Journal of Aerospace Engineering, Volume 31, Issue 1, January 2018
DOI:10.1061/(ASCE)AS.1943-5525.0000796

Finding precursory ATM Safety metrics using Exploration of Trajectory radar tracks

Rocío Barragán Montes; Fernando Gómez Comendador;
Francisco Javier Sáez Nieto; and Luis Pérez Sanz

Authors

Rocío Barragán Montes (rocio.barragan@upm.es), Assistant Lecturer at Technical University of Madrid (UPM) -ETSIAE, Pza. Cardenal Cisneros 3, 28040 Madrid, Spain (corresponding author).

Dr. Fernando Gómez Comendador (fernando.gcomendador@upm.es), Professor at Technical University of Madrid (UPM) - ETSIAE, Pza. Cardenal Cisneros 3, 28040 Madrid, Spain.

Dr. Francisco Javier Sáez Nieto (p.saeznieto@cranfield.ac.uk), Reader at Cranfield University, Cranfield MK43 0AL, Bedfordshire, United Kingdom.

Dr. Luis Pérez Sanz (l.perez@upm.es), Professor at Technical University of Madrid (UPM) - ETSIAE, Pza. Cardenal Cisneros 3, 28040 Madrid, Spain.

Abstract

The definition of a set of precursory safety metrics is critical to detect when an airspace is degrading in terms of safety and thus undesired effects are more likely. Furthermore, safety metrics are paramount to the measurement of the impact of new operational procedures or technical improvements on the Air Traffic Control system. The study presented in this paper introduces three safety metrics (Reaction Time Performance Indicator, Time to Closest Point of Approach Performance Indicator, and Time to Closest Point of Approach Critical Limit Ratio) derived from a given airspace and a sizable, assorted traffic sample extracted from traffic surveillance track data. The metrics are used to characterize the airspace as a function of the safety outcome, which can be continuously oversighted. The final goal of the safety metrics is to be used as an airspace safety warning system, where precursory metrics would signal the need to act to maintain the Air Traffic Control system safety target on the face of operational, organizational, technical and/or legal changes.

Introduction

The seemingly correlation between Air Traffic Controllers (ATCo)'s reaction times upon the detection of a possible Loss of Separation (LoS) minima and the time until it is avoided, appears to be a measure which can bring information about the safe behavior of the Air Traffic Management (ATM) system. Were time remaining until the incident long or did it take little to solve upon detection, the system would be performing adequately in terms of safety. Additionally, in a standardized airspace, once the LoS is detected ahead, times to the closest point of approach at the moment of the avoidance maneuver should remain bounded, regardless of the ATCo's providing service, the active sector configuration, the traffic density or any other factor.

Air Traffic Control (ATC) has an impact on the flight trajectories: non-programmed maneuvers in the flight plan correspond to clearances and actions taken by ATC on the ground to synchronize trajectories and avoid losses of separation among aircraft. The goal of this research work is double: first, establishing that reaction times and times remaining to closest point of approach assessed using radar tracks vary depending on traffic circumstances; and second, determining that those times until the potential conflict fluctuate according to factors which affect safety. Therefore, some metrics derived from flight trajectories could be provided such that they could alert about a system degradation in terms of safety.

The need of objective, measurable and independent metrics to assess safety has triggered the research question which is: **Can precursory safety metrics be inferred from post-flight trajectories extracted from radar tracks?**

Definitions

Some of the terms used in this paper are defined as follows:

Conflict: Actual LoS minima between two aircraft in both dimensions, vertical and horizontal.

Potential conflict: Any situation where a LoS ahead is foreseen unless a corrective change is provided. It will last for as long as the action taken, which changes the evolution of any of the two aircraft involved, provides the required separation minima.

It is important to point out here, that Potential conflict in this paper is equivalent to the term Conflict defined in (ICAO 2005) Section 2.7. Conflict Management, as "any situation, involving aircraft and hazards, in which the separation minima may be compromised". This convention has been followed to continue with the same nomenclature as used in previous referenced papers.

Closest Point of Approach (CPA): Physical moment when the minimum distance between two aircraft is reached.

Look Ahead Time (LAT): Exploratory time in search of potential conflicts.

Initial Time To CPA (iTTCPA): Time to the CPA upon the detection of a potential conflict. Were two flights following a uniform movement on conflict course, iTTCPA would be equal to LAT. However, some maneuvers during the flight provoke potential conflicts between aircraft in a temporal horizon smaller than the LAT. Thus, it can be stated that, for each potential conflict: $iTTCPA \leq LAT$.

Reaction Time (RT): Time elapsed between potential conflict detection and the perceivable change in the trajectory uniform movement. RT comprises the following time lapses since the potential conflict detection: ATCo recognition, analysis of possible resolutions, most suitable action decision, instruction issuing, crew taking it and aircraft evolution, changing the current drift and avoiding being on conflicting course any longer. Hence, the change of trajectory in terms of direction, speed or rate of climb or descend will be considered the "reaction moment".

Time To CPA (TTCPA): Remaining time to the CPA upon the corrective tactical action is completed. Due to the definitions of terms, it is true that: $iTTCPA = RT + TTCPA$.

Fig. 1 represents the relationship between the last three concepts which will be thoroughly used during this paper. At t_0 a potential conflict between Aircraft 1 and Aircraft 2 is identified. If no corrective change is provided, the CPA will occur at t_{CPA}, and hence, the initial time to CPA is calculated as $(t_{CPA} - t_0)$ and so-called $iTTCPA$. At t_1 the corrective tactical action is completed changing the drift of Aircraft 1. The latency time between t_0 and $t_1$is the RT $(RT = t_1 - t_0)$ while the remaining time from t_1 to t_{CPA} is $TTCPA$ $(TTCPA = t_{CPA} - t_1)$.

Initial considerations

Taking a further look to the previously defined concepts of RT and TTCPA, when observing a sample of stored radar tracks for a given airspace, the sample exhibits a well-defined distribution of the RT with respect to TTCPA when it resembles the one shown in Fig. 2. As expected, the majority of potential conflicts are solved far from 0 min of TTCPA and, although it takes some time (RT) to solve them, there are just a few whose TTCPA falls below 1 min. In terms of safety, it can be considered that the worst potential conflicts would be those whose TTCPA is below 1 min. Additionally, if they also exhibit high RT, it means a long latency of the hazard (potential conflict); although there was enough time to take a positive corrective action, it was not taken until the actual LoS was on the verge to occur.

A more stressed example of the same airspace is shown in Fig. 3, where the sample shows the peak distribution of TTCPA falling 1 min below the previous Fig. 2.

A decrease in the TTCPA average does not necessarily mean a flaw in safety. Although risk probability increases as corrective actions are postponed in a congested airspace or degraded conditions for instance, when more potential conflicts arise, the tactical tasks for the controllers increase and the potential conflicts cannot be solved as fast as in a quieter environment. Nevertheless, if they are managed quickly and in a normalized way, airspace should not be considered unsafe.

Human beings tend to procrastinate and delay taking actions as long as the situation is completely under control (Ferrari 2010). Not only may procrastination cause a delay in the issuing of ATC clearances, but also the balance between flight efficiency and capacity requirements impacts on ATCo's RT's. Finally, ATCo's may delay the tactical actions while maintaining the safety levels when considering the uncertainties associated to trajectory prediction and conflict detection. Some potential conflicts may arise with a large time to conflict and if the traffic is being managed with ease, the action to solve it may be delayed. Thus, long RT's do not necessarily mean a decrease in safety.

This research assesses the relationship between the TTCPA's and RT's distribution vs. the airspace and traffic spatial-temporal characterization based on stored surveillance tracks. It is here postulated that safety performance has an impact on TTCPA's and RT's distribution for a given scenario. Then, the backwards effect is outlined: overseeing those distributions it will be possible to trigger a precursory safety metric.

This paper is structured in six sections: first, this "Introduction" section which brings information about the terms used and the main ideas that trigger this research work. Second, a "Literature review" section which shows the main industry motivation and other scientific work previously published on this topic. Then, a "Methodology and Tools" section provides information about the process followed, the software and statistical tools utilized, data availability and main operational conditions and assumptions considered. "Data Exploration" section specifies the analysis of the random variables studied, trying to find patterns of behavior, individually and jointly. The conclusions of the section drive the formulation of hypotheses and metrics proposition, presented in "Proposed new Precursory Safety Metrics" section. Finally, the "Verification exercise" section includes the validation of the metrics using the data available and assessing if their outcomes are consistent with the airspace daily situation.

Literature review

Several recommendations from International Civil Aviation Organization (ICAO) (ICAO 2013), (ICAO 2013) and European Aviation Safety Agency (EASA) (EASA 2014) point at shifting the traditional approach to the assessment of safety performance in ATM. The current prescriptive rules accounting for incidents and accidents reported by ATM actors achieve good results when the number of incidents per flying hour is small. It has been widely used for the last years and published in (SESAR-JU P16 01 01 2014), using and Accidents and Incidents Model (AIM) based on (Perrin, et al. 2007) Integrated Risk Picture (IRP) model. That appraisal seems now a subjective and non-standardized performance assessment upon which it is difficult to take proactive measures to be shared among States. Thus, it is acknowledged in (Eurocontrol Performance Review Commission 2016) and the yearly publication (Eurocontrol 2016). In the latter, along with (EASA 2014), key concepts of this approach are described, like Risk-Based Oversight (RBO) and Performance-Based Oversight (PBO), which are designed within a Performance Based Environment (PBE).

The current dependency of the reported safety events on the involved actors, makes the present Key Performance Indicators (KPI's) subjective while they should be independent, measurable and objective.

Safety assessment based on available radar tracks determining the position and time of potential conflicts, potential collisions and safety barriers was presented in the thesis dissertation by former UPM professor (García González 2013) and disseminated in the following papers:

- In (García González, Sáez Nieto and Izquierdo 2007), trajectory segmentation based on radar tracks to foresee safety events was presented. Also, the theory of potential conflicts assessed by its time to Closest Point of Approach (CPA) instead of distance was divulged. Thus, aircraft evolution became a parameter for the conflict estimation.
- In (Sáez Nieto, Arnaldo Valdés, et al. 2010) time of detection of the potential conflict and ATCo's reaction time applied to safety assessment were introduced in the development of a three-dimensional Collision Risk Model (CRM) using Matlab®: it was called 3D-CRM. Moreover, the algorithms to detect the tactical actions taken by ATCo's from surveillance systems were defined, accounting ATCo's tactical intervention, and the appearance of alert systems: Short Term Conflict Alert (STCA), Traffic Collision Avoidance System (TCAS) both Traffic Advisory (TA) and Resolution Advisory (RA), based on the time remaining until the potential conflict. The tactical action taken on the aircraft in en-route phase of flight was determined by the beginning of a new

segment after the detection of the potential conflict. This paper grouped the metrics in two categories:

- Risk context metrics, which provide information about the events that could drive to potential conflicts;
- Safety metrics, which indicate effectiveness and the stress level of safety barriers.

A validation of 3D-CRM was included in the referenced paper with data from one month of traffic from Maastricht Upper Airspace Centre (MUAC). 3D-CRM has been adapted for the work presented in this paper as explained below.

Other papers related to this topic are:

- In (Cózar Maldonado, Sáez Nieto and Ricaud Álvarez 2015) a further step in segmentation was presented, approaching radar tracks with mixed types of straight lines and curved segments, using cubic spline functions. The data used in this work were radar tracks from Madrid Terminal Maneuvering Area (TMA).
- Another approach to safety assessment was presented in (Pozzi, et al. 2011). Based on Eurocontrol's previous work on Automatic Safety Monitoring Tool (ASMT), carried out by either big data analysts or ATM experts, it introduced the requirement of an intermediate viewpoint between ATM and data analysis experts, the so-called 'Information Design Perspective', acknowledging that the future of Safety Analysis relies on translating data into information and information into knowledge.

These research work are the baseline from where the current research starts, taking advantage of the available material but aiming at ICAO's objectives related to safety metrics, shifting from traditional safety performance indicators based on risk probabilities to PBO.

Methodology and tools

Methodology

For a better understanding of variables RT and TTCPA, a mathematical data exploration has been carried out. The search of correlation, patterns, relationships among variables, etc., under the most feasible causes of safety degradation is presented in the "Data Exploration" section and constitute the first step of the methodology. Although this analysis did not lead to the expected results under the most stressing traffic conditions, a common distribution shape of two derived random variables was observed.

The data analysis does enable the formulation of a set of hypotheses which are introduced in the "Proposed new Precursory Safety Metrics" section. This hypotheses are founded on the Information Design Perspective as defined in (Pozzi, et al. 2011). It provides the definition of three suggested performance indicators which create an airspace classification in terms of its safety performance.

Finally, the hypotheses and metrics are tested in the "Verification exercise" section, where everyday situation is matched against the metrics outcomes, validating if the performance indicators do conclude consistent results.

Data available for the assessment

Post-flight trajectory information for academic research is a scarce resource as it is commercially sensitive for Airlines and Air Navigation Service Providers (ANSP)'s. The accessible information is the set of multi-radar tracks provided by Maastricht Automated Dataprocessor (MADAP) derived from Primary and Secondary Surveillance Radar stations (PSR) and (SSR). Thus, the data provided is the set of filtered points p for all flights f detected and calculated by the radar system every 4.8 seconds:

$$\{(t_{fp,}, x_{fp}, y_{fp}, h_{fp}, v_{fp}, \dot{h}_{fp}), \forall p, \forall f\}$$

where

t_{fp} is the time at which the point p of the flight f is detected,

(x_{fp}, y_{fp}) are the 2D-coordenates of the flight f at t_{fp},

h_{fp} is the altitude of the aircraft at t_{fp},

v_{fp} is the 2-D speed vector calculated by the surveillance system at t_{fp},

$\dot{h}_{fp}$ is the vertical speed calculated by the surveillance system at t_{fp},

The inception of this research work is the set of radar points from all flights overflying MUAC airspace, over Flight Level (FL) 245 during seven days of January 2007. A sample of 8,793 potential conflicts has been identified and analyzed.

Tools

Software tools: 3D-CRM

Present research uses algorithms previously developed in the doctoral theses (García González 2013) and (Cózar Maldonado 2015). In their research, starting from a set of radar points defined as in the previous section, the named authors developed two CRMs. Both analyzed aircraft pairs in course collision but, whilst the latter was implemented for TMA, the former was applied to en-route airspace using the previously introduced 3D-CRM software tool. The requirements met by this tool are the identification of all proximate events based on radar data, a complete classification of all proximate events using clear and consistent criteria, obtaining detailed information on the evolution of each proximate event and calculating a collision risk estimate.

The tool bases the analysis of radar tracks by rebuilding the aircraft trajectories as an ordered sequence of segments. Thus, for aircraft number i, its flight is determined by the following set of n segments:

$$R_i = \{[t_{i0}, x_{i0}, y_{i0}, h_{i0}, v_{i0}, \dot{h}_{i0}], [t_{i1}, x_{i1}, y_{i1}, h_{i1}, v_{i1}, \dot{h}_{i1}], \ldots, [t_{in}, x_{in}, y_{in}, h_{in}, v_{in}, \dot{h}_{in}]\}$$

where

t_{ik} is the starting time of the segment k,

(x_{ik}, y_{ik}) are the 2D-coordenates of the aircraft at t_{ik},

h_{ik} is the altitude of the aircraft at t_{ik},

v_{ik} is the 2-D speed vector, which is supposed constant at all the points of the segment k,

$\dot{h}_{ik}$ is the vertical speed, which is supposed constant at all the points of the segment k,

(García González 2013), Annex B, analyzes the accuracy of the radar traces transformation into trajectories through this method. A sliding window is used to smooth the trajectory and filter out data errors. The sliding window is small enough to detect existing short segments of flight, but must content a significant number of data samples to infer a homogeneous behavior. It was proven in the cited publication above through a sensitivity analysis, that optimum sliding window size in the vertical segmentation, i.e., the minimum number of errors was 43 seconds. In the horizontal plane, using the defined 43 seconds sliding window, the maximum deviation in both, longitudinal and lateral dimensions, was 400m.

Along time, a projection of the current position over a determined LAT is computed assuming uniform movement throughout each segment, exploring for potential conflicts with the rest of flying aircraft.

Although 3D-CRM keeps all the information regarding each potential conflict, for this research only times and regions in which it occurs have been accounted.

Statistical tools: Box & Whiskers and Bags & Bolsters

With the objective to infer results from the pairwise RT and TTCPA observations, different ways of representing the information have been developed.

The classical 'Box-and-Whiskers' plot has been used to explore the behavior of the random variables RT and TTCPA for different spatial-temporal scenarios. Their median and spread have been studied independently.

The concept of bagplot (Rousseeuw, Ruts and Tukey 1999) was introduced and called as 'bag-and-bolster' plot as an analogy to 'Box-and-Whiskers' plots for bi-dimensional representations. It is based on Tukey's half space location depth, $ldepth$, presented in (Tukey 1975) which is a measure of inversely distance between any point in the $\mathbb{R}^2$ domain and the center of the sample. To better represent the skewness and spread of the sample and evaluate the half space location depth, (Rousseeuw, Ruts and Tukey 1999) recommends that both dimensions should follow a normal distribution.

Let us define convex sets of samples $D_k = \{z_i: ldepth(z_i, Z) \geq k\}$ being $Z = \{z_1, z_2, \dots, z_n\}$ the set of 2-dimensional samples. Note that the definition of D_k implies that $D_{k+1} \subseteq D_k\ \forall k > 0$. The depth median would be the z_m with the highest $ldepth$, i.e., $D_{m+1} = \emptyset$, while the bag of the sample, equivalent to the box of the 'Box-and-Whiskers' graph, is: $D_{n/2} = \{z_i: ldepth(z_i, Z) \geq n/2\}$. The bag is plotted darker than the rest of data and it is enclosed in a borderline. The fence is calculated following (Rousseeuw, Ruts and Tukey 1999), multiplying the bag relative to the depth median by a ρ factor of 3. Some authors calculate the fence with other ρ factors, for instance (Hyndman and Shang 2010). In that case, the ρ value of 2.58 was chosen as it was the one that best fitted a projected bivariate score following a normal distribution, which is not the case of the present study. The fence is plotted in lighter color and does not have a borderline. The points outside the fence are set as outliers. See example shown in Fig. 4.

Main applicable operational conditions and assumptions

The lack of correlated flight plans on the stored surveillance tracks or the lack of evidence of ATCo's intervention, have driven the assumption of considering that any significant change in the aircraft uniform trajectory has been triggered by an ATCo's corrective tactical action -this assumption was justified in (Sáez Nieto, Arnaldo Valdés, et al. 2010) and both thesis dissertations (García González 2013) and (Cózar Maldonado 2015).

The en-route Reduced Vertical Separation Minima (RVSM) airspace subject to this research rules its separation minima as 5 NM in the horizontal dimension and 1,000 ft (10 FL's) in the vertical dimension. Conventional mode C radar provides tracks altitude with a 100 ft. resolution (ICAO - APAC 2007). Therefore, the existence of a potential loss of separation in vertical plane is detected when the vertical separation is lower than 850 ft., avoiding the detection of crossing flights at contiguous flight levels as potential conflicts.

Aligned with the work previously developed in (Sáez Nieto, Arnaldo Valdés, et al. 2010), based on operational factors, the LAT has been established in 10 minutes. It is neither too small to account most of ATCo's corrective tactical actions nor too long to create false potential conflict detections.

For each potential conflict, the RT interval has been computed from the time at which the potential conflict was detected, i.e., t_0 in Fig. 1, until the time at which the evolution of the aircraft has changed and there is no longer a potential conflict ahead, i.e., t_1 in Fig. 1. It has been assumed that the change is driven by an ATCo tactical action regardless of the condition that could have triggered it. However, by considering the expected latency between the potential conflict identification and the implementation of the issued ATCo instruction, it has been assumed that RT, can never be smaller than 10 secs. Changes in the trajectory happening within less than 10 secs upon the potential conflict detection have been discarded. Therefore, the moment chosen to be considered the potential conflict resolution is the subsequent change in the evolution of the aircraft.

Data exploration

Reaction Time distribution

The RT of processes involving Human-Automation Interaction which entails information display, situation assessment, decision making and action taking, follows approximately a log normal distribution, as stated in (Sheridan 2006). An example of the distribution of reaction times for day 1 is shown in Fig. 5, which substantially follows a log normal distribution.

To mathematically confirm the log normal distribution of RT for each day, a Kolmogorov-Smirnov test was carried out on the random variable "logarithm of the RT". The posing of the test was:

$$H_0: log(RT) follows\ a\ normal\ distribution$$

The results for the tests are shown in

Table **1** for every day dataset. The hypothesis tests results failure to reject the null hypothesis at a significance level α of 0.01 for every day dataset but for day 6. Graphically, the normal fitting of log(RT) is shown in Fig. 6 for day 1. These results can be taken as "generally log(RT) follows a normal distribution" which is sufficient for the sake of its exploration using Bags & Bolsters graphs and the calculation of depth distance specifically.

TTCPA Distribution

Analogously, it has been also proven that random variable "TTCPA" generally follows a normal distribution. The results of the Kolmogorov-Smirnov test applied on every day data are shown in Table 2, and graphically displayed in Fig. 7.

Thus, stored raw data were used to compute the random variables RT, TTCPA and log(RT) and then used as coordinates in the different plot representations.

RT and TTCPA representation

Statistical representations were obtained to assess and evaluate the behavior of the random variables RT and TTCPA, both independently and pairwise. The evolution of the mean and median of each of them were compared under different situations: high density and hotspots regions, different hours of the day or grouping potential conflicts into different time of detection windows.

The retained information from each set of plots contains bi-dimensional and unidimensional representations for both random variables (TTCPA and RT, or log(RT)) for each situation to ease the analysis of the relationship between these variables and the conditions of the scenario. In Fig. 8 the different graphs are showing:

- The first upper plot presents the bi-dimensional histogram of events having RT and TTCPA pairs for the identified potential conflicts among the selected data.
- The histograms of events containing RT and TTCPA are displayed in upper plots 2 and 3.
- The fitted probability density function of the distribution for log(RT) is shown in the first lower plot, including its characteristics parameters, μ and σ. Not only is its expectation μ important, but also the variance σ provides information about the spread of the variable log(RT).
- Classical Box and Whiskers is placed next. The median of the data is marked with the straight horizontal red line while the mean is marked with a green diamond. Thus, 50% of data, what from now on will be called "range", is included in the boxes, between 25% and 75% of the samples and 90% of data are contained between the whiskers. Finally, outliers are red crosses out of the boundary of the whiskers. Note that both random variables behavior is represented, one for RT and one for TTCPA.
- Finally, the last lower plot on the right shows the Bags and Bolsters graph, with the depth median marked with a cross, the bag highlighted in dark blue and the fence surrounding it in light blue. The metrics for the bag area, bag width, bag height, fence area and fence width have been included in the plot. As stated in previous sections, to better show the distribution, the variable presented

on abscise is in this case log(RT). While bag height units do not need to be transformed, as they are expressed in minutes, for a better understanding of the results, the bag and fence widths information is transformed into minute's interval calculating $e^{\log(\max(RT))} - e^{\log(\min(RT))}$ for the 50% interval data sample and the fence size sample, respectively. Bag and Fence sizes are calculated in two dimensional minutes times log(minutes). By convention, the units used in this paper for bag and fence areas are expressed as [area units].

The same type of representation is contained in Fig. 9, providing the results, in this case for day 4, over the high-density area, defined by the red rectangle displayed in Fig. 10 and for the highest levels of traffic occurred between 07:00 and 10:00. The density map highlights white stars as points where the potential conflicts for that time interval occurred. Matching the density map information with the traffic flows for the day 4 depicted in Fig. 11, it is shown that the potential conflicts tend to concentrate in the highest density areas, where north-south traffic flow crosses east-west traffic flow, marked as red and pink 7.5NM colored density squares.

Exploration conclusions

An assorted set of scenarios have been analyzed, and the following initial conclusions have been drawn. When traffic increases:

- The RT's tend to be smaller, and
- The TTCPA's tend to shorten.

For instance, to illustrate this, by comparing the first lower plots on log(RT) shown in Fig. 8, for all day 4 data, and Fig. 9, for peak hour and high density region data, it can be seen how both the expectation μ (-0.18898 vs. -0.35278), and the variance σ (0.708 vs. 0.51319), are smaller during the peak hour. That means that RT exhibits a clear tendency to decrease under higher traffic, in this example from 50 sec to 42 sec, whereas the variance for the RT distribution tend to be also reduced, having more potential conflicts solved around the mean, indicating that the controllers reaction upon the detection of potential conflicts under higher traffic conditions tends to be more standardized.

This reasoning is also supported by the second lower plots in the same figures Fig. 8 and Fig. 9. Accordingly, the RT boxplot shortens its height under supposedly more stressed conditions while the mean tends to decrease. Having a look at TTCPA box and whiskers plot, the median for the day is larger than the median for the peak hour, 5.2283 min vs. 5.0027 min. However, the range of TTCPA middle 50% sample, remains the same. In some other days and other peak hours, it was observed that results differed without a clear tendency, so no robust conclusions could be derived.

The last lower plot produces visual information about the bi-dimensional representation for RT and TTCPA samples: both bag and fence slenderize under high density (more stressed) conditions, which means RT range is smaller. It is then again, the same conclusion achieved by analyzing RT box and whiskers plot. Regarding the median value calculated following the bag characteristic dimensions, for the whole day it is (-0.1824, 5.342) and for the peak hour/region (-0.3698, 5.205), showing a slightly smaller TTCPA while RT clearly shortens. Other metrics extracted from the bagplot: in peak hour and high density region, bag area falls to 4.5977 [area units], whereas for the day average is 7.3442 [area units]; bag width falls to 0.89802 min from 1.7283 min; and bag height falls to 5.39 min from 5.5708 min.

More scenarios were analyzed to consolidate the above described clear behavior pattern of the key random variables under comparable scenarios. Similar hour, for same regions and different days did not produce the expected outcomes, akin to the presented above for day 4. Besides, small high density areas outcomes were strongly biased by outliers, whereas bigger regions did not show uniform behavior.

Even though it is not possible to extract firm conclusions from the high-density traffic airspace data, it can be affirmed that RT and TTCPA random variables change depending on factors to be determined. These results have brought some valuable information and the keys for further research by performing a more holistic and integrated approach using new random variables statistics to characterize the system behavior and its evolution.

Proposed new Precursory Safety Metrics

Formulation of Hypotheses

The relationship between the three statistics derived from the key random variables: (TTCPA and RT medians, and the size of the TTCPA range), is now assessed by considering the relative behavior of the last two (RT median and TTCPA range size) over the first one (median TTCPA). The random variable log(RT), used for the bags & bolsters analysis, to normalize the lognormal behavior of RT, is no longer needed after the conclusions obtained from the data exploration analysis.

Fig. 12 shows four integrated bar diagrams displaying the comparison from all available data median values for the TTCPA and RT and their ranges (50% of the central values of the samples). Fig. 13 displays the same comparison for the derived ratio RT median / TTCPA median and TTCPA range / TTCPA median values. iTTCPA is used to define 2-minute intervals to group random variables in both figures. Light blue bars stand for times above 8 minutes to CPA, orange bars represent those having between 6-8 minutes, grey bars those between 4-6, yellow bars those between 2-4 and finally, dark blue bars those having less than 2 minutes. These graphs are analyzed to expose how the ATCo's reaction time changes depending on how close to the CPA the potential conflicts were identified.

Fig. 12 shows that both, RT and TTCPA medians tend to decrease when the potential conflict is closer to the CPA. Although the TTCPA range decreases significantly, RT range does not follow a clear tendency. Moreover, Fig. 13 shows that the ratio between RT median and TTCPA median monotonically increases significantly when the potential conflict is getting closer to the CPA. Furthermore, the ratio between TTCPA range and TTCPA median rises as well, showing an exponential behavior. These results, built from medians, lead us towards three hypotheses:

- When the ratio between the reaction time median and the remained time to CPA median for given air traffic scenarios grows above a certain threshold, then a stressed situation has been reached. It would represent a situation in which RT's are large in average compared with typical TTCPA for the sample. It is highly unlikely that with a high TTCPA the RT is big, as $TTCPA + RT = iTTCPA \leq LAT$. Nevertheless, if the RT/TTCPA ratio is big having small TTCPA (i.e., potential conflicts solved near its CPA), it would mean a high RT compared to its TTCPA, which would identify a stressed situation.
- When the ratio between the TTCPA range and the TTCPA median for given traffic scenarios grows above a certain threshold, then a non-normalized situation has been reached. It would represent situations in which 50% of the variability in TTCPA's is significant compared with the TTCPA median. The normalized behavior of the airspace would be keeping a small variability when TTCPA median is small, while allowing a more spread variability when TTCPA median is large. Otherwise, the situation would be non-normalized.
- When the ratio between the range size and the remained TTCPA median for given air traffic scenarios grows above 1, then the above nominal stressed situation has been reached. This is a case of the previous scenario, in which there is larger variability than the value of TTCPA median. Whereas the previous hypothesis highlights non-normalized behavior through a significant variability in TTCPA with respect to TTCPA median, this limit features bigger variability in TTCPA than the TTCPA median. Samples over this critical limit mean that the disparity in the 50% mid values of TTCPA is bigger than the median of that sample. It points out that some

potential conflicts are brought close or equal to their CPA and therefore risky situations are more likely to occur.

The assessment of the evolution of the medians in the conflict resolution processes is dependent on the probability of concurrent proximate events which are more likely under high-traffic density. Additionally, the workload of an ATC sector at a particular time is closely related to its complexity (Suárez Tetzlaff, et al. 2014). Considering complexity by its Single European Sky ATM Research (SESAR) definition consolidated in (SESAR-JU 2016) as “the number of simultaneous or near-simultaneous interactions of trajectories in a given volume of airspace”, it is trivial to state that complexity increases with traffic density or occupancy. Thus, high density airspace and low density airspace are here analyzed separately. On the other hand, isolated high ratios of RT/TTCPA do not necessarily mean a stressed situation or lack of safety. Hence, stored potential conflict data have been grouped into two hours intervals to reduce the effect on statistics values being biased by outliers. Additionally, when these two hours intervals contain one sample, they have been dismissed.

Once the above consideration has been applied, the ratio RT median / TTCPA median vs. TTCPA median for all available data is shown in Fig. 14. The parameters for the interpolated exponential curve approach, for the set of points derived from gathered data, high density and non-high density regions, for each two-hour interval and performed for every day facts, is shown in Table 3. The last column addressed as Crossing Points contents the coordinates of the intersection of the TTCPA median line and the approached exponential curve. In an analogous manner, the ratio between TTCPA range / TTCPA median vs. TTCPA median for the set of points extracted from every day data, different density regions and the two-hour intervals, and its interpolated exponential curve approach is shown in Fig. 15, whereas the exponential fit parameters are detailed in Table 4. Additionally, in Fig. 15, a horizontal red line has been added to highlight the situation where the sample TTCPA range equals the sample median TTCPA when the potential conflict was solved. This value is here interpreted as the admissible limit ratio and it is considered as value 1. The data samples above it have been emphasized with red ellipses. They will be furtherly analyzed below.

Proposed metrics

RT Performance Indicator

As it has been shown above, the experimental results have been interpolated by an exponential curve for the ratio RT median / TTCPA median vs. TTCPA median. The shape of this exponential curve changes on each particular day, the traffic density, the time interval and possibly other factors.

Fig. 16 plots the dispersion diagram containing the median values for the pairs RT median / TTCPA median vs. TTCPA median derived from two-hour intervals data samples and different density regions. The correspondent fitted exponential curve is also included as a continuous dark blue curve, as well as the TTCPA median as a straight vertical black line. The intersection point of these two lines calculated based on all available data can be expressed as follows:

$$\begin{cases} x_D = median(TTCPA_D) \\ y_{RT,D} = a_{RT,D} \cdot e^{-b_{RT,D} \cdot x_D} \end{cases} \tag{1}$$

where $a_{RT,D}$ and $b_{RT,D}$ are the parameters of the exponential distribution fitted from all available data (D) dispersion diagram of the derived random variable: RT median / TTCPA median over TTCPA median.

Overlaid in Fig. 16, the different daily exponential approaches have been plotted using dashed lines, each one of them with a different color. Also daily TTCPA medians have been plotted with dashed lines in its correspondent daily color. Consider the daily representative "crossing point" as the intersection point between those two lines. The analytical expression for RT Crossing point is a pair in the 2D coordinates, expressed as:

$$\begin{cases} x_d = median(TTCPA_d) \\ y_{RT,d} = a_{RT,d} \cdot e^{-b_{RT,d} \cdot x_d} \end{cases} \tag{2}$$

where $a_{RT,d}$ and $b_{RT,d}$ are the parameters of the exponential distribution fitted from daily data (d) dispersion diagram of derived random variable: RT median / TTCPA median over TTCPA median.

The difference between (1) and (2) is that (1) considers D as all available data, describing the average scenario. The larger available dataset, the better approximation to an average behavior is obtained. Nevertheless, equation (2) is calculated based on a one day dataset (d) whose safety performance will be evaluated.

The hypothesis establishes that daily crossing points expressed by equation (2) may fall within one of these four areas:

1. When the crossing point falls into the right and upper side of the graph ($x_d > x_D, y_{RT,d} > a_{RT,D} e^{-b_{RT,D} \cdot x_d}$), to zone numbered as 1, it indicates that the average time to potential conflict has increased which can be interpreted as a safety gain. However, the ratio RT median / TTCPA median has also increased, even with higher TTCPA average, which indicates that, in average, the actions were not taken as expeditiously as in the reference conditions. This zone of the graph has been named as "Relaxed Zone".

2. If the crossing point falls into the right but the lower side of the graph ($x_d > x_D, y_{RT,d} < a_{RT,D} e^{-b_{RT,D} \cdot x_d}$), the zone 2, the TTCPA median is again above the reference while the RT's has shortened. It has been considered that the further the crossing point falls into this region, the better performance improvement. This zone is called "Favorable Zone".
3. When the crossing point moves towards the left and the lower part of the graph ($x_d < x_D, y_{RT,d} < a_{RT,D} e^{-b_{RT,D} \cdot x_d}$) to zone 3, it means that in average there was less time remaining to the potential conflict whilst RT's were small (due to the fact that RT median / TTCPA median is smaller than the average) Actions were taking expeditiously after the potential conflict was detected even though the system was more stressed. This region is called "Stressed Zone".
4. When the crossing point moves to the upper and left side of the graph ($x_d < x_D, y_{RT,d} > a_{RT,D} e^{-b_{RT,D} \cdot x_d}$), to zone 4, potential conflicts were solved over a longer period than average after being detected, because TTCPA was short and RT's did not remain small. This zone is called "Critical Zone".

In summary, the first metric proposed is "RT performance indicator", which evaluates the location of crossing points obtained from equation (2) with respect to the global dispersion diagram and zones delimited by its overall RT median / TTCPA median ratio exponential approach and TTCPA median, defined in equation (1).

TTCPA Performance Indicator

The previous analysis used RT median /TTCPA median as a dependent random variable vs. TTCPA median. It is supplemented by the analogous graphic showing TTCPA range / TTCPA median ratio vs. TTCPA median as shown in Fig. 15. The same four regions should help determining the frontier between a standard and a non-nominal behavior.

A similar reasoning for calculating the analytical 2D coordinates defines all data TTCPA range crossing point and daily TTCPA range crossing point in equations (3) and (4):

$$\begin{cases} x_D = median(TTCPA_D) \\ y_{TTCPA,D} = a_{TTCPA,D} \cdot e^{-b_{TTCPA,D} \cdot x_D} \end{cases} \quad (3)$$

where $a_{TTCPA,D}$ and $b_{TTCPA,D}$ are the parameters of the exponential distribution fitted from all available data (D) dispersion diagram of derived random variable: TTCPA range / TTCPA median over TTCPA median.

Consider daily crossing point as the point at which TTCPA median intersects TTCPA range / TTCPA median vs. TTCPA median exponential approach for every day, two-hour intervals and different density regions.

$$\begin{cases} x_d = median(TTCPA_d) \\ y_{TTCPA,d} = a_{TTCPA,d} \cdot e^{-b_{TTCPA,d} \cdot x_d} \end{cases} \quad (4)$$

where $a_{TTCPA,d}$ and $b_{TTCPA,d}$ are the parameters of the exponential distribution fitted from daily data (d) dispersion diagram of derived random variable: TTCPA range / TTCPA median over TTCPA median.

The four regions support the same reasoning as RT performance indicator though applied to reaction variability. In a similar way, as it has been previously stated, everyday crossing points obtained from equation (4) compared with the overall average for TTCPA range / TTCPA median ratio over TTCPA median in equation (3) could lead to conclusions regarding the variation of the sample behavior. The same zones may be applicable for the position of these crossing points, whereas Relaxed, Favorable, Stressed and Critical Zones regarding RT median / TTCPA median ratio, points out the closeness to the conflict, TTCPA range / TTCPA median ratio zones bring more information about the standardized responses in average. This performance metric has been named “TTCPA performance indicator”.

TTCPA critical limit ratio

Finally, the limit line representing TTCPA range / TTCPA median with ratio 1 in Fig. 15, is taken as the critical limit among the data samples. Considering a representative sample, the points which fall over the limit line may imply a highly non-standard behavior, since their sample time range (variability) in the remaining TTCPA is bigger than their TTCPA median. The sample points above the limit are highlighted in Fig. 15 using red ellipses. The higher the TTCPA range / TTCPA median ratio, the more critical the represented interval. Nevertheless, if $x_s > x_D$, considering x_s the TTCPA median of the critical sample, it is interpreted as a period of laxity, since its TTCPA median is bigger than the overall median. On the contrary, when $x_s < x_D$, a lack of safety is spotted, since the sample TTCPA median is smaller than the overall median. This performance metric is called “TTCPA critical limit ratio”. If for one day, two consecutive samples over the same region fall over the TTCPA critical limit ratio, an important safety issue could have taken place.

Other airspaces may have different overall crossing points regions, but since Favorable Zone falls below the tail of the exponential fit, it is highly unlikely that the critical limit falls below it as well. Thus, it is improbable that a sample over the critical limit falls in the TTCPA Favorable Zone.

Traffic samples considerations

Samples with a small number of potential conflicts are highly biased by the disparity inside the set and the use of medians instead of means. An example of this statement is shown in the sample from day 5 (high density region between 4:00 AM and 6:00 AM). The three potential conflicts occurred are detailed in Table 5 and are separated by more than 20 minutes between the potential conflicts. Considering the LAT is 10 minutes, it can be affirmed that the potential conflicts are not concurrent.

In Table 5, iTTCPA, TTCPA and RT have been rounded to three decimals so derived data could have minimum decimal differences. The median of each random variable is highlighted in bold letters. The disparity in the relevant variables in the three potential conflicts makes the sample statistics very sensitive to peculiarity. Hence, the variability in iTTCPA and RT, and consequently in TTCPA, is big (6.435 secs) with respect to the median value (1.948 secs). The results of the overall statistics in Table 5 and the derived random variables RT median / TTCPA median and TTCPA range / TTCPA median show that the results of the metrics applied to this sample are not representative of the overall system behavior in this timeframe. Were this sample representative of the whole timeframe and region, TTCPA performance indicator would grow dramatically due to a little TTCPA median, and a big variability in the three TTCPA's, which are reported into 50% of the data (not dismissing any outlier). This is shown in the sample RT and TTCPA crossing points included in Table 5. $x_s = 1.948$ min is far below $x_D = 5.2705$ min while $y_{RT,s} = 0.670$ vs. $y_{RT,D} = 0.1522$ and $y_{TTCPA,s} = 3.302$ vs. $y_{TTCPA,D} = 0.6168$. This implies that RT and TTCPA crossing points of the sample behave much worse than the average, exhibiting even a sample whose TTCPA range / TTCPA median is over the critical limit ratio.

Thus, it is here recommended to account only for the samples with more than a few potential conflicts included or, if little populated samples are too frequent, define wider regions or timeframes were more traffic implies that potential conflicts are more likely.

Verification exercise

The hypotheses established and the safety metrics proposed have been validated through a verification exercise which has been carried out calculating the derived daily crossing points, for the data available, and assessing whether the metrics presented provides a reasonable interpretation. The different zones described in the "Proposed metrics" subsection were analyzed individually.

In order to verify the defined metrics, different information was considered. On the one hand, the metrics, using this information:

- Fig. 17, which zooms on each day exponential fitted curve and its crossing point (using as dependent variable RT median / TTCPA median) with respect to the average fit of all data available. Every day curves and crossing points are marked in a different color. The 2D coordinates $(x_d, y_{RT,d})$ are shown in the last column of Table 3, along with the average coordinates in the first row $(x_D, y_{RT,D})$.
- Fig. 18, which likewise shows the comparison of TTCPA range / TTCPA median ratio over TTCPA median and the relative position of the daily crossing points. The daily curves and crossing points are represented using the same colors as in Fig. 17. The 2D coordinates $(x_d, y_{TTCPA,d})$ are shown in the last column of Table 4, along with the average coordinates in the first row $(x_D, y_{TTCPA,D})$.
- The information regarding the sample points which fall above the TTCPA critical limit ratio: the number of potential conflicts involved, values of TTCPA median, RT median, help determining the criticality of the sample. Those points highlighted in Fig. 15 with red ellipses, are detailed in Table 6. The table is ordered by their sample TTCPA median, x_s, revealing the points surrounded by ellipses in Fig. 15 from left to right. As it has been explained before, samples with little group count, which stands for the number of potential conflicts enclosed, will not be considered. The most remarkable samples, those which will be analyzed, will be the samples enclosing more than 3 potential conflicts.

On the other hand, the daily situation and conditions were outlined using this information:

- The exponential approach details presented in Table 3 and Table 4, which have been explained earlier, particularly the daily adjusted coefficient of determination, or adjusted R-squared and the daily and overall crossing points.
- The daily number of potential conflicts, which can easily be used to roughly calculate the average occupancy of the airspace. The information is shown in Table 7.
- When necessary, a more detailed look into daily data samples and exponential approaches will be provided.

A summary of the distribution of daily crossing points (for RT median / TTCPA median and TTCPA range / TTCPA median) is shown in Fig. 19.

Relaxed Zone

The days whose crossing points fall into Relaxed Zone are day 1 and day 2, meaning a better than average safety performance, but exhibiting a non-orthodox behavior. Backing this reasoning, the number of potential conflicts those two days were significantly less than the rest of the analyzed days, see Table 7, having 60% and 25% (respectively) of less potential conflicts than the busiest day: day 7. However, for day 1, there is one significant sample falling over TTCPA critical limit ratio, between 12:00 and 14:00 in the high-density region, as can be seen in Table 6, and exhibiting a TTCPA median, $x_s = 3.421$ min, well below the overall TTCPA median $x_D = 5.2705$ min. Thus, the conclusion derived from the proposed metrics is that day 1 and day 2 had a more relaxed behavior than the average, although certain unwanted, unsafe laxity did arise.

Favorable-Average Zone

The day falling into Favorable-average Zone is day 4. As has been shown before, there is a slight improvement in the RT performance indicator (5.3067,0.1474) over the TTCPA performance indicator, (5.3067,0.6089), which lays a response close to the standard.

The number of potential conflicts on day 4 was on the average of the available data: 14% from the total. Nevertheless, the exponential fits parameters in Table 4 show through the adjusted coefficient of determination (adjusted R-square equals to 0.02294) that the TTCPA range / TTCPA median vs. TTCPA median does not properly follow an exponential behavior. That can be depicted as random behavior in terms of the time remaining until the potential conflict. However, in Table 3, it is shown that RT median / TTCPA median vs. TTCPA median can be interpolated by an exponential function exhibiting a higher degree of confidence: adjusted R-square equals to 0.4301. In Fig. 20, there is another representation of the dispersion and approach for day 4 data. Two plots and two groups of lines in each plot have been added. Same colors as used in Fig. 14 and Fig. 15 are utilized to maintain coherency: dashed light green lines refer to day 4 samples while continuous black lines refer to all available samples. The straight vertical lines correspond to the TTCPA medians and the curves represent the interpolated exponential approaches. The upper plot refers to RT median / TTCPA median ratio over TTCPA median while the lower plot refers to TTCPA range / TTCPA median ratio over TTCPA median. The derived two-hour intervals and different density regions from the available data samples for day 4 regarding RT median / TTCPA median vs. TTCPA median for the first plot and TTCPA range / TTCPA median vs. TTCPA median in the second plot, are shown as a dispersion diagram. It can be seen that the TTCPA range samples (below) cannot be approached by any particular fitting while RT samples (above) do follow an exponential behavior.

Also, TTCPA median, $x_4 = 5.3067$ min, falls above the overall median, $x_D = 5.2705$ min, and there are no samples shown in the dispersion diagram in Fig. 20 whose median is dramatically low. Regarding sample points above the TTCPA critical limit ratio in Table 6, day 4 exhibits one, between 14:00 and 16:00 in the high-density region, which is translated into a period of laxity. Despite the critical point x_s falls on 4.356 min, well under the overall average, the daily TTCPA is bigger than the global TTCPA. Hence, in light of the proposed metrics, it can be stated that day 4 behaves in a standard way, marginally better than the average, however exhibiting an interval of time with unsafe indulgence. It is coherent with the information available for day 4.

Stressed Zone

Days 5, 6 and 7 are the ones falling in the Stressed Zone in both RT and TTCPA performance indicators. Days 6 and 7 exhibit a high number of potential conflicts, while day 5 stays on the average. Although the TTCPA medians were smaller than the average, $x_5 = 5.0702$ min, $\mathrm{x}_6 = 5.1774$ min, $x_7 = 5.0868$ min vs $x_D = 5.2705$ min, RT's were kept low, (demonstrated by the RT and TTCPA zone they fall into). Also the RT median / TTCPA median and TTCPA range / TTCPA median samples follow a well-behaved exponential distribution, as it is shown in Table 3 and Table 4. Only day 5 has a significant sample over the TTCPA critical limit ratio, $(x_s, y_s) = (3.496, 1.179)$, in high–density traffic area from 10:00 to 12:00. Based on the proposed metrics information, it can be stated that days 5, 6 and 7 retained a more stressed than the average situation, although, day 5 had a critical period. Thus, the results of the metrics are plausible.

Critical Zone

Finally, day 3 falls in Critical Zone. Table 3 shows that for day 3, the adjusted coefficient of determination of RT median / TTCPA median over TTCPA median distribution (its adjusted R-square) is very low, even negative: -0.05449, and that the distribution of the medians does not follow an exponential curve. Still Table 4 shows that the adjusted R-square for TTCPA range / TTCPA median vs. TTCPA median is marginally bigger. Deepening into this idea, Fig. 21 shows the same information as Fig. 20, for day 3 and still conserving the coherency between Fig. 14 and Fig. 15, including its data plotted in light blue. For both plots, it can be seen that the dispersion diagrams cannot be fitted by any particular function. Thus, it can be affirmed, that the ratios follow an erratic distribution. However, although $x_3 = 5.1735$ min is worse than the average $x_D = 5.2705$ min, all intervals' TTCPA medians are kept above 4 minutes, which is close to the typical TTCPA for the complete sample. Also, there are no significant intervals for day 3 in Table 6 exhibiting a critical value above the limit. Thus, according to the metrics and day 3 samples, it is inferred that day 3 was more critical than the average and that it followed an anarchic behavior.

Conclusions and further work

Recalling the research question presented in the "Introduction" section - **Can precursory safety metrics be inferred from post-flight trajectories extracted from radar tracks?** -, it can be stated that the main goal of this research work - to find evidence of safety performance through post-flight trajectory information - has been successfully achieved. The perceivable relationship between controllers' reaction times upon a potential conflict detection and the buffer remaining until it was expected to occur within a safe airspace has been scrutinized.

After thoroughly performing data exploration, the hypothesis set about the "relationship of times to potential conflicts" and "ATCo's reaction times" (as extracted from radar tracks) has been proven. Moreover, precursory metrics have been defined and tested with the available data. Three performance indicators summarize the operation of the airspace in terms of safety, highlighting:

- Timeframes when stressed situations happen and whether they were satisfactorily tackled or reactions could have been improved.
- Timeframes when non-normalized behavior arise and, depending on the gap to potential conflicts, they could have meant an unsafe situation or a laxity period.
- Additionally, a critical limit is defined, above which individual situations are highly recommended to be monitored.

Whenever stressed situations are not successfully solved or non-normalized behavior occurs, incidents are more likely to appear.

It can be affirmed that in this paper the traditional safety metrics, based on hazards and risks and their probabilities, have been revisited through a new PBO based approach.

More data is needed to better understand the newly presented performance metrics' nature and to complete their definitions. Several facets could be improved:

- Enhancing the performance indicators with the actual sector configuration data and support ATCo's intervention detection with other source of information which would allow detailed actions per ATC position;
- Finding direct causes of RT and TTCPA fluctuation and its impact on the defined performance indicators;

- Further testing the metrics' behavior under different scenarios;
- Establishing targets for the metrics which are expected to be different for diverse scenarios.

Future work could define the methods to determine the triggers for an alert when the system is degrading in terms of safety. They would be based on the infringement of the thresholds that should be established for each airspace.

Several promising future applications are now possible:

- They could be used by each ANSP to oversee safety performance, continuously monitoring the metrics to detect when the system is degrading.
- Including these metrics as KPI's in SESAR performance framework, to test the impact of new SESAR solutions on Safety Key Performance Area (KPA).
- Using them as a performance indicators of the impact on safety of new operational changes or technological developments.
- Including them as ANSP information to be provided to the Safety Regulation Commission as Safety KPI's and as such, publishing them in Annual Reports.
- Other KPA's could be furtherly explored, using the developed metrics as an input to make decisions regarding:
  - ✓ Cost-efficiency, which could be enhanced by scrutinizing relaxed days and proposing an improved sector distribution.
  - ✓ ATCo's workload, which could also be assessed if the information of the active operational configuration of sectors and the working shifts are available.
  - ✓ Resilience, calculating the time to return to normal safety performance after non-nominal situations.

Acknowledgments

The authors would like to express their gratitude to Cranfield University for its openness and hospitality, providing the corresponding author the opportunity to join its team and work along the research group on ATM. Along with the continuous support from UPM Air Navigation Research Group (GINA-UPM) an excellent environment has been created for the developing of this research work, hopefully the first of many collaboration between both research groups.

Also, the authors would like to thank Consejo Social of UPM for its financial support to Rocío Barragán during her stay at Cranfield University.

Special thanks to Nicolás Suárez Tetzlaff and Alan Ross Groskreutz for their selfless help building the paper, particularly reviewing the use of English.

References

Cózar Maldonado, Juan Alberto. «Desarrollo de un Modelo de Riesgo de Colisión en espacios aéreos de Área Terminal.» Doctoral Thesis, 2015.

Cózar Maldonado, Juan Alberto, Francisco Javier Sáez Nieto, and Enrique Ricaud Álvarez. "Radar track segmentation with cubic splines for collision risk models in high density terminal areas." *Journal of Aerospace Engineering* (Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering) 229, no. 8 (06 2015): 1371-1383.

EASA. «A Harmonised European Approach to a Performance Based Environment (PBE).» 2014.

Eurocontrol. "Annual Report 2014." 2015.

Eurocontrol. "Annual Report 2015." 2016.

Eurocontrol Performance Review Commission. "Performance Review Report 2015 - An Assessment of Air Traffic Management in Europe during the Calendar Year 2015." 2016.

Ferrari, Joseph R. *Still procrastinating - The no-regrets guide to getting it done.* John Wiley & Sons, 2010.

García González, Eduardo José. "Development of a 3-Dimensional Mathematical Collision Risk Model based on recorded aircraft trajectories to estimate the Safety Level in high density en-route airspaces." Thesis dissertation, 2013.

García González, Eduardo José, Francisco Javier Sáez Nieto, and María Isabel Izquierdo. "Identification and analysis of proximate events in high density enroute airspaces." *Proceedings of ATM Seminar.* 2007.

Hyndman, Rob J, and Han Lin Shang. "Rainbow plots, bagplots and boxplots for functional data." *Journal of Computational and Graphical Statistics*, 2010: Issue 1, 29-45, doi: 10.1198/jcgs.2009.08158.

ICAO - APAC. "Guidance Material on Comparison of Surveillance Technologies (GMST)." 2007.

ICAO. "Annex 19 - Safety Management - First Edition." 2013.

ICAO. «Annex 2 - Rules of the Air.» 2005.

ICAO. "Doc 9854 - Global Air Traffic Management Operational Concept - First edition." 2005.

ICAO. "Doc 9859 - Safety Management Manual - Third Edition." 2013.

Perrin, Eric, Barry Kirwan, Ron Stroup, Eurocontrol, and FAA. "A systemic Model of ATM Safety: The Integrated Risk Picture." *Proceedings of ATM Seminar.* 2007.

Pozzi, Simone, et al. "Safety monitoring in the age of big data." *Ninth USA/Europe air traffic management research and development seminar (ATM2011).* Berlin, Germany, 2011.

Rousseeuw, Peter J, Ida Ruts, and John W Tukey. "The Bagplot: A Bivariate Boxplot." *The American Statistician*, 1999: Vol. 53, Issue 4, 382-387, DOI: 10.1080/00031305.1999.10474494.

Sáez Nieto, Francisco Javier. "The long journey toward a higher level of automation in ATM as safety critical, sociotechnical and multi-Agent system." *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering* (ASCE) 230, no. 9 (07 2016).

Sáez Nieto, Francisco Javier, Rosa María Arnaldo Valdés, Eduardo José García González, Gerard McAuley, and María Isabel Izquierdo. "Development of a Three-Dimensional Collision Risk Model Tool to Assess Safety in High Density En-Route Airspaces." *Proceedings of the Institution of Mecanical Engineers, Part G: Journal of Aerospace Engineering 224(10)*, 2010: 1119-29. doi: 10.1243/09544100JAERO704.

SESAR CONSORTIUM. “CONOPS Acronyms and Definitions, Task 2.2.2 - Milestone 3, DLT-0612-222-02-00.” 2007.

SESAR-JU P16 01 01. "D17-Validation / Verification of the SESAR Accident Incident Model (AIM)." 2014.

SESAR-JU. «Transition ConOps SESAR 2020 - Consolidated deliverable with contribution from Operational Federating Projects.» 2016.

Sheridan, Thomas B. "Next generation air transportation systems: Human-automation interaction and organizational risks." *Proceedings of the 2nd Symposium on Resilience Engineering.* 2006.

Suárez Tetzlaff, Nicolás, Patricia López de Frutos, Eva Puntero Parla, and Sara Rodríguez Gómez. "Quantifying Air Traffic Controller Mental Workload." *Proceedings of 4th Sesar Innovation Days.* 2014.

Tukey, John W. «Mathematics and the Picturing of Data.» *Proceedings of the International Congress of Mathematicians.* Vancouver, 1975. 523-231.

Tables

Table 1. Kolmogorov-Smirnov test results on log(RT) in seven different days

	h_0	α	*p-value*	*Test statistic*	*Critical value*
Day 1	0	0.01	0.44416471	0.03374228	0.06376231
Day 2	0	0.01	0.03773152	0.04109814	0.04749497
Day 3	0	0.01	0.08638465	0.03408912	0.04429998
Day 4	0	0.01	0.02291743	0.04174479	0.04546320
Day 5	0	0.01	0.15084251	0.03189048	0.04571427
Day 6	1	0.01	0.00135703	0.04855194	0.04136260
Day 7	0	0.01	0.08174111	0.03186330	0.04104491

Table 2. Kolmogorov-Smirnov test results on TTCPA in seven different days

	h_0	α	*p-value*	*Test statistic*	*Critical value*
Day 1	0	0.01	0.44152573	0.03340433	0.06299056
Day 2	0	0.01	0.08748313	0.03647212	0.04749497
Day 3	0	0.01	0.01085822	0.04395369	0.04429999
Day 4	1	0.01	0.00871405	0.04605146	0.04546319
Day 5	1	0.01	0.00493712	0.04867106	0.04571427
Day 6	1	0.01	0.00629645	0.04313449	0.04136259
Day 7	0	0.01	0.01113024	0.04062718	0.04104491

Table 3. Exponential fit of RT median / TTCPA median vs. TTCPA median on different days

Day	Exponential Fit	SSE	R-square	Adjusted R-square	RMSE	Crossing points $(x_d, y_{RT,d})$
All	$1.276 \cdot e^{-0.4034x}$	0.259	0.5879	0.5846	0.0438	(5.2705,0.1522)
Day 1	$1.768 \cdot e^{-0.454x}$	0.01506	0.8166	0.8051	0.03068	(5.3729,0.1543)
Day 2	$1.746 \cdot e^{-0.4538x}$	0.03209	0.4768	0.4477	0.04223	(5.4690,0.1460)
Day 3	$0.1983 \cdot e^{-0.01535x}$	0.0685	0.001011	-0.05449	0.06169	(5.1735,0.1832)
Day 4	$0.7273 \cdot e^{-0.3008x}$	0.02001	0.4586	0.4301	0.03245	(5.3067,0.1474)
Day 5	$1.621 \cdot e^{-0.4663x}$	0.00742	0.9736	0.9721	0.0203	(5.0702,0.1524)
Day 6	$0.4517 \cdot e^{-0.2086x}$	0.0088	0.5734	0.5483	0.02275	(5.1774,0.1534)
Day 7	$0.4151 \cdot e^{-0.2016x}$	0.02835	0.2433	0.1988	0.04084	(5.0868,0.1489)

Table 4. Exponential fit of TTCPA range / TTCPA median vs. TTCPA median on different days

Day	Exponential Fit	SSE	R-square	Adjusted R-square	RMSE	CROSSING POINTS $(x_d, y_{TTCPA,d})$
All	$6.02 \cdot e^{-0.4323x}$	5.023	0.5823	0.5792	0.1929	(5.2705,0.6168)
Day 1	$6.149 \cdot \mathrm{e}^{-0.4261\mathrm{x}}$	0.2325	0.8147	0.8031	0.1206	(5.3729,0.6232)
Day 2	$1.981 \cdot \mathrm{e}^{-0.2043\mathrm{x}}$	0.1364	0.4681	0.4385	0.08705	(5.4690,0.6481)
Day 3	$1.678 \cdot \mathrm{e}^{-0.166\mathrm{x}}$	0.552	0.1802	0.1346	0.1751	(5.1735,0.7109)
Day 4	$1.112 \cdot \mathrm{e}^{-0.1135\mathrm{x}}$	0.5516	0.07179	0.02294	0.1704	(5.3067,0.6089)
Day 5	$8.655 \cdot \mathrm{e}^{-0.5261\mathrm{x}}$	1.214	0.8309	0.8215	0.2597	(5.0702,0.6011)
Day 6	$1.727 \cdot \mathrm{e}^{-0.1943\mathrm{x}}$	0.3387	0.4177	0.3834	0.1411	(5.1774,0.6316)
Day 7	$2.706 \cdot \mathrm{e}^{-0.2823\mathrm{x}}$	0.4516	0.4072	0.3724	0.163	(5.0868,0.6437)

Table 5. Detail of potential conflicts of Day 5, high density region, between 04:00 and 06:00

Time of detection	iTTCPA (min)	TTCPA (min)	RT (min)	RT / TTCPA	Overall results
04:45:12	3.100	**1.948**	1.152	**0.591**	
05:13:22	4.625	1.633	2.991	1.831	
05:35:57	9.373	8.068	**1.305**	0.161	
RANDOM VARIABLES		RT median / TTCPA median			0.670
		TTCPA range / TTCPA median			3.302
METRICS		$(x_s, y_{RT,s})$			(1.948, 0.670)
		$(x_s, y_{TTCPA,s})$			(1.948, 3.302)

Table 6. Statistical details of samples whose TTCPA range over TTCPA median ratio is above limit

Day	Density	Time interval	Number pot. confl.	iTTCPA median (min)	TTCPA median (min)	RT median (min)	RT median / TTCPA median	TTCPA range / TTCPA median
Day 5	High	04:00-06:00	3	4.626	1.949	1.306	0.670	3.302
Day 1	High	12:00-14:00	13	4.824	3.421	1.382	0.404	1.456
Day 5	High	10:00-12:00	20	4.795	3.496	1.037	0.296	1.179
Day 4	High	14:00-16:00	24	5.329	4.356	0.807	0.185	1.128
Day 3	Low	02:00-04:00	3	5.891	5.504	1.939	0.352	1.21
Day 5	Low	02:00-04:00	3	7.545	6.007	0.775	0.129	1.031

Table 7. Number of potential conflicts per day

Day	Number of potential conflicts	% of potential conflicts from total
Day 1	646	7%
Day 2	1167	13%
Day 3	1342	15%
Day 4	1274	14%
Day 5	1260	14%
Day 6	1540	18%
Day 7	1564	18%
TOTAL	8793	100%

2017-10-23

Finding precursory air traffic management safety metrics using exploration of trajectory radar tracks

Barragán Montes, Rocío Barragán

American Society of Civil Engineers

Rocío Barragán Montes, Fernando Gómez Comendador, Francisco Javier Sáez Nieto, and Luis Pérez Sanz. Finding precursory air traffic management safety metrics using exploration of trajectory radar tracks. Journal of Aerospace Engineering, Volume 31, Issue 1, January 2018, article number 04017085

http://dx.doi.org/10.1061/(ASCE)AS.1943-5525.0000796