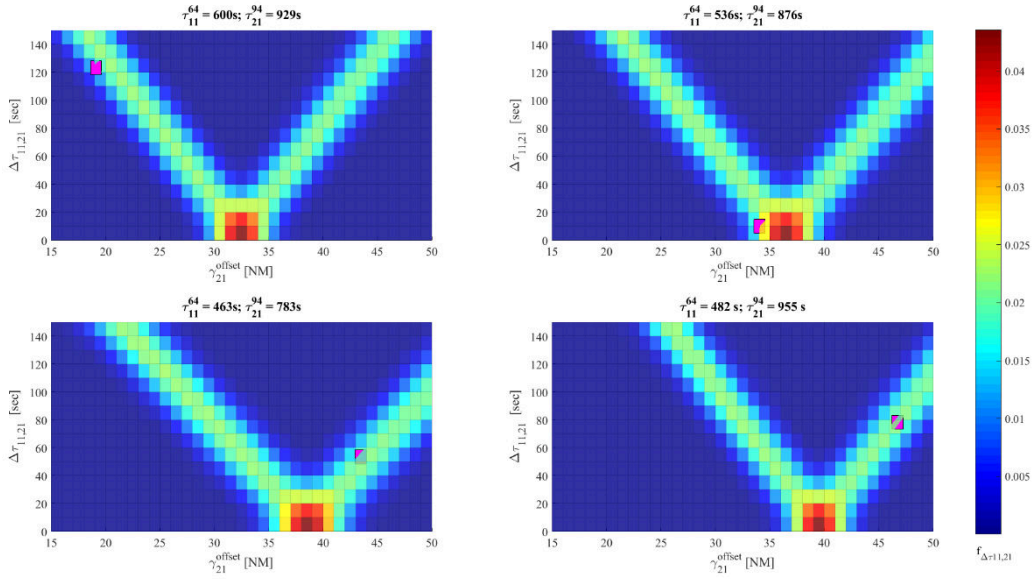


Figure 13 Probability Density Function of the predicted time separation for four different cases. The purple box in each subplot is indicative of the actual time separation and the actual offset.

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

These four cases are illustrative of the results derived from Eq. (21). However, these four cases may not be representative of the whole sample. Figure 14 plots the residuals of the predicted time separation as a function of the actual time separation for those pairs of aircraft that flew following these recurrent patterns for the training sample. The critical cases (potential conflicts) would be those where the actual values are below the temporal threshold (120 s). 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. 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.

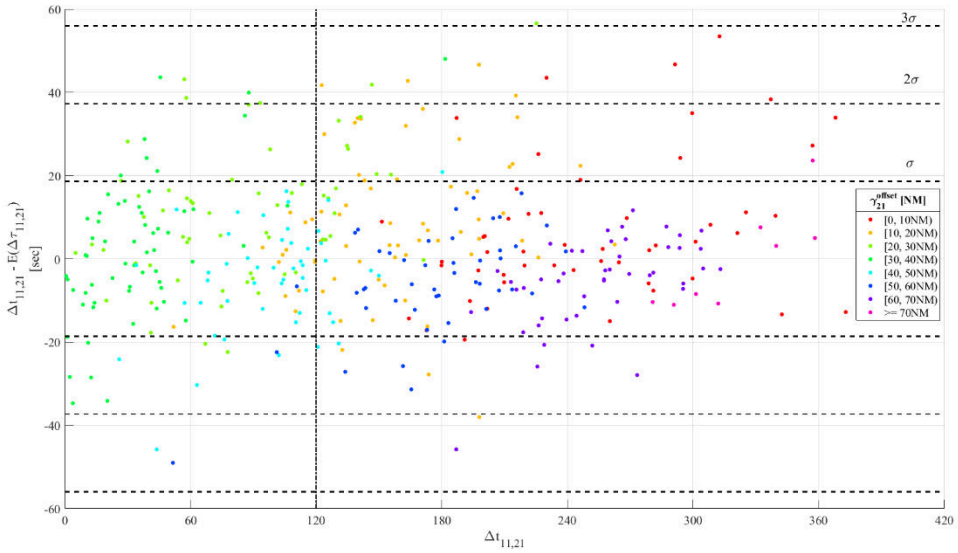


Figure 14 Residual distributions as a function of the actual time separation. Different colours represent different offset values.

3.4. Horizontal Interdependency

Eq. (23) 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 arrive at 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. (23) are illustrated by means of a series of figures. Figure 15 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, corresponding 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 their 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 $\Pi^{11,21}_{\Delta t_{Thres},D_H}$ 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.

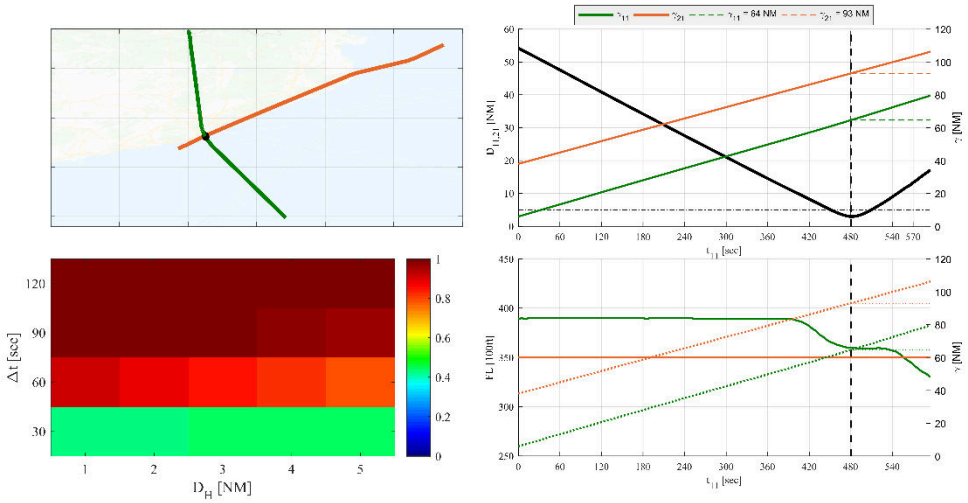


Figure 15 Horizontal Interdependency Measure. Case of Study 1.

The case illustrated in the Figure 15 was selected based on the values of $\Pi^{11,21}_{\Delta t_{Thres},D_H}$. The interdependency values were computed when the aircraft corresponding to α^0_{11} (green) entered the sector. Then, a set of highly interdependent pairs of aircraft were 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 (Δt_{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 15 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.

Appendix I details three additional cases for further discussion about the interdependency measure (Figure 26, Figure 27, Figure 28). Those cases are not discussed within the current section for easing the readability of the paper. 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.

3.5. 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. (24), 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. (31) modelled the trajectory vertical profile as function of the TOD location. It was observed in Figure 15 and Figure 26-Figure 28 that conflicting scenarios impacted on the TOD locations for these specific cases. Figure 17 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 16 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 conflicting flights, the TOD locations are spread over the γ_{11} range. It is observed a remarkable number of flights with early CFL instructions (γ_{11} between 0NM and 35NM), a circumstance that was negligible for non-conflicting flights.

It is 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 γ_{11} = 65NM, observing more frequent TOD locations after 65NM for the conflicting aircraft.

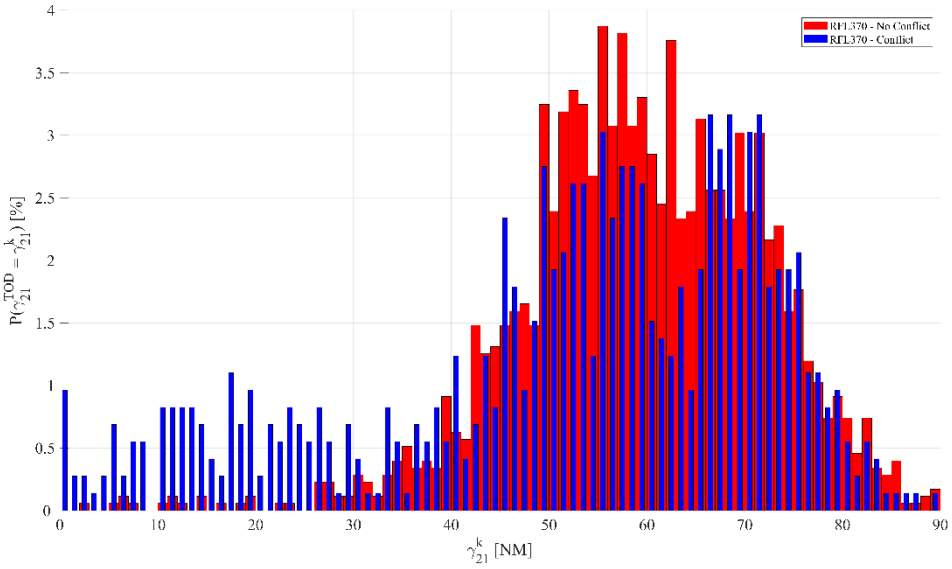


Figure 16 TOD Location depending on whether the encounter was classified as a conflict or not. The red colour indicates non-conflicting encounters, whereas blue columns indicate the opposite.

An example of the first case (early TOD) is shown in Figure 27, whereas two examples of the latter (TOD after 65NM) are represented in Figure 26 and Figure 28. Figure 15 illustrated a third case, where the flight was cleared to descend to an intermediate FL, until the potential conflict disappeared.

Figure 16 show that the hypothesis formulated in Eq. (32) 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. (32), and therefore, the independence of the TOD position to the categorisation by the conflict measure can be rule 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 again.

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.

3.6. Trajectory Vertical Profile Characterisation

Figure 17 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 20 represents similar information, but for conflicting flights.

The first row corresponds to $\rho = 50$FL, the second row to $\rho = 100$FL and the third row to $\rho = 150$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.

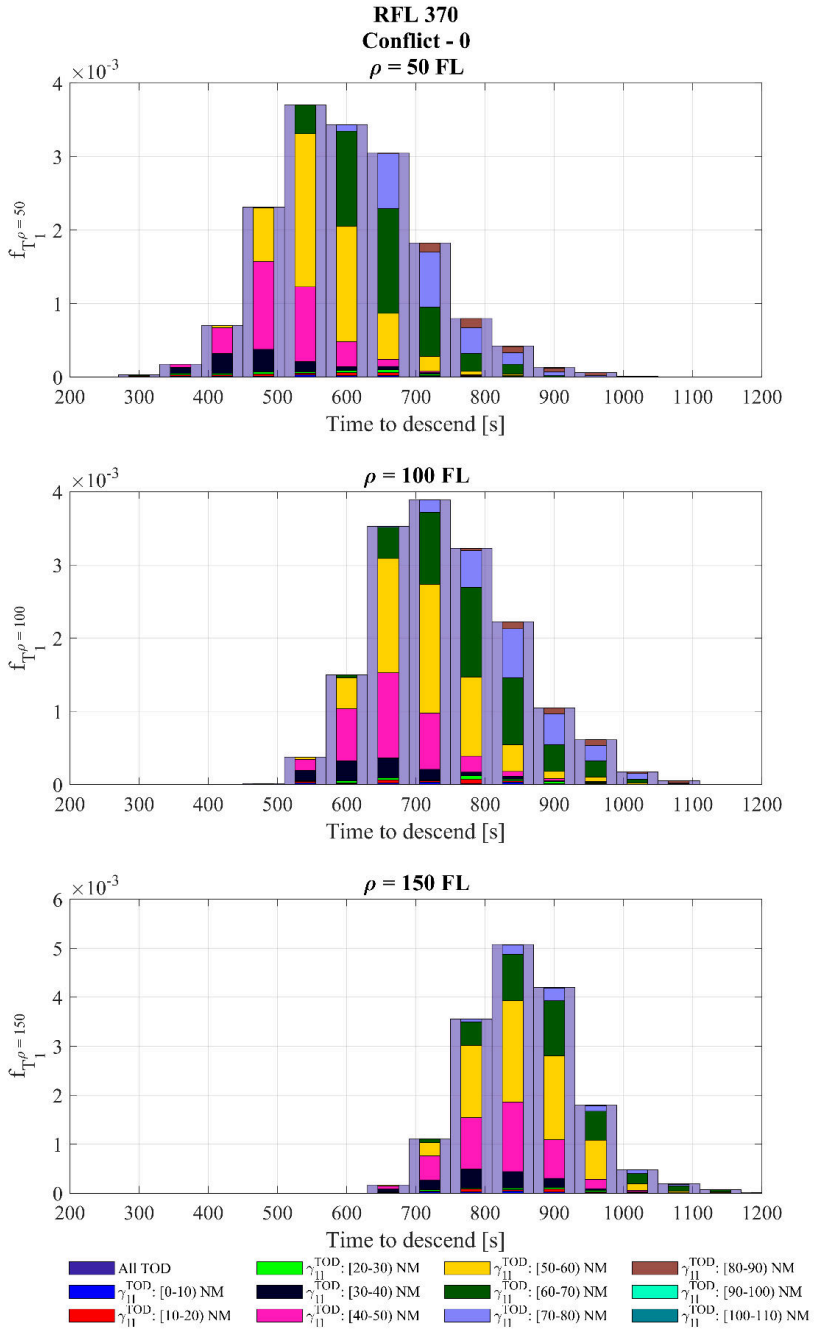


Figure 17 Time to descend 50, 100 and 150FL Density Probability for non-conflicting flights with RFL370, for different TOD locations grouped in 10NM-width bins.

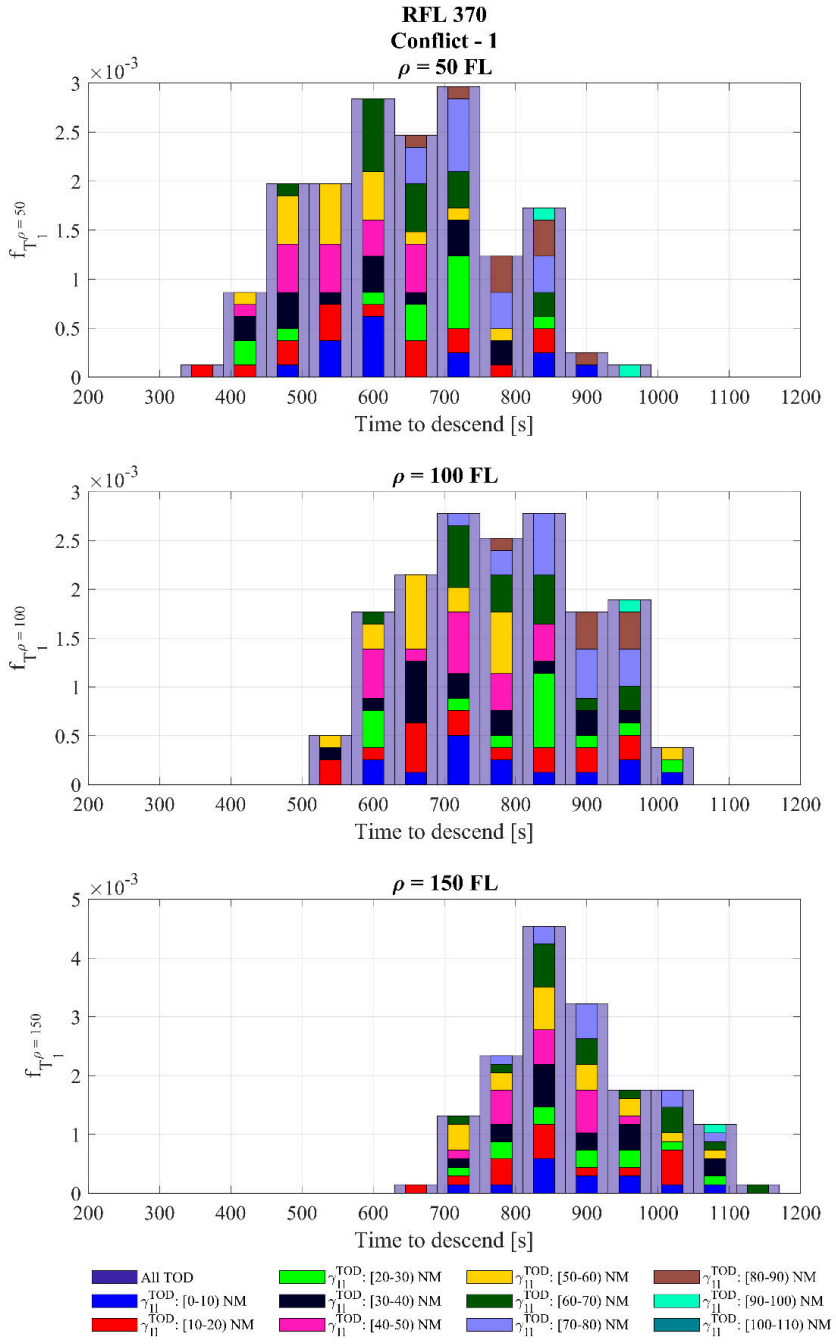


Figure 18 Time to descend 50, 100 and 150FL Density Probability for conflicting flights with RFL370, for different TOD locations grouped in 10NM-width bins.

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 $\rho = 100$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 19. It splits the second row of Figure 17 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 20).

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. (29) and (Verdonk Gallego et al., 2018b; Wang et al., 2018). 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 $\gamma_{1l} = 0$.

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.

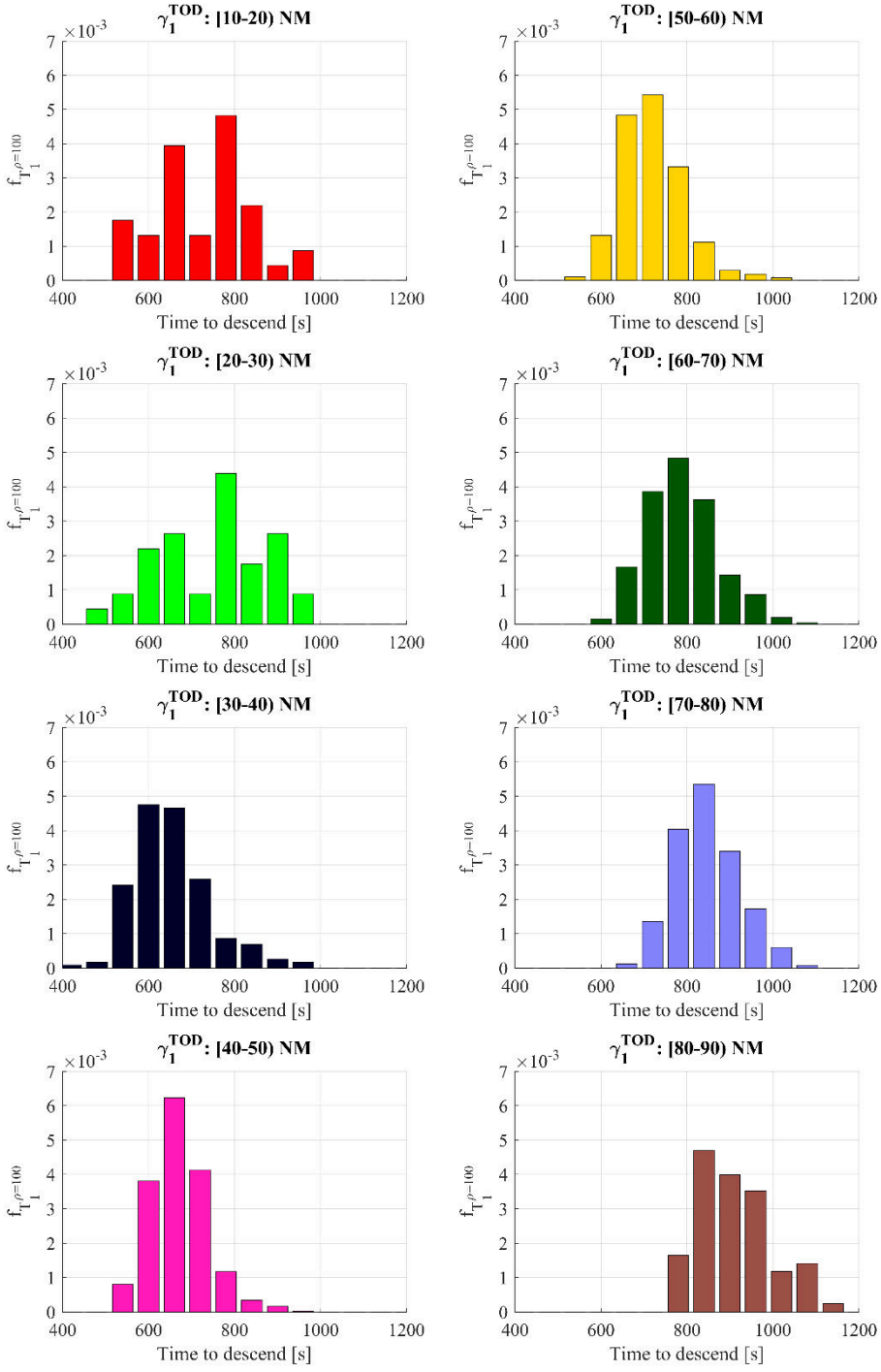


Figure 19 Vertical profile Density Probability for descending 100 FL for non-conflicting aircraft with RFL370, for different locations of the TOD.

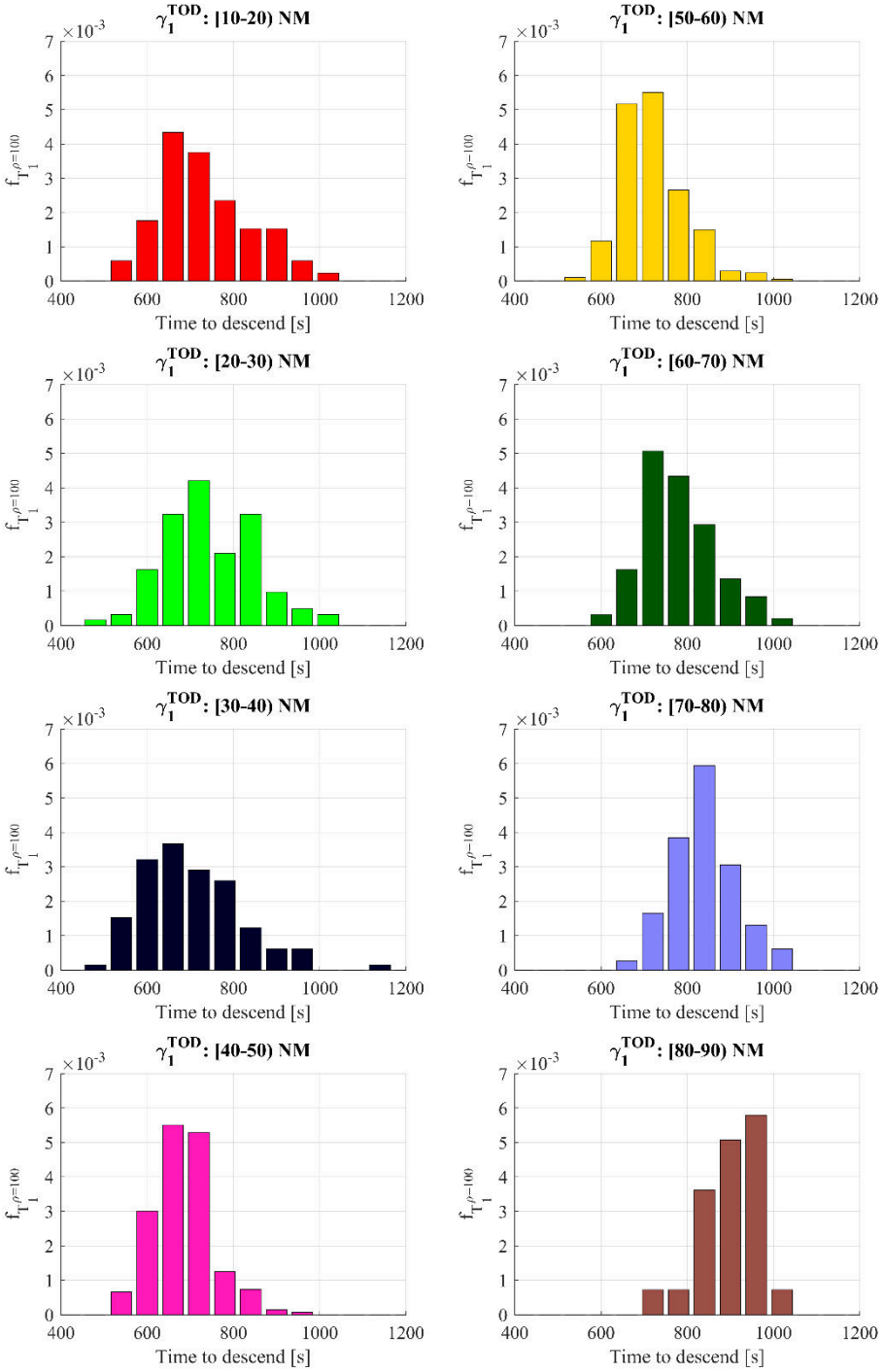


Figure 20 Vertical profile Density Probability for descending 100 FL for conflicting aircraft with RFL370, for different locations of the TOD.

3.7. Vertical Profile Trajectory Predictor Results

Following Eq. (31), 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. All the predictions were carried out when aircraft entered the sector.

Figure 21 illustrates these curves for the training sample for the non-conflicting cases. 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.

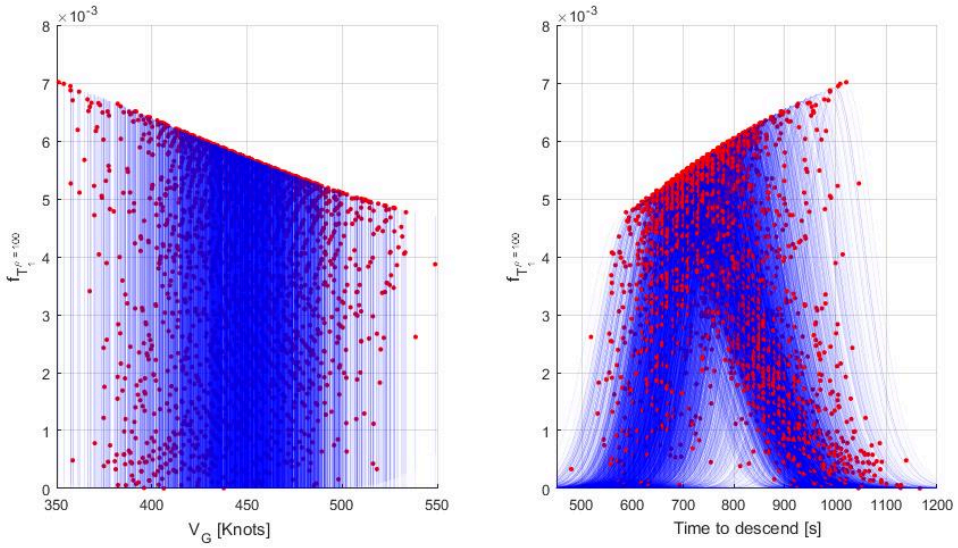


Figure 21 $f_{t_1^\rho}$ distributions as a function of the time to descend $t_1^{\rho=100}$ 100FL for flights in non-conflicting scenarios. Each red point is the $f_{t_1^\rho}$ value corresponding to an actual time to descend value.

Figure 21 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. It also shows that 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 13. 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. It was observed that conflicting and non-conflicting aircraft show a similar pattern.

Figure 22 specifies the difference between the actual value of $t_1^{\rho=100}$ and the median value of the theoretical probability density function $f_{t_1^\rho} = 100$. Columns discriminate the residuals depending on whether the aircraft were conflicting with other traffic (right-hand side) or not (left-hand side). The figure introduces a histogram of the distribution of the residuals, based on 10-second bins.

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.

(Verdonk Gallego et al., 2018b) 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.

LEPA airport has two parallel runways, which can be operated in westerlies (24R/L) or easterlies (06R/L) configurations. Figure 23 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 $f_{t_1^\rho}$.

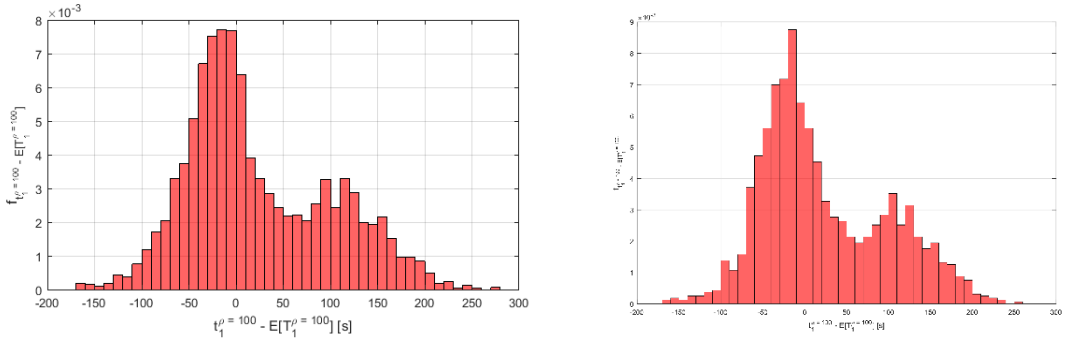


Figure 22 Vertical TP Residuals Distribution for non-conflicting flights (left-hand side) and conflicting flights (right-hand side). The top-row provides the residual distribution whereas the bottom row provides the values of $f_{t_1^\rho}$ *corresponding to actual time to descend values.*

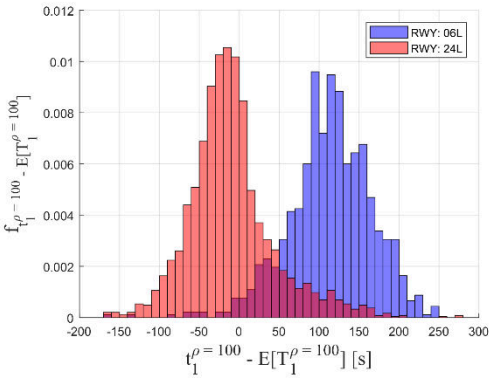


Figure 23 Vertical TP Residuals Distribution for non-conflicting flights as a function of LEPA Runway Configuration.

Eq. (31) introduced $f_{T_{il}^{\rho}}$ as a function of the probability mass function of the TOD location along the horizontal central trajectory. If now we train a model for the TOD location as a function of different parameters, such as Vg, RFL, the wake vortex and the airport configuration, we can generate k probability mass distributions 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.

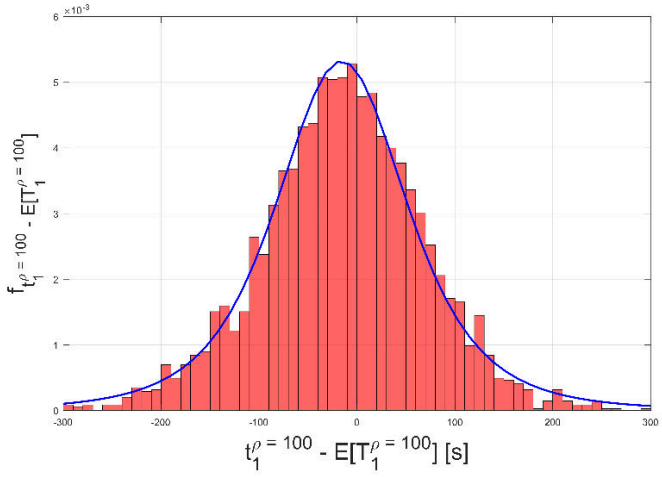


Figure 24 Vertical TP Residuals Distribution for non-conflicting after including the airport configuration as a factor for predicting the TOD position. The top-row provides the residual distribution whereas the bottom row provides the values of $f_{t_1^{\rho}}$ corresponding to actual time to descend values. The top-row figure includes blue line corresponding to a fitted t-student distribution with parameters $\mu = -15$ sec and $\upsilon = 3.5$.

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 24. 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 $\upsilon = 3.5$ (degrees of freedom). The RMSE of the prediction was 108 seconds.

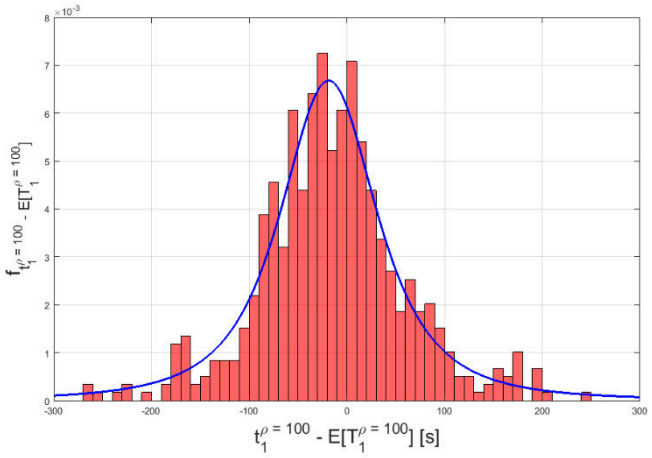


Figure 25 Vertical TP Residuals Distribution for non-conflicting after including the airport configuration as a factor for predicting the TOD position for the testing dataset. The top-row provides the residual distribution whereas the bottom row provides the values of $f_{t_1^\rho}$ corresponding to actual time to descend values. The top-row figure includes blue line corresponding to a fitted t-student distribution with parameters $\mu = -18.5$ sec and $\upsilon = 2.19$.

Figure 25 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 $\upsilon = 2.19$ (degrees of freedom), showing heavier tails than for the training dataset. The Root Mean Square Error (RMSE) of the prediction was 109.92 seconds

4. Discussion

The results have been organised in seven sections. The present section discusses key findings from each of them. Section 3.1 detailed the training and testing datasets used in this study. The datasets corresponded to an en-route sector with evolution flows which may conflict with cruise flows for two different years and in different seasonal periods.

Section 3.2 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 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 3.2 also discussed the normality assumption for the TP's residual distributions. Figure 9 characterised the dependency of the residuals with the arc-length parameter along the horizontal central trajectory for one of the studied recurrent patterns. 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 14). Further work could explore the statistical dependency of the

residuals of the prediction for one aircraft with the presence of another aircraft to analyse the mutual impact on the horizontal profile.

Section 3.3 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 as modelled. 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 Section 3.4. 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 provided 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 also involving *operational* vertical conflicts.

The novel contribution of the approach 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 (Ayhan et al., 2018; Erzberger et al., 1997; Hernández-Romero et al., 2019; Liu and Hwang, 2011; Paielli and Erzberger, 2008; Piroddi and Prandini, 2010; Schefers et al., 2017). In addition, the modelling of the time separation as a folded normal distribution was similar to the one presented by (Irvine, 2001), 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 (Hong and Lee, 2015), but with other operational purpose.

Based on the insights gained from Section 3.5, 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 verified 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 (Histon and Hansman, 2008) and in (Hong and Lee, 2015), where the arrival times are predicted based on the situation of a precedent aircraft within the same terminal flow.

Section 3.6 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 3.7 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 previous research work (Verdonk Gallego et al., 2018b), 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 aircraft model, intrinsic intention features such as the expected guidance mode (Dalmau Codina, Ramon; Pérez-Batlle, Marc; Prats Menéndez, 2018) or the aircraft inferred mass (Alligier and Gianazza, 2018). The integration of richer datasets may reduce errors on the prediction, with longer look-ahead times.

Trajectory predictors based on state-dependent hybrid mode transitions, such as (Dalmau Codina, Ramon; Pérez-Batlle, Marc; Prats Menéndez, 2018; Huang et al., 2007; Liu and Hwang, 2011) could benefit from this work to characterise the probability vector governing the transitions between states.

In addition, the incorporation of weather-based information such as wind direction and intensity, temperature or pressure is crucial to improve the accuracy of the trajectory predictors. Previous research (de Leege et al., 2013; Verdonk Gallego et al., 2018b) identified the ground speed and the initial altitude at the TOD as key explanatory variables. The ground speed partially captures the wind impact on the trajectory and how the aircraft reacts to achieve a desired performance and guidance.

Moreover, it has been identified that the TP should not only consider local weather factors, but a global view of the meteorological conditions along the remaining planned trajectory or derived factors, such as the airport runway configuration. It could be explored the extension to the tactical TP the 4D data cubes proposed enriched with weather data by Ayhan in (Ayhan and Samet, 2016) aligned with the route structure and recurrent patterns.

The conclusion of this paper should be aimed at unveiling if the traffic information shall be included in the trajectory predictors. Table II 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 17 and Figure 18. 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 II 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. Conclusions

The objective of this paper was to consider surrounding air traffic as a contextual factor for trajectory prediction. To that end, the paper 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 paper applied known 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 predictors 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 for the specific case study which was studied.

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 when the crossing point is located around the normal TOD position. 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, and the position of the TOD may be crucial for handling the uncertainty of the time of arrival.

The paper has focused in a specific Spanish airspace sector. Future work could address the generalisation of this 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.

Finally, 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 may help to develop new trajectory management metrics and procedures to manage complex operational environments.

Acknowledgement

The authors express their most sincere gratitude to CRIDA A.I.E. in Spain for proportionating the data and support to conduct this research.

References

Alligier, R., Gianazza, D., 2018. Learning aircraft operational factors to improve aircraft climb prediction: A large scale multi-airport study. Transp. Res. Part C Emerg. Technol. 96, 72–95. doi:10.1016/j.trc.2018.08.012

Alligier, R., Gianazza, D., Durand, N., 2015. Machine Learning and Mass Estimation Methods for Ground-Based Aircraft Climb Prediction. IEEE Trans. Intell. Transp. Syst. 16, 3138–3149. doi:10.1109/TITS.2015.2437452

Alligier, R., Gianazza, D., Durand, N., 2013. Learning the aircraft mass and thrust to improve the ground-based trajectory prediction of climbing flights. Transp. Res. Part C Emerg. Technol. 36, 45–60. doi:10.1016/j.trc.2013.08.006

Andrienko, G., Gianazza, D., Alligier, R., Hurter, C., Puechmorel, S., Andrienko, N., 2014. Wind parameters extraction from aircraft trajectories. Comput. Environ. Urban Syst. 47, 28–43. doi:10.1016/j.compenvurbsys.2014.01.005

Ayhan, S., Costas, P., Samet, H., 2018. 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. ACM Press, pp. 239–248. doi:10.1145/3274895.3274947

Ayhan, S., Samet, H., 2016. 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. ACM Press, San Francisco, California, USA, pp. 21–30. doi:10.1145/2939672.2939694

Besada, J.A., Frontera, G., Crespo, J., Casado, E., Lopez-Leones, J., 2013. Automated aircraft trajectory prediction based on formal intent-related language processing. IEEE Trans. Intell. Transp. Syst. 14, 1067–1082. doi:10.1109/TITS.2013.2252343

Casado Magaña, E.J., 2016. Trajectory Prediction Uncertainty Modelling for Air Traffic Management.

Chaloulos, G., Lygeros, J., 2007. Effect of Wind Correlation on Aircraft Conflict Probability. J. Guid. Control. Dyn. 30, 1742–1752. doi:10.2514/1.28858

Dalmau Codina, Ramon; Pérez-Batlle, Marc; Prats Menéndez, X., 2018. Real-time Identification of Guidance Modes in Aircraft Descents Using Surveillace Data, in: 37th IEEE/AIAA Digital Avionics Systems Conference (DASC'18). London, pp. 1295–1304.

de Leege, A., van Paassen, M., Mulder, M., 2013. A Machine Learning Approach to Trajectory Prediction, in: AIAA Guidance, Navigation, and Control (GNC) Conference. Boston, MA. doi:10.2514/6.2013-4782

Delahaye, D., Puechmorel, S., 2010. Air traffic complexity based on dynamical systems. Proc. IEEE Conf. Decis. Control 2069–2074. doi:10.1109/CDC.2010.5718004

Delahaye, D., Puechmorel, S., 2000. Air traffic complexity: towards intrinsic metrics, in: 3rd USA/Europe Air Traffic Management R&D Seminar. pp. 1–11.

Delahaye, D., Puechmorel, S., Hansman, R.J., Histon, J.M., 2003. Air Traffic Complexity Map Based on Non Linear Dynamical Systems, in: 5th USA/Europe ATM R&D Seminar. Budapest, Hungary.

Delahaye, Daniel, Puechmorel, Stéphane, Tsiotras, Panagiotis, Féron, E., Delahaye, D, Puechmorel, S, Tsiotras, P, Feron, E., 2014. Mathematical Models for Aircraft Trajectory Design : A Survey. Lect. notes Electr. Eng. 290, 205–247. doi:10.1007/978-4-431-54475-3_12

Djokic, J., Lorenz, B., Fricke, H., 2010. Air traffic control complexity as workload driver. Transp. Res. Part C Emerg. Technol. 18, 930–936. doi:10.1016/j.trc.2010.03.005

ENAIRE, A., 2016. Aeronautical Information Publication [WWW Document]. Insignia. URL https://ais.enaire.es/insignia/index_en.aspx (accessed 10.29.18).

Erzberger, H., Paielli, R.A., Isaacson, D.R., Eshow, M.M., 1997. Conflict detection and resolution in the presence of prediction error, in: 1st USA/Europe Air Traffic Management R&D Seminar, Saclay, France. Citeseer, pp. 17–20.

EUROCONTROL, 2017. EUROCONTROL Specification for Trajectory Prediction. Brussels.

Flynn, G.M., 2004. European Organisation for the Safety of Air Navigation Eurocontrol Experimental Centre Cognitive Complexity in Air Traffic Control a Literature Review. EEC note.

Franco, A., Rivas, D., Valenzuela, A., 2018. Probabilistic aircraft trajectory prediction in cruise flight considering ensemble wind forecasts. Aerosp. Sci. Technol. 82–83, 350–362. doi:10.1016/j.ast.2018.09.020

García González, E.J., 2013. 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.

Gariel, M., Srivastava, A.N., Feron, E., 2011. Trajectory clustering and an application to airspace monitoring. IEEE Trans. Intell. Transp. Syst. 12, 1511–1524. doi:10.1109/TITS.2011.2160628

González-Arribas, D., Soler, M., Sanjurjo-Rivo, M., 2017. Robust Aircraft Trajectory Planning Under Wind Uncertainty Using Optimal Control. J. Guid. Control. Dyn. 41, 673–688. doi:10.2514/1.g002928

Hernández-Romero, E., Valenzuela, A., Rivas, D., 2019. A probabilistic approach to measure aircraft conflict severity considering wind forecast uncertainty. Aerosp. Sci. Technol. 86, 401–414. doi:10.1016/j.ast.2019.01.024

Hernandez, E., Valenzuela, A., Rivas, D., 2016. Probabilistic Aircraft Conflict Detection Considering Ensemble Weather Forecast. 6th SESAR Innov. Days.

Histon, J.M., Hansman, R.J., 2008. Mitigating Complexity in Air Traffic Control: The Role of Structure-Based Abstractions. MIT Int. Cent. Air Transp. Dep. Aeronaut. Astronaut. Massachusetts Inst. Technol.

Hong, S., Lee, K., 2015. Trajectory Prediction for Vectored Area Navigation Arrivals. J. Aerosp. Inf. Syst. 12, 490–502. doi:10.2514/1.I010245

Huang, H., Roy, K., Tomlin, C., 2007. Probabilistic Estimation of State-Dependent Hybrid Mode Transitions for Aircraft Arrival Time Prediction, in: AIAA Guidance, Navigation and Control Conference and Exhibit. South Carolina.

International Civil Aviation Organization, 2005. Doc 9854 Global Air Traffic Management Operational Concept. ICAO Tech. Publ.

Irvine, R., 2001. A geometrical approach to conflict probability estimation. Encounter 10, 1–15. doi:10.1109/DASC.2001.964214

Liu, W., Hwang, I., 2011. Probabilistic Trajectory Prediction and Conflict Detection for Air Traffic Control. J. Guid. Control. Dyn. 34, 1779–1789. doi:10.2514/1.53645

Niarchakou, S., Simon Selva, J., 2017. EUROCONTROL ATFCM Operations Manual. Brussels.

Paielli, R.A., Erzberger, H., 2008. Conflict Probability for Free Flight. J. Guid. Control. Dyn. 20, 588–596. doi:10.2514/2.4081

Pang, Y., Yao, H., Hu, J., Liu, Y., 2019. A Recurrent Neural Network Approach for Aircraft Trajectory Prediction with Weather Features From Sherlock. AIAA Aviat. 2019 Forum 1–14. doi:10.2514/6.2019-3413

Piroddi, L., Prandini, M., 2010. A geometric approach to air traffic complexity evaluation for strategic trajectory management. Proc. IEEE Conf. Decis. Control 2075–2080. doi:10.1109/CDC.2010.5718055

Prandini, M., Piroddi, L., Puechmorel, S., Brázdilová, S.L., 2011. New Generation Air Traffic Management Systems. Transportation (Amst). 12, 809–818. doi:10.1109/TITS.2011.2113175

Radanovic, M., Piera Eroles, M.A., Koca, T., Ramos Gonzalez, J.J., 2018. Surrounding traffic complexity analysis for efficient and stable conflict resolution. Transp. Res. Part C Emerg. Technol. 95, 105–124. doi:10.1016/j.trc.2018.07.017

Sáez Nieto, F.J., 2016. 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. 230, 1533–1547. doi:10.1177/0954410015596763

Schefers, N., Angel, M., José, J., Nosedal, J., 2017. Causal Analysis of Airline Trajectory Preferences to Improve Airspace Capacity. Procedia - Procedia Comput. Sci. 104, 321–328. doi:10.1016/j.procs.2017.01.141

Schuster, W., Ochieng, W., 2014. 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. 35, 92–101. doi:10.1016/j.jairtraman.2013.11.005

Suárez, N., López, P., Puntero, E., Rodriguez, S., 2014. Quantifying Air Traffic Controller Mental Workload. Fourth SESAR Innov. Days 1–6.

Sun, J., Vu, H., Ellerbroek, J., Hoekstra, J.M., 2018. Weather field reconstruction using aircraft surveillance data and a novel meteo-particle model. PLoS One 13, e0205029. doi:10.1371/journal.pone.0205029

Tastambekov, K., Puechmorel, S., Delahaye, D., Rabut, C., 2013. Aircraft trajectory forecasting using local functional regression in Sobolev space. Transp. Res. Part C Emerg. Technol. 39, 1–22. doi:10.1016/j.trc.2013.11.013

Verdonk Gallego, C.E., Gómez Comendador, V.F., Sáez Nieto, F.J., Martínez García, M., 2018a.

Discussion On Density-Based Clustering Methods Applied for Automated Identification of Airspace Flows, in: 37th IEEE/AIAA Digital Avionics Systems Conference (DASC'18). London.

Verdonk Gallego, C.E., Gómez Comendador, V.F., Sáez Nieto, F.J., Orenga Imaz, G., Arnaldo Valdés, R.M., 2018b. Analysis of air traffic control operational impact on aircraft vertical profiles supported by machine learning. Transp. Res. Part C Emerg. Technol. doi:10.1016/j.trc.2018.03.017

Verdonk Gallego, C.E., Javier, F., Nieto, S., 2016. Discussion on Complexity and TCAS Indicators for Coherent Safety Net Transitions, in: Schaefer, D. (Ed.), 6th SESAR Innovation Days. Delft, p. 8.

Wang, Z., Liang, M., Delahaye, D., 2018. A hybrid machine learning model for short-term estimated time of arrival prediction in terminal manoeuvring area. Transp. Res. Part C Emerg. Technol. 95, 280–294. doi:10.1016/j.trc.2018.07.019

Wickens, C.D., 2008. Multiple Resources and Mental Workload. Hum. Factors J. Hum. Factors Ergon. Soc. 50, 449–455. doi:10.1518/001872008x288394

Acronyms

AMAN	Arrival Manager
ANN	Artificial Neural Network
ATC	Air Traffic Control
ATCC	Air Traffic Control Centre
ATCo	Air Traffic Controller
ATFCM	Air Traffic Flow and Capacity Management
ATM	Air Traffic Management
CFL	Cleared Flight Level
CP	Crossing Point
FL	Flight Level
HIM	Horizontal Interdependency Matrix
LAT	Look-Ahead Time
ML	Machine Learning
NM	Nautical Mile
PCA	Principal Component Analysis
PDF	Probability Density Function
PMF	Probability Mass Function
RDBSCAN	Recursive Density-Based Clustering Algorithm with Noise
RFL	Requested Flight Level
RMSE	Root Mean Squared Error
STCA	Short-Term Conflict Alert
TBO	Trajectory-Based Operations
TOD	Top of Descent
TP	Trajectory Prediction / Trajectory Predictor

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
υ	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
$\hbar_{il}^0$	Horizontal Central Trajectory of the recurrent pattern α_{il}^0
γ_{il}	Arc-length parameter associated to $\hbar_{il}^0$
Γ_{il}	Arc-length space associated to $\boldsymbol{\hbar_{il}^0}$
L_{il}	Total length of $\hbar_{il}^0$
$\underline{X}$	Feature vector for the horizontal trajectory predictor
$\widetilde{\eta_{r,\imath}}$	Set of predicted trajectories for the aircraft *r* for a major airspace flow *i*
$\widetilde{\eta_{r,il}}$	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 $\hbar_{il}^0$
$\mathcal{T}_{il}^{\gamma_{il}}$	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 $\hbar_{il}^0$ 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 $\hbar_{il}^0$
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 $\hbar_{il}^0$

Appendix I

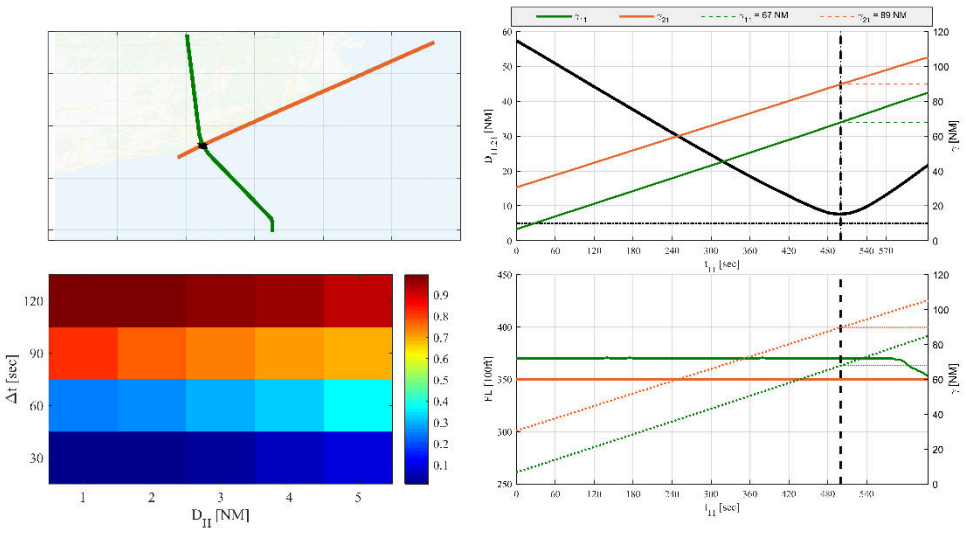


Figure 26 Horizontal Interdependency Measure. Case of Study 2

The second case of study is illustrated in the Figure 26. 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.

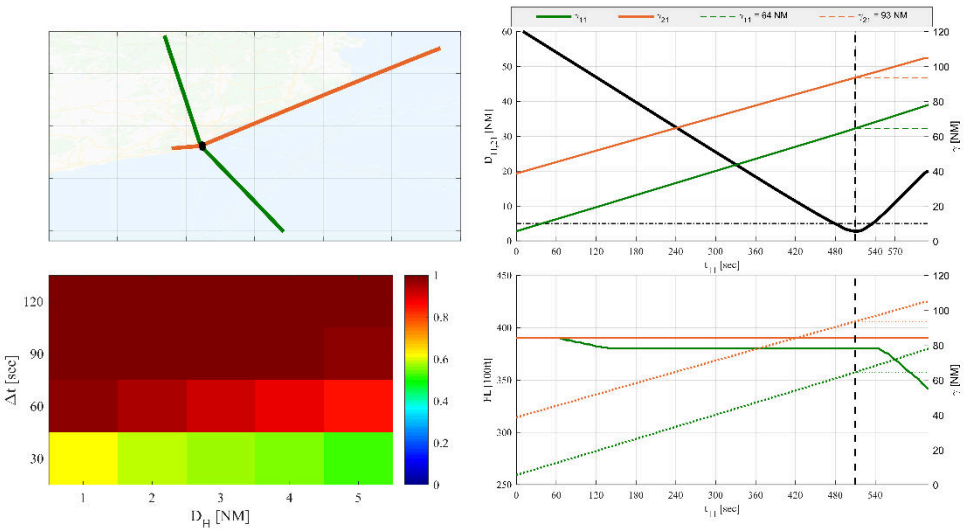


Figure 27 Horizontal Interdependency Measure. Case of Study 3

The third case of study is illustrated in the Figure 27. 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 instruction one minute after entering the sector, to prevent the conflict.

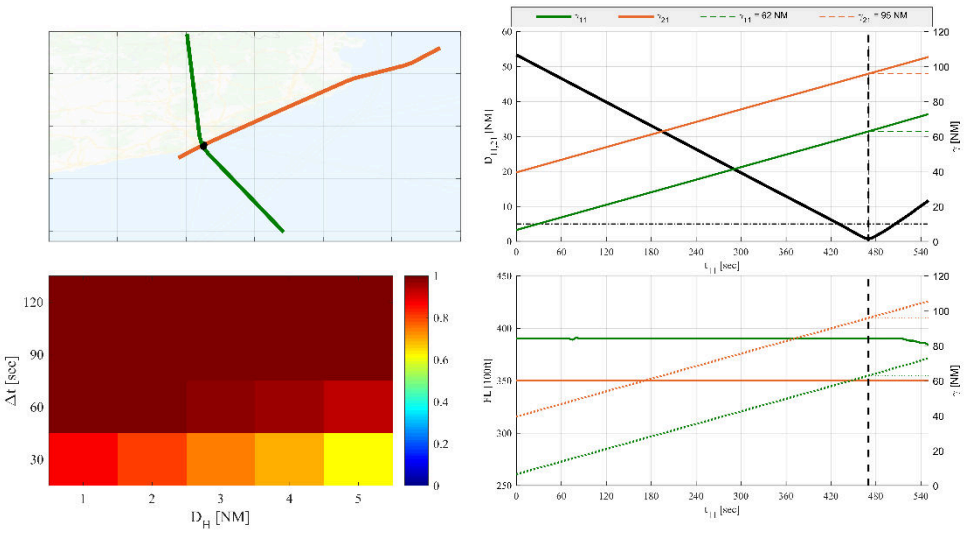


Figure 28 Horizontal Interdependency Measure. Case of Study 4

The fourth case of study is illustrated in the Figure 28. 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.



https://dspace.org/

A machine learning approach to air traffic interdependency modelling and its application to trajectory prediction

Verdonk Gallego, Christian Eduardo; Gómez Comendador, Victor Fernando; Carmona, Manuel Angel Amaro; Valdés, Rosa María Arnaldo; Nieto, Francisco Javier Sáez ...

2019-08-27
Elsevier
https://doi.org/10.1016/j.trc.2019.08.015
Verdonk Gallego CE, Gomez Comendador VF, Amaro Carmona MA, et al., (2019) A machine learning approach to air traffic interdependency modelling and its application to trajectory prediction. Transportation Research Part C: Emerging Technologies, Volume 107, October 2019, pp. 356-386.

Downloaded from a DSpace repository. https://dspace.org/