



*International Civil Aviation Organization*

**Middle East Regional Monitoring Agency Board**

**Eighteenth Meeting (MIDRMA Board/18)**  
*(Doha, Qatar, 19 – 20 September 2022)*

---

---

**Agenda Item 4: RVSM Monitoring and Related Technical Issues**

**PROGRESS REPORT ON THE UPGRADE OF  
THE MID RISK ANALYSIS SOFTWARE (MIDRAS)**

*Presented by the MIDRMA*

**SUMMARY**

This paper presents in-depth the progress and ongoing works on the upgrade and development of new features in MIDRAS software incorporating Artificial Intelligence (MIDRAS AI).

Action by the meeting is at paragraph 3.

**REFERENCES**

- MIDRMA Board/17 Report
- MIDRAS AI Project Proposal (BB-QUOT-2022-002 V1.1)
- Cost-Benefit Analysis of the MIDRAS Upgrade Project

**1. INTRODUCTION**

1.1 Given the continued growth in air transportation, one of the key challenges faced by Air Navigation Service Providers (ANSPs) and airlines is: how to increase airspace capacity without compromising on safety? New air traffic management (ATM) paradigms by ICAO aim for doubling the airspace capacity (2x) while increasing the safety by a factor of 10 by 2030.

1.2 To achieve such ambitious targets, development of new operational concepts, safety measures and safety performance indicators in the air traffic system are not only expected but also necessary. Reduced Vertical Separation Minimum (RVSM) airspace which ranging vertically from 29,000 feet (FL290) to 41,000feet (FL410), reduces the vertical separation from 2000 feet to 1000 feet, adding 6 extra flight levels.

1.3 To maintain the safety and integrity of airspace stringent procedures by ICAO are in place. To achieve these measures, MIDRMA has developed MIDRAS software for collision risk assessment of Middle East airspace.

1.4 This software system uses ICAO RGCSP Vertical Collision Risk Model for collision risk computation. The ICAO model is based on knowledge of the traffic flows along a given route

structure. The software computes Collision Risk equation parameters and process flight data for each member state and computes Collision Risk. The MIDRAS software also provide an interactive interface for collision risk visualization, simulation, modelling of scenarios and Hot-Spot analysis.

## 2. DISCUSSION

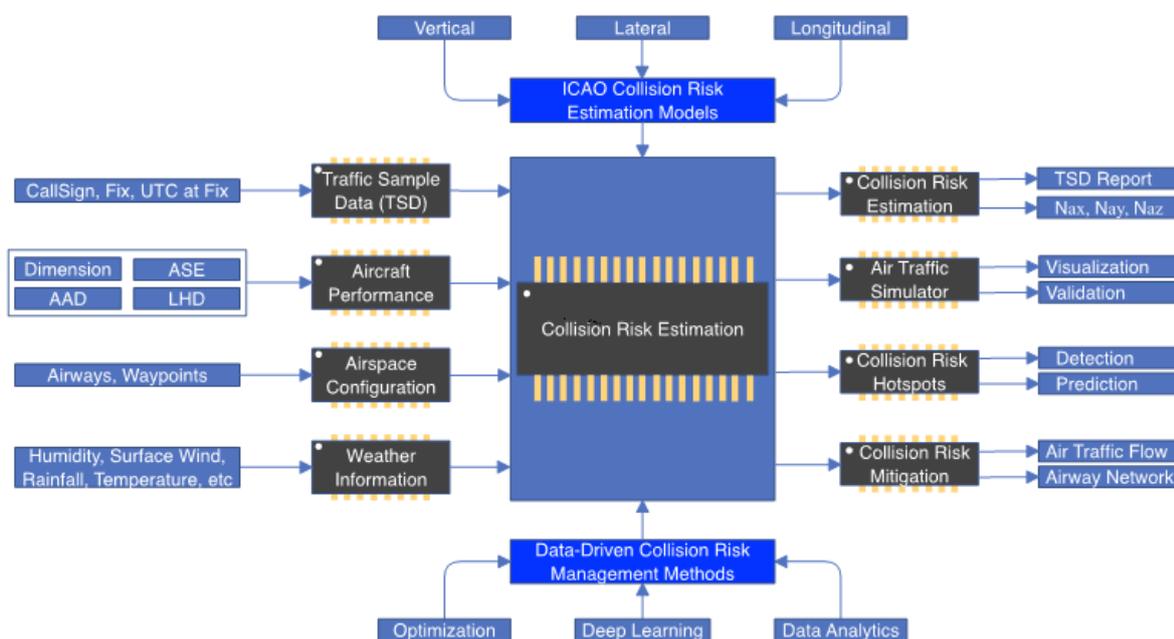
2.1 MIDRAS is a Collision Risk Assessment software system with interactive features for collision risk analysis and visualization for Middle East airspace region. The MIDRAS software integrates the ICAO standard models for collision risk calculations and provides an interactive interface for Collision Risk analysis, scenario planning, Hot-Spot analysis and fast-time air traffic simulation.

2.2 With the changing dynamics of Middle East air traffic flow, airspace reconfiguration, new aircraft and increased congestion, the MIDRAS software needs significant upgrades in terms of new Hot-Spot modelling, congestion analysis as well as several other features that are requires to serve the emerging needs of airspace users and ANSPs in the Middle East region.

2.3 This report presents the progress and ongoing work on the upgrades and development of new features in MIDRAS software incorporating Artificial Intelligence (MIDRAS AI). These new features and upgrades will enable MIDRMA in making effective analysis of collision risk in the region and gain new insights into emerging traffic dynamics to better manage airspace congestion.

### 2.2 MIDRAS New Architecture

2.2.1 The new software architecture for MIDRAS is presented in Figure 1, which illustrates various components and data module which have been integrated in MIDRAS. Such changes allow for the Hot-Spot modelling, congestions analysis, and Top of Descent identification in the air traffic data.

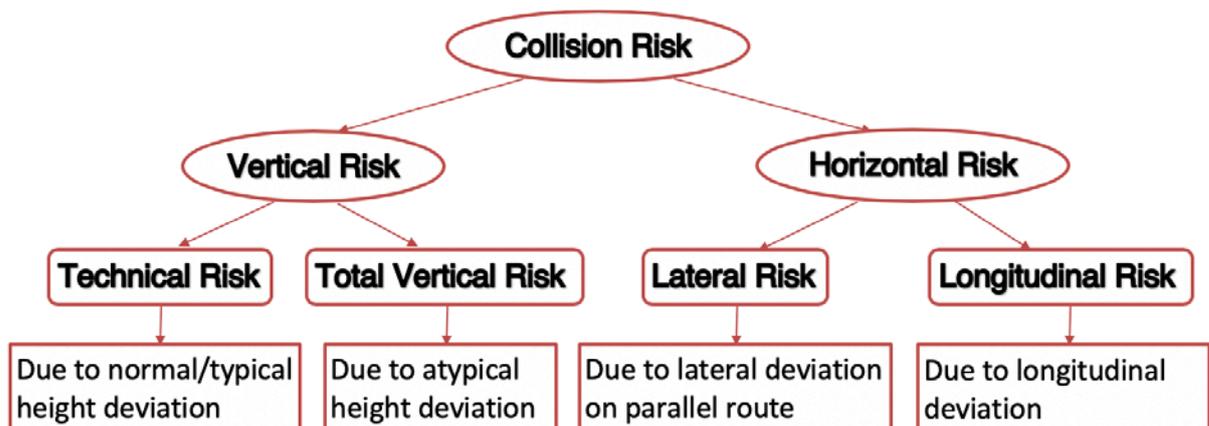


**Figure 1:** MIDRAS Architecture design with integrated Machine Learning Module.

### 2.3 Collision Risk Model

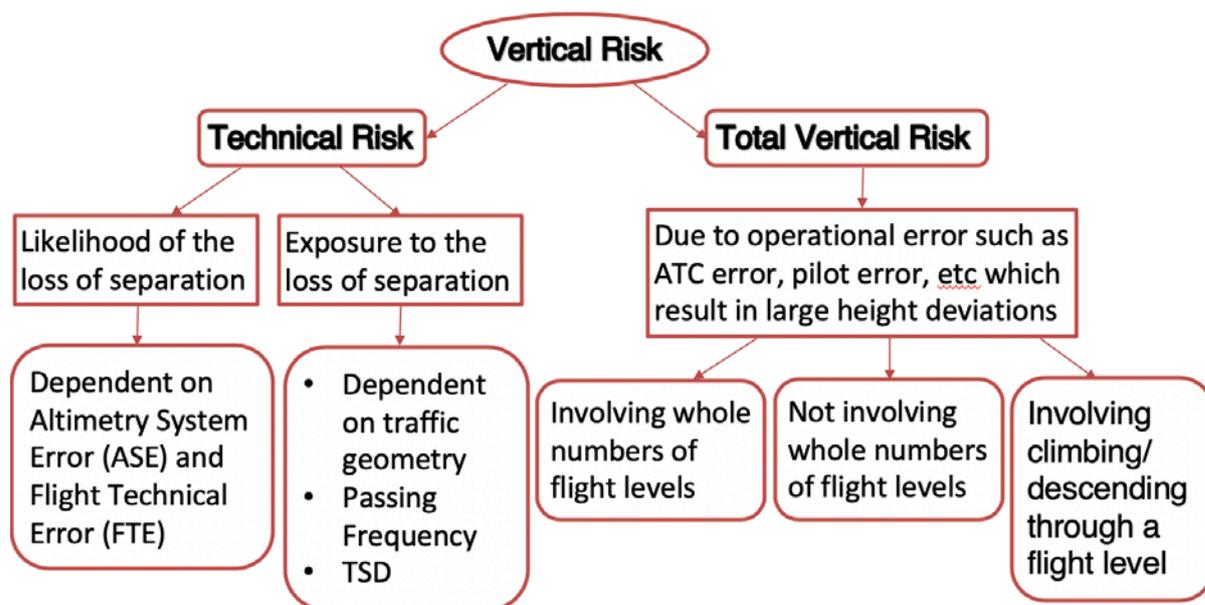
2.3.1 According to ICAO, collision risk is defined as “the expected number of mid-air aircraft accidents in a prescribed volume of airspace for a specific number of flight hours due to loss of planned separation”. Collision risk provides a holistic view for the safety level of the traffic within a given airspace for a given period.

2.3.2 Based on ICAO regulation, there are two specific safety objectives for collision risk assessment, namely an assessment of the technical vertical risk against a TLS of  $2.5 \times 10^{-9}$  fatal accidents per flight hour (fafh), and an assessment of the total vertical risk against a TLS of  $5 \times 10^{-9}$  fafh. The horizontal risk that consists of lateral risk and longitudinal risk is also assessed against a TLS of  $5 \times 10^{-9}$  fafh. According to FAA, the 2015 annual flight hours for the USA was 9.8 million flight hours. Therefore, a vertical TLS of  $5 \times 10^{-9}$  fafh equates to an acceptable value of risk of roughly 1 fatal accident every 20 years resulting from a loss of vertical separation. Figure 2 illustrate the Collision Risk components.

**Figure 2:** Collision Risk components comprising of vertical risk and horizontal risk.

### 2.4 Vertical Risk

2.4.1 In MIDRAS, the ICAO collision risk model is used to compute the vertical collision risk. Fig 3 illustrates the concept diagram of the vertical risk. In this model, the important parameters that have to derive are  $P_z(S)_z$  (Probability of vertical overlap),  $P_y(0)$  (Probability of lateral overlap) and  $n_z(equiv)$  (Equivalent passing frequency). Before these parameters can be computed, the flight data and the supplementary data has to be pre-processed first.



**Figure 3: Vertical Collision Risk components comprising of Technical Risk and Total Vertical Risk**

## 2.5 Calculation of Lateral Overlap Probability

2.5.1 According to ICAO documents, there are two models for calculating the lateral overlap probability, i.e.,  $P_y(S_y)$ . Those two models are respectively shown below.

The first model (M1) calculates  $P_y(S_y)$  using the following equation:

$$P_y(S_y) = \int_{-\lambda_y}^{\lambda_y} \int_{-\infty}^{\infty} f_y(y_1) f_y(S_y + y_1 - y) dy_1 dy \approx 2\lambda_y \int_{-\infty}^{\infty} f_y(y_1) f_y(S_y + y_1) dy_1 \quad (1)$$

where  $f_y(y_1)$  is the probability density function for the lateral deviation  $y_1$ . The probability distribution of  $f_y(y_1)$  is calculated as

$$f_y(y_1) = (1 - \alpha) \frac{1}{2a_1} e^{-\frac{|y_1|}{a_1}} + \alpha \frac{1}{2a_2} e^{-\frac{|y_1|}{a_2}} \quad (2)$$

in which  $\alpha$  represents the percentage of aircraft that experience anomalies caused by aircraft navigation system errors.

In the above equation, the value of  $\alpha$  is estimated as  $\alpha = 1 - 0.05^{1/n}$  with  $n$  being the annual number of flights [49]. The parameter  $a_1$  is determined by the RNP specification and is estimated as  $a_1 = -\text{RNP}/\ln 0.05$ , while parameter  $a_2$  is estimated as  $a_2 = S_y$ .

The second model (M2) calculates  $P_y(S_y)$  using the following equation:

$$P_y(S_y) = \left( \frac{2\alpha\lambda_y}{\beta_2} \right) e^{-\frac{S_y}{\beta_2}} \quad (3)$$

where  $\alpha$  is the weight of the a-typical lateral error of the overall distribution and it is estimated as

-5-

$$\alpha = \frac{\lambda(k)}{M \cdot e^{\frac{15}{s_y}}} \quad (4)$$

in which M is defined as the cumulative 12-month traffic counts.  $\lambda(k)$  is the intensity parameter and is dependent of parameter  $k$  which is the cumulative 12-month total number incident of large lateral deviation. The relationship between  $\lambda$  and  $k$  is as follow:

$$0.95 = e^{-\lambda} \frac{\lambda_k}{k!} \quad (5)$$

## 2.6 Calculation of Vertical Overlap Probability

The probability of vertical overlap of aircraft nominally flying at the same flight level on laterally adjacent flight paths, viz.,  $P_z(0)$ , is calculated as

$$P_z(0) = \int_{-\lambda_z}^{\lambda_z} \int_{-\infty}^{\infty} f_z(z_1) f_z(z_1 - z) dz_1 dz \approx 2\lambda_z \int_{-\infty}^{\infty} f_z(z_1) f_z(z_1 - z) dz_1 \quad (6)$$

where  $f_z(z_1)$  is the probability distribution function of the vertical deviation  $z_1$ . Note that the vertical deviation is mainly caused by ASE and AAD. Let  $f^{ASE}(z)$  and  $f^{AAD}(z)$  respectively be the probability distribution functions of the ASE and AAD. Then  $f_z(z_1)$  is calculated as

$$f_z(z_1) = \int_{-\infty}^{\infty} f^{ASE}(z) f^{AAD}(z_1 - a) da \quad (7)$$

where  $f^{AAD}(z)$  follows a double exponential distribution given as  $f^{AAD}(z) = \frac{1}{\sigma_{AAD}\sqrt{2}} \exp(-|z|\sqrt{2}/\sigma_{AAD})$ .

In order to estimate  $f^{ASE}(a)$  involved in equation 4, it is assumed that there are  $n_{AT}$  different types of aircraft operating in a given RVSM airspace. Each type of aircraft have their own parameters subject to the ASE variability of all the airframes and each type have their own ASE probability distributions  $f_i^{ASE}(a), i = 1, \dots, n$ . An overall ASE probability density  $f^{ASE}(a)$  for the full RVSM aircraft population is calculated as a weighted mixture of the ASE distribution by different types of aircraft, i.e.,

$$f^{ASE}(a) = \sum_{i=1}^{n_{AT}} \beta_i f_i^{ASE}(a) \quad (8)$$

in which the weighing factor  $\beta_i$  is the proportion of flight time contributed by each aircraft type.

Both the weighting factors and the ASE probability distributions required are derived from the monitoring data with regards to a given RVSM airspace. The data can be collected by ground-based Height Monitoring Units (HMUs) or by air portable GPS Monitoring Units (GMUs). Based on monitoring data,  $f_i^{ASE}(a)$  is observed to follow the gaussian double exponential distribution given as

$$f^{ASE}(a) = (1 - \alpha_i) \frac{1}{\sigma_{1i}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{a-\mu_i}{\sigma_{1i}}\right)^2} + \alpha_i \frac{1}{\sigma_{2i}\sqrt{2\pi}} e^{-\frac{|a-\mu_i|\sqrt{2}}{\sigma_{2i}}} \quad (9)$$

## 2.7 Calculation of Occupancy

There are 2 types of occupancy, namely same direction occupancy and opposite direction occupancy. To determine the occupancy, the number of proximate aircraft pairs for pairs of parallel routes have to be determined first. Every flight on its route is compared with other flights that are on the adjacent parallel route by using the reporting points. The reporting points used must be points that are at a right angle to the plane of the parallel routes so as to compare the passing time of aircraft on one route to the passing time of aircraft on the other route and they can be the waypoints or interpolated points between the waypoints that have timestamps

If two flights are at the same flight level, then they are counted as an approximate pair as long as the time difference between any two reporting points for those two flights is less than or equal to 15 minutes (a time duration taken by a flight to fly  $S_x$ ).

The number of proximate aircraft pairs for all route pairs is summed up after considering all the flights. The purpose of computing the number of proximate pairs is to determine  $E_y$  (*lateral occupancy*). The occupancy is calculated as  $E_y = 2n_y/n$ , where  $n_y$  is the number of proximate aircraft pairs and  $n$  is the total number of aircrafts in the parallel route pairs.

### 2.7.1 Calculation of Horizontal Overlap Probability

The horizontal overlap probability for a pair of aircraft with speeds  $V_1$  and  $V_2$  at a given time point  $t$  is calculated as

$$HOP(t|V_1, V_2) = \frac{\pi\lambda_{xy}^2}{16\lambda^2} e^{-|D_x(t)|/\lambda} \left( \frac{|D_x(t)|}{\lambda} + 1 \right) \quad (10)$$

where  $D_x(t) = (V_1 - V_2)t - \hat{d}_1^0 + \hat{d}_2^0$  is the distance between two aircraft and  $\lambda$  is the scale parameter of the along-track and cross-track error distributions. The along-track and cross-track errors are assumed to follow a double exponential distribution.

### 2.8 Distribution of Speed Variation

The notations of  $f_1(V_1)$  and  $f_2(V_2)$  in the longitudinal risk model are the probability distribution functions for the aircraft speed variations. It is assumed that both  $f_1(V_1)$  and  $f_2(V_2)$  follow a double exponential distribution with scale parameter  $\lambda_v = 5.82 \text{ knots}$ . The assumed average aircraft ground speed of 480 knots is used as the location parameter  $V_0$ . The double exponential distribution is truncated at 100 knots on either side of the location parameter, and then normalized to equal 1. Thus,  $f_1(V_1)$  and  $f_2(V_2)$  have the same form as follows

$$f_1(V_1) = \frac{1}{2\lambda_v} e^{-\frac{|V-V_0|}{\lambda_v}}, -100 \leq V_1 \leq 100 \quad (11)$$

### 2.9 Mathematical model to compute the Collision Risk

#### 2.9.1 Technical Risk

The model that is used for the assessment of the technical vertical collision risk is shown in Equation 12.

$$N_{az} = 2P_z(S_z)P_y(0)n_z(\text{equiv}) \left\{ 1 + \frac{|\bar{y}|}{2\bar{V}} + \frac{\lambda_{xy} |\bar{z}|}{\lambda_z 2\bar{V}} \right\} \quad (12)$$

The definition of the parameters is shown in Table 1.

Table 1: Parameters and definition of Technical Risk

Parameter	Definition
$N_{az}$	The expected number of fatal aircraft accidents per flight hour due to the loss of vertical separation
$S_z$	The vertical separation minimum
$P_z(S_z)$	The probability of vertical overlap for aircraft nominally flying on adjacent flight levels
$P_y(0)$	The probability of lateral overlap for aircraft nominally flying at the same route
$\bar{V}$	The average ground speed of a typical aircraft
$ \bar{y} $	The average of the absolute value of the relative cross-track speed between two typical aircraft flying at adjacent flight levels of the same route
$ \bar{z} $	The average of the absolute value of the relative vertical speed between two typical aircraft which have lost $S_z$ feet of vertical separation
$\lambda_x$	The average length of a typical aircraft
$\lambda_y$	The average width of a typical aircraft
$\lambda_z$	The average height of a typical aircraft
$\lambda_{xy}$	The average diameter of a typical aircraft (Maximum of height and width)

To determine the  $P_z(S)_z$  (Probability of vertical overlap), we have to first model the TVE distribution as shown in equation 13:

$$f^{TVE}(z) = \int_{-\infty}^{\infty} f^{ASE}(a)f^{AAD}(z - a)da \quad (13)$$

Whereby the ASE and AAD distribution have to be determine first. The equation for ASE and AAD is as follow,

$$f^{ASE}(a)f^{ASE}(a) = \sum_{i=1}^{n_{AT}} \beta_i f_i^{ASE}(a) \quad (14)$$

$$f_i^{ASE}(a)_{GDE} = (1 - \alpha_i) \frac{1}{\sigma_{1i}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{a-\mu_i}{\sigma_{1i}}\right)^2} + \alpha_i \frac{1}{\sigma_{2i}\sqrt{2}} e^{-\frac{|a-\mu_i|\sqrt{2}}{\sigma_{2i}}} \quad (15)$$

$$f^{AAD}(a)_{GDE} = \frac{1}{\sigma_{AAD}\sqrt{2}} e^{\frac{|a|\sqrt{2}}{\sigma_{AAD}}} \text{ where } \sigma_{AAD} = 39.8 \text{ ft} \quad (16)$$

Thereafter, we can determine the  $P_z(S)_z$  (Probability of vertical overlap) as

$$P_z(S_z) = \int_{-\lambda_z}^{\lambda_z} \int_{-\infty}^{\infty} f_1^{TVE}(z_1) f_2^{TVE}(S_z + z_1 - z) dz_1 dz \quad (17)$$

and it can be approximated by

$$P_z(S_z) \approx 2\lambda_z \int_{-\infty}^{\infty} f_1^{TVE}(z_1) f_2^{TVE}(S_z + z_1 - z) dz_1 \quad (18)$$

To determine the  $P_y(0)$  (Probability of lateral overlap), the following equation is used,

$$P_y(0) = \int_{-\lambda_y}^{\lambda_y} f_{y_{12}}(y) dy \quad (19)$$

where  $\lambda_y$  denotes the average width of the aircraft and  $f_{y_{12}}(y)$  denotes the probability density of the lateral distance  $y_{12}$  of the two aircraft having a deviation of  $y_1$  and  $y_2$  and it can be approximate by,

$$P_y(0) \approx 2\lambda_y f_{y_{12}}(0) \quad (20) \text{ where,}$$

$$\begin{aligned} f_{y_{12}}(y) = & (1 - \alpha)^2 \frac{1}{\sigma_{VOR/DME} \sqrt{2} \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{y}{\sigma_{VOR/DME} \sqrt{2}} \right)^2} \\ & + 2\alpha(1 - \alpha) \frac{1}{\sqrt{\sigma_{VOR/DME}^2 + \sigma_{GNSS}^2} \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{y}{\sqrt{\sigma_{VOR/DME}^2 + \sigma_{GNSS}^2}} \right)^2} \\ & + \alpha^2 \frac{1}{\sigma_{GNSS} \sqrt{2} \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{y}{\sigma_{GNSS} \sqrt{2}} \right)^2} \quad (21) \end{aligned}$$

Based on the parameter from the ICAO,  $\sigma_{VOR/DME} = 0.3 \text{ NM}$  and  $\sigma_{GNSS} = 0.06123 \text{ NM}$  and based on the assumption that 75% of the aircraft are using GNSS navigation equipment. This parameter can be change in the parameter field of the vertical risk interface depending on different assumptions.

Next, the  $n_z(equiv)$  (Equivalent passing frequency) is computed using the equation 11.

$$n_z(equiv) = \left[ n_z(opp) + \frac{1}{P_y(0)} \sum_{i=1}^n n_z(\theta_i) \frac{\left\{ 1 + \frac{\pi}{2} \frac{\lambda_{xy}}{V_{rel}(\theta_i)} \frac{|\dot{z}|}{2\lambda_z} \right\}}{\left\{ 1 + \frac{|\dot{y}|}{2V} + \frac{\lambda_{xy}}{\lambda_z} \frac{|\dot{z}|}{2V} \right\}} \right] \quad (22)$$

Whereby the following default parameters used:

$|\dot{x}|$ : 15 knots

$|\dot{y}|$ : 20 knots

$\overline{|\dot{z}|}$ : 1.5 knots

Proximity Scaling is in used where  $\frac{\lambda_{xy}}{\overline{S_x}}$  where  $\overline{S_x} = 5$  NM. The  $n_z(opp)$  opposite passing frequency can be compute by:

$$n_z(opp) = \text{Number of Opposite Crossing/Total Flight Hour} \quad (23)$$

Finally, the kinematic factor of the technical risk is computed by using equation 13 and the relevant parameter into the equation as follows:

$$\left\{ 1 + \frac{\overline{|\dot{y}|}}{2\overline{V}} + \frac{\lambda_{xy} \overline{|\dot{z}|}}{\lambda_z 2\overline{V}} \right\} \quad (24)$$

The important parameters of Technical Vertical Risk are summarized in Table 2 as follows:

Table 2: Parameter of the Technical Vertical Risk

Parameter
$S_z$
$P_z(S_z)$
$P_y(0)$
$n_z(equiv)$
$\left\{ 1 + \frac{\overline{ \dot{y} }}{2\overline{V}} + \frac{\lambda_{xy} \overline{ \dot{z} }}{\lambda_z 2\overline{V}} \right\}$

## 2.10 Total Vertical Risk

2.6.1 The total vertical risk is the risk that are due to large a-typical height deviation and it is known as the Large Height Deviation (LHD). These LHD can be due to operational error such as pilot error, ATC error, etc or exceptional technical error. The vertical collision risk due to LHD can be classified into three cases:

- Vertical collision risk due to large height deviations not involving whole numbers of flight levels
- Vertical collision risk due to large height deviations involving aircraft climbing or descending through a flight level without a proper clearance
- Vertical collision risk due to large height deviations involving aircraft levelling off at a wrong flight level

For computing the risk that involve LHD not involving whole numbers of flight levels, the following model that is similar to the technical risk model was used:

$$N_{az}^* = 2P_z(S_z)^* P_y(0) n_z(equiv) \left\{ 1 + \frac{\overline{|\dot{y}|}}{2\overline{V}} + \frac{\lambda_{xy} \overline{|\dot{z}|}}{\lambda_z 2\overline{V}} \right\} \quad (25)$$

Whereby  $P_z(S_z)^*$  is different from  $P_z(S_z)$  used in the technical risk.  $P_z(S_z)^*$  can be calculated by taking the AAD distribution to be

$$f^{AAD}(a) = (1 - \alpha) \frac{1}{\sigma_1^{AAD} \sqrt{2}} e^{-\frac{|a|\sqrt{2}}{\sigma_1^{AAD}}} + \alpha \frac{1}{\sigma_2^{AAD} \sqrt{2}} e^{-\frac{|a|\sqrt{2}}{\sigma_2^{AAD}}} \quad (26)$$

And the ASE distribution remain the same as technical risk.

For computing the risk that involve LHD due to aircraft climbing or descending through a flight level without a proper clearance, the following model that is similar to the technical risk model was used:

$$N_{az}^{cl/d} = 2P_z(S_z)^{cl/d} P_y(0) n_z(equiv) \left\{ 1 + \frac{|\bar{y}|}{2\bar{V}} + \frac{\lambda_{xy} |\bar{z}|}{\lambda_z 2\bar{V}} \right\} \quad (27)$$

Whereby

$$P_z(S_z)^{cl/d} = \frac{n^{cl/d} \times \frac{2\lambda_z}{|\dot{z}_c|}}{T} \quad (28)$$

The definitions of the parameter in equation 28 is shown in Table 3.

Table 3: Definitions of parameter for LHD involving aircraft climbing or descending through a flight level without a proper clearance

Parameter	Definition
$N_{az}^{cl/d}$	Expected number of fatal accidents per flight hour due to aircraft climbing or descending through a flight level without proper clearance
$P_z(S_z)^{cl/d}$	Probability of vertical overlap due to aircraft climbing or descending through a flight level without proper clearance
$n^{cl/d}$	Number of aircraft climbing or descending through a flight level without a proper clearance during a period of time with $T$ flying hours
$ \dot{z}_c $	Average climb or descend rate for aircraft climbing or descending through a flight level without a proper clearance.
$T$	Amount of flying time during the period of time the incident data were collected

For computing the risk that involve LHD due to aircraft levelling off at a wrong flight level, the following model that is similar to the technical risk model was used:

$$N_{az}^{wl} = 2P_z(S_z)^{wl} P_y(0) n_z(equiv) \left\{ 1 + \frac{|\bar{y}|}{2\bar{V}} + \frac{\lambda_{xy} |\bar{z}|}{\lambda_z 2\bar{V}} \right\} \quad (29)$$

Whereby

$$P_z(S_z)^{wl} = \frac{P_z(0) \times n^{wl} \times \bar{t}^{wl}}{T} \quad (30)$$

The definitions of the parameters in equation 30 is shown in Table 4.

Table 4: Definitions of parameter for LHD involving aircraft levelling off at a wrong flight level

Parameter	Definition
$N_{az}^{wl}$	Expected number of fatal accidents per flight hour due to aircraft levelling off at a wrong flight level
$P_z(S_z)^{wl}$	Probability of vertical overlap due to aircraft levelling off at a wrong flight level
$P_z(0)$	Probability of vertical overlap for aircraft nominally flying at the same flight level
$n^{wl}$	Number of aircraft levelling off at a wrong flight level during a period of time with T flying hours
$\bar{t}^{wl}$	Average sojourn time (hours) of an aircraft at a wrong flight level after incorrectly levelling off

### 2.11 Lateral Risk

The model that is used for the assessment of the lateral collision risk is shown in equation 31.

$$N_{ay} = P_y(S_y)P_z(0) \frac{\lambda_x}{S_x} \left\{ E_y(\text{same}) \left[ \frac{|\bar{x}|}{2\lambda_x} + \frac{|\bar{y}(S_y)|}{2\lambda_y} + \frac{|\bar{z}|}{2\lambda_z} \right] + E_y(\text{opp}) \left[ \frac{|\bar{v}|}{\lambda_x} + \frac{|\bar{y}(S_y)|}{2\lambda_y} + \frac{|\bar{z}|}{2\lambda_z} \right] \right\} \quad (31)$$

The definition of the parameters is shown in Table 5.

Table 5: Parameters and definition of Lateral Risk

Parameter	Definition
$N_{ay}$	The expected number of accidents (two per each aircraft collision) per flight hour due to the loss of lateral separation between aircraft flying on tracks with nominal spacing $S_y$ .
$S_y$	Minimum standard lateral separation
$P_y(S_y)$	The probability of lateral overlap of aircraft nominally flying on laterally adjacent paths at the same flight level
$P_z(0)$	The probability of vertical overlap of aircraft nominally flying at the same flight level.
$E_{y\text{same}}$	The same direction lateral occupancy, i.e. the average number of same directions aircraft flying on laterally adjacent tracks at the same flight level within segments of length $2S_x$ centred on the typical aircraft.
$E_{y\text{opposite}}$	the opposite direction lateral occupancy, i.e. the average number of opposites direction aircraft flying on laterally adjacent tracks at the same flight level within segments of length $2S_x$ centred on the typical aircraft.
$S_x$	Length of the longitudinal window used in the calculation of occupancies
$ \bar{v} $	Average groundspeed of the aircraft
$ \bar{y} $	The average lateral cross-track speed between aircraft that have lost their lateral separation
$ \bar{z} $	The average relative vertical speed of aircraft flying at the same flight level
$\lambda_x$	The average length of a typical aircraft
$\lambda_y$	The average width of a typical aircraft
$\lambda_z$	The average height of a typical aircraft
$\lambda_{xy}$	The average diameter of a typical aircraft (Maximum of height and width)

The number of proximate pair for each route pair is sum up after considering all the flight. The purpose of computing the number of proximate pair is to determine the  $E_{y_{opposite}}$  (opposite direction lateral occupancy). The equation used to compute the occupancy is as follow:

$$E_y = \frac{2n_y}{n} \quad (32)$$

where

$n_y$ : No of proximate pairs

$n$  : Total number of aircrafts in the parallel route pair

To determine the  $P_y(S_y)$  (Probability of lateral overlap), the probability of lateral overlap of aircraft nominally flying on parallel flight path separated by  $S_y$  is defined by:

$$P_y(S_y) = \int_{-\lambda_y}^{\lambda_y} \int_{-\infty}^{\infty} f^y(y_1) f^y(S_y + y_1 - y) dy_1 dy \quad (33)$$

Whereby the  $f^y(y_1)$  (overall probability density of lateral navigation errors) have to be determine first. The equation is defined by:

$$f_y(y_1) = (1 - \alpha) \times f_1(y_1) + \alpha \times f_2(y_1) \quad (34)$$

Where:

$f_1(y_1)$  : probability density function due to navigation error arising from typical deviation of the aircraft navigation system

$f_2(y_1)$  : probability density function due to navigation error arising from human error, equipment failure

$\alpha$  : percentage of aircraft that has probability density function due to navigation error arising from human error, equipment failure

The probability density function  $f_1(y_1)$  and  $f_2(y_1)$  can be expressed as follows:

$$f_1(y_1) = \frac{1}{2a_1} \exp - \left| \frac{y_1}{a_1} \right| \quad (35)$$

$$f_2(y_1) = \frac{1}{2a_2} \exp - \left| \frac{y_1}{a_2} \right| \quad (36)$$

Whereby

$$a_1 = - \frac{RNP}{\log 0.05} \quad (37)$$

$$a_2 = S_y \quad (38)$$

Therefore,  $f^y(y_1)$  (overall probability density of lateral navigation errors) can be expressed as:

$$f_y(y_1) = (1 - \alpha) \frac{1}{2a_1} \exp - \left| \frac{y_1}{a_1} \right| + \alpha \frac{1}{2a_2} \exp - \left| \frac{y_1}{a_2} \right| \quad (39)$$

Thereafter, we can determine the  $\alpha$  as

$$\alpha = 1 - 0.05^{1/n} \quad (40)$$

## 2.12 Longitudinal Risk

The model that are used for the assessment of the longitudinal collision risk is as follows:

$$CR(t_0, t_1) = 2NP \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{t_0}^{t_1} HOP(t|V_1, V_2) P_z(h_z) \left( \frac{2V_{rel}}{\pi\lambda_{xy}} + \frac{|\dot{z}|}{2\lambda_z} \right) f_1(V_1) f_2(V_2) dt dV_1 dV_2 \quad (41)$$

$$HOP(t|V_1, V_2) = \frac{\pi\lambda_{xy}^2}{16\lambda^2} e^{-|D_x(t)|/\lambda} \left( \frac{|D_x(t)|}{\lambda} + 1 \right) \quad (42)$$

The definition and value of the parameters are defined in Table 7.

Table 6: Definition and value of Longitudinal Risk

Parameter	Definition	Value
$D_x(t)$	Distance between the two aircraft	$abs(V_1 - V_2) \times t + 50\text{NM}$
$RNP$	Required Navigation Performance	10
$\lambda$	The scale parameter of the along-track and cross-track error	$-\frac{RNP}{\log(0.05)} = 3.338$
$\lambda_{xy}$	Average diameter of aircraft	0.02605 NM
$\lambda_z$	Average height of aircraft	0.00744 NM
$\lambda_p$	Scale factor for speed error distribution	5.82
$V_1$	Average ground speed of aircraft 1	480 knots
$V_2$	Average ground speed of aircraft 2	480 knots
$V_{ref}$	Relative velocity	$abs(V_1 - V_2)$
$NP$	No of a/c per hour	1
$P_z(0)$	Probability of vertical overlap	0.538
$ \dot{z} $	Average relative vertical speed of a co altitude aircraft pair assigned to the same route	1 knot
$t_0, t_1$	ADS periodic report	0, 27 min
$\tau$	Controller Intervention Buffer	4 min, 10.5 min and 13.5 min
$f_{DE}(V)$	Variation in aircraft speed	$\frac{1}{2\lambda_p} e^{-\frac{ V-V_0 }{\lambda_p}}$

## 2.13 Collision Risk Hot spot

## 2.13.1 Collision Risk Hot-spot Definition (Qualitative / Mathematical):

- a) **Qualitative Definition:** Collision risk hotspots are sub-regions of an airspace where multiple aircraft may frequently pass each other on airway segments at adjacent flight levels, as precursor events to airborne collision due to the loss of separation.
- b) **Mathematical Definition:** Consider an airspace volume  $A \in R^3$  and a set of crossing points CP within A. The crossing points are calculated based on a given one-month TSD. Let us define a convex **polytope**  $P_i \in A$  with its height being  $S_z$  ( $S_z = 1000$  feet). Let  $C_{sam}$ ,  $C_{iopp}$  and  $C_{ioth}$  respectively be the collision risk of the same direction crossings, opposite direction crossings, and other angle direction crossings within airspace  $P_i$ . Then airspace  $P_i$  is a collision risk hot-spot if the following condition is satisfied.

$$C_{sam}^i + C_{iopp}^i + C_{ioth}^i \geq \frac{1}{m} \sum_{i=1}^m (C_{sam}^i + C_{iopp}^i + C_{ioth}^i)$$

where  $m$  is the total number of convex polytopes in A. Note: For a given one-month TSD, let TFH be the total flight hours. For a given crossing point  $cp_j \in CP$ , assume that it is a same direction crossing. Then its collision risk is calculated as:

$$n_{cp_j}^{sam} = 2P_z(S_z)P_y(0) \left( 1 + \frac{|\bar{y}|}{|\bar{x}|} + \frac{\lambda_{xy} |\bar{z}|}{\lambda_z |\bar{x}|} \right) / \text{TFH}$$

Analogously, the collision risk for an opposite and other angle crossing points are respectively calculated as

$$n_{cp_j}^{opp} = 2P_z(S_z)P_y(0) \left( 1 + \frac{|\bar{y}|}{2V} + \frac{\lambda_{xy} |\bar{z}|}{\lambda_z 2V} \right) / \text{TFH}$$

$$n_{cp_j}^{oth} = 2P_z(S_z) \left( 1 + \frac{|\bar{y}|}{2V} + \frac{\lambda_{xy} |\bar{z}|}{\lambda_z 2V} \right) / \text{TFH}$$

Consequently,  $C_{sam}$ ,  $C_{iopp}$  and  $C_{ioth}$  are calculated in a generic way as follows

$$C_{type}^i = \sum_{j=1}^{n_{type}^i} n_{cp_j}^{type}$$

In which  $n_{type}$  is the number of the certain type of crossing points.

## 2.13.2 Collision Risk Hot-spot Detection

The mid-air collision risk is caused because of the loss of planned separation. Regarding the vertical collision risk, the contributing factor is the aircraft crossings. As stated earlier, each crossing corresponds to a pair of aircraft with two GPS coordinates. Based on our definition of collision risk hot-spot, a hot-spot could be a cluster of the crossing points. In light of this, in this work we propose the following three-stage method for collision risk hot-spot detection.

**Stage 1 – Clustering** At the first stage, clustering algorithms are applied to group the crossing points into clusters. These clusters are the potential collision risk hot-spots.

**Stage 2 – Identification** At the second stage, the potential number of clusters are being determined by using the halving of clusters ( $k$ ) method to determine the optimal  $k$  so that the size of the cluster is not too big nor too small.

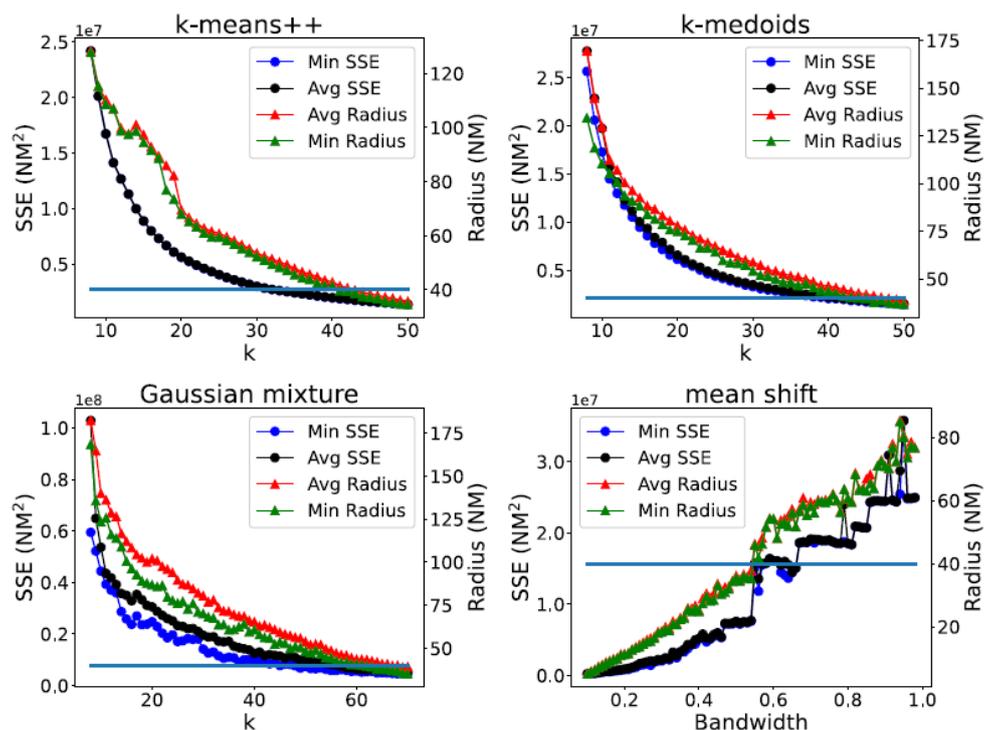
**Stage 3 – Determination** At the third stage, the hotspot clusters is determined based on the criterion given in Eq. 2.

### 2.13.3 Methodology for Hot-spot Detection

#### a) Crossing Points Clustering:

There are myriad clustering algorithms in the literature, and amongst which is the  $k$ -means algorithm. The term “ $k$ -means” was first used by James MacQueen in 1976 [42] but the original idea was credited to Hugo Steinhaus in 1956, while the standard algorithm was first developed by Stuart Lloyd with Bell Labs in 1957. The  $k$ -means algorithm has one intrinsic parameter  $k$ . A standard  $k$ -means algorithm randomly chooses  $k$  observations from the input data and set them as the initial means (centroids). As a consequence, the  $k$ -means algorithm is not stable.

Meanwhile, it has been found that the worst-case running time for the  $k$ -means algorithm is super-polynomial to the size of the input data. Regarding these two drawbacks, many variations of the  $k$ -means algorithm have been developed by scientists, and amongst which are the  $k$ -means++ algorithm, the  $k$ -medoids algorithm and the Gaussian Mixture algorithm. Apart from those variants, the mean shift and DBSCAN clustering algorithms are also briefly discussed in what follows.



**Figure 4:** Different settings of the parameters of the four clustering algorithms and their corresponding performance in terms of the SSE Radius metrics when applied to the crossings at all airspace layers for 30 independent runs.

#### b) $k$ -means++ Clustering Algorithm

In data mining, k-means++ is an algorithm for choosing the initial values (or "seeds") for the k-means clustering algorithm. It was proposed in 2007 by David Arthur and Sergei Vassilvitskii, as an approximation algorithm for the NP-hard k-means problem—a way of avoiding the sometimes-poor clusterings found by the standard k-means algorithm. It is similar to the first of three seeding methods proposed, in independent work, in 2006 by Rafail Ostrovsky, Yuval Rabani, Leonard Schulman and Chaitanya Swamy. The k-means++ algorithm addresses the second of these obstacles by specifying a procedure to initialize the cluster centers before proceeding with the standard k-means optimization iterations. With the k-means++ initialization, the algorithm is guaranteed to find a solution that is  $O(\log k)$  competitive to the optimal k-means solution.

### c) **k-medoids Clustering Algorithm**

The k-medoids problem is a clustering problem similar to the k-means. The name was coined by Leonard Kaufman and Peter J. Rousseeuw with their PAM algorithm. Both the k-means and k-medoids algorithms are partitional (breaking the dataset up into groups) and attempt to minimize the distance between points labelled to be in a cluster and a point designated as the centre of that cluster. In contrast to the k-means algorithm, k-medoids chooses actual data points as centers (medoids or exemplars), and thereby allows for greater interpretability of the cluster centres than in k-means, where the centre of a cluster is not necessarily one of the input data points (it is the average between the points in the cluster). Furthermore, k-medoids can be used with arbitrary dissimilarity measures, whereas k-means generally requires Euclidean distance for efficient solutions. Because k-medoids minimizes a sum of pairwise dissimilarities instead of a sum of squared Euclidean distances, it is more robust to noise and outliers than k-means.

k-medoids is a classical partitioning technique of clustering that splits the data set of  $n$  objects into  $k$  clusters, where the number  $k$  of clusters assumed known a priori (which implies that the programmer must specify  $k$  before the execution of a k-medoids algorithm). The "goodness" of the given value of  $k$  can be assessed with methods such as the silhouette method. The medoid of a cluster is defined as the object in the cluster whose average dissimilarity to all the objects in the cluster is minimal, that is, it is a most centrally located point in the cluster.

### d) **Gaussian Mixture Models**

Clustering Algorithm Gaussian Mixture Models (GMMs) assume that there are a certain number of Gaussian distributions, and each of these distributions represent a cluster. Hence, a Gaussian Mixture Model tends to group the data points belonging to a single distribution together. Gaussian Mixture models work based on an algorithm called Expectation-Maximization, or EM. When given the number of clusters for a Gaussian Mixture model, the EM algorithm tries to figure out the parameters of these Gaussian distributions in two basic steps. The E-step makes a guess of the parameters based on available data. Data points are assigned to a Gaussian cluster and probabilities are calculated that they belong to that cluster. The M-step updates the cluster parameters based on the calculations from the E-step. The mean, covariance, and density are calculated for clusters based on the data points in the E step. The process is repeated with the calculated values continuing to be updated until convergence is reached.

#### 2.13.4 Mean Shift Clustering Algorithm

Basic mean shift clustering algorithms maintain a set of data points the same size as the input data set. Initially, this set is copied from the input set. Then this set is iteratively replaced by the mean of those points in the set that are within a given distance of that point. By contrast, k-means restricts this

updated set to  $k$  points usually much less than the number of points in the input data set, and replaces each point in this set by the mean of all points in the input set that are closer to that point than any other (e.g. within the Voronoi partition of each updating point). A mean shift algorithm that is similar then to  $k$ -means, called likelihood mean shift, replaces the set of points undergoing replacement by the mean of all points in the input set that are within a given distance of the changing set. One of the advantages of mean shift over  $k$ -means is that the number of clusters is not pre-specified, because mean shift is likely to find only a few clusters if only a small number exist. However, mean shift can be much slower than  $k$ -means, and still requires selection of a bandwidth parameter.

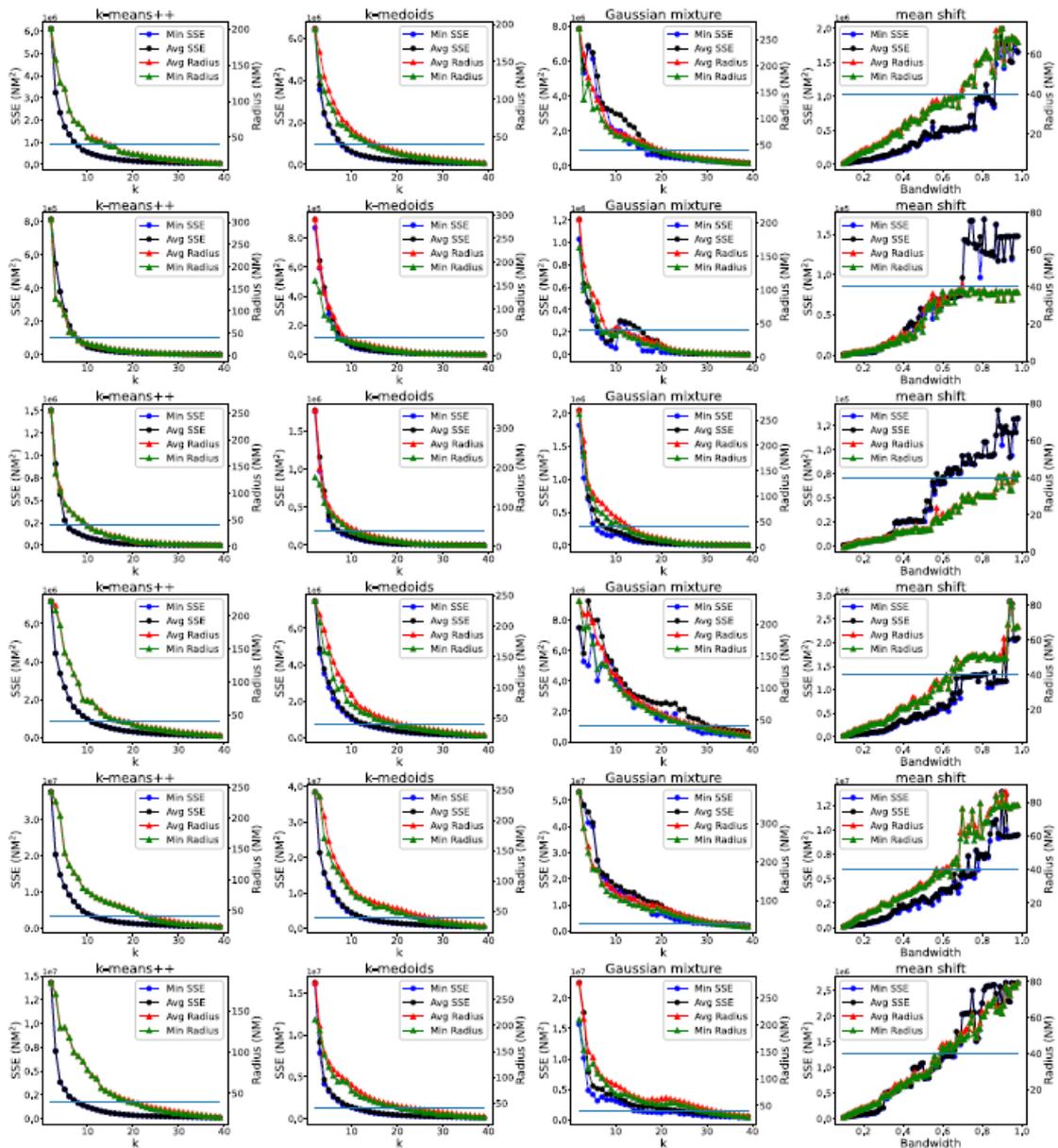


Figure 5: Different settings of the parameters of the four clustering algorithms and their corresponding performance in terms of the SSE and the Radius metrics when applied to the crossings at airspace layers L1~L6 (by row) for 30 independent runs

2.13.5 DBSCAN Clustering Algorithm

Density-based spatial clustering of applications with noise (DBSCAN) is a data clustering algorithm proposed by Martin Ester, Hans-Peter Kriegel, J Sander and Xiaowei Xu in 1996. It is a density-based clustering non-parametric algorithm: given a set of points in some space, it groups together points that are closely packed together (points with many nearby neighbors), marking as outliers points that lie alone in low-density regions (whose nearest neighbors are too far away). DBSCAN requires two parameters:  $\epsilon$  (eps) and the minimum number of points required to form a dense region (minPts). It starts with an arbitrary starting point that has not been visited. This point's  $\epsilon$ -neighborhood is retrieved, and if it contains sufficiently many points, a cluster is started. Otherwise, the point is labeled as noise. Note that this point might later be found in a sufficiently sized  $\epsilon$ -environment of a different point and hence be made part of a cluster. If a point is found to be a dense part of a cluster, its  $\epsilon$ -neighborhood is also part of that cluster. Hence, all points that are found within the  $\epsilon$ -neighborhood are added, as is their own  $\epsilon$ -neighborhood when they are also dense. This process continues until the density-connected cluster is completely found. Then, a new unvisited point is retrieved and processed, leading to the discovery of a further cluster or noise.

2.13.6 Performance Evaluation Metrics for Hot Spot

a) The SSE Metric

In statistics, the residual sum of squares (RSS), also known as the sum of squared estimate of errors (SSE), is the sum of the squares of residuals (deviations predicted from actual empirical values of data). It is a measure of the discrepancy between the data and an estimation model, such as a linear regression. A small SSE indicates a tight fit of the model to the data. It is used as an optimality criterion in parameter selection and model selection. As for the crossing point clustering, the SSE with respect to a given clustering result is calculated as follows:

$$SSE = \sum_{i=1}^k \sum_{j=1}^{k_i} \|c_i - x_j\|^2 \tag{7}$$

where  $c_i$  is the centroid of the  $i$ -th cluster  $C_i$  with  $x_j$  being the  $j$ -th crossing point within cluster  $C_i$  containing  $k_i$  points, and  $\|\cdot\|$  represents the geographical distance (great circle distance between two GPS locations).

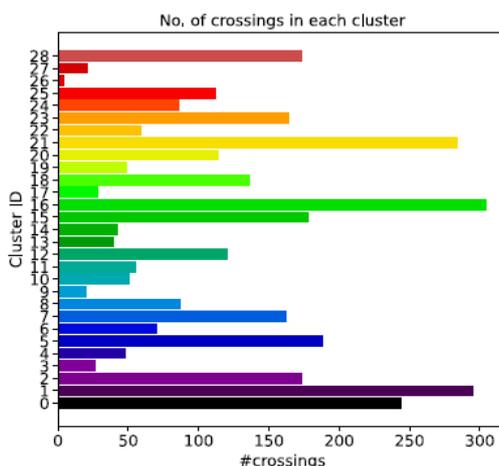
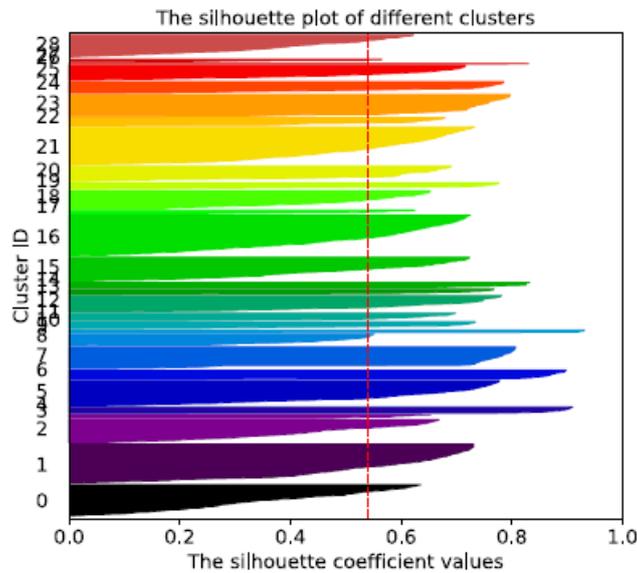


Figure 6: Sum of the squares of residuals (deviations predicted from actual empirical values of data).

b) The Silhouette Metric

Elbow Method is an empirical method to find the optimal number of clusters for a dataset. In this method, we pick a range of candidate values of  $k$ , then apply K-Means clustering using each of the values of  $k$ . Find the average distance of each point in a cluster to its centroid, and represent it in a plot. Pick the value of  $k$ , where the average distance falls suddenly. Silhouette refers to a method of interpretation and validation of consistency within clusters of data. The technique provides a succinct graphical representation of how well each object has been classified.

The silhouette value is a measure of how similar an object is to its own cluster (cohesion) compared to other clusters (separation). The silhouette ranges from  $-1$  to  $+1$ , where a high value indicates that the object is well matched to its own cluster and poorly matched to neighboring clusters. If most objects have a high value, then the clustering configuration is appropriate. If many points have a low or negative value, then the clustering configuration may have too many or too few clusters. The silhouette can be calculated with any distance metric, such as the Euclidean distance or the Manhattan distance.



**Figure 7:** Silhouette plot of different clusters.

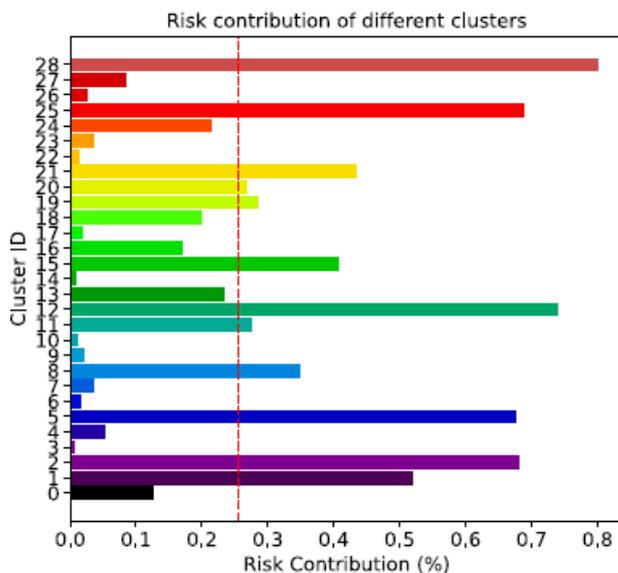
### c) The Radius Metric

Keep in mind that the clustering task is to cluster the crossing points into groups each of which could be a possible hot-spot. A hot-spot cannot be too large in its airspace volume. This is because that a hot-spot with a size of an airspace sector does not make any sense to decision makers as it gives no insights for airspace design and ATFM. Similarly, a hot-spot cannot be too small in its size. For example, a hot-spot at a waypoint level helps nothing as crossing exactly at a waypoint is a rare event. Considering the above factors, here we introduce the radius metric. For a given cluster  $C_i$ , its radius is calculated as follows:

$$\text{Radius} = \max \{ r_i = \|c_i - x_j\|, j \in [1, k_i] \} \quad (8)$$

A clustering algorithm can output a set of clusters of the crossing points. However, it is not necessary for each cluster to be a collision risk hot-spot. Even a dense cluster, i.e., a cluster with many crossing points, may not need to be a hot-spot. In order to identify the possible hot-

spots, we first calculate the collision risk of the clusters by summing up the collision risk of the crossing points within the clusters. If the collision risk of a cluster satisfies the criterion given in Eq. 2, then the cluster is identified as a hot spot.



**Figure 8: Risk contribution of different clusters.**

2.13.7 Final Hot-spot Visualization

Finally, after determining the hotspots, the hotspots can be visualised as polytopes with different filled color based on the intensity factor of the hotspot. The intensity factor defined as the risk contribution of each cluster divided by the average risk contribution of all the clusters. A higher intensity factor indicates a darker filled colour in the polytope. This is work in progress

2.13.8 Deliverables and timelines

The work-packages start is concurrently depending upon the architecture of the software and related activities in respective work packages. Given the nature of the project, there can be significant overlaps in the work packages, which may require the output of one work package as input to others. Each work package is not envisioned in a silo but as interacting and influencing components. *Appendix A* shows the deliverable items with the planned delivery date.

**3. ACTION BY THE MEETING**

3.1 The meeting is invited to note the information in this IP.

-----

	Item / Module / Tool name	Action	Status	Delivery date
1	BADA Database (Aircraft Performance Tables)	New aircraft data have been combined with BADA performance table and imported to the Program	Tested with new aircraft	Week No. (2-4) April 2022
2	Update Aircraft Dimensions file	Update Aircraft Dimensions Data is sourced from FAA.	Aircraft Dimensions file is updated. Now there are total 163 aircraft including previous types.	
3	ASE Mapping (Updated every 2 year from (Eurocontrol))	Request sent to Eurocontrol for the latest up-to-date Files	Waiting for ASE Mapping and ADD files	<b>Expected Week No (4) of Sep. 2022</b>
4	Assigned Altitude Deviation Sample Data (Eurocontrol)	Assigned Altitude Deviation Sample Data		
5	Removal of Procedural Waypoints	Procedural and mapping removed.	Tested	Week No. (1-4) of May & June 2022
6	Errors in Traffic data records	Updated Error Summary	On-going process	
7	Simulation speed controller with speed scaling option and zoom pan.	Code changed and updated	Tested	
8	Auto Upload MIDRAS Session files	Code is Updated	Tested	Week No. (1-4) of July & August 2022
9	MIDRAS New Database Architecture.	Data modules integrations	On-going process and testing	
10	<b>Collision Risk Graphical Analysis</b>	<b>New functionality to be developed and tested</b>	<b>On-going process</b>	<b>Week No. (1-4) of Sep 22</b>
11	<b>Congestion Analysis and Visualization</b>			
12	<b>Plot TOD points</b>			
13	<b>Airways occupancy Statistics plotting</b>	<b>New functionality to be developed and tested</b>	<b>-</b>	<b>Week No. (1-4) of Oct 22</b>
14	<b>Hot -Spot region Visualization</b>			
15	<b>Help File Update</b>			