

## Article

# Trajectory Predictor and Conflict Detection Figures of Merit for a Performance-Based Adaptive Air Traffic Monitoring System

Chen Xia <sup>1,2,\*</sup> , Christian Eduardo Verdonk Gallego <sup>2</sup> , Adrián Fabio Bracero <sup>3</sup>,  
Víctor Fernando Gómez Comendador <sup>1</sup>  and Rosa María Arnaldo Valdés <sup>1</sup> 

<sup>1</sup> Aeronautical Systems, Air Transport and Airports Department, Universidad Politécnica de Madrid, 28040 Madrid, Spain; fernando.gcomendador@upm.es (V.F.G.C.); rosamaria.arnaldo@upm.es (R.M.A.V.)

<sup>2</sup> CRIDA A.I.E., 28022 Madrid, Spain; ceverdonk@e-crida.enaire.es

<sup>3</sup> EUROCONTROL Innovation Hub (EIH), 91222 Bretigny-Sur-Orge, France; adrian.fabio@eurocontrol.int

\* Correspondence: cxia@e-crida.enaire.es

**Abstract:** This paper investigates the impact of trajectory predictor performance on the encounter probability generated by an adaptive conflict detection tool and examines the flexibility of the tool dependent on its adjustable thresholds, using historical radar track data. To achieve these objectives, two figures of merit were proposed: System Dynamic Range and System Tuning Envelope. To examine the conflict detection's performance variability under different uncertainty levels and predictor types, simple multi-horizon trajectory predictors trained with two machine learning techniques of different characteristics are assessed at various look-ahead times: extreme gradient boosting with a discrete nature and a multi-layer perceptron regressor with a continuous nature. The results highlight the interdependence between the performances of the trajectory predictor and the conflict detector, and the quantification of this relationship can be represented through a sigmoid function. In addition, the two proposed figures of merit are effective for selecting suitable operating points in an adaptive conflict detector, based on dynamic thresholds and the performance requirements necessary for the trajectory predictors to achieve the expected detection performance at different look-ahead time.

**Keywords:** adaptive conflict detection; trajectory prediction; uncertainty; encounter probability; adjustable thresholds; machine learning; air traffic management; air traffic control



**Citation:** Xia, C.; Verdonk Gallego, C.E.; Fabio Bracero, A.; Gómez Comendador, V.F.; Arnaldo Valdés, R.M. Trajectory Predictor and Conflict Detection Figures of Merit for a Performance-Based Adaptive Air Traffic Monitoring System. *Aerospace* **2024**, *11*, 155. <https://doi.org/10.3390/aerospace11020155>

Academic Editor: Roberto Sabatini

Received: 27 December 2023

Revised: 31 January 2024

Accepted: 13 February 2024

Published: 15 February 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The accommodation of increasing levels of air traffic has long been limited by the insufficient capacity of the ATM (Air Traffic Management) system, leading to a significant number of delays caused by air traffic flow and management capacity constraints. This not only results in longer flight times but also contributes to increased fuel consumption and emissions [1].

The COVID-19 pandemic caused a significant drop in air traffic, but as the global situation improved, traffic now is rising once again to pre-pandemic levels [2]. To meet the increasing demand, airspace capacity must increase. Capacity increments can be approached from various perspectives. The main driver of capacity limits is the workload of air traffic controllers (ATCos). The workload of ATCos is tied to the tasks they perform during their shifts, and it varies based on the number of flights within their sector and the level of complexity each flight presents [3]. For example, in an en-route sector, a flight may only require minimal attention if it is simple to monitor, or it may require a significant amount of attention if it interacts with other flights and potentially causes conflicts that require additional clearances [4,5]. Therefore, employing additional supporting tools for ATCos to handle potential conflicts has long stood as one of the solutions to reduce workload and increase sector capacity.

Efforts are underway, particularly through the SESAR 3 Joint Undertaking programme in Europe and the NextGen program in the United States, to address this challenge. In par-

ticular, a project of the SESAR programme, PJ18-W2 4DSkyways, developed a solution named “Improved Ground Trajectory Predictions enabling future automation tools”, which aimed to increase the quality of separation management service. The primary objective of the solution was to validate enhanced conflict detection and resolution (CDR) tools, rooted in the improvement of the performance of ground trajectory predictors (TPs), thereby enhancing the overall quality of service [6]. The present paper was developed within this project. Therefore, this paper presents metrics to assess the interrelation between the performance of TPs and CDs (Conflict Detectors), to facilitate establishing performance requirements for TP services to support the separation management function.

The sections of this paper are arranged as follows: Section 1 includes a brief introduction and Section 1.1 presents a literature review of the related topic; the methodology adopted for the construction of the TP and CD tools and proposed figures of merit is described in Section 2, including the outline of the selected use case for the analysis; the results of the figures of merit are displayed in Section 3; and finally, the findings of this work and the next steps are summarised in Section 4.

### 1.1. The State of the Art

A study carried out by EUROCONTROL Experimental Centre using fast time simulation demonstrated that, when controllers are equipped with conflict detection tools such as medium-term conflict detection (MTCD) and a tactical controller tool with a perfect accuracy level, their task load is reduced by up to 21%, leading to a more balanced distribution of task load between planner and tactical controllers [7]. Consequently, one viable approach to reducing air traffic controller workload and increasing capacity is to provide them with more advanced decision support tools (DSTs) for CDR [8].

However, CDR functionality is dependent on an accurate trajectory prediction. As a result, the performance of the CDR tool and the overall ATC (Air Traffic Control) system is directly impacted by the accuracy of the TP [8].

Trajectory prediction was defined in [9] as the process that estimates a future trajectory of an aircraft through computation, and it is performed by a trajectory predictor. The TP can adopt various methodological approaches [10]. Ref. [11] conducted a general review of 282 papers, with a detailed examination of 20 of them. These selected studies showcased diverse methodologies, including mathematical models such as point-mass models with the BADA (Base of aircraft data) database [12], kinematic models, and kinetic models, among others. Recent studies have facilitated the integration of novel input sources and methodologies. Regarding data, the inclusion of onboard information through extended projected profiles enables prediction enhancement [13]. Regarding the method, the utilisation of data-driven predictors trained using machine learning (ML) algorithms has become widespread in present times [14,15]. These ML techniques have exhibited superior performances to conventional methods, given an appropriate dataset and training pipeline [16].

Predictions are always subject to error. The inaccuracies of predictions stem from diverse sources, leading to a high level of uncertainty [17–20], and their applicability depends on the operational context [21]. The consideration of this uncertainty is crucial in decision-making processes within ATM. High uncertainty leads to an amplification of the workload, due to the increasing need to monitor potential conflicts, an increase in both false and missed alerts [19,22], and results in more drastic clearances at a small look-ahead time (LAT) [23]. On the contrary, low uncertainty implies high reliability, enabling anticipated resolutions.

Ehrmanntraut [22] highlighted the importance of considering the uncertainty of trajectory prediction in conflict detection tools. CDR tools generally operate by identifying the breach of minimum separation between two aircraft, indicated by a specific measure falling below a predetermined threshold. The commonly used approach involves projecting the trajectory over a specified look-ahead time and determining if the two aircraft will lose

separation by evaluating the probability of their relative distance being below a previously set threshold for each time point in the series [24].

Regarding the requirements of TP and CDR, basic metrics to evaluate TP performance encompass longitudinal, lateral, and vertical prediction error [25]. In [26], Mondoloni introduced a framework to define metrics, providing a list of metrics for assessing TP performance and presenting the system operating characteristics (SOC) curve, which depicts the trade-off between false alert (FA) and missed alert (MA) of a CD tool. Additionally, Schuster et al. [8] conducted a review on the requirements of TP and CDR for each element of the SESAR Concept of Operations and proposed performance metrics such as TP accuracy, TP integrity, CD false alert, and CD integrity.

Concerning the link between TP and CDR, ref. [27] proposed a methodology for the functional verification and performance validation of TPs. The proposed process categorises TP requirements into two groups: direct and indirect. Direct requirements are directly related to the TP itself, such as prediction accuracy, while indirect requirements relate to the impact on the overall system performance, like meeting the required CDR performance. Ref. [27] highlighted the difficulty in validating indirect TP requirements using recorded radar track data, primarily due to controller-issued clearances for deconfliction, which complicate the computation of false alerts and positive alerts. Therefore, ref. [27] proposed a validation approach relying on simulations that have a high level of fidelity, although the cost is expensive.

Many studies have used simulations to assess the impact of TP accuracy on CD performance. Ref. [28] evaluated the relationship between different levels of longitudinal error and the performance of a conflict probe using Monte Carlo simulations, and ref. [29] used simulations with a time shift from [30] to emulate conflict events, evaluating several scenarios of different TP accuracy levels using the hit rate and the FA rate. The latter study led to the conclusion that the conflict probe performance depends on TP accuracy, showcasing reduced false alarm rates and increased true positive rates as TP accuracy improves. Furthermore, the CD performance was estimated using a linear regression function based on TP accuracy.

Based on the SOC curve [26] and the conflict probability [24], this study proposes quantitative figures of merit using historical traffic data to evaluate the influence of TP performance on the encounter probability of a dynamic CD tool, without the need for expensive simulation effort. Apart from revealing their relation, these metrics inform the selection of the operating point of the CD for different LATs, based on the trade-off between false alerts and missed alerts, establishing requirements for TP performance and CD settings.

In this context, an encounter is considered to be a state which may develop into a loss of separation between two flights in the absence of ATCos intervention, and dynamic CD refers to a detection tool with dynamic variables in the setting capable of adjusting its operational performance to meet task-specific requirements.

Two TPs trained with different ML algorithms are selected for comparative evaluation regarding TP accuracy and prediction stability: extreme gradient boosting (XGB), a member of boosting algorithm family, and a multi-layer perceptron regressor (MLP), a type of artificial neural network.

This paper focuses on analysing the dependency between the performance of TPs and CDs, considering performance in terms of their accuracy and stability. The impact of accuracy is quantitatively assessed with two figures of merit, each supporting the selection of operating points in adaptive CD and defining performance requirements.

## 2. Methodology

This section provides an overview of the trajectory predictor and conflict detector tools developed for the analysis, including the corresponding figures of merit to assess their performance and their relationships. As the analysis focuses on the horizontal plane,

the predictor is referred to as the horizontal trajectory predictor (HTP). To facilitate comprehension, the nomenclature used in this paper is summarised in Tables 1–4.

**Table 1.** Nomenclature—Relevant Abbreviations.

| Abbreviation | Definition                        | Abbreviation | Definition                                     |
|--------------|-----------------------------------|--------------|--|
| ANN          | Artificial Neural Network         | MLP          | Multi-Layer Perceptron Regressor               |
| AoI/R        | Area of Interest/Responsibility   | NM           | Nautical Miles                                 |
| ATCo         | Air Traffic Controller            | OSED         | Operational Service and Environment Definition |
| CA/P         | Critical Area/Point               | RP           | Real positive                                  |
| CDR          | Conflict Detection and Resolution | s            | Seconds  |
| DCB          | Demand and Capacity Balancing     | SDR          | System Dynamic Range                           |
| DST          | Decision Support Tool             | STE          | System Tuning Envelope                         |
| FA/MA        | False Alert/Missed Alert          | TA           | True Alert                                     |
| FL           | Flight Level                      | TN           | True negative                                  |
| HTP          | Horizontal Trajectory Predictor   | TP           | Trajectory Predictor                           |
| LAT          | Look-Ahead Time                   | WV           | Wake Vortex                                    |
| ML           | Machine Learning                  | XGB          | eXtreme Gradient Boosting regressor            |

**Table 2.** Notation—Indexes.

| Index | Definition                  | Index | Definition                            |
|-------|-----------------------------|-------|---------------------------------------|
| $i$   | Recurrent pattern of flow 1 | $k$   | Critical point within a critical area |
| $j$   | Recurrent pattern of flow 2 | $p$   | Present or current                    |

**Table 3.** Notation—Trajectory Predictor.

| Notation                    | Definition  | Notation | Definition                                |
|-----------------------------|---|----------|---|
| $\gamma$                    | Arc-length along the trajectory of each flow pattern [NM]   | $t_e$    | Elapsed time at sector entry [s]          |
| $\gamma_p$                  | Current position of the flight, measured as elapsed distance $\gamma$ of the flight from the entry point [NM] | $t_p$    | Elapsed time from entry to $\gamma_p$ [s] |
| $\delta\gamma$              | Target distance from the current location to make prediction [NM]   | $FL_p$   | Flight level at $\gamma_p$ [FL]           |
| $\hat{t}_{\gamma,\gamma_p}$ | Predicted required time to reach the target point $\delta\gamma$ from the current position $\gamma_p$ [s]     | $v_p$    | Ground speed at $\gamma_p$ [knots (kts)]  |

**Table 4.** Notation—Conflict Detector.

| Notation         | Definition  | Notation   | Definition   |
|------------------|---|--|--|
| $f_1$            | Flight 1 of an interacting flight pair  | $\gamma_p$   | Arc-length of evaluation, indicating $\gamma_{p2}$   |
| $f_2$            | Flight 2 of an interacting flight pair  | $\gamma_{CP_k}^{i,f_1}(\gamma_{CP_k}^{j,f_2})$                                       | $\gamma$ of pattern i (j) of $f_1$ ( $f_2$ ) to reach the critical point k [NM]  |
| $d_{THR}$        | Horizontal distance threshold to identify CP [NM]                               | $CA_{i,j}$   | Critical point k between patterns [i, j], composed of the location of $f_1$ and $f_2$ along patterns i and j, respectively, [NM] |
| $\Delta t_{THR}$ | Time difference threshold to obtain real conflict and encounter probability [s] | $CP_k^{i,j}$   | Critical point k between patterns [i, j], composed of the location of $f_1$ and $f_2$ along patterns i and j, respectively, [NM] |
| $P_{THR}$        | Probability threshold and operation threshold to decide whether notify an alert | $\hat{t}_{\gamma_p,\gamma_{CP_k}}^{i,f_1}(\hat{t}_{\gamma_p,\gamma_{CP_k}}^{j,f_2})$ | Time predicted to reach $\gamma_{CP_k}^{i,f_1}(\gamma_{CP_k}^{j,f_2})$ at $\gamma_p$ through pattern i (j) [s]                   |
| $\Delta t_{min}$ | Real-time difference [s]  | $\Delta \hat{t}_{res,\gamma_p,CP_k}^{i,j}$   | Difference between the predicted time to reach $CP_k^{i,j}$ of both flights for the patterns [i, j] [s]                          |
| $\Delta d_{min}$ | Real minimum horizontal separation [NM]   | $\sigma_{\gamma_p,\gamma_{CP_k}}^{i,f_1}(\sigma_{\gamma_p,\gamma_{CP_k}}^{j,f_2})$   | Standard deviation of HTP to predict $\gamma_{CP_k}^{i,f_1}(\gamma_{CP_k}^{j,f_2})$ at $\gamma_p$ through pattern i (j) [s]      |
| $\Delta h_{min}$ | Real minimum vertical separation [ft]   | $\sigma_{res,\gamma_p,CP_k}^{i,j}$   | Combined standard deviation of the predictions of both flights for $CP_k$ [s]  |

Table 4. Cont.

| Notation                   | Definition                               | Notation                                   | Definition   |
|----------------------------|--|--|--|
| $\mu$                      | Mean                                     | $P_{\gamma_p, CP_k}^{i,j, \Delta t_{THR}}$ | Encounter probability at $CP_k$ of the patterns [i, j] at $\gamma_p$ |
| $\sigma$                   | Standard deviation                       | $P_{\gamma_p, CA}^{i,j, \Delta t_{THR}}$   | Encounter probability of the patterns [i, j] of the flight pair      |
| $\gamma_{p1}(\gamma_{p2})$ | Current position of $f_1$ ( $f_2$ ) [NM] | $P_{\gamma_p}^{\Delta t_{THR}}$            | Encounter probability of the flight pair                             |

In this study, the measures are derived from the time separation at a specified location where minimum separation will be violated if both aircraft arrive concurrently, based on one of the cognitive abstractions of air traffic control operators (ATCOs) [4], who concentrate on critical points in an air traffic sector. Consequently, the probabilistic definition of that metric between two aircraft also considers uncertainty in the temporal dimension instead of the spatial dimension. This perspective was explored in [23].

Hence, the primary reference variable for the models is  $\gamma$ , which represents the arc-length along the trajectory of each pattern [23].

### 2.1. Horizontal Trajectory Predictor

The HTPs trained in this study are termed *multi-horizon* HTPs; this term refers to trajectory predictors that can update the prediction in accordance with variations in current position information by updating the inputs accordingly. This unique characteristic enables the assessment of performance at various positions without requiring the training of multiple individual predictors.

The objective of the HTP is to predict the time a flight requires ( $\hat{t}_{\gamma, \gamma_p}$ ) to reach a given distance  $\gamma$  from the current position  $\gamma_p$ . The inputs include the following:

- aircraft characteristics: wake vortex (WV);
- information collected at the current position: current position ( $\gamma_p$ ), current flight level ( $FL_p$ ), current speed ( $v_p$ ) and elapsed time from the entry ( $t_p$ );
- target point: distance from current location to the target distance for prediction ( $\delta\gamma$ ).

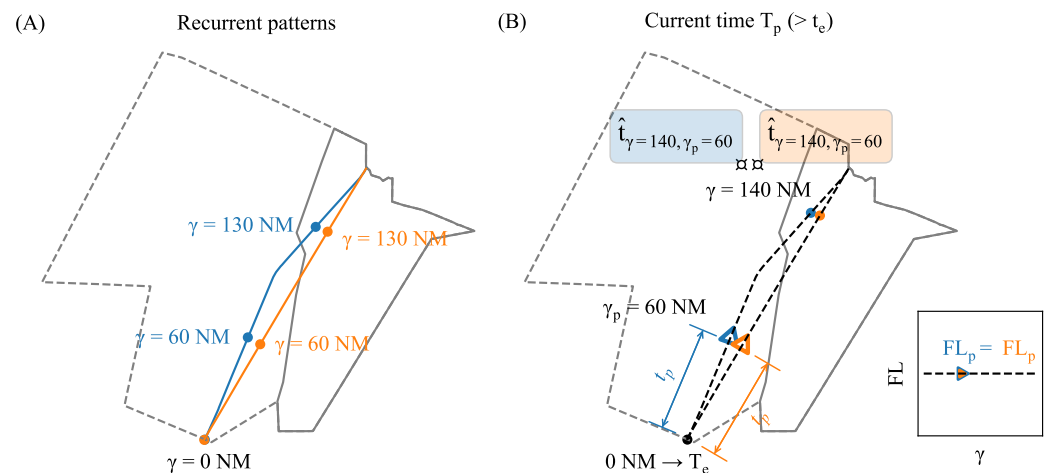
Regarding training, each identified traffic flow possesses its own predictor. As pointed out in [31], each flow exhibits recurring patterns—with the most commonly flown trajectories within it—with the planned route being one of them. Hence, it is worth noting that the same  $\gamma$  can refer to different geographic coordinates, depending on the specific pattern and flow that are being considered, as shown in Figure 1. However, for each recurrent pattern, the relation between the coordinates and  $\gamma$  is bijective.

The left-hand side of Figure 1 illustrates the concept of traffic flow, with two recurrent patterns for a single flow, and depicts the bijective relation between  $\gamma$  and the geographical coordinates, varying according to the pattern. The right hand side of Figure 1 shows the concept of *multi-horizon* HTP, which depends on the current position and the target point.

The entry to the sector is considered to be the starting point of the trajectory of analysis. When a flight reaches its entry point at instance  $T_e$ , its current position or elapsed distance within the sector  $\gamma_p$  is 0 NM, and its associated elapsed time of entry  $t_e$  is also 0 s. As the flight progresses, its position  $\gamma_p$  changes accordingly, along with the associated elapsed time  $t_p$ .

To present performance differences and their implications, two *multi-horizon* HTPs were trained using different types of ML algorithms that used the same features:

- Extreme gradient boosting regressor: a tree-based algorithm capable of providing reasonable performance with a small training sample. This HTP is referred to as XGB-HTP.
- Multi-layer perceptron regressor: a feed-forward artificial neural network (ANN). This HTP is referred to as MLP-HTP.



**Figure 1.** (A) Example of differences between patterns of the same flow. The same  $\gamma$  represents different geographical locations depending on the pattern. (B) Prediction when the flight has flown 60 NM ( $\gamma_p$  is 60 NM) and the target point is at 140 NM.

## 2.2. Conflict Detector

The aim of the CD tool is to detect flight pairs that could evolve into an encounter without ATCo intervention. Each identified flight pair is considered as potential conflict. This paper only considers flight pairs that belong to different traffic flows, excluding potential conflicts within the same flow.

The CD used for the study has three user-defined thresholds, namely distance ( $d_{THR}$ ), time ( $t_{THR}$ ), and probability ( $P_{THR}$ ), for its functioning. The latter two parameters are always modifiable, while  $d_{THR}$  is immutable during the functioning of the tool; it can only be set at the configuration phase. The incorporation of the two dynamic variables within the system architecture endows the CD tool with adaptive capabilities, enabling it to adjust its responses to meet different operational needs.

The CD tool identifies potential conflicts by calculating the time separation between two flights at their critical area (CA), a specific spatial location where an infringement in horizontal separation occurs if two aircraft arrive simultaneously or with small time separation. This is inherent information for the CD tool that is calculated based on the recurrent patterns of flow pairs. Thus, a critical area between the pattern  $i$  of a flow and the pattern  $j$  of another flow ( $CA_{i,j}$ ) is an area that contains a set of tuples of two values of  $\gamma$  - one for each flow - where the horizontal separation is below the selected distance threshold ( $d_{THR}$ ). Each tuple is a critical point (CP),  $CP_k^{i,j}$ , and the number of CPs depends on the convergence between the two pattern trajectories.

Let us consider two flights: Flight 1 ( $f_1$ ) as the first flight entering the sector of analysis, and Flight 2 ( $f_2$ ) as the second. The mathematical representation of the critical area is provided as follows in Equation (1):

$$CA_{i,j} = \begin{bmatrix} CP_1^{i,j} \\ CP_2^{i,j} \\ \dots \end{bmatrix} = \begin{bmatrix} (\gamma_{CP_1}^{i,f_1}, \gamma_{CP_1}^{j,f_2}) \\ (\gamma_{CP_2}^{i,f_1}, \gamma_{CP_2}^{j,f_2}) \\ \dots \end{bmatrix} \quad (1)$$

The identification of the potential conflicts relies on comparing between the time separation estimated at CA and the minimum time separation ( $\Delta t_{THR}$ ).

As the estimated time separation is the difference between the two predictions—which has an associated uncertainty—the output of the CD tool is also a probabilistic variable. The formula for estimating the probability of conflict, presented in Equations (2)–(4), was demonstrated in [23], which resembles those demonstrated in [24] for distance-based



thresholds. The probability, as estimated using Equation (2) as follows, is derived by applying the error function, denoted as  $erf$ .

$$P(x \leq \Delta t_{THR}) = \frac{1}{2} \left[ erf\left(\frac{\Delta t_{THR} + \mu}{\sqrt{2}\sigma}\right) + erf\left(\frac{\Delta t_{THR} - \mu}{\sqrt{2}\sigma}\right) \right] \quad (2)$$

Based on the previous work of [23,24], when the prediction data follows a normal distribution centred around the prediction  $\mu$  with its corresponding standard deviation  $\sigma$ , the encounter probability is defined as the probability of the time separation falling below  $\Delta t_{THR}$ .

Due to the probabilistic nature of the output, a probability threshold ( $P_{THR}$ ) is set to select the sample of infringement notifications; flight pairs whose encounter probability is below  $P_{THR}$  will be notified and require attention. Therefore, given the same encounter probability, the notifications vary according to  $P_{THR}$ .

### Conflict Detection Process

Given two flights, the evaluation of their encounter probability begins when the second flight enters the sector. Hence, the arc-length of evaluation depends on the location of  $f_2$  ( $\gamma_{p2}$ ). For facilitation, the location of the evaluation is named  $\gamma_p$  in the sections of CD ( $\gamma_p = \gamma_{p2}$ ).

The detection process at the arc-length of evaluation  $\gamma_p$  encompasses the selection of potential interacting flight pairs and the calculation of the encounter probability.

The steps to identify the flights that potentially interact with  $f_2$  are as follows (see Figure 2A):

1. Select the concurrent flights of  $f_2$ ; these are the flights that are also in the sector at  $\gamma_p$ . Figure 2A shows an example of the instant when  $f_2$  enters the sector. The triangles represent the flights and the three coloured lines represent the recurrent patterns of  $f_2$ 's flow; these are possible trajectories that the flight might take.  $f_2$  has five concurrent flights at the entry (A, B, D, E, and F), and flight C is discarded because it is still out of the sector.
2. Select from the concurrent flights those that belong to flows that interact with  $f_2$ 's flow and do not belong to the same flow as  $f_2$ . In the example of Figure 2A, A and F are discarded because A belongs to  $f_2$ 's flow and F does not interact with  $f_2$ .
3. Estimate the present location of the concurrent flights:  $\gamma_{p1}$ .
4. Select those flights that have not reached any of their CA with  $f_2$ 's flow, i.e., flights whose  $\gamma_{p1} < \min(\gamma_{CP_1}^{i,f_1}, \gamma_{CP_2}^{i,f_1}, \dots)$ . The flights that have already overflown the critical areas will not lose separation with  $f_2$ . In the example of Figure 2A, B has already overflown CA.

For each remaining flight  $f_1$ , the encounter probability ( $P_{\gamma_p, CA}^{i,j, \Delta t_{THR}}$ ) is evaluated for each pattern combination  $[i, j]$ , because the pattern is unknown in advance.

1. For each  $CP_k^{i,j}$  of the  $CA_{i,j}$  (see Figure 2B):
  - (a) Predict the time each flight crosses  $CP_k^{i,j}$  for both flights:  $\hat{t}_{\gamma_p, \gamma_{CP_k}}^{i,f_1}$  and  $\hat{t}_{\gamma_p, \gamma_{CP_k}}^{j,f_2}$ . Figure 2B takes the pattern 1 (blue lines) of flight D and  $f_2$ , as an example. The HTP of each flow will predict the time each flight crosses the corresponding  $\gamma_{CP_k}^{i,f_1}$  ( $\gamma_{CP_k}^{j,f_2}$ ) of each  $CP_k^{i,j}$ .
  - (b) Determine the standard deviation of the following predictions:  $\sigma_{\gamma_p, \gamma_{CP_k}}^{i,f_1}$  and  $\sigma_{\gamma_p, \gamma_{CP_k}}^{j,f_2}$ .
  - (c) Obtain  $\Delta \hat{t}_{res, \gamma_p, CP_k}^{i,j}$  and the corresponding prediction error  $\sigma_{res, \gamma_p, CP_k}^{i,j}$  (see Equations (3) and (4), as follows).

$$\mu = \Delta \hat{t}_{res, \gamma_p, CP_k}^{i,j} = \hat{t}_{\gamma_p, \gamma_{CP_k}}^{i,f_1} - \hat{t}_{\gamma_p, \gamma_{CP_k}}^{j,f_2} \quad (3)$$

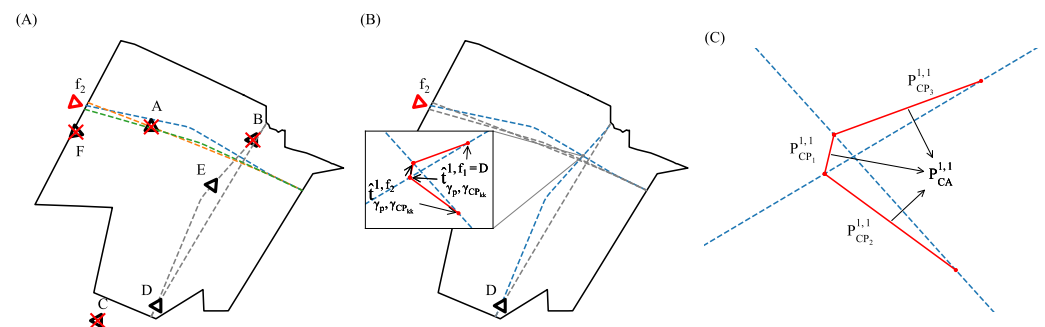
$$\sigma = \sigma_{res, \gamma_p, CP_k}^{i,j} = \sqrt{(\sigma_{\gamma_p, \gamma_{CP_k}}^{i, f_1})^2 + (\sigma_{\gamma_p, \gamma_{CP_k}}^{i, f_2})^2} \quad (4)$$

(d) Based on the selected  $\Delta t_{THR}$ , evaluate  $P_{\gamma_p, CP_k}^{i,j, \Delta t_{THR}}$  using Equation (2).

$$P_{\gamma_p, CP_k}^{i,j, \Delta t_{THR}} = P(x \leq \Delta t_{THR}) \quad (5)$$

2. The final encounter probability between patterns [i,j] ( $P_{\gamma_p, CA}^{i,j, \Delta t_{THR}}$ ) is calculated using the mean of the probabilities obtained from each  $CP_k^{i,j}$  (see Figure 2C).

Finally, flight pairs whose  $P_{\gamma_{p2}, CA}^{i,j, \Delta t_{THR}}$  exceeds  $P_{THR}$  are considered to be detected potential conflicts that could lead to horizontal infringement if the flights continue to follow their patterns [i,j].



**Figure 2.** (A) An example of an instantaneous capture of the situation when  $f_2$  enters the sector featuring candidate interacting flights (E and D) and non-interacting flights (the rest). The red triangle symbolizes  $f_2$ , while the black ones indicate other flights; those marked with a cross are non-interacting flights. (B) Each red line connects a pair of points, representing a critical point. Alongside each red line, the associated predicted time of the corresponding  $\gamma_{CP_k}^{1, f_1 \text{ or } f_2}$  of each  $CP_k^{1,1}$  between pattern combination [1,1] is shown. (C) Encounter probability between  $f_2$  and  $f_1$  ( $P_{CA}^{1,1}$ ), if they follow their planned trajectory.

### 2.3. Evaluation Metrics and Figures of Merits

For the assessment of HTP accuracy, the prediction errors are used. The target point for prediction is always a critical point, enabling the analysis of the encounter probability between flights.

Once evaluated, the impact of different HTP accuracies on the CD functioning is presented by two figures of merit. Beyond the quantification of the impact, each exhibits a specialised application.

The figures of merits are both defined in the Operational Service and Environment Definition (OSD) document of the project PJ18-W2 4DSkyways:

- System dynamic range (SDR): an element to support the possible encounter classification and to prioritise resolution, as well as evaluating the performance of the CD tool. It mainly represents the impact of the HTP accuracy on the outputs of the CD tool.
- System tuning envelope (STE): an element to facilitate the tuning of the CD tool using the parameters  $\Delta t_{THR}$  and  $P_{THR}$ . It assesses the performance of the CD tool, supporting the selection of operating points.

These metrics are further detailed in Sections 2.3.1–2.3.3.

#### 2.3.1. Horizontal Trajectory Predictor Performance Metrics

The performance of the HTP considered in this paper includes mainly the prediction accuracy, but the stability of the prediction is also analysed, with stability defined as the tendency to maintain the predictions' consistencies over time.



The performance of the trajectory predictors is assessed by analysing the residuals of the test sample, which are the differences between the prediction ( $\hat{t}_{\gamma_p, \gamma}$ ) and the real value ( $t_{\gamma_p, \gamma}$ ), as follows:

$$res = \hat{t}_{\gamma_p, \gamma} - t_{\gamma_p, \gamma} \quad (6)$$

Based on the residuals obtained from the test sample, the metrics that facilitate the visualisation of the performance are as follows:

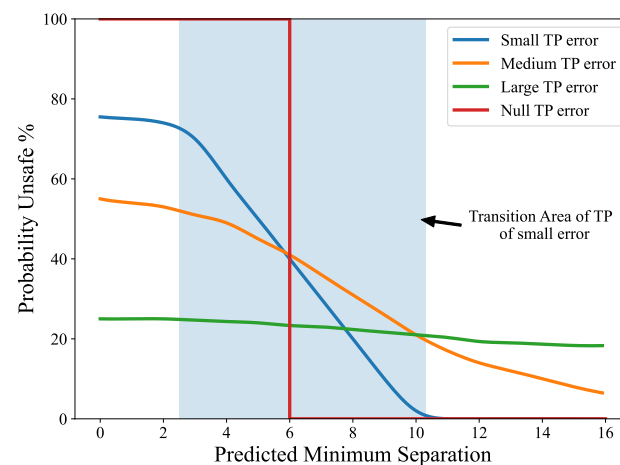
- Heatmap of residuals' mean ( $\mu_{res, \gamma_p, \gamma}$ ), which represents the residual mean for each current position and possible target point. Ideally the mean is zero for every combination of  $\gamma$  and  $\gamma_p$ .
- Heatmap of residuals' standard deviation ( $\sigma_{res, \gamma_p, \gamma}$ ): ideally the deviation is zero for every combination of  $\gamma$  and  $\gamma_p$ .

### 2.3.2. System Dynamic Range

Quoting the definition from the OSED: "... the dynamic range can be considered as the rates at which the conflict probability decreases as the predicted minimum separation increases. ...". Hence, the SDR shows the relationship between the outputs of the TP and the CD, which are the predictions and the encounter probability, respectively. It is an extension of the system operating characteristic curves [26].

The graphical representation of the SDR is the encounter probability  $P_{\gamma_p, CA}^{i,j, \Delta t_{THR}}$  versus the peak probability-predicted time separation  $\Delta \hat{t}_{\gamma_p, CP_{maxP}}^{i,j}$ .

Figure 3 shows a draft of possible SDR behaviours, adapted from the OSED. Each curve represents a SDR curve using predictors of a given accuracy, from the perfect TP in red to the green curve with a large TP error. The coloured area from the highest to the lowest probability of the SDR is named transition area.



**Figure 3.** Unsafe probability with given predicted minimum separation—draft.

If the trajectory prediction error is large—indicating large uncertainty—the CD performance shows a slow transition from high to low encounter probabilities with a broader transition area, suggesting a low dynamic range and a shallow curve. On the contrary, high-performance predictors will exhibit steep curves and consequently, a large dynamic range and a quick transition. In the case of a perfect detection tool without error, the curve will resemble a step function, with an encounter probability equal to 100% for predictions below the "safe" time-over-difference threshold, and an abrupt drop to zero thereafter.

When the uncertainty follows a normal distribution and when considering the absolute predicted time difference, the resulting probability has a folded normal distribution [24]. Hence, the representation of the SDR should follow a S-shaped function. Based on this

theoretical property, the shape of the SDR distribution can be approximated with the following logistic function:

$$f(x) = \frac{L}{1 + e^{-k(x-x_0)}} + b \quad (7)$$

The parameters that define the logistic functions are as follows:

- L: the maximum value of the curve ( $b = 0$ ) when the logistic curve is increasing, or the minimum value when it is decreasing.
- k: the steepness of the curve.
- b: the displacement of the curve over the y-axis.
- $x_0$ : the x-axis value of the midpoint.

### 2.3.3. System Tuning Envelope

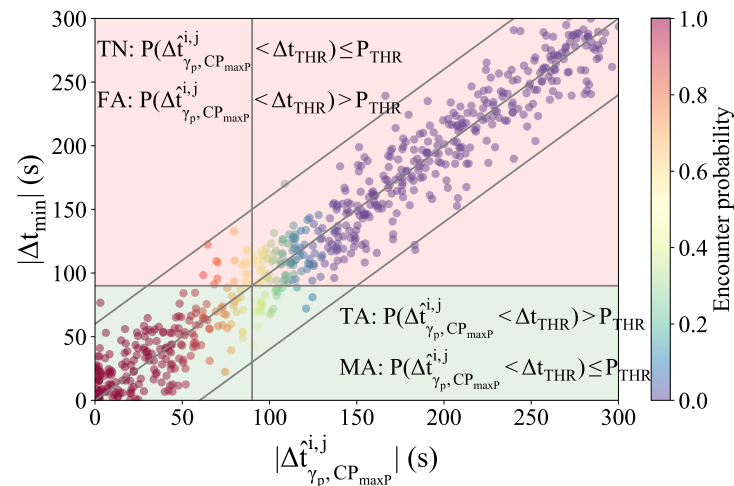
Due to the presence of uncertainty, conflict detection also has associated errors. Hence, a notification from the tool can be genuine or a false alert (FA), and a hidden instance can be correctly hidden or a missed alert (MA). These elements reflect the accuracy and reliability of the CD tool. STE, as a figure of merit, enables the assessment of CD accuracy using the proportion of FAs and MAs.

The adaptive conflict detector provides conflict alerts based on the estimation of the encounter probability, which depends on  $\Delta t_{THR}$  and  $d_{THR}$ , and the operation threshold  $P_{THR}$ . Every possible interaction with an estimated probability higher than  $P_{THR}$  will be notified, while those with an estimated probability lower than  $P_{THR}$  will be filtered out.

However,  $P_{THR}$  is a dynamic threshold. Therefore, a low  $P_{THR}$  will result in a large number of notifications, including most of the genuine infringements, in exchange for overloading the controller with excessive nuisance notifications. This increases the workload and diminishes user confidence in the tool. Conversely, a high  $P_{THR}$  implies fewer false alerts, but risks missing critical notifications.

The conditions to detect the nuisances and missed notifications are visualised in Figure 4. The horizontal and vertical grey lines indicate the  $\Delta t_{THR}$ . All instances falling into the green area represent real infringements, and those in the red area do not. Those on the left side of the vertical line are notified alerts, and the rest are hidden. Based on Figure 4, the following variables related to CD accuracy are derived and used for the representation of STE:

- False alerts or *nuisance*: incorrectly notified alerts. Flight pairs whose real minimum separation is **above** the threshold ( $|\Delta t_{min}| > \Delta t_{THR}$ ) but are notified as potential conflicts because the estimated probability is **above** the decision threshold ( $P_{\gamma p, CP_{maxP}}^{i,j} > P_{THR}$ ).
- Missed alerts or *missed*: incorrectly hidden alerts. Flight pairs whose real minimum separation is **below** the threshold ( $|\Delta t_{min}| < \Delta t_{THR}$ ) but the estimated probability is **below** the decision threshold ( $P_{\gamma p, CP_{maxP}}^{i,j} \leq P_{THR}$ ).
- True alerts (TA): correctly notified alerts. Flight pairs whose real minimum separation is **below** the threshold ( $|\Delta t_{min}| < \Delta t_{THR}$ ) and estimated probability is **above** the decision threshold ( $P_{\gamma p, CP_{maxP}}^{i,j} > P_{THR}$ ).
- True negative (TN): correctly filtered-out alerts. Flight pairs whose real minimum separation is **above** the threshold ( $|\Delta t_{min}| > \Delta t_{THR}$ ) and estimated probability is **below** the decision threshold ( $P_{\gamma p, CP_{maxP}}^{i,j} \leq P_{THR}$ ).
- Real positive (RP) or real conflict: flight pairs whose real minimum separation is below the threshold ( $|\Delta t_{min}| < \Delta t_{THR}$ ).



**Figure 4.** Conditions to detect nuisance and missed alerts based on predicted and real time separation. The grey lines represent the  $\Delta t_{THR}$  set at 90 s.

The proportions of nuisance and missed alerts used in STE are calculated using the following equations:

$$\text{Nuisance \%} = \frac{FA}{TN + FA} \quad (8)$$

$$\text{Missed \%} = \frac{MA}{TA + MA} \quad (9)$$

The graphical representation of STE illustrates the trade-off between nuisance and missed alerts for every  $P_{THR}$ . These are reflected as proportions of FAs and MAs, respectively. Changes in the output notifications resulting from adjustments to  $P_{THR}$  are captured in the STE, which helps balance the trade-off between nuisance and missed alerts. Consequently, STE supports the selection of the most suitable operating point for the CD tool at different LATs to align with specific operational needs.

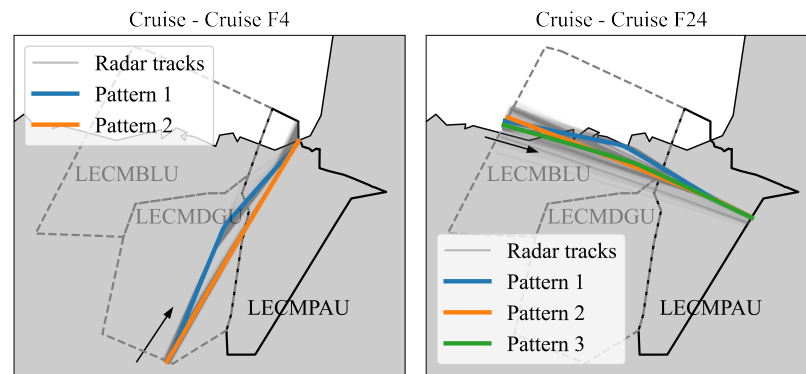
#### 2.4. Use Cases

To demonstrate the applicability of the proposed methodology, we have selected a use case that includes a complex crossing point. This use case has been chosen based on its relevance and adequacy to highlight our methodology.

The selected use case is located in the Spanish airspace. The area of responsibility (AoR) for the analysis is LECMPAU (Pamplona Upper Sector), where the interactions between flows take place. For the planning controller of LECMPAU, adjacent sectors should also be considered. Therefore, the collapsed sector LECMBDP is selected as the area of interest (AoI) for the analysis to anticipate possible encounters. LECMBDP comprises LECMPAU, LECMBLU (Bilbao Upper Sector), and LECMDGU (Domingo Upper Sector).

Two cruising flows are considered, named F4 and F24 (see Figure 5). F4 is a northward flow with 2890 flights, and F24 is eastbound with 1432 flights. F24 interacts with F4 within the AoR, close to the boundary.

This use case was selected because the traffic densities are significant and the critical points between the flows are close to the sector boundary. Clearances issued to the flights during their crossing through BLU or DGU could induce a conflict at PAU. An incident between a similar pair of flows occurred in 2018 [32].



**Figure 5.** Recurrent patterns of the two flows of analysis (pattern 1 refers to the planned route). The two flows interact in LECMPAU.

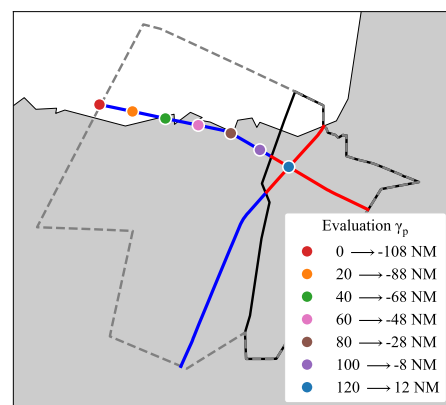
### Evaluation Contextual Parameters

The following four parameters are considered to analyse the adaptiveness and corresponding variability in performance: the independent variable  $\gamma_p$  and the three adjustable parameters  $d_{THR}$ ,  $\Delta t_{THR}$ , and  $P_{THR}$ .

For the analysis, the following values are selected:

- $d_{THR}$ : context threshold that influences the resulting critical area.
  - 1 NM.
- $\Delta t_{THR}$ : internal parameter of CD that impacts the encounter probability.
  - 10 s, 30 s, 60 s, 90 s, and 120 s.
- $P_{THR}$ : internal parameter of CD that affects the notified alerts.
  - from 0.1 to 0.99, incremented by 0.1.
- $\gamma_p$ : independent variable that indicates LAT and consequently, the level of prediction error.
  - 0, 20, 40, 60, 80, 100, and 120 NM from the entry point in the sector of analysis.

Due to the characteristics of the two flows, flight pairs with the second flight belonging to F24 are more likely to evolve into a conflict. Therefore, the analysis is focused on these instances and along the trajectory of F24. The arc-length of evaluation ( $\gamma_p$ ) is referred to LECMPAU's boundary; that is, the entry at LECMPAU is considered to be  $\gamma_p = 0$  NM. The planned trajectory of F24 from entry at LECMBDP to LECMPAU is 108 NM long, which equates to approximately 15 min, considering the mean required time of the traffic sample (see Figure 6).



**Figure 6.** Selected arc-length of evaluation  $\gamma_p$ .

### 3. Results

This section presents the results of the study, which are divided into two subsections. Section 3.1 analyses and compares the performance of the two HTPs, and Section 3.2 details the analysis of the metrics SDR and STE.

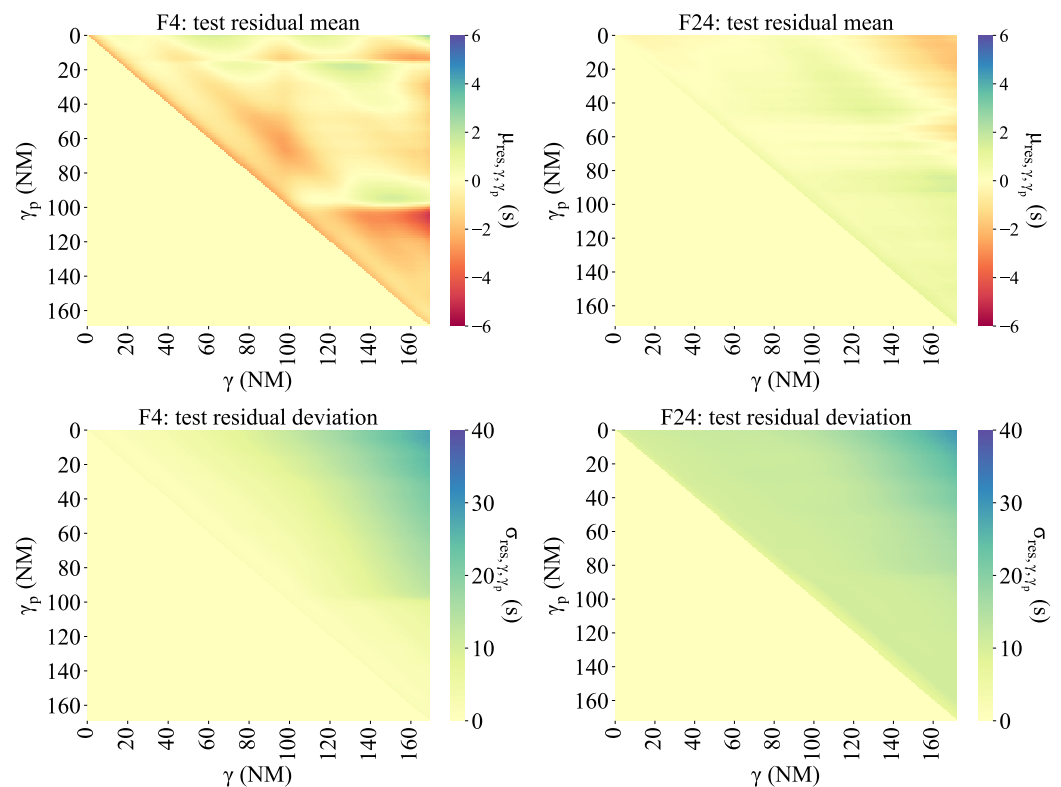
#### 3.1. Horizontal Trajectory Predictor Performance

This subsection presents the performance comparison between the four trained HTPs for the flows F4 and F24.

For both flows, the data were partitioned, with 75% reserved for training and the remaining 25% designated as the test sample, to assess the models' performance on unseen data. The XGB-HTPs were trained by fine-tuning their parameters, setting the depth of trees to five for F24 and six for F4, with the learning rate set to 0.055 in both cases. On the other hand, the network architecture of the MLP-HTPs consisted of two hidden layers, each with 12 neurons, and used the ReLU (Rectified Linear Unit) activation function.

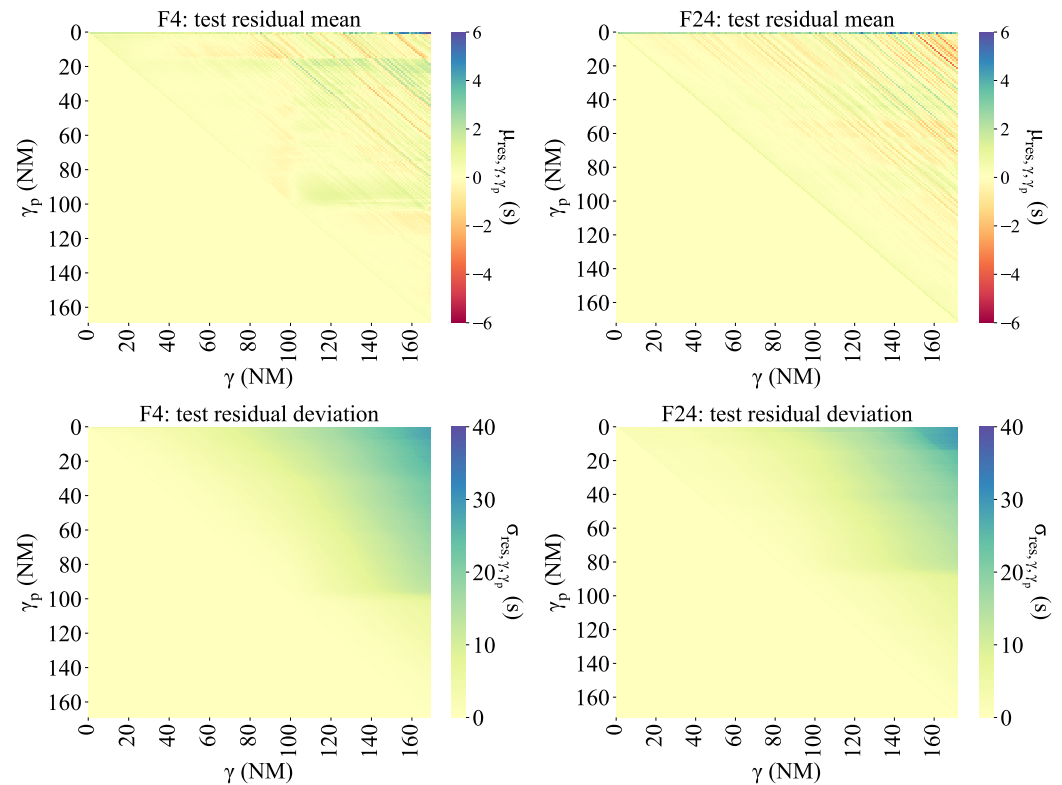
Regarding the HTP accuracy, it is dependent on the current position of the flight and the prediction point; a smaller  $\delta\gamma$  implies reduced uncertainty and increased predictability. Details of the prediction errors of the MLP and XGB-HTPs are shown in Figures 7 and 8.

In each figure, columns represent flows; the top row illustrates the mean of the residuals, and the second row shows the standard deviation. Each heatmap displays a matrix containing the corresponding value for each pair of  $\gamma_p$  and  $\gamma$ . As the target point  $\gamma$  should be larger than  $\gamma_p$ , only values in the upper triangular area are available.



**Figure 7.** Residual statistics of MLP models for the flows of analysis.

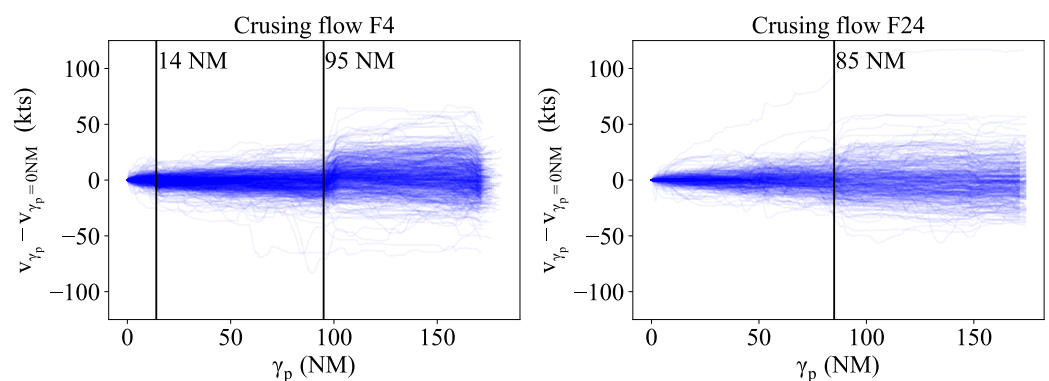
For both flows,  $\sigma_{res,\gamma_p,\gamma}$  escalates with an increase in the distance between  $\gamma_p$  and  $\gamma$  because the uncertainty increases with  $\gamma$ , whereas  $\mu_{res,\gamma_p,\gamma}$  predominantly fluctuates around zero. Statistically, the maximum value of  $\sigma_{res,\gamma_p,\gamma}$  across the four HTPs is comparable, approximately 28 s, while  $\mu_{res,\gamma_p,\gamma}$  varies within a range of −6 s to 6 s.



**Figure 8.** Residual statistics of XGB models for the flows of analysis.

In comparison, the difference in  $\sigma_{res, \gamma_p, \gamma}$  for F4's HTPs is minimal. However, for F24's HTPs, particularly when the  $\gamma_p$  of F24 is high,  $\sigma_{\gamma_p, \gamma}$  is notably lower in the XGB than in the MLP model. Additionally, the deviation of  $\mu_{res, \gamma_p, \gamma}$  from zero is more pronounced in F4.

Diverse regions are discernible in the  $\mu_{res, \gamma_p, \gamma}$  heatmaps of the four HTPs. These areas are related to trajectory patterns and common manoeuvres of the aircraft. A contributing factor to the formation of these regions is the speed variation. Figure 9 illustrates these speed variations from the entry point to the subsequent points ( $V_{\gamma_p} - V_{\gamma_p} = 0$  NM) for each flight of the test sample.



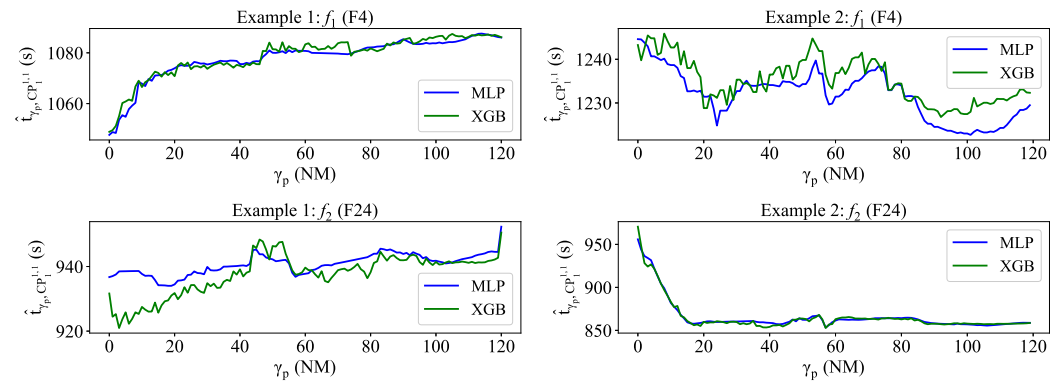
**Figure 9.** Speed variation in relation to the entry speed-test sample. The vertical lines indicate the positions of region splits in the performance heatmaps.

Regarding the prediction stability, a clear difference between the XGB-HTPs and MLP-HTPs is evidenced by the presence of diagonal bands in XGB-HTPs heatmaps (see Figure 8), indicating prediction instability over  $\gamma$ . This phenomenon stems from the discrete nature of decision tree algorithms, where similar inputs ranges lead to identical prediction allocation, despite differences in the actual input values.



However, when analysing the prediction stability of particular flights, predictions of MLP-HTPs also exhibit fluctuation over  $\gamma$ , despite the continuous nature of the algorithm. The cause of the fluctuation could be due to various reasons, such as the non-linearity of the problem.

Figure 10 shows two flight pairs selected as examples, which present the predictions of the total time from the entry to their corresponding  $CP_1^{1,1}$  at each  $\gamma_p$ .



**Figure 10.** Comparison of  $\hat{t}_{\gamma_p, CP_1^{1,1}}$  variation by  $\gamma_p$  of the flights between MLP and XGB models—examples 1 and 2.

In most cases, the predictions for XGB-HTPs exhibit more variation, although MLP-HTPs also show fluctuation.

### 3.2. Conflict Detection

One derivation from Equation (2) is that the performance of the CD is impacted by the uncertainty of the associated predictors, which increases over  $\gamma$ . Consequently, the performance tendency of the CD varies for each pair of patterns.

This subsection is focused on the horizontal potential conflicts between the pattern pair 1–1 of the flow pair F4–F24, where the second flight belongs to F24 (i.e., the flight of F4 is already within LECMBDP).

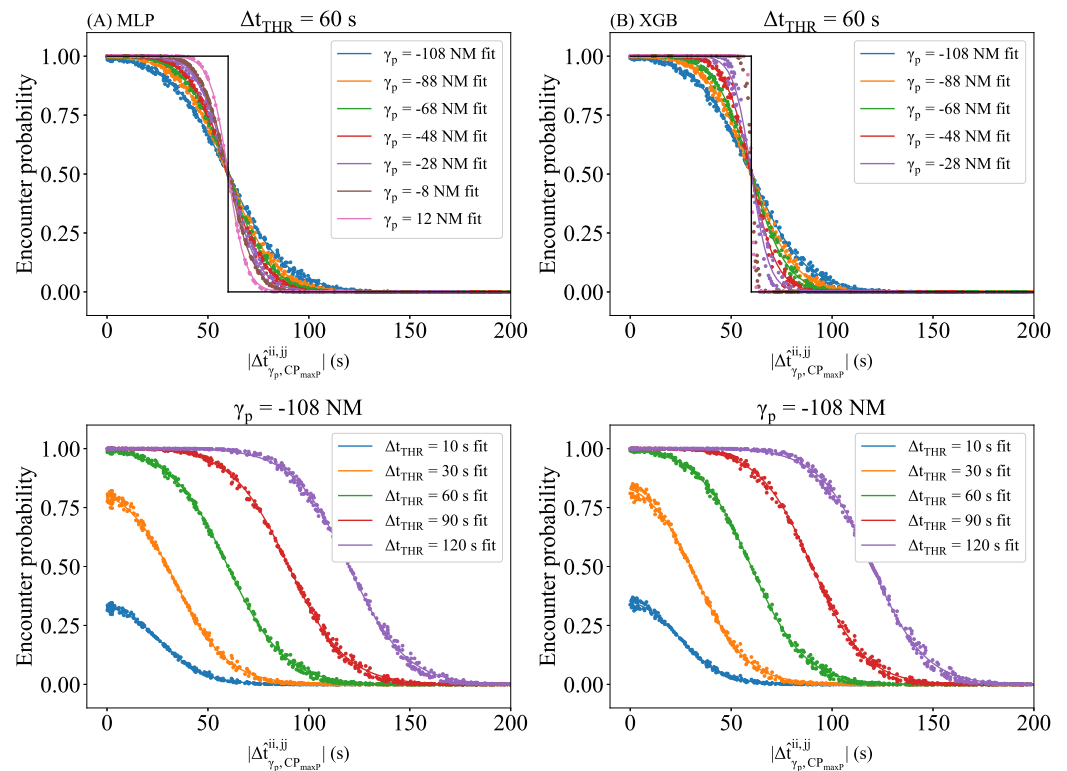
The CD performance is evaluated from two perspectives. Firstly, the HTP performance impact assessment considers the two differently modeled HTPs, alongside a selection of  $\gamma_p$ , which are selected as they represent different levels of uncertainty. Secondly, the examination of the effects of dynamic thresholds involves the  $P_{THR}$ , which sets the notification threshold, and  $\Delta t_{THR}$ , which influences the estimation of encounter probability.

The analyses of the metrics SDR and STE are detailed in Sections 3.2.1 and 3.2.2.

#### 3.2.1. System Dynamic Range

The estimated probability, and consequently the CD performance, are intimately related to the performance of the trajectory predictors used in the detection process. This dependency can be observed in the SDR graphs, which compare the predicted time separation and the estimated encounter probability.

Figure 11 presents the SDR graphs of the pair F4–F24 using different HTPs, time difference thresholds, and evaluation arc-lengths. It contains four graphs. The figures of column A are obtained using MLP-HTPs whereas those for column B are obtained using XGB-HTPs. The two upper graphs show the SDR curves across  $\gamma_p$  when  $\Delta t_{THR} = 60$  s, and the lower two show the variation in  $\Delta t_{THR}$  when  $\gamma_p = -108$  NM (entry point at LECMBDP).



**Figure 11.** (A) The SDR and logistic curve fit variation by  $\gamma_p$  when  $\Delta t_{THR} = 60$  s and variation by  $\Delta t_{THR}$  when  $\gamma_p = -108$  NM. The grey step line represents the ideal performance of the CD tool for the selected  $\Delta t_{THR}$ —MLP-HTPs for F4–F24. (B) The same graphs using XGB-HTPs for F4–F24.

In the four graphs, each point represents a flight pair whose encounter probability  $P_{\gamma_p, CP_{maxP}}^{1,1, \Delta t_{THR}}$  is greater than zero. Points of the same colour form the SDR distribution of the same evaluation arc-length  $\gamma_p$  or threshold  $\Delta t_{THR}$ . Each distribution has its corresponding logistic fit in the same colour.

Regarding the variation by  $\gamma_p$ , all logistic fits intersect with the black step function, which represents the ideal CD performance, around a midpoint. At this intersection, the  $|\Delta \hat{t}_{\gamma_p, CP_{maxP}}^{1,1}|$  is approximately  $\Delta t_{THR}$  and the encounter probability is about 0.5. Due to the characteristic of the probability estimation formula, Equation (2) simplifies when  $|\Delta \hat{t}_{\gamma_p, CP_{maxP}}^{1,1}| = \Delta t_{THR}$ , as shown in Equation (10), as follows:

$$P(x \leq \Delta t_{THR}) = \frac{1}{2} \left[ \text{erf}(0) + \text{erf}\left(\frac{2\Delta t_{THR}}{\sqrt{2}\sigma}\right) \right] = \frac{1}{2} \text{erf}\left(\frac{2\Delta t_{THR}}{\sqrt{2}\sigma}\right) \quad (10)$$

As error function is bounded within the range of  $-1$  and  $1$  and  $\Delta t_{THR}$  is positive, the resulting probability is  $0.5$  at this point in the distribution. Therefore, the transition area is centred around this midpoint; as  $|\Delta \hat{t}_{\gamma_p, CP_{maxP}}^{1,1}|$  increasingly deviates from  $\Delta t_{THR}$ , the resulting probabilities asymptotically approach the extremities of zero and one.

In addition, the prediction uncertainty diminishes as  $\gamma_p$  increases, because the flights are approaching the critical point. Consequently, the SDR distribution progressively approximates ideal behaviour, characterised by a steeper slope.

Regarding the variation by  $\Delta t_{THR}$ , it is noteworthy that the maximum probability of the SDR escalates with increasing  $\Delta t_{THR}$  until it reaches  $1$ . The maximum encounter probability occurs when  $|\Delta \hat{t}| = 0$  s, where the probability estimation (Equation (2)) simplifies to the following form:

$$P(x \leq \Delta t_{THR}) = \frac{1}{2} \left[ \text{erf}\left(\frac{\Delta t_{THR}}{\sqrt{2}\sigma}\right) + \text{erf}\left(\frac{\Delta t_{THR}}{\sqrt{2}\sigma}\right) \right] = \text{erf}\left(\frac{\Delta t_{THR}}{\sqrt{2}\sigma}\right) \quad (11)$$

As result, the encounter probability will equal the error function of  $\frac{\Delta t_{THR}}{\sqrt{2}\sigma}$ . The ideal result will be  $P = 1$  for  $|\Delta \hat{t}| = 0$  s. Given the characteristics of the error function, the larger  $\frac{\Delta t_{THR}}{\sqrt{2}\sigma}$  is, the closer to one the output of the error function will be. Therefore, the maximum probability achievable is determined by the selected threshold and the performance of the predictor and  $\Delta t_{THR} > \sqrt{2}\sigma$  is a necessary condition to approach optimum performance.

To check the condition, the highest  $\sigma_{res,\gamma_p,\gamma}$  of both flows for the first critical point are selected, which correspond to the prediction errors at their corresponding entry point (see Table 5). The  $\sigma_{res,\gamma_p,\gamma}$  of MLP-HTPs for F4 and F24 are 20.78 and 17.73 s, respectively, resulting in a  $\sigma$  (using Equation (4)) of 27.31 and  $\sqrt{2}\sigma \approx 38.63$ . On the other side, the  $\sigma_{res,\gamma_p,\gamma}$  of XGB-HTPs are 21.15 and 15.56 s, respectively, resulting in  $\sqrt{2}\sigma \approx 37.13$ . Thus, the results of both predictors reached a probability of one when the time separation equals 60 s. However, for a  $\Delta t_{THR}$  of 10 and 30 s, the maximum probability achievable is less than one, because  $\sqrt{2}\sigma > \Delta t_{THR}$ , resulting in a distribution with a reduced maximum value.

**Table 5.** Variation of  $\sigma$  in both HTPs for different  $\gamma_p$ .

| $\gamma_{p1}, \gamma_{p2}$ | $f_1$ MLP<br>$\sigma_{res,\gamma_{p1},\gamma}$ | $f_2$ MLP<br>$\sigma_{res,\gamma_{p2},\gamma}$ | $f_1$ XGB<br>$\sigma_{res,\gamma_{p1},\gamma}$ | $f_2$ XGB<br>$\sigma_{res,\gamma_{p2},\gamma}$ | MLP $\sigma$ | XGB $\sigma$ | MLP $\sigma$ –XGB $\sigma$ |
|----------------------------|--|--|--|--|--------------|--------------|----------------------------|
| 0, 0                       | 20.78  | 17.73  | 21.15  | 15.56  | 38.63        | 37.13        | 1.50                       |
| 20, 20                     | 17.35  | 14.27  | 17.74  | 11.47  | 31.77        | 29.88        | 1.89                       |
| 40, 40                     | 14.06  | 12.91  | 14.30  | 9.69   | 27.00        | 24.42        | 2.57                       |
| 60, 60                     | 11.48  | 12.08  | 11.54  | 6.64   | 23.57        | 18.82        | 4.74                       |
| 80, 80                     | 9.07   | 11.41  | 9.05   | 5.34   | 20.61        | 14.86        | 5.75                       |
| 100, 100                   | 4.18   | 10.65  | 3.87   | 1.64   | 16.18        | 6            | 10.23                      |

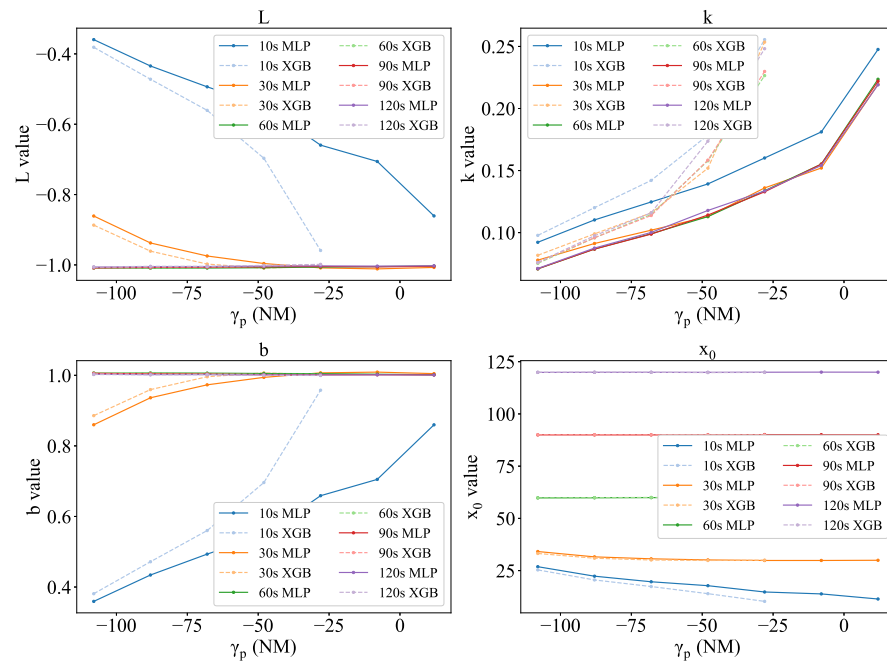
Although the overall deviation of the predictors is slightly lower in MLP-HTPs for F4 for low  $\gamma_{p1}$  and  $\gamma_{p2}$ , the low  $\sigma_{res,\gamma_p,\gamma}$  in XGB-HTPs for F24 compensates for the error and leads to a set of lower  $\sigma$ , as reflected in Table 5. As a result, the XGB-HTPs' distributions exhibit a more pronounced slope than those of the MLP-HTPs.

The analysis of the SDR graphs is further detailed by examining the parameters of the logistic functions (see Figure 12). The values that lead to an optimum SDR are:

- L:  $-1$ . The negative sign indicates the direction of the function.
- $x_0$ :  $\Delta t_{THR}$ .
- k: **the higher the better**.
- b:  $1$ .

In Figure 12, the optimal values of L, k, and b are achieved in the CD tool for the  $\Delta t_{THR}$  values of 60, 90, and 120 s, and they remain constant. This constancy is because they all satisfy the condition of  $\Delta t_{THR} > \sqrt{2}\sigma$  for any given  $\gamma_p$ , whereas for  $\Delta t_{THR}$  values of 10 s and 30 s, the parameters are far away from optimal. As uncertainty decreases with increasing  $\gamma_p$ , the parameters of these curves approach the optimal. Specifically, for 30 s, the SDR curve meets the optimal values for L, k, and b at  $-28$  NM. For the same reason, the logistic parameters of CD using XGB-HTPs reach the optimal values earlier than those using MLP-HTPs.

As result, the SDR and its associated logistic parameters facilitate setting the performance requirements for the trajectory predictors used in the detection process. By selecting the appropriate logistic parameters, the SDR curve that aligns with the expected CD output distribution and performance (to be analysed in Section 3.2.2) can be constructed, and the key points of the SDR will delineate the required TP accuracy. In the adaptive CD tool, where the dynamic parameters are adjustable during operation, the TP requirements may vary accordingly, to sustain CD performance across various CD configurations. This approach ensures flexibility in selecting the CD configuration while consistently upholding CD performance.

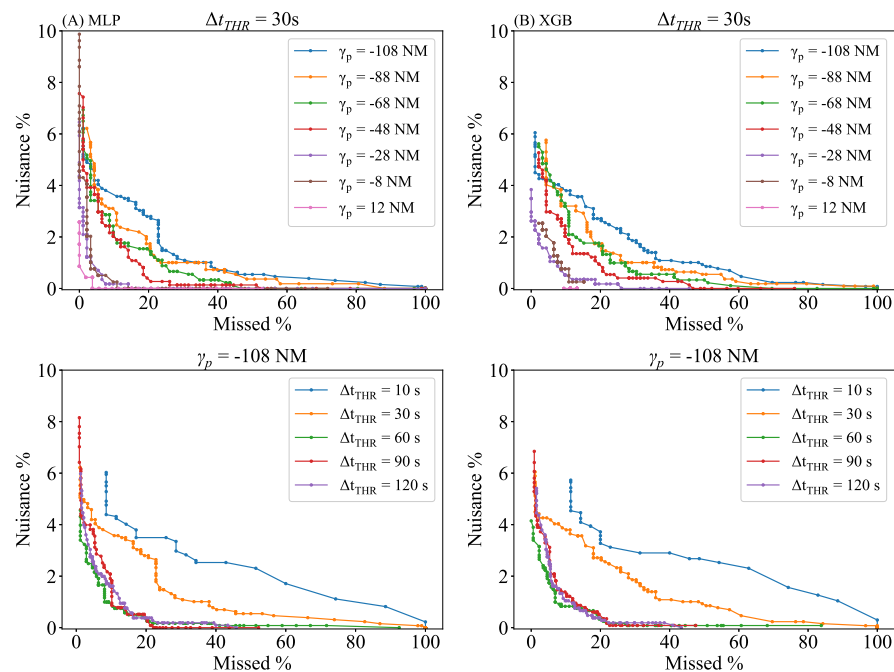


**Figure 12.** Variation of the four parameters of the sigmoid function by  $\Delta t_{THR}$  and  $\gamma_p$ .

### 3.2.2. System Tuning Envelope

For the variable STE, the CD performance variations, mainly the accuracy, arising from HTP performance and dynamic thresholds are examined.

The resulting STE curves are presented in Figure 13. Each curve is composed of several points, which represent the proportion of nuisances and missed alerts for a given  $P_{THR}$ . The two graphs on the first row present the variation of the STE curves by  $\gamma_p$  for a fixed  $\Delta t_{THR} = 30$  s, and the other two illustrate the variation by  $\Delta t_{THR}$  for a fixed  $\gamma_p = -108$  NM. In this subsection, the selected  $\Delta t_{THR}$  is 30 s to visualise distinct variations, because it is the closest value to the maximum  $\sigma$ , as presented in Section 3.2.1.

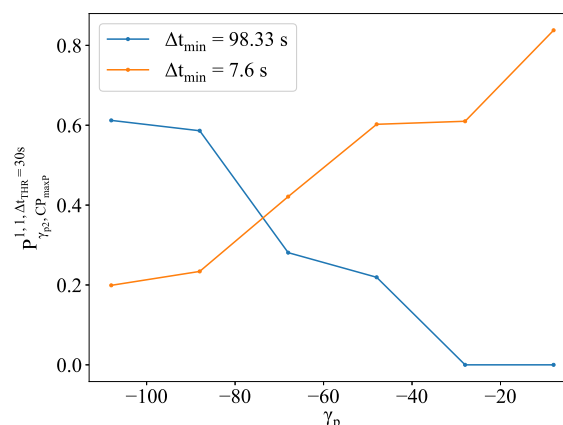


**Figure 13.** (A) STE variation by  $\Delta t_{THR}$  when  $\gamma_p = -108$  NM, and STE variation by  $\gamma_p$  when  $\Delta t_{THR} = 30$  s for MLP-HTPs. (B) Same graphs for XGB-HTPs.

In general, despite the fluctuations in the curves, the STE curves exhibit a clear trend across  $\gamma_p$ , where the proportion of nuisances and missed alerts decreases as  $\gamma_p$  increases, approaching the critical point. This phenomenon is due to the reduction in the size of the sample of potential conflicts; as  $\gamma_p$  increases, many flights have already flown over the critical point, and the potential interaction is cleared.

Regarding the HTP-induced variation, XGB-HTPs generally exhibit a higher proportion of FAs, whereas MLP-HTPs show a greater proportion of MAs. This phenomenon, taking into account the smaller standard deviation of XGB-HTPs' residuals, is caused by the use of the error function ( $\text{erf}\left(\frac{\Delta t_{THR} + \mu}{\sqrt{2}\sigma}\right)$  and  $\text{erf}\left(\frac{\Delta t_{THR} - \mu}{\sqrt{2}\sigma}\right)$ ) in the probability estimation, using Equation (2). The error function is a sigmoid function in nature, whose slope depends on  $\sigma$ . A lower  $\sigma$  in XGB-HTPs causes the function to have a steep slope, leading to a rapid transition to extreme values as the predicted time separation deviates from  $\Delta t_{THR}$ , resulting in a sharp probability transition and thereby an increase in notified positives and FAs. In contrast, MLP-HTPs, with higher  $\sigma$  values, induce a gradual slope in the error function, manifesting a slower probability transition over the predicted separation range and a tendency towards more MAs due to a less pronounced probability change.

The CD performance variation due to the larger disparity of the predictors' accuracy is assessed by comparing results at different evaluation points. As presented in Figure 13, the FA and MA proportions both decrease with an increase in  $\gamma_p$ . This reduction is caused by two factors. Firstly, a higher  $\gamma_p$  leads many flights to overfly the critical point, making the interaction impossible and consequently reducing the interaction sample size. Second, the uncertainty diminishes as  $\gamma_p$  increases, enhancing the accuracy of probability estimations and resulting in higher rates of True Positive (TP) and True Negative (TN). Figure 14 provides two flight pair examples, showcasing significant probability variation over  $\gamma_p$ , converting what initially appear as FAs and MAs into TNs and TAs, respectively. In the blue one, although it is a conflict-free flight pair, whose  $\Delta t_{min}$  lies above  $\Delta t_{THR} = 30$  s, the estimated probability at  $\gamma_p = -108$  NM is relatively high at 0.61, making it possible to be notified and then classified as an FA. However, as the uncertainty decreases, this probability decreases to zero when  $\gamma_p$  reaches  $-28$  NM, allowing the tool to accurately identify this case as conflict-free. The second flight pair presents the opposite case: a RP with an initial low estimated probability of 0.2 that escalates to 0.84 at  $\gamma_p = -8$  NM, close to the sector boundary of PAU.



**Figure 14.** Examples of flight pairs with probability variation by  $\gamma_p$ .

Concerning the variation by  $\Delta t_{THR}$  at  $\gamma_p = -108$  NM, the observed trend is inconsistent. The FA and MA proportions initially decrease with increasing  $\Delta t_{THR}$  up to 60 s, followed by a slight increase at 90 s and 120 s. The sample sizes and the error function again influence the results. RP and CD notifications are both dependent on  $\Delta t_{THR}$ ; the size of both samples increases with less strict  $\Delta t_{THR}$ . In terms of the error function, unlike the variation by  $\gamma_p$ , changing the  $\Delta t_{THR}$  does not affect the overall steepness of the curve; it changes

the point along the existing sigmoid curve. Given the same  $\sigma$  and predicted separation, the difference between  $\Delta t_{THR}$  and the predicted separation increases with  $\Delta t_{THR}$ , where the estimated point is likely to be located at the extreme steep regions. Table 6 shows an example, illustrating changes in analysis sample size with different  $\Delta t_{THR}$ .

**Table 6.** Number of flight pairs with  $\Delta t_{min} < \Delta t_{THR}$  at  $\gamma_p = -108$  NM and  $P_{THR} = 0.5$ .

| $\Delta t_{THR}$ | No. RP | No. FA | No. MA |
|------------------|--------|--------|--------|
| 10               | 35     | 0      | 35     |
| 30               | 92     | 26     | 21     |
| 60               | 173    | 12     | 15     |
| 90               | 226    | 21     | 23     |
| 120              | 325    | 18     | 28     |

These graphs could assist the user in selecting the most appropriate CD operating point for the operation, depending on LAT, by choosing suitable dynamic thresholds. First, this involves selecting the most appropriate  $\Delta t_{THR}$  that exhibits the lowest proportions of FAs and MAs.

Secondly, the selection of  $P_{THR}$  determines the notification of predicted alerts, representing the trade-off between FAs and MAs. The dynamic nature of  $P_{THR}$  is beneficial during the operation. A low  $P_{THR}$  implies a higher proportion of nuisances and fewer missed alerts, while a high  $P_{THR}$  implies fewer nuisances and more missed alerts. The trade-off is tailored to the specific roles and tasks of the executive and planner controller, as well as to the look-ahead time.

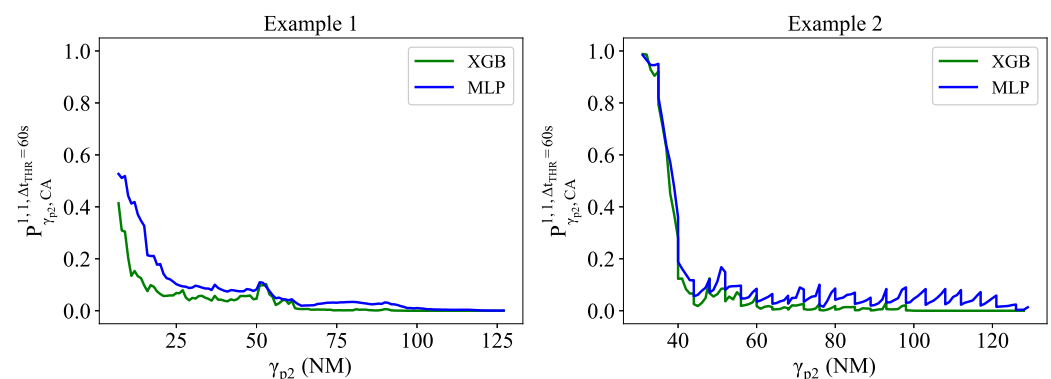
On one hand, a large LAT allows for the consideration of a low rate of MAs and a reduction in the number of missed alerts. This is considering the greater TP prediction error and more available time for the controller to further evaluate the situation at a large LAT. Consequently, this implies selecting a low  $P_{THR}$ .

On the other hand, as flights approach the critical point, a higher operating point that prioritises highly probable encounters becomes more appropriate. This is aimed at short LATs, to reduce false alerts by notifying interactions that require immediate actions and thus avoiding overwhelming the controllers.

Hence, the STE facilitates the selection of different operating points for the adaptive CD tool, depending on the corresponding LAT.

### CD Stability

The stability of the estimated probability cannot be directly discerned from the previous figures of merit; it requires the examination of specific instances. The analysis, based on the two flight pairs selected in Section 3.1, demonstrates that CD stability is influenced by the stability of the predictions, as depicted in Figure 15. Moderate fluctuations in the predictions result in smooth variation in the estimated probability with  $\gamma_p$ , while large irregularities result in high-frequency fluctuations.



**Figure 15.** Stability of estimated encounter probability—examples.



#### 4. Discussion

In this paper, the primary objectives included assessing the impact of trajectory predictor performance on the encounter probability generated by an adaptive conflict detection tool and examining the flexibility of the CD tool dependent on its adjustable thresholds. Two figures of merit were proposed to address these objectives, which also assist in selecting the most suitable trajectory predictors and dynamic CD thresholds. These selections ultimately determine the operating point necessary to meet operational requirements.

For the analysis conducted, trajectory predictors were used to compute the time separation at critical points where minimum separation might be infringed. The conflict detection tool then evaluated the difference between the predicted time and the actual time to compute the encounter probability.

Two simple machine-learning-based horizontal trajectory predictors were employed for each flow of analysis: one trained using extreme gradient boosting, and the other using a multi-layer perceptron regressor. By leveraging these predictors and multiple evaluation points with varying uncertainty, changes in the figures of merit due to the prediction error could be visualised.

Flexibility was attributed to the conflict detection tool through its three adjustable thresholds: the distance threshold, defining the critical point; the time difference threshold, influencing the resulting encounter probability; and the probability threshold, determining the notification of alerts. The latter two thresholds were examined with various values to illustrate their effects on alert notifications and the implications for different operating points.

To assess the impact, predictor performance was quantified using the mean and standard deviation of prediction errors, evaluated at each location and for every subsequent segment. Distinct regions were identified in the metrics' heatmaps, with locations of splits typically corresponding to ground speed variations.

Furthermore, the relationship between trajectory predictors and the CD tool was demonstrated using two figures of merit: system dynamic range and system tuning envelope. System dynamic range illustrates the variation of encounter probability with predicted time separation. It demonstrates that the rate at which the encounter probability decreases with increasing predicted separation is higher for high-performance predictors, and it declines with the deterioration of TP performance. The distribution of the system dynamic range follows a logistic function, and the analysis of its parameters aids in selecting appropriate horizontal trajectory predictors.

The remaining figure of merit, system tuning envelope, illustrates the trade-off between the proportion of nuisances and missed alerts. It demonstrated that high-performance predictors provide a smaller proportion of nuisances and missed alerts. However, its main utility lies in the selection of operating points, allowing for the fine-tuning of the system to better align with specific requirements for managing the airspace safely. The notifications are mainly impacted by the probability threshold; lower thresholds result in a larger proportion of nuisances and a lower proportion of missed alerts, while higher thresholds enhance the reliability of notified alerts at the expense of increased missed alerts. The trade-off decision rests with the controllers and depends on their specific tasks.

The selection of the time difference threshold depends on the performance of the trajectory predictors. It has been observed that when the ratio of the time separation threshold for conflict detection to the prediction errors exceeds  $\sqrt{2}$ , the behaviour of the conflict detector converges towards the ideal, following a stepped function.

In conclusion, this study demonstrated that the performance of the trajectory predictor and the conflict detector are intrinsically related, and the figures of merit are useful in selecting appropriate trajectory predictors and operating points for an adaptive conflict detection tool. Additionally, given a specified operating point, these figures of merit facilitate the specification of prediction performance requirements and conflict detection configurations to achieve the desired operational outcomes.

Building upon the findings of this study, the next stages can be divided into three primary blocks. The first block encompasses further research and development to enhance trajectory predictors, aligned with operational requirements that may be elicited from the application of this study. This part can involve extending investigations to vertical predictors, as well as assessing other impacts on the proposed figures of merit, such as flight Intent variations. To achieve this, it may entail the exploration of advanced machine learning algorithms and the integration of additional input sources, such as airborne information through extended projected profiles.

The second block focuses on the further development of the conflict detector, expanding the detection process to encompass not only interactions between flights within the same flow but also those between singular flights that do not belong to any standard flows, and between singular flights.

Following this, the third block will focus on integrating the figures of merit into the decision making process within the existing separation management workflow. The results of this study can contribute to the evolution of human–machine interactions of air traffic controllers, enabling the dynamic selection of operating points of conflict detectors, based on their operational needs. Additionally, incorporating uncertainty management and improving trajectory predictors could enhance the controller’s trust in these tools, thus underpinning the deployment of future concepts such as a reduction of the separation minima, which would bring significant performance benefits [33].

**Author Contributions:** Conceptualization, C.X. and C.E.V.G.; methodology, C.X. and C.E.V.G.; software, C.X.; validation, C.X. and C.E.V.G.; formal analysis, C.X.; investigation, C.X.; data curation, C.X.; writing—original draft preparation, C.X.; writing—review and editing, C.X.; visualization, C.X.; supervision, C.E.V.G., V.F.G.C., R.M.A.V. and A.F.B.; project administration, C.E.V.G.; funding acquisition, C.E.V.G. and A.F.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** This project received funding from the European Union’s H2020 research and innovation programme under Grant Agreement no. 872320. This publication reflects only the authors’ views and the SESAR 3 Joint Undertaking is not responsible for any use that may be made of the information it contains.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author, due to restrictions.

**Conflicts of Interest:** Adrián Fabio Bracero was employed by the company EUROCONTROL Innovation Hub. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

1. Performance Review Comission. *PRR 2021 Performance Review Report, An Assessment of Air Traffic Management in Europe*; Technical report; EUROCONTROL: Brussels, Belgium, 2022.
2. EUROCONTROL. *EUROCONTROL Forecast Update 2022–2028. European Flight Movements and Service Units—Three Scenarios for Recovery from COVID-19*; Technical report; EUROCONTROL: Brussels, Belgium, 2022.
3. Djokic, J.; Lorenz, B.; Fricke, H. Air traffic control complexity as workload driver. *Transp. Res. Part Emerg. Technol.* **2010**, *18*, 930–936. [\[CrossRef\]](#)
4. Histon, J.; Hansman, R.J. Air Traffic Controller Operating Modes and Cognitive Complexity Regulation. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **2011**, *55*, 345–349. [\[CrossRef\]](#)
5. Radanovic, M.; Piera Eroles, M.A.; Koca, T.; Ramos Gonzalez, J.J. Surrounding traffic complexity analysis for efficient and stable conflict resolution. *Transp. Res. Part Emerg. Technol.* **2018**, *95*, 105–124. [\[CrossRef\]](#)
6. PJ18-W2 4D Skyways Grant N° 872320. Available online: <https://cordis.europa.eu/project/id/872320> (accessed on 15 August 2023).
7. EUROCONTROL. *Conflict Detection Tools Impact on Controller Taskload—Fast Time Study*; Technical report; EUROCONTROL: Brétigny-sur-Orge, France, 2010.

8. Schuster, W.; Ochieng, W. 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.* **2014**, *35*, 92–101. [CrossRef]
9. Common Trajectory Predictor Structure and Terminology in Support of Sesar and NextGen. Available online: [https://www.academia.edu/9410077/Common\\_Trajectory\\_Predictor\\_Structure\\_and\\_Terminology\\_in\\_support\\_of\\_Sesar\\_and\\_NextGen](https://www.academia.edu/9410077/Common_Trajectory_Predictor_Structure_and_Terminology_in_support_of_Sesar_and_NextGen) (accessed on 20 November 2023).
10. Zeng, W.; Chu, X.; Xu, Z.; Liu, Y.; Quan, Z. Aircraft 4D Trajectory Prediction in Civil Aviation: A Review. *Aerospace* **2022**, *9*, 91. [CrossRef]
11. Musialek, B.; Munafo, C.F.; Ryan, H.; Paglione, M. *Literature Review of Trajectory Predictor Technology*; DOT/FAA/TC-TN11/1; U.S. Department of Transportation Federal Aviation Administration: Washington, DC, USA, 2010.
12. Poles, D.; Nuic, A.; Mouillet, V. Advanced aircraft performance modeling for ATM: Analysis of BADA model capabilities. In Proceedings of the 29th Digital Avionics Systems Conference, Salt Lake City, UT, USA, 3–7 October 2010; pp. 1.D.1-1–1.D.1-14. [CrossRef]
13. Bronsvort, J.; McDonald, G.; Torres, S.; Paglione, M.; Young, C.; Hochwarth, J.; Boucquey, J.; Vilaplana, M. Use of the Extended Projected Profile (EPP) in Trajectory Management. In Proceedings of the 16th AIAA Aviation Technology, Integration, and Operations Conference, Washington, DC, USA, 13–17 June 2016.
14. Wang, Z.; Liang, M.; Delahaye, D. Short-term 4d trajectory prediction using machine learning methods. In Proceedings of the SID 2017, 7th SESAR Innovation Days, Belgrade, Serbia, 28–30 November 2017.
15. Shi, Z.; Xu, M.; Pan, Q. 4-D Flight Trajectory Prediction With Constrained LSTM Network. *IEEE Trans. Intell. Transp. Syst.* **2021**, *22*, 7242–7255. [CrossRef]
16. Alligier, R.; Gianazza, D. Learning aircraft operational factors to improve aircraft climb prediction: A large scale multi-airport study. *Transp. Res. Part Emerg. Technol.* **2018**, *96*, 72–95. [CrossRef]
17. Rudnyk, J.; Ellerbroek, J.; Hoekstra, J. Trajectory Prediction Sensitivity Analysis Using Monte Carlo Simulations. In Proceedings of the 2018 Aviation Technology, Integration, and Operations Conference, Atlanta, GA, USA, 25–29 June 2018. [CrossRef]
18. Rudnyk, J.; Ellerbroek, J.; Hoekstra, J. Trajectory Prediction Sensitivity Analysis Using Monte Carlo Simulations Based on Inputs' Distributions. *J. Air Transp.* **2019**, *27*, 1–18. [CrossRef]
19. Mondoloni, S.; Rozen, N. Aircraft trajectory prediction and synchronization for air traffic management applications. *Prog. Aerosp. Sci.* **2020**, *119*, 100640. [CrossRef]
20. Graas, R.; Sun, J.; Hoekstra, J. Quantifying accuracy and uncertainty in data-driven flight trajectory predictions with gaussian process regression. In Proceedings of the 11th SESAR Innovation Days, Virtual Event, 7–9 December 2021.
21. Dalmau, R.; Melgosa, M.; Vilardaga, S.; Prats, X. A Fast and Flexible Aircraft Trajectory Predictor and Optimiser for ATM Research Applications. In Proceedings of the ICRAT 2018—8th International Conference for Research in Air Transportation, Barcelona, Spain, 26–29 June 2018; pp. 1–8.
22. Ehrmanntraut, R. Full Automation of Air Traffic Management in High Complexity Airspace. Ph.D. Thesis, Technischen Universität Dresden, Dresden, Germany, 2010.
23. Verdonk Gallego, C.E.; Gómez Comendador, V.F.; Amaro Carmona, M.A.; Arnaldo Valdés, R.M.; Sáez Nieto, F.J.; García Martínez, M. A machine learning approach to air traffic interdependency modelling and its application to trajectory prediction. *Transp. Res. Part Emerg. Technol.* **2019**, *107*, 356–386. [CrossRef]
24. Irvine, R. A Geometrical Approach to Conflict Probability Estimation. *Air Traffic Control. Q.* **2002**, *10*, 85–113. [CrossRef]
25. EUROCONTROL. *EUROCONTROL Specification for Trajectory Prediction*; Technical report; EUROCONTROL: Brussels, Belgium, 2017.
26. Mondoloni, S.; Swierstra, S.; Paglione, M. Assessing trajectory prediction performance—metrics definition. In Proceedings of the 24th Digital Avionics Systems Conference, Washington, DC, USA, 30 October–3 November 2005; Volume 1. [CrossRef]
27. Vivona, R.; Paglione, M.; Hughes, W.; Cate, K.; Enea, G. Definition and Demonstration of a Methodology for Validating Aircraft Trajectory Predictors. In Proceedings of the AIAA Guidance, Navigation, and Control Conference, Toronto, ON, Canada, 2–5 August 2010. [CrossRef]
28. Mondoloni, S. Trajectory-based operations—Robust planning under trajectory uncertainty. In Proceedings of the 2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC), Sacramento, CA, USA, 25–29 September 2016; pp. 1–10. [CrossRef]
29. Paglione, M.; Young, C.M.; Torres, S.; Hochwarth, J.K.; McDonald, G.; Bronsvort, J.; Boucquey, J. Operational impact of trajectory prediction accuracy on air traffic automation tools. In Proceedings of the 2017 IEEE/AIAA 36th Digital Avionics Systems Conference (DASC), St. Petersburg, FL, USA, 17–21 September 2017; pp. 1–10. [CrossRef]
30. Paglione, M.; Oaks, R.; Bilimoria, K. Methodology for Generating Conflict Scenarios by Time Shifting Recorded Traffic Data. In Proceedings of the AIAA's 3rd Annual Aviation Technology, Integration, and Operations (ATIO) Forum, Denver, CO, USA, 17–19 November 2003. [CrossRef]
31. Verdonk Gallego, C.E.; Gómez Comendador, V.F.; Sáez Nieto, F.J.; García Martínez, M. Discussion On Density-Based Clustering Methods Applied for Automated Identification of Airspace Flows. In Proceedings of the 2018 IEEE/AIAA 37th Digital Avionics Systems Conference (DASC), London, UK, 23–27 September 2018; pp. 1–10. [CrossRef]

32. *Comisión de Investigación de Accidentes e Incidentes de Aviación Civil*; Report IN-039/2018; Technical report; Ministerio de Transportes, Movilidad y Agenda Urbana, Gobierno de España: Madrid, Spain, 2018.
33. Li, T.; Wan, Y. A fuel savings and benefit analysis of reducing separation standards in the oceanic airspace managed by the New York Air Route Traffic Control Center. *Transp. Res. Part Logist. Transp. Rev.* **2021**, *152*, 102407. [[CrossRef](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.