Fifteenth USA/Europe Air Traffic Management Research and Development Seminar (ATM2023)

Data-Driven Airborne Collision Risk Modelling using a Probability Density Function

Benoit Figuet, Raphael Monstein
Zurich University of Applied Sciences (ZHAW)
Centre for Aviation
Winterthur, Switzerland
and
SkAI Data Services
Zürich, Switzerland

Steven Barry
Airservices Australia
Canberra, Australia

***Abstract*—This paper introduces a novel data-driven mid-air collision risk model for an aircraft flying through a flow of aircraft, modelled using a probability density function to describe position, and a speed vector. The proposed model is, compared to traditional Monte-Carlo simulations, computationally efficient and, thus, facilitates exploration of risks as a function of key parameters, such as aircraft performance, or with different scenarios. Compared with traditional collision risk models, the proposed solution can handle more complex trajectories and traffic flows. The usefulness of the novel model is illustrated on a real-world example by applying it to the terminal airspace of Zurich airport, Switzerland. Specifically, the probability of collisions between go-arounds on Runway 14 and departures on Runway 16 is quantified. The results of the model were validated through comparison with Monte-Carlo simulations, with comparable outcomes but significantly lower computational costs.**

***Keywords*—collision risk modelling, data-driven, probability density function, kernel density estimation, safety**

I. Introduction

When designing new air traffic procedures and regimes, or assessing existing procedures, one key factors is the estimation of the mid-air collision risk (CR). This should be below accepted Target Levels of Safety (TLS), below the risk for equivalent accepted systems, or As Low as Reasonably Practical (ALARP).

The TLS for CR in en-route airspace is usually 1.5×10^{-8} fatal accidents per flight hour, with one collision being two fatal accidents. This is sub-divided into a risk of 5×10^{-9} per physical dimension (lateral, longitudinal and vertical), when separation standards are defined by ICAO. The TLS in the terminal area is less defined, with 'per flight hour' not always appropriate, and replaced by 'per movement' (arrival or departure). The overall aircraft risk has historically been divided into the three similarly weighted phases of take-off, en-route and landing; with equipment failure and collision with terrain / obstacles dominating the risk. The total aircraft risk is of the order of 10^{-7} per flight, with the mid-air collision TLS designed to be in the order of 1/10 of this value. The TLS for terminal-area (TA) collisions is not formalised, with values of orders 10^{-8} to 5×10^{-10} per movement being used, depending on context. Since the terminal-area risk is not neatly divided into lateral, longitudinal or vertical risk, a convenient measure is then 1.5×10^{-8} fatal accidents per movement, being similar to the en-route TLS. [1], [2] give good reviews of TLS in aviation.

This research was partially funded by the Swiss Federal Office of Civil Aviation grant number SFLV 2018-037.

Aviation collision risk can be estimated using many different models, often in combination: assuming traffic distributions along predefined routes based on required or observed navigational performance [3]; considering distributions for normal (core) and ab-normal (tail) operations, particularly used to define separation standards for parallel tracks [4]; considering operational errors such as large deviations at given angles [5]; doing pure Monte-Carlo simulations of traffic [6]–[8]; and, considering traffic densities in small defined grids. Different versions of these models have historically been used by ICAO to define new separation standards (usually assuming defined route structures), and by approved airspace authorities (Regulators or Air Navigational Service Providers - ANSP).

This paper introduces a novel, data-driven model to estimate the collision risk in complex airspace. The model calculates the probability of collision for an aircraft flying through a flow of aircraft. The flow of aircraft is modelled using a probability density function (PDF) derived from historical trajectories, and a speed vector. The aircraft flying through the flow is represented as a 4-dimensional trajectory, with the usual 3-dimensional (3-D) position and time. By using historical data to estimate the probability density function of the positions of the flow, the proposed model eliminates the need for unreliable assumptions about traffic behaviour, while preserving the stochastic nature of the process that generates the PDF. This method is also computationally efficient, reducing the processing time by orders of magnitude compared to traditional Monte-Carlo simulations. Furthermore, the model is capable of handling more complex aircraft trajectories, such as turns, which are not addressed by conventional models such as Anderson-Hsu [9] or by the Reich model [3], [10] .

The model is illustrated with a case study from Zurich

airport (LSZH), Switzerland. Specifically, the collision risk model (CRM) is used to quantify the risk between go-arounds (GAs) on Runway 14 and take-offs (TOs) on Runway 16 if they were operated independently. Fig. 1 illustrates the the situation in Zurich laterally, and Fig. 2 shows the altitude distribution of the aircraft when they cross the tracks. Currently, the runways are not operated independently, avoiding simultaneous operation to mitigate the risk. However, the novel method demonstrated in this paper may be of use in defining the acceptable thresholds for separating an approach to Runway 14 and a departure on Runway 16.

The paper is organised as follows: In Section II key concepts related to collision risk modelling such as Monte Carlo simulations, the *crossing track* problem, and position PDF estimation using Kernel Density Estimation are introduced. Section III provides a mathematical description of the developed model, its validation and its limitations. The proposed model's application to estimate collision probabilities between GAs and TOs at Zurich airport is the focus of Section IV. Finally, Section V concludes the paper and highlights potential areas for future research.

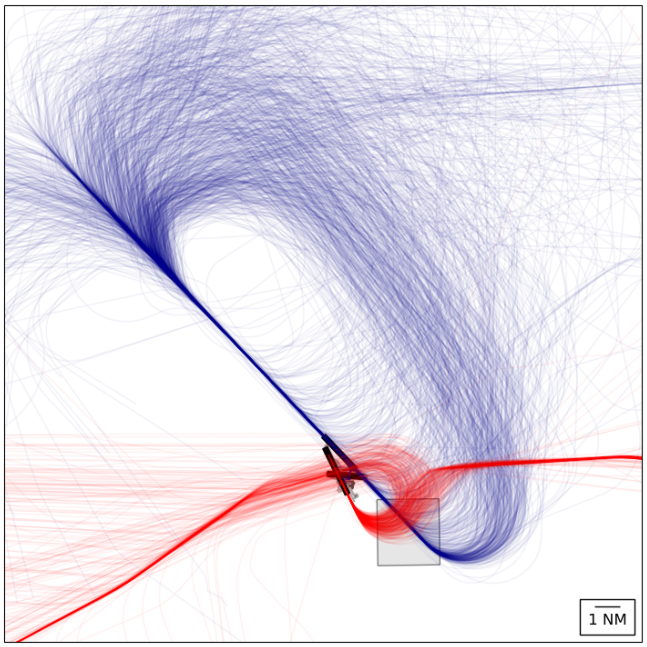

Figure 1: Overview of the modelled scenario at the Zurich airport. Blue lines: 620 go-arounds on Runway 14. Red lines: 500 randomly sampled departures on Runway 16. The grey area indicates one area of potential conflict and a one nautical mile reference line is shown.

II. Theoretical Background

A. Collision risk modelling with Monte-Carlo simulations

Monte-Carlo simulations are useful to, among other things, estimate the probability of an event happening when the system under study cannot be modelled analytically, such as the complex terminal-airspace. In the past, collision risk modelling in TAs were often done using Monte-Carlo simulated trajectories e.g., in [6]–[8].

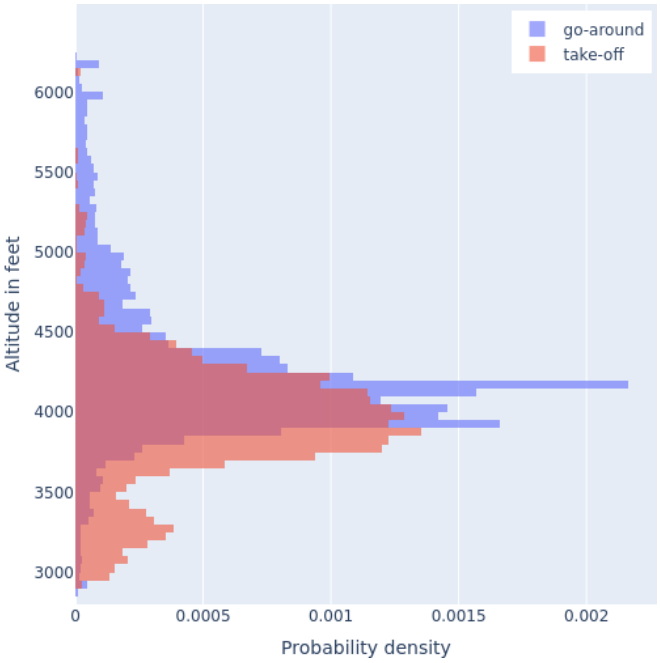

Figure 2: Histogram of the altitude when crossing tracks.

Monte-Carlo simulations can be used to estimate the expectation of a probability density by leveraging the empirical mean of a random variable sampled multiple times. For instance, in the context of assessing the collision risk between two aircraft, the event of a collision can be described by a random variable following a Bernoulli distribution. It is then possible to randomly generated n pairs of these two aircraft trajectories and estimating the collision probability by counting the number of observed collision such that the probability of a collision, $\mathrm{P_{col}}$, is

$$\mathrm{P_{col}} = \frac{n_{\text{observed collisions}}}{n}. \tag{1}$$

If historical aircraft trajectories are available, Monte-Carlo simulations can be performed by sampling from those trajectories. In cases where such data is not available, simulated trajectories can be used. However, the accuracy of the collision risk estimation is directly impacted by the quality of the simulated trajectories. Particularly in terminal areas, factors such as aircraft type, weather conditions, and air traffic control and pilot behaviour can lead to significant variations in these trajectories and, thus, the results of the simulations. Additionally, the slow convergence of Monte-Carlo estimators requires large numbers of simulation runs to accurately estimate small probabilities, such as in the case of estimating the probability of a mid-air collision. Using the central limit theorem, it is possible to define confidence interval on the probability estimation. For example, for estimating a probability of 10^{-9} with an absolute error at 95% lower than 10^{-10} would require about 4×10^{11} runs. Variance reduction techniques, such as importance splitting, can help to overcome this issue [11].

B. Crossing track collision probability

The *crossing track* collision risk model computes the probability of collision in a 2-D space for an aircraft crossing

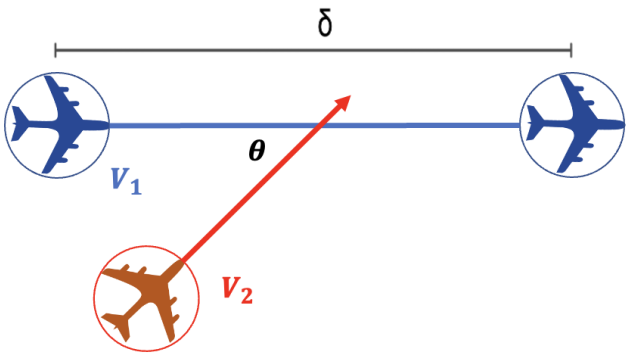

Figure 3: A route where aircraft$_1$ are flying at a speed V_1 and separated by a distance δ. aircraft$_2$ flying at a speed V_2 crosses the route with an angle θ.

a flow of aircraft (such as a route or standard procedure). This is a classic situation of interest for both procedure designers, as well as those monitoring an airspace. For example in [5] and [12], this model is used on deviating aircraft from a track, inadvertently crossing an adjacent route: here the question is what separation distance meets the TLS given a known deviation rate (an operational error or blunder) and density of traffic. Similarly, the monitoring of Reduced Vertical Separation Minima (RVSM) airspace risks by Regional Monitoring Agencies requires the estimation of risk for an operational error of an aircraft climbing through another flight level [13]. Airspace design often consists of two streams of aircraft crossing each other, either separated vertically for procedural separation, or at similar altitudes where it is assumed the air traffic controller will manage conflicts. In both cases, the modelling considers the rate of typical errors (aircraft at the wrong altitude, pilot/controller errors in managing conflicts) and then the risk associated with the aircraft crossing through the other route.

Fig. 3 shows a stream of aircraft$_1$ of speed $\vec{V_1}$, all separated by a distance δ and aircraft$_2$ crossing the flow at a speed $\vec{V_2}$ with an intersection angle θ. The aircraft are assumed to have a circular shape with a diameter λ.

This type of *crossing model* has been used within ICAO [5], [12] to define new oceanic separation standards for parallel tracks with space-based ADS-B surveillance, but satellite-based communications. Unlike traditional ICAO models for parallel tracks [3], [4] which are based on assumed core and tail lateral distributions, this model include a rate of deviating aircraft, and then the risk associated with this aircraft crossing the track of the adjacent route. This conservatively modelled the aircraft as squares, with Barry 2022 [14] describing the model for circular aircraft. This type of model, although fundamentally equivalent to Reich-style models [3], is more intuitive.

The novel model proposed in this paper can be seen as a generalization of the *crossing track* model. Thus, the derivation of the *crossing track* model is presented in this section since its concepts are used for the development of the proposed model. Furthermore, this model will also be utilized later on as a mean to validate the new method.

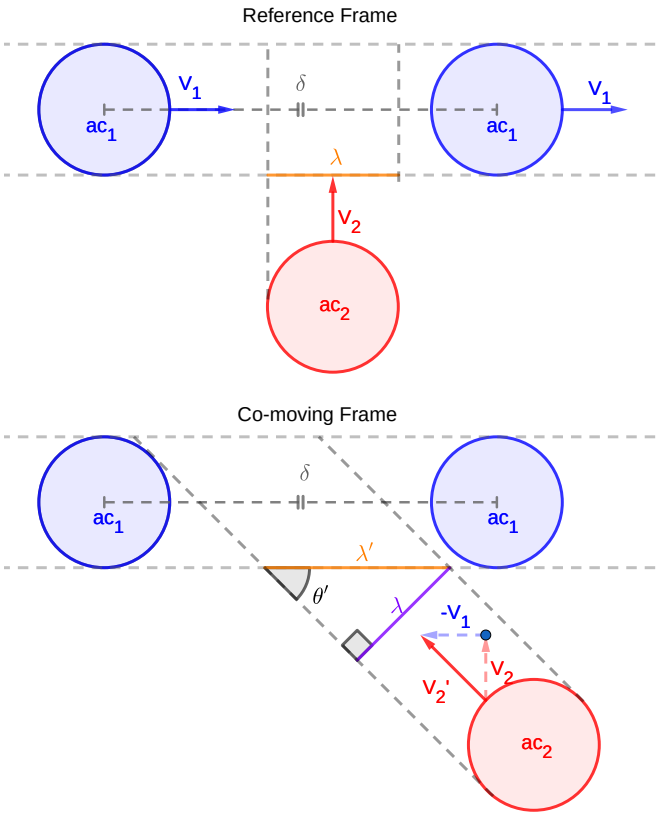

Figure 4: Aircraft (in red) crossing a stream (in blue) in the reference frame (top) and in the co-moving frame (bottom).

To simplify the analysis of the situation shown in Fig. 3, a co-moving frame fixed to aircraft$_1$ is introduced. aircraft$_1$ will then have zero speed, as shown in Fig. 4. Hence, the respective aircraft speeds in the co-moving frame, can expressed as

$$V_1' = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, V_2' = \begin{bmatrix} ||\vec{V_2}|| \cos\theta - ||\vec{V_1}|| \\ ||\vec{V_2}|| \sin\theta \end{bmatrix} \equiv \begin{bmatrix} V_2 \cos\theta - V_1 \\ V_2 \sin\theta \end{bmatrix}. \tag{2}$$

The cross-section is the distance between two aircraft$_2$, where one would scrape the left side of aircraft$_1$ and the other aircraft$_2$ would scrape the right side of this same aircraft$_1$, as depicted in Fig. 5.

Since aircraft$_1$ are separated by a distance δ, the collision probability is expressed by the ratio $\frac{2\lambda'}{\delta}$ where λ' is the aircraft$_2$ cross-section in the co-moving frame as shown in Fig. 4 and Fig. 5. The probability of a collision, $\mathrm{P_{col}}$,extended to

$$\mathrm{P_{col}} = \frac{2\lambda'}{\delta} = \frac{2\lambda}{\delta \sin\theta'} \tag{3}$$

with,

$$\begin{aligned} \sin\theta' &= \frac{V_2 \sin\theta}{\sqrt{(V_2\cos\theta - V_1)^2 + (V_2 \sin\theta)^2}} \\ &= \frac{1}{\sqrt{1 + \left(\frac{V_2\cos\theta - V_1}{V_2 \sin\theta}\right)^2}}. \end{aligned} \tag{4}$$

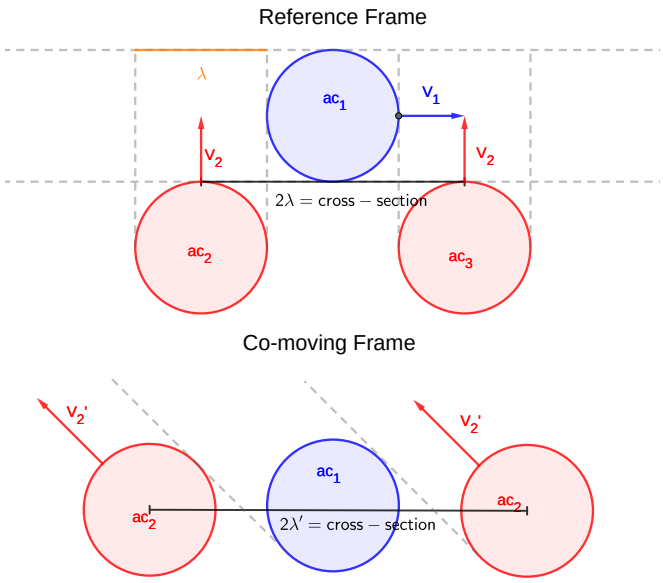

Figure 5: Cross sections in the reference frame (top) and in the co-moving frame (bottom).

Combining (3) and (4) results in

$$\mathrm{P_{col}} = \frac{2\lambda}{\delta}\sqrt{1+\left(\frac{V_2\cos\theta - V_1}{V_2\sin\theta}\right)^2}. \tag{5}$$

Note that $\mathrm{P_{col}}$ can, in some specific circumstances, be larger than one. This is due to the fact that it allows for an aircraft crossing the stream to have multiple collisions e.g., if an aircraft is very slow, it could be hit by multiple aircraft in the stream. This is a rather theoretical concern, though, and can be mitigated by limiting the probability such that $\{\mathrm{P_{col}}|0 \leq \mathrm{P_{col}} \leq 1\}$.

It is then possible to extend (5) to account for different speeds, spacings and crossing angles using probability densities:

$$\mathrm{P_{col}} = \iiiint \frac{2\lambda}{\delta}\sqrt{1+\left(\frac{V_2\cos\theta - V_1}{V_2\sin\theta}\right)^2} \\ f(\delta)f(V_1)f(V_2)f(\theta)\,d\delta\,dV_1\,dV_2\,d\theta \tag{6}$$

where $f(\delta)$, $f(V_1)$, $f(V_2)$ and $f(\theta)$ are respectively the probability density functions describing aircraft$_1$ spacings on the route, aircraft$_1$ and aircraft$_2$ speeds, and the crossing angle.

C. Kernel density estimation

The PDFs in (6) are usually estimated from data. In some cases, this can be done using classical, parameteric PDFs, but in many practical cases parameteric PDFs are insufficient. In these cases, a kernel density estimation (KDE) can be used. KDE is a non-parametric method to estimate the probability density function based on a finite sample. The method is well established and was created by Rosenblatt [15] and Parzen [16]. KDE became widely popular with the advent of machine learning and many implementations in different programming languages exist, both for univariate and multivariate data.

In the scope of this study, a KDE is fitted to the 3-D positions of the trajectories of one of the two traffic flows. The probability density function obtained by the KDE can be interpreted as the function $\rho\colon \mathbb{R}^3 \to [0,1]$, which describes the probability of being at a given position.

The KDE has one free parameter, the bandwidth h, and the selection of a suitable value is essential, as illustrated in Fig. 6. In the figure, three KDEs with different values for h

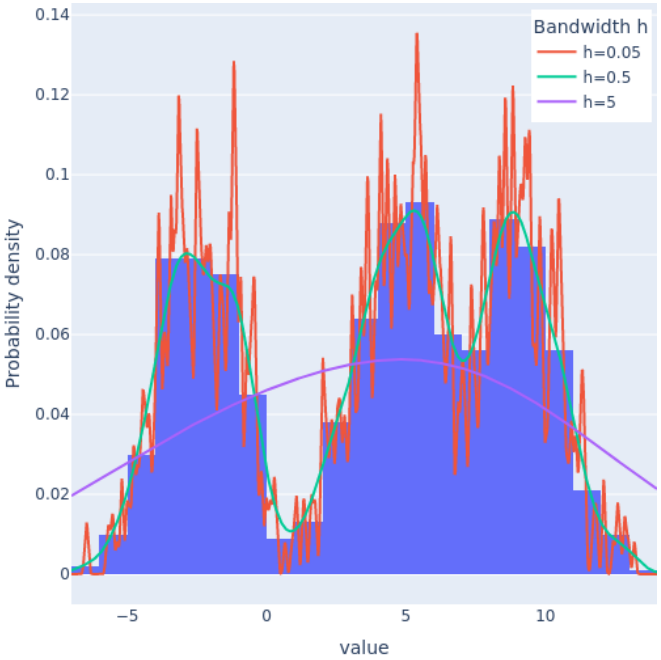

Figure 6: The histogram (blue bars) and fitted KDE with different bandwidth h (colored lines) on simulated univariate data illustrate the importance of selecting a suitable bandwidth.

are fitted to simulated univariate data. The KDE with $h = 5$ (violet line) is oversmoothing the underlying distribution, $h = 0.05$ (red line) is overfitting, and $h = 0.5$ seems visually appropriate. But this visual try-and-error approach to estimate a suitable bandwidth is subjective and is not practical for multivariate (i.e. higher dimensional) data. To avoid this issue, multiple methods have been proposed to estimate a value for h. Using a cross-validation method is popular and results in robust estimates, but in the case of relatively large datasets it becomes computationally expensive and infeasible. The consensus in the literature is that *plug-in selectors* are valid alternatives to cross-validation, while being significantly less computationally expensive (see e.g. [17]). The selector used in this project is the multivariate plug-in selector of Wand and Jones [18].

In contrast with the univariate example in Fig. 6, the bandwidth of a multivariate KDE is a matrix, not a scalar. In this project, the matrix is assumed to be a diagonal matrix with one parameter for each of the dimensions and zeros for the off-diagonal elements: the three physical dimensions (x, y and z) are independent and with different scales. For example,

it is expected that the horizontal dispersion would be different than the vertical (altitude).

The PDF obtained by fitting a KDE to the position is subsequently used in estimating the probability of an overlap in the position, as described in the following sections.

III. Proposed Method: Probability of Collision using Line Integration on a Density

The proposed method extends the prior model presented in Sec. II-B by computing the collision probability for an aircraft flying through a flow of aircraft and offers several key advantages over the *crossing track* model. First, it allows for the consideration of aircraft crossing the flow with complex, non-linear trajectories, rather than just straight paths. This improved accuracy provides a better understanding of the risk in real-world scenarios where aircraft may not follow a single 1-D path. Second, the model addresses the challenge of insufficient data through the use of KDE smoothing as introduced in Sec. II-C, compensating for missing or sparse data.

A. Mathematical formulation

The basic idea of the proposed method is to integrate the fitted probability density function along the path of aircraft$_2$. The formulation of the collision risk model is presented in the 2-D horizontal plane, with the capability to be extended to three dimensions without loss of generality. Therefore, aircraft are represented as circular shapes with diameter λ.

The flow of aircraft$_1$ is not static (see Fig. 3) and it is, thus, not possible to directly integrate the PDF along the path s of aircraft$_2$ without considering the direction and the speed of the flow of aircraft$_1$. Nevertheless, it is possible to use the co-moving frame introduced in Sec. II-B and Fig. 4. Since aircraft$_1$ are static in the co-moving frame, the probability of collision is obtained by integrating along the path of aircraft$_2$ in the co-moving frame s',

$$\mathrm{P_{col}} = \int_s \text{cross-section}(s')\rho(s')\,ds' \tag{7}$$

Unfortunately, integrating in the co-moving frame is not as straight forward as integrating in the reference frame. Therefore, next steps show how to compute $\mathrm{P_{col}}$ by integrating in the frame of reference using a speed factor.

For an interval of time Δ_t, aircraft$_2$ travels a distance $||\vec{V_2}'||\Delta_t$ in the co-moving frame, while it travels a distance $||\vec{V_2}||\Delta_t$ in the reference frame. We define the speed factor k_V as the ratio of the two speeds:

$$k_V = \frac{||\vec{V_2}'||}{||\vec{V_2}||} = \frac{\sqrt{(V_2\cos\theta\ - V_1)^2 + (V_2\sin\theta\ )^2}}{V_2} \tag{8}$$

This factor can be used to compute the collision probability by integrating along aircraft$_2$ path in the reference frame:

$$\begin{aligned}
\mathrm{P}_{col} &= \int_s \text{cross-section}(s)\rho(s)k_V(s)\,ds \\
&= \int_s 2\lambda\rho(s)\frac{\sqrt{(V_2(s)\cos\theta(s)\ - V_1)^2 + (V_2(s)\sin\theta(s)\ )^2}}{V_2(s)}\,ds \\
&= 2\lambda \int_0^{+\infty} \frac{\sqrt{(V_2(t)\cos\theta(t)\ - V_1)^2 + (V_2(t)\sin\theta(t)\ )^2}}{V_2(t)} \\
&\quad \rho\left(x_2(t), y_2(t)\right)\ \sqrt{(V_2(t)\cos\theta(t))^2 + (V_2(t)\sin\theta(t))^2}dt \\
&= 2\lambda \int_0^{+\infty} \sqrt{(V_2(t)\cos\theta(t)\ - V_1)^2 + (V_2(t)\sin\theta(t)\ )^2} \\
&\quad \rho\left(x_2(t), y_2(t)\right)\ dt
\end{aligned} \tag{9}$$

where $(x_2(t), y_2(t))$ represent the 2-D coordinates of aircraft$_2$ at time t, and $\rho(x, y)$ is the probability of an aircraft$_1$ being at the coordinates (x, y). Note that this model assumes that the PDF ρ is locally flat i.e., constant within the given cross-section of the aircraft (2λ). In real-world applications this is not a problem, since the risk will not change dramatically within, say, 100 meter.

B. Model Validation

The novel model in (9) can be partially validated by applying it to a case where an aircraft is crossing a flow with constant density, as shown in Fig. 7.

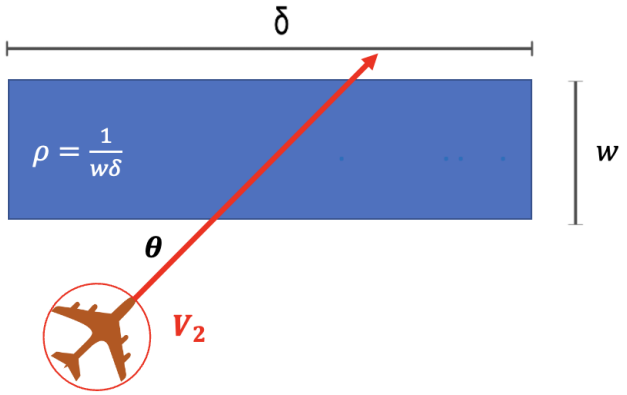

Figure 7: An aircraft (in red) is crossing at a speed V_2 and an angle θ through a flow of aircraft going at a speed V_1.

This scenario can be described defining the density $\rho = \frac{1}{w\delta}$ with w being the track width and delta the longitudinal spacing between aircraft. Since the different speeds and the

crossing angle are constant, (9) can be expressed as:

$$\begin{aligned} \mathrm{P_{col}} &= \frac{2\lambda}{w\delta} \frac{\sqrt{(V_2\cos\theta\ -V_1)^2 + (V_2\sin\theta\ )^2}}{V_2} \int_s ds \\ &= \frac{2\lambda}{w\delta} \frac{\sqrt{(V_2\cos\theta\ -V_1)^2 + (V_2\sin\theta\ )^2}}{V_2} \frac{w}{\sin\theta} \\ &= \frac{2\lambda}{\delta} \sqrt{1 + \left(\frac{V_2\cos\theta - V_1}{V_2\sin\theta}\right)^2} \end{aligned} \tag{10}$$

Equation (10) is equal to (5) and, therefore, the novel model is shown to be equivalent in the special case of the *crossing track* model with constant PDF.

C. Model limitations

The proposed collision risk model does not account for variations in speed and heading within the aircraft flow; it assumes that all the aircraft in the flow are flying with the same speed vector. The effect of this assumption on the risk can be accounted for by an appropriate sensitivity analysis (i.e. analyse of the varying results with a range of speeds and headings), which is possible given the computational efficiency of the method.

IV. Case Study

The scenario presented in the following case study was mainly chosen because it is relatively simple and can be validated with Monte-Carlo simulations with a reasonable effort. Of course, the proposed model can be applied to many different scenarios. This includes the estimation of collision risk between manned aircraft and Unmanned Aerial Vehicles (UAV) or between UAVs and UAVs.

A. Scenario Description

At Zurich airport, the intersecting flight paths of take-offs from Runway 16 and go-arounds from Runway 14 may pose a risk of mid-air collision, shown in Fig. 1 and 2. This risk depends on the aircraft departing and going around simultaneously (within a relatively short time window). To mitigate the risk, Runways 14 and 16 cannot operate independently; with departures held until an approach has landed. This negatively impacts the capacity of the airports. In this study, the collision risk is evaluated assuming that the two runways operate independently, and that there is always an aircraft approaching Runway 14 (uniformally positioned between 0 and 5 nautical miles from the threshold) when a take-off on Runway 16 is initiated, as illustrated in Fig. 8.

B. Dataset

This case study is based on historical Automatic Dependent Surveillance-Broadcast (ADS-B) data from the OpenSky Network [19]. From the historical data, 6300 departures were identified on Runway 16 from January to November 2022. After removing observations that were incomplete or noisy, about 5100 trajectories were eventually used. Since the number of observed go-arounds is significantly smaller, a longer period is needed to obtain sufficient observations. In the period from 2017 to 2019, 620 go-arounds were detected

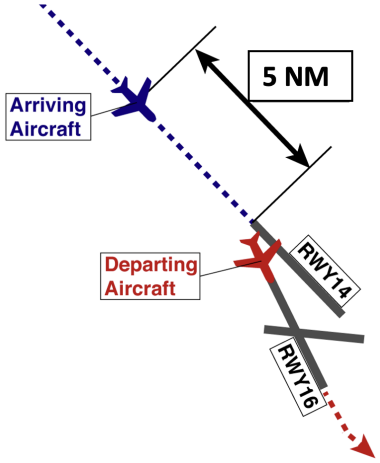

Figure 8: Simultaneous operation concept: a landing is approaching (between 0 and 5 nautical miles from the threshold) when a take-off on Runway 16 is initiated.

on Runway 14. Taking the conservative assumption that all go-arounds fly through the volume of potential collision (i.e., removing trajectories that turn left early), 237 go-arounds remained for modelling.

C. Modelling

In this study, the collision risk is modelled with the Bernoulli random variable C, where $\mathrm{P}(C=1)$ (denoted as $\mathrm{P}(C)$ for simplicity) is the probability of a collision occurring and $\mathrm{P}(C=0)$ is the probability that no collision occurs.

The novel method is illustrated on one of the potential conflict zones highlighted with the grey box in Fig. 1. For a collision to occur in this scenario, a Runway 14 GA and a Runway 16 TO have to be in the highlighted volume V at the same time. The probability of a collision, per Runway 16 TO, is calculated for simultaneous operation of approaches to Runway 14 and Runway 16 take-offs. This definition of *simultaneous operation* is conservative and means that for every aircraft departing on Runway 16, an aircraft is on the approach within five nautical miles from the threshold. The probability of a collision, given a simultaneous GA, can be expressed as

$$\begin{aligned} \mathrm{P}(C|\mathrm{GA}) =& \mathrm{P}\left(C|(\mathrm{GA\ in\ V}|\mathrm{TO\ in\ V})\right) \\ &\cdot \mathrm{P}(\mathrm{GA\ in\ V}|\mathrm{TO\ in\ V}). \end{aligned} \tag{11}$$

The terms in this equation are:

- $\mathrm{P}(C|\mathrm{GA})$: Probability of a take-off to have a collision, given a simultaneous GA on Runways 14.
- $\mathrm{P}\left(C|(\mathrm{GA\ in\ V}|\mathrm{TO\ in\ V})\right)$: Probability of a take-off to have a collision, given that a GA is in the volume V whenever the TO is in the volume V.
- $\mathrm{P}(\mathrm{GA\ in\ V}|\mathrm{TO\ in\ V})$: Probability of a GA being in the volume V, given that a take-off is in the volume V.

The probability of a collision of a take-off $\mathrm{P}(C)$, assuming (i.e., given) simultaneous operation of Runways 14 and 16, can calculated:

$$\mathrm{P}(C) = \mathrm{P}(C|\mathrm{GA}) \cdot \mathrm{P}(\mathrm{GA}). \tag{12}$$

with P(GA) being the probability of a landing to perform a go-around.

The probability of a take-off to have a collision, P(C), can also be interpreted as the probability of a collision *per* take-off. Knowing that P(C) is the probability of a Bernoulli random variable, the expected value is E[C] = P(C). With this in mind, the collision risk R, namely the expected number of fatal collisions per take-off, can be expressed as

$$\mathrm{R} = 2 \cdot \mathrm{P}(C). \tag{13}$$

The factor of 2 is due to the fact that a collision event involves two aircraft and, thus, counts twice. The risk R, and the probability of a collision P(C), are in this scenario the same for a landing and a departing aircraft, since for each aircraft landing there is one aircraft departing. In scenarios where the traffic is not balanced, R and P(C) can analogously be expressed per landing.

To estimate the risk of a collision, all the terms in (11) and (12) can be estimated from the data. The GA rate P(GA) in (12) is based on data from a large, publicly available dataset of go-arounds [20]. This data suggests that the GA probability of an aircraft landing on Runway 14 in Zurich is about 3.2 per thousand i.e., P(GA) = 0.0032.

The term P(GA in V|TO in V) in (11) can be estimated from the data by using a simple Monte-Carlo simulation. This simulation uses historical trajectories for the GAs and the TOs. In each simulation run, a randomly selected GA is randomly placed within five nautical miles from the threshold of Runway 14. At the same time, a randomly selected aircraft is taking-off on Runway 16, as depicted in Fig. 8. With the two historical trajectories evolving over time, the number of samples within the volume V can then be counted (presuming an equal, constant, and sufficiently large sampling rate of the trajectories). The probability of a GA being in the volume V, given that a take-off is in the volume V, can be calculated by

$$\mathrm{P}(\text{GA in V}|\text{TO in V}) = \frac{n_{\text{GA in V} \cap \text{TO in V}}}{n_{\text{TO in V}}} \tag{14}$$

with $n_{\text{GA in V} \cap \text{TO in V}}$ being the number of samples where both the GA and the TO are in the volume at the same time and $n_{\text{TO in V}}$ being the number of samples where the TO is in the volume.

The term P$(C|(\text{GA in V}|\text{TO in V}))$ in (11) is shortened in the following to P($C|V$) to increase readability. This term describes the novel model proposed in Sec. III and represents the probability of an overlap in the position of the two aircraft (i.e., a collision). The following describes how this can be calculated.

1) Fitting the probability density function: In a first step, a probability density function, namely a KDE, is fitted to the observed trajectories of the go-arounds. The KDE applied in this study uses a Gaussian kernel and the bandwidth h is estimated using the multivariate plug-in selector of Wand and Jones (see Sec. II-C). The estimated bandwidth matrix is

$$h = \begin{pmatrix} 11.1 & 0 & 0 \\ 0 & 11.2 & 0 \\ 0 & 0 & 1.8 \end{pmatrix}. \tag{15}$$

As stated earlier, the bandwidth h is a diagonal matrix and with the elements on the diagonal representing the bandwidth for each axis. The estimated matrix h has very similar estimates for the first two axis, which represent the lateral position in x and y. The third row represents the bandwidth of the altitude and is much smaller. This matches the expectation that the dispersion in the horizontal plane (x and y) would be similar, while the one in the vertical axis would be different. A slice through the 3-D PDF at about 3900 feet is shown in Fig. 9. The areas in darker red show a higher probability

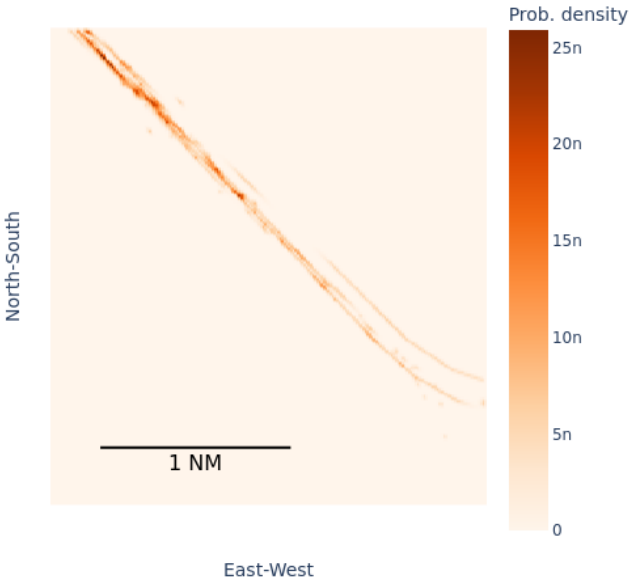

Figure 9: Visualisation of a horizontal slice at around 3900 ft through the fitted 3-D probability density function.

that an aircraft performing a go-around will be at this position. As expected, the probability is higher in areas where multiple trajectories were crossing through the volume.

2) Computing the probability of overlap: To estimate the probability of collision when both a take-off and a go-around are simultaneously in the volume, the 3-D version of (9) has been applied to 5100 Runway 16 departures. All aircraft have been modelled as cylinders with a diameter $\lambda = 70$m and a height $\lambda_z = 10$m. The GAs have been modelled with a constant speed of 216 Knots, a climb rate of $809\,\mathrm{ft/min}$ and a track angle of 137°, with these values being the mean of the corresponding data as extracted from the historical trajectories. For a given take-off TO$_i$, the probability is

$$\begin{aligned} \mathrm{P}(C|V)_{\mathrm{TO}_i} = 4\lambda\lambda_z \int_{t=0}^{t=t_{\text{end}}} \Big(& (V_2(t)\cos\theta(t) \; - V_1)^2 \\ & + (V_2(t)\sin\theta(t)\;)^2 + (Vz_2(t) - Vz_1\;)^2 \Big)^{1/2} \\ & \rho\left(x_2(t), y_2(t), z_2(t)\right)\, dt \end{aligned} \tag{16}$$

where V_1 is the average ground speed of the GAs, Vz_1 the average vertical rate of the GAs, and $x_2(t), y_2(t), z_2(t)$ the

3-D position of the considered TO_i at time t. Equation (16) integrates along the path of TO_i from the moment it enters the volume ($t = 0$) until it leaves it again ($t = t_{\text{end}}$). This integral can be computed numerically relatively easily. An illustration on a random sample of take-offs is shown in Fig. 10. The

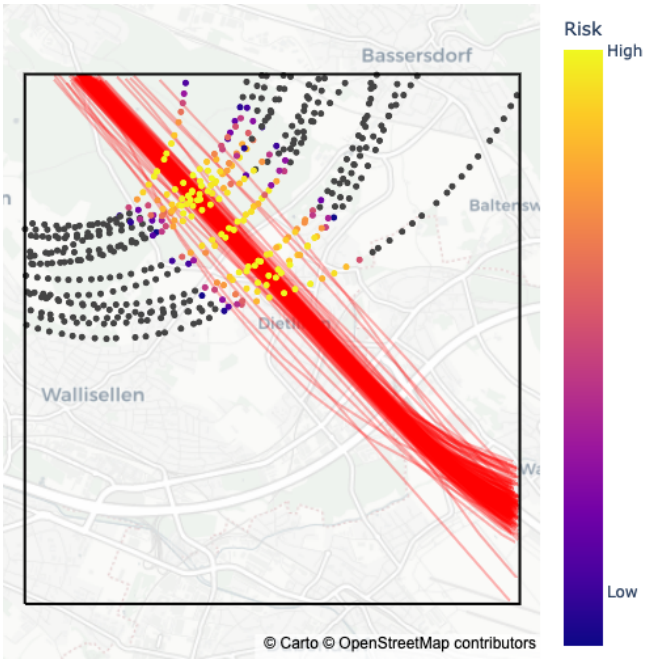

Figure 10: Risk of collision for 15 randomly chosen take-offs inside the area of interest. The risk varies in function of the position of each take-off and is indicated by the color of the markers. The red lines corresponds to the GAs trajectories.

color for each observation of the trajectories indicates the probability of a collision. The integral in the equation above integrates along one such trajectory.

The estimate of the probability of an overlap in position, $\text{P}(C|V)$, can be obtained by averaging the probability of each individual take-off:

$$\text{P}(C|V) = \frac{1}{n}\sum_{\text{i=1}}^{n} \text{P}(C|V)_{\text{TO}_i} \tag{17}$$

D. Results

1) Probability of an overlap in position estimation: Fig. 11 shows the histogram of the $\text{P}(C|V)_{\text{TO}_i}$ for each take-off. It can be seen that the probability of an overlap in position for an individual take-off is $0 < \text{P}(C|V)_{\text{TO}_i} < 0.085$. Averaging over all take-offs results in:

$$\text{P}(C|V) = 2.42 \times 10^{-3} \tag{18}$$

2) P(GA in V|TO in V) estimation: The probability of a GA being in the volume V, given that a take-off is in the volume V, was computed according to (14). Performing 1,280,000 Monte-Carlo simulation runs resulted in

$$\text{P}(\text{GA in V}|\text{TO in V}) = 0.334 \tag{19}$$

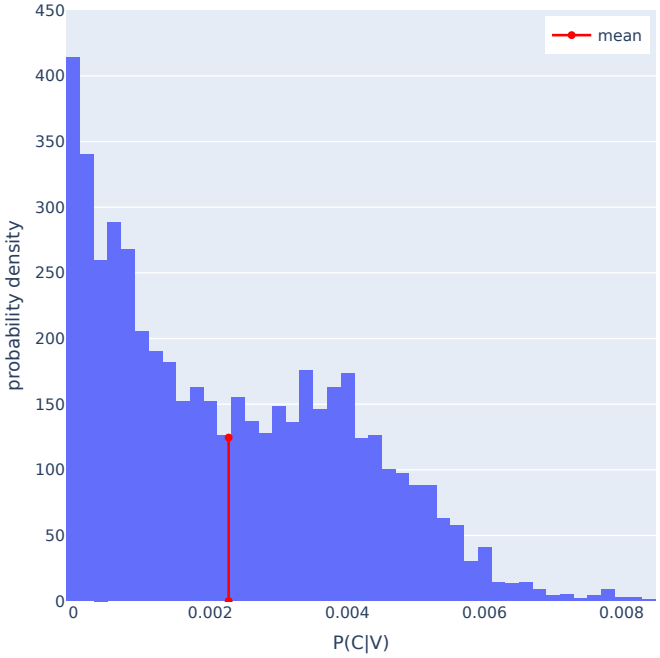

Figure 11: $\text{P}(C|V)$ estimation using 5100 take-offs. The red line corresponds to the mean value.

3) Probability of collision: Combining all the obtained probabilities and a go-around rate of $\text{P}(\text{GA}) = 3.2 \times 10^{-3}$, the probability of a collision, given simultaneous operation, is

$$\begin{aligned}\text{P}(C) &= \text{P}(C|V) \cdot \text{P}(\text{GA in V}|\text{TO in V}) \cdot \text{P}(\text{GA}) \\ &= 2.42 \times 10^{-3} \cdot 0.334 \cdot 3.2 \times 10^{-3} \\ &= 2.58 \times 10^{-6}\end{aligned} \tag{20}$$

Therefore, the probability of a take-off to collide with an aircraft performing a GA, assuming simultaneous operation, is 2.58×10^{-6}. This value is well above the target level of safety, as would be expected, since the two runways under study are not operated independently. This result is validated with the following Monte-Carlo simulation.

4) Comparison with Monte-Carlo simulations: A Monte-Carlo simulation of 1,280,000 runs, where TOs and GAs where sampled independently, has been conducted to compare the results obtained with the newly proposed method. For the Monte-Carlo simulation, it is possible to estimate the probability of collision using:

$$\begin{aligned}\text{P}(C)_{MC} &= \text{P}(C|\text{GA})_{MC} \cdot \text{P}(\text{GA}) \\ &= 1.32 \times 10^{-4} \cdot 3.2 \times 10^{-3} \\ &= 4.22 \times 10^{-6}.\end{aligned} \tag{21}$$

The results obtained with the Monte-Carlo simulation are: $\text{P}(C)_{MC} = 4.22 \times 10^{-6}$ while a result of $\text{P}(C) = 2.58 \times 10^{-6}$ was obtained by integrating TO trajectories in the density. The slight discrepancy between the results obtained form the Monte-Carlo simulation and the proposed method can be attributed to the smoothing effect of the KDE on the estimated probability density. While a perfect match between the results

was not expected , the close proximity of the results suggests that the proposed method is reliable.

E. Using the model to gain insight into the system

The proposed method can also be used to gain insight into the system under study. For example, the same model could be applied to compute the probability of a near mid-air collision or close proximity event; a procedure may have a collision risk well below a TLS, but still have an unacceptable rate of close-proximity events. The method can also be used to determine which types of aircraft have a greater risk of collision as shown in Fig 12. In the case study shown, aircraft with lower climb performance have a smaller probability to collide with a go-around, since they do not reach the altitude of the go-around procedure. This type of insight can be useful to ANSPs where procedures may be defined for specific aircraft types (for example, wake-turbulence separation by categories).

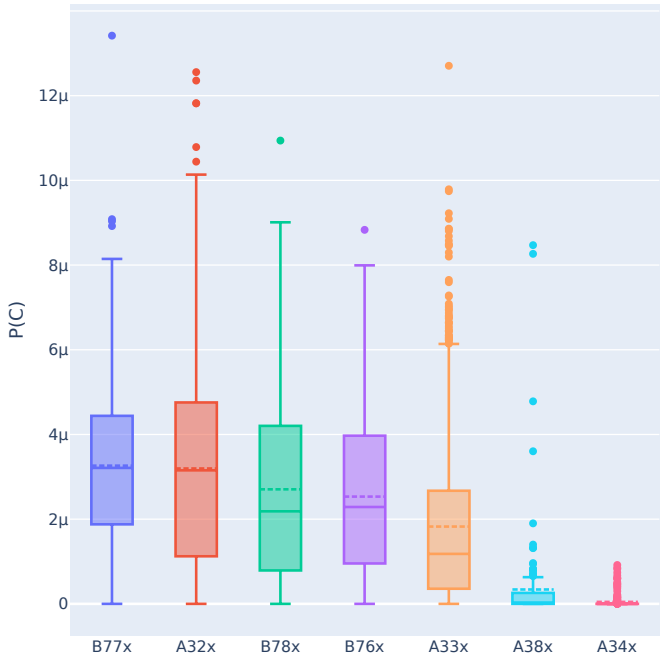

Figure 12: Box plot of collision probabilities for the 7 most common aircraft type in the historical data. The plain horizontal lines from top to bottom correspond to the minimum, the first, second and third quartiles and finally the maximum respectively. The dashed line corresponds to the mean value.

Thanks to the comparably low computational cost, this method is suitable to quantify the risk of multiple different scenarios. In the presented case study, the model could be used to develop an operational concept that meets the target level of safety, while maximising the airport throughput.

V. Conclusions and Outlook

The novel collision risk model proposed in this paper is based on a data-driven approach. It mathematically describes an aircraft crossing a flow of aircraft, with the latter being modelled as a probability density function. This approach is, compared to Monte-Carlo simulations, computationally more efficient while, in comparison to classical analytical models, suited for the complex traffic flows encountered in a terminal airspace.

A real-world application of the model is illustrated on a case-study from the airport of Zurich, Switzerland. The novel model was used to estimate the collision risk between go-arounds on Runway 14 and take-offs on Runway 16, which have intersecting flight path. The same system was in parallel modelled with Monte-Carlo simulations to validate the results of the new method. A comparison of the two methods shows similar results, suggesting that the new method is reliable in such applications.

Compared to other methods, the advantage of the novel model is that it is able to model complex traffic patterns, while still being relatively computationally light-weight. This allows the model to be used in gaining insight into the system under study, for example to efficiently identify aircraft types that, due to their performance, create lower or higher collision risks. Such insights can be helpful in finding a balance between meeting the target level of safety and maximising the throughput of an airport.

while the proposed model is a useful and efficient tool to model the collision risk in complex airspaces, it also has its limitations that could be addressed in future research. The main limitation in the current model is the assumption of a constant speed and heading of all the traffic modelled by the probability density function. Developing a model that can handle distributions for these parameters would give the model yet more flexibility and would open up further applications. A more ambitious extension to the model would be to estimate the collision risk based on two probability density functions that represent the positions of the traffic. If successful, this would be a significant step to estimate the collision risk independently of its geometry and would significantly simplify collision risk modelling and monitoring.

We are convinced that the proposed model can be a useful tool in modelling collision risk in complex airspaces and we are looking forward to seeing it applied in real-world applications, both in the world of manned aviation and in unmanned aviation.

Acknowledgement

The authors are grateful to Martijn Weterings for his helpful contribution to solving some issues with the mathematical formulation of the problem.

References

[1] X. Lin, N. Fulton, and M. Westcott, “Target level of safety measures in air transportation – review, validation and recommendations,” in *Proceedings of the IASTED International Congress on Advances in Management Science and Risk Assessment (AMSRA 2009)*, pp. 662–222, 10 2009.

[2] Eurocontrol, “Safety minima study: Review of existing standards and practices,” 2000.

[3] ICAO, “Unified framework for collision risk modelling in support of the manual on airspace planning methodology with further applications, circular 319-an/181 ed,” *International Civil Aviation Organization, Montreal, Canada*, 2008.

[4] ICAO, “Doc 9689, manual on airspace planning methodology for the determination of separation minima,” *International Civil Aviation Organization, Montreal, Canada*, 1998.

[5] ICAO, "Doc 10116, manual for ats surveillance separation using rcp240 communications," *International Civil Aviation Organization, Montreal, Canada*, 2008.
[6] M. Henry, S. Schmitz, N. Revenko, and K. Kelbaugh, *A Monte Carlo Simulation for Evaluating Airborne Collision Risk in Intersecting Runways*, pp. 1–21. AIAA, 2013.
[7] H. Fricke, S. Förster, and M. Vogel, "Using agent-based modeling to determine collision risk in complex tma environments: the turn-onto-ils-final safety case," *Aeronautics and Aerospace Open Access Journal*, vol. 2, 06 2018.
[8] M. L. Williams, L. C. Wood, and B. J. Nelson, "Safety study of closely spaced parallel operations utilizing paired approach," in *2019 IEEE/AIAA 38th Digital Avionics Systems Conference (DASC)*, pp. 1–10, 2019.
[9] D. Hsu, "The evaluation of aircraft collision probabilities at intersecting air routes," *The Journal of Navigation*, vol. 34, no. 1, pp. 78–102, 1981.
[10] P. G. Reich, "Analysis of long-range air traffic systems: Separation standards—i," *The Journal of Navigation*, vol. 19, no. 1, p. 88–98, 1966.
[11] J. Morio, T. Lang, and C. Le Tallec, "Estimating separation distance loss probability between aircraft in uncontrolled airspace in simulation," *Safety Science*, vol. 50, p. 995–1004, 04 2012.
[12] S. Barry, "Manual for ats surveillance separation using RCP240 communications: Chapter 3 - mathematical modelling," in *ICAO, Separation Airspace Safety Panel*, vol. SASP-WGH-33-2019-Nov-Montreal, pp. 1–10, ICAO, Nov 2019.
[13] ICAO, "Doc 9937, operating procedures and practices for regional monitoring agencies in relation to the use of a 300 m (1 000 ft) vertical separation minimum between fl 290 and fl 410 inclusive," *International Civil Aviation Organization, Montreal, Canada*, 2010.
[14] S. Barry, "Mathematics for crossing a track," in *ICAO, Separation Airspace Safety Panel*, vol. SASP-WGH-36-2022-May-Virtual-IP-13, pp. 1–14, ICAO, May 2022. SASP-WGH-36-2022-May-Virtual-IP-13.
[15] M. Rosenblatt, "Remarks on Some Nonparametric Estimates of a Density Function," *The Annals of Mathematical Statistics*, vol. 27, no. 3, pp. 832 – 837, 1956.
[16] E. Parzen, "On Estimation of a Probability Density Function and Mode," *The Annals of Mathematical Statistics*, vol. 33, no. 3, pp. 1065 – 1076, 1962.
[17] A. Gramacki, *Bandwidth Selectors for Kernel Density Estimation*, pp. 63–83. Cham: Springer International Publishing, 2018.
[18] M. P. Wand and C. Jones, "Multivariate plug-in bandwidth selection," *Computational Statistics*, vol. 9, no. 2, pp. 97–116, 1994.
[19] M. Schäfer, M. Strohmeier, V. Lenders, I. Martinovic, and M. Wilhelm, "Bringing up opensky: A large-scale ads-b sensor network for research," in *IPSN-14 Proceedings of the 13th International Symposium on Information Processing in Sensor Networks*, pp. 83–94, 2014.
[20] R. Monstein, B. Figuet, T. Krauth, M. Waltert, and M. Dettling, "Large landing trajectory dataset for go-around analysis," *Engineering Proceedings*, vol. 28, no. 1, 2022.