Composite risk index: the new Safety Performance Indicator of risk exposure

Tamara Pejovic[1],[*][†]

[1]EUROCONTROL, Aviation Intelligence, Performance Review Services, Belgium

Abstract. Conceptually, all safety programmes desire accurate safety risk quantification in order to provide a meaningful expression of risk. As there are typically, multiple safety risks associated with a system or event, the quantification of total safety risk is a major challenge. One possible way to define and accept the total safety risk of any system is using the concept of a composite risk estimate. This paper represents development of the new safety performance indicator and overall methodology that could be used to measure the performance of European ATM systems as a whole and its individual entities. It describes the computation of the Composite Risk Index (CRI), logic behind it, its use (on the example EUROCONTROL Member States) and limitations and areas of potential improvement. CRI represents a cumulative risk value calculated aggregating all reported, assessed and severity classified key safety-related incidents to form an index. This measure of risk exposure is based on probability and severity that considers the human perception of equivalent risk. Overall idea behind CRI is that the performance of safety system can be analysed within three important broad categories: the quality of reporting system with reporting entity, measured risks within the system, and human perception of risk.

1 Introduction

Risk is the potential for mishaps or other adverse variation in the cost, schedule, or safety performance of ATM system. Safety risk therefore can be explained as the potential for mishaps that could result in injury, fatality, equipment or system damage or total loss.

Conceptually, all safety programmes desire accurate safety risk quantification in order to provide a meaningful expression of risk. As there are typically multiple safety risks associated with a system or event, the quantification of total safety risk is a major challenge.

One possible way to define the total safety risk of any system is using the concept of a composite risk estimate [1]. Current methods of obtaining this composite risk estimate use summing techniques to add the individual risks and produce a single number [1, 2]. This method seems natural, however, it is often difficult to determine particular occurrence probabilities (e.g. when historical information is of limited time series) or to quantify their

---

[*] Corresponding author: tamara.pejovic@eurocontrol.int

[†] Disclaimer: The views expressed in this paper are the authors' own and do not represent a policy or position of EUROCONTROL.

severity (e.g. when information in safety databases is missing). That makes the additive computation of risk difficult or impossible.

Moreover, although risk in general can be quantified, as it represents combination of probability and severity of specific occurrence happening, the human perception of risk often influences how risk is addressed [3]. For example, on the level of decision makers the risk perception does not necessarily map directly to probability and severity in a linear fashion. That makes computation of total risk additionally difficult and subjective.

For all these reasons, the concept of a Composite Risk Index (CRI) that could measure the performance of European ATM systems as a whole or also its individual entities (service providers or Member States) is proposed. In simple words, CRI presents a cumulative risk value calculated aggregating all reported, assessed and severity classified safety-related incidents to form an index. This measure of risk exposure is based on probability and severity that considers the human perception of equivalent risk.

1.1 Data

In order to calculate composite risk, each historical, reported occurrence had to have assigned severity and probability. Safety information about reported events was acquired through EUROCONTROL Annual Summary Template (AST) reporting system [4]. The AST reporting mechanism captures information on Air Traffic Management (ATM) related occurrences, both ATM operational and technical occurrences. The safety data, related to the reported occurrences in the AST, included occurrence category (accident or incident) and its severity reported by the States and calculated using severity classification risk assessment methodology (RAT).

For definition of the Weights, that would explain the human perception of risk for each type of safety occurrence, and also overall CRI methodology development in generall, the AST data was provided by EUROCONTROL AST Team for period 2015-2018. Nevertheless, it has to be noted that modelling of Weights can be customised additionally to a local environment, it can be performed using different source of safety occurrences data, as long as the input satisfies the minimum data requirements (This includes, the total number of occurences per each type of incident, separately for each severity class that is modelled. In addition exposure data in terms of flight hours is needed as well.).

The classification scheme for safety occurrences in ATM specifies six severity categories for ATM related occurrences impacting the safe operations of the aircraft [5]. They are as follows: Accident, Serious Incident (AA / A), Major incident (B), Significant incident (C), Not determined (D), No safety effect (E).

The RAT classification scheme [6] specifies five qualitative frequency categories (repeatability), however, these values are not commonly reported through the AST. Moreover, each State in principle should develop their own quantitative boundaries, which should consider national traffic volumes and specific operating conditions of the national ATM system. As these values were not available the occurrence probability was calculated using historical data (frequency of occurrences over all available years was used as a proxy for probability) from the past three years separately for each State in order to simulate/take into consideration local conditions.

1.2 Methodology

As a proxy of safety risk within certain airspace or a State, at the preliminary stage of development, it was decided to base CRI calculations on the following (Figure 1):
- Accidents,

- Operation occurrences - OPS (high/medium risk incidents with Severity A to C): runway incursions (RI), separation minima infringements (SMI), unauthorized penetrations of airspace (UPA),
- Other operational occurrences - OTHER;
- Technical occurrences – TECH (high/medium risk incidents with Severity AA/A to C).
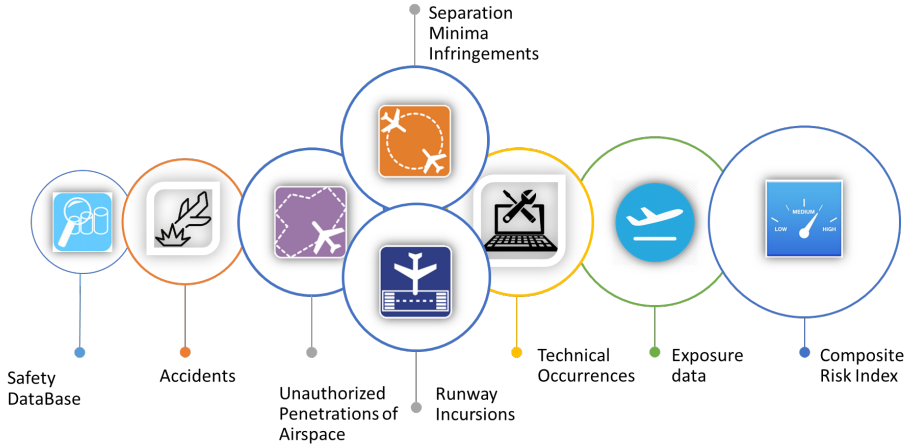
Figure 1.   Iconographic showing different elements of data input for calculation of CRI

To take into account the local conditions within each Entity, to have an objective comparison across small and large States/entities, scaling of variables by an appropriate size measure, in this case the total number of flight hours within each State, was used as an additional input (CRI normalised results).

1.2.1 Data input for calculation of CRI

Missing data often hinder the development of robust composite indicators. Data can be missing in a random or non-random fashion. In case of AST safety data available, considering the type of safety information collected, the missing values do not depend on the variable of interest or on any other observed variable in the data set. In other words, the missing values in Severity classification would be of the missing completely at random type: i.e. Severity classification has no correlation with type of occurrence or with reporting entity.

There are two general methods for dealing with missing data: case deletion, or imputation. No imputation model is free of assumptions and the imputation results should be thoroughly checked for their statistical properties, such as distributional characteristics, as well as heuristically for their meaningfulness[2].

Data imputation could lead to the minimisation of bias and the use of 'expensive to collect' data that would otherwise be discarded by case deletion. The uncertainty in the imputed data should be reflected by variance estimates. This makes it possible to take into account the effects of imputation in the course of the analysis. The multiple imputation method, which provides several values for each missing value, can more effectively represent the uncertainty due to imputation [7].

For all these reasons, all severity unclassified/not assessed events (Severity Category D) were distributed into groups A to E based on historical distribution (determined using the last three years of AST data). The probability of occurrence being assigned to specific severity category was calculated using historical data, separately for each State in order to simulate/take into consideration local conditions.

1.2.2 Estimated numbers of occurrences

The formula how the numbers of occurrences of specific type were estimated is presented below.

$$No\_occ\_est_i = \big((No\_D_i * P_i) + No\_occ_i\big) \tag{1}$$

$No\_occ\_est_i$ – *estimated number of safety occurrence type i*
No_D_i – *number of reported occurrences of Severity Classification D (not determined) for occurrence type i*
P_i – *probability of safety occurrence type i*
No_occ_i - *number of reported safety occurrences of type i*

Probability (probability is taken as proxy for frequency as explained before) of each type of occurrence was calculated using the simple principle:

$$P_i = \frac{No_{occ_i}}{Total\, No_{occ_i}} \tag{2}$$

No_{occ_i}– *number of reported safety occurrences type i in a year*
$Total\, No_{occ_i}$ - *total number of reported safety occurrences type i in a group in all years*

1.2.3 Human perception of risk

In order to add human perception of risk to the CRI index, certain values/weights had to be utilised in order to attribute personal perception of risk to its values.

Overall, when used in a benchmarking framework, weights can have a significant effect on the overall composite indicator and the entity rankings. A number of weighting techniques exist however, regardless of which method is used, weights are essentially value judgements. While some analysts might choose weights based only on statistical methods, others might reward (or punish) components that are deemed more (or less) influential, depending on expert opinion, to better reflect policy priorities or theoretical factors. In case of CRI both statistical/optimisation technique and expert judgement are used.

Accepted methods of quantifying severity include monetary amounts. However, although expressing severity in terms of cost establishes consistency, it is still difficult to put an amount on human life or injuries, or failure or loss of certain functionalities of the system. Furthermore, perception of what constitutes "high" risk may vary from entity to entity and State to State.

Therefore, introduction of Weights to express severity of event allows their description in non-monetary terms which have meaningful and easy understandable explanation in human perception.

Each weight value for specific Severity category was determined using optimisation technique, with the aim to select combination of weights that will not disturb the computation of the CRI from year to year if significant changes in reporting are introduced. In other words, the goal was to determine which combination of weights would result in the lowest standard deviation of CRI values between the years for each State.

Due to a large number of variables involved and enormous number of combination possible, optimisation and selection of Weights was done in several stages:
1. Selection of Weights for accident, all OPS and TECH occurrences (Base Weights);
2. Selection of Weights for OPS occurrences based on their type (RI, SMI, UPA, OTHER), taking into consideration overall OPS Weight determined in Step 1 (Occurrence type Weights);

3. Selection of Weights for OPS and TECH occurrences based on their severity (AA/A, B, C), taking into consideration overall OPS Weight determined in Step 2 (Occurrence severity Weights).

Overall optimal solution in each step, was the one that results in the minimum mean value of CRI standard variations for all entities in a single configuration (combination) of Weights.

In addition, each type of weight selection had predefined weight ranges (based on EUROCONTROL experts judgement, using techiques such as brainstorming and voting) to allow for incremental Severity classification order based on human perception of risk (from accident to Severity C incident, i.e. from high risk to low risk). In other words, each range had an expectation value associated with it.

The following ranges used for selection of Weights in different steps are presented below:
- Base Weights selection:
  - Accidents weight: w_acc = (0.5 : 0.7)
  - Operational occurrences weight: w_ops = (0.3 : 0.5)
  - Technical occurrences weight: w_tech = (0.05 : 0.4)
- Occurrence type Weights selection:
  - UPA weight: w_ops_upa = (0.15 * w_ops : 0.25 * w_ops)
  - SMI weight: w_ops_smi = (0.4 * w_ops : 0.55 * w_ops)
  - RI weight: w_ops_ri = (0.25 * w_ops : 0.35 * w_ops)
  - OTHER operational weight: w_ops_other = (0.05 * w_ops : 0.1 * w_ops)
- Occurrence severity Weights selection (where the letters in the name represent occurrence type and Severity class respectively):
  - w_ops_ri_a = (0.3 * w_ops_ri : 0.7 * w_ops_ri)
  - w_ops_ri_b = (0.1 * w_ops_ri : 0.5 * w_ops_ri)
  - w_ops_ri_c = (0.05 * w_ops_ri : 0.3 * w_ops_ri)
  - w_ops_upa_a = (0.3 * w_ops_upa : 0.7 * w_ops_upa)
  - w_ops_upa_b = (0.1 * w_ops_upa : 0.5 * w_ops_upa)
  - w_ops_upa_c = (0.05 * w_ops_upa : 0.3 * w_ops_upa)
  - w_ops_smi_a = (0.3 * w_ops_smi : 0.7 * w_ops_smi)
  - w_ops_smi_b = (0.1 * w_ops_smi : 0.5 * w_ops_smi)
  - w_ops_smi_c = (0.05 * w_ops_smi : 0.3 * w_ops_smi)
  - w_tech_aa = (0.4 * w_tech : 0.5 * w_tech)
  - w_tech_a = (0.3 * w_tech : 0.4 * w_tech)
  - w_tech_b = (0.1 * w_tech : 0.3 * w_tech)
  - w_tech_c = (0.04 * w_tech : 0.08 * w_tech)

Optimisation results indicated that, for this setup, the best combination of Weights was as in Table 1.

Using estimated number of occurrences and adding human perception of their risk it was possible to calculate CRI for each EUROCONTROL Member State separately. The simple formula to calculate CRI is presented below.

$$CRI_i = No_{occ\_est\_i} * w_i \qquad (3)$$

$$CRI = \sum_i \frac{CRI_i}{Total\ N_{occ}} \qquad (4)$$

CRI_i – CRI index for occurrence type i
W_i – weight (based on severity and human perception) assigned to specific type of safety occurrence
$Total\ N_{occ}$ – total number of all occurrences in a year

In the formula above *Weights* added to each equation represent additional human perception of risk for specific event, introduced so that the CRI can at the end consider the human perception of equivalent risk.

Finally, to allow applicability of CRI to airspaces with different traffic levels, the CRI was normalised by flight hours for each State. The CRInorm was calculated based on the following formula:

$$CRInorm = \frac{CRI}{Number\ of\ Flight\ Hours} \tag{5}$$

| Type of occurrence	Weight index	Value
Weight for accident	w_acc	0.54646
Weight for OPS incident	w_ops	0.38844
Weight for TECH incident	w_tech	0.06509
Weight for RI incident	w_ops_ri	0.09711
Weight for SMI incident	w_ops_smi	0.21364
Weight for UPA incident	w_ops_upa	0.05827
Weight for OTHER OPS incident	w_ops_other	0.01942
Weight for serious RI incident (A)	w_ops_ri_a	0.06798
Weight for major RI incident (B)	w_ops_ri_b	0.02428
Weight for significant RI incident (C)	w_ops_ri_c	0.00486
Weight for serious SMI incident (A)	w_ops_smi_a	0.11196
Weight for major SMI incident (B)	w_ops_smi_b	0.09100
Weight for significant SMI incident (C)	w_ops_smi_c	0.01068
Weight for serious UPA incident (A)	w_ops_upa_a	0.03046
Weight for major UPA incident (B)	w_ops_upa_b	0.02453
Weight for significant UPA incident (C)	w_ops_upa_c	0.00328
Weight for serious TECH incident (AA)	w_tech_aa	0.02980
Weight for serious TECH incident (A)	w_tech_a	0.02597
Weight for major TECH incident (B)	w_tech_b	0.00651
Weight for significant TECH incident (C)	w_tech_c	0.00281

Table 1. Weights based on severity and human perception

2 Preliminary Results

Using methodology described in previous section, the CRI for all EUROCONTROL Member States for 2018 (for which data was available, 39 Member States) is calculated and shown at Figure 2. It shows that two thirds of the EUROCONTROL Member States in 2018 had higher risk exposure in comparison to the EUROCONTROL average. Nevertheless, this calculation did not take into account the local conditions. Therefore, CRI was normalised, (CRInorm), taking into account the total number of flight hours within each State (Figure 3.), as it is assumed that amount of traffic could impact level of risk exposure within specific airspace.
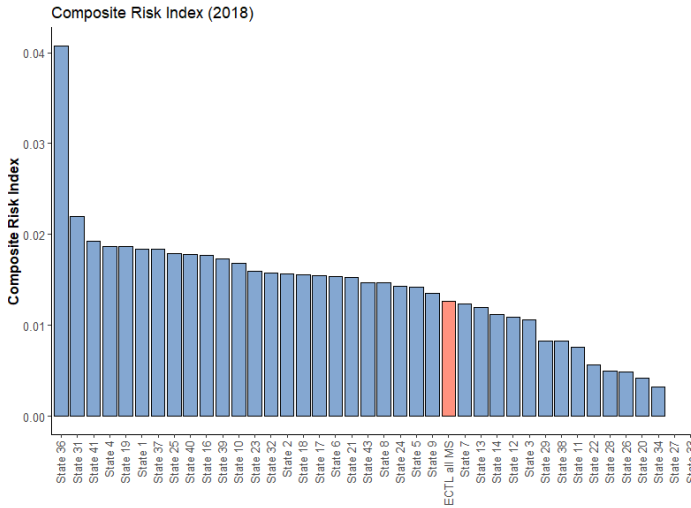
**Figure 2.** CRI in EUROCONTROL Member States (2018)
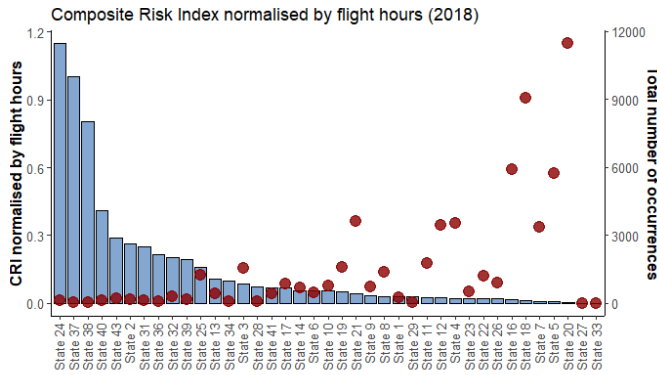
Figure 3. CRInorm vs Total Number of Occurrences in EUROCONTROL Member States (2018)

The further analysis of 2018 CRInorm results indicate that 25% (10 States) of EUROCONTROL Member States have CRInorm above 0.18 (treshhold taken from boxplot of CRInorm, upper quartile of 75% of population). With only three States having high CRInorm (boxplot outliers – i.e. above 0.5). One possible reason for this overall positive result could be that this is somehow related to the reporting culture of the State. Therefore, CRI normalised was correlated with the total number of reports reported by each State. Figure 3. shows that the States with a higher number of reports, which could indicate a good reporting culture (red dots - total number of reports), tend to have a low CRI normalised (blue bars).

2.1 CRI trend

Using CRI index, it is possible to follow the trend of safety performance, as CRI can be used as a quick indicator of the status of either safety performance based on the type and severity of historical reported occurrences but also as indicator of reporting culture. Figure 4. shows how the CRInorm has changed in a period 2015-2018 in EUROCONTROL area.
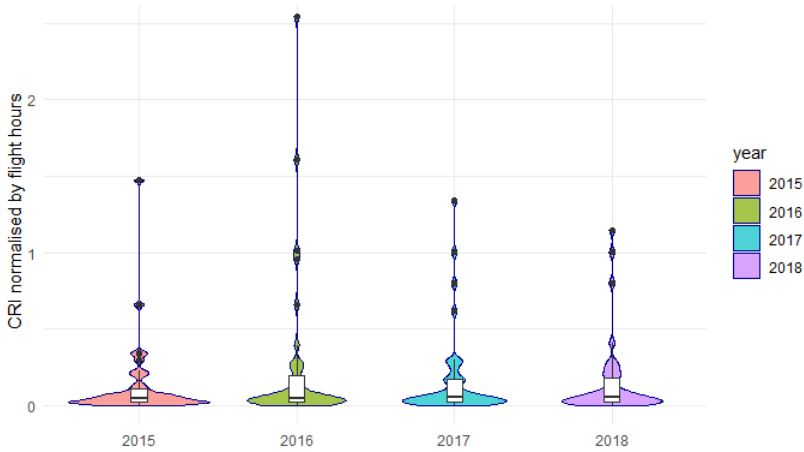
Figure 4.   Normalised CRI in EUROCONTROL area (2015-2018)

In 2016 CRInorm has increased comparing to the previous year, which was mainly influenced by very high CRInorm of several States. In 2017, regardless of the massive increase in the number of OPS occurrences by 65% (which have higher Weight values) and overall increase in all reported occurrences by approximately 38%, the CRI scores for EUROCONTROL area have decreased. This trend continued in 2018 as well. Furthermore, on the positive note, the number of States with extreme CRInorm has decreased (number of outliers has decresed).

Overall, using selected weights, further analysis showed that variation in reported number of occurrences does not have adverse effect on CRI computation and does not solely depend on the number of reported events. This is very important as the nature of CRI computation also allows calculation and monitoring of CRI of a single specific type of occurrences, e.g. the key risk occurrences within an airspace or organisation.

3 Conclusion

Idea behind CRI is that the performance of safety system can be analysed within three important broad categories: the quality of reporting system within reporting entity, measured risks within the system, and human perception of risk. Therefore it was considered that the concept of a CRI, as a cumulative risk value calculated aggregating all reported, assessed and severity classified safety-related incidents, has potential to become a proxy of exposure to risk within certain airspace for top management information and decision making.

Preliminary analysis shows that CRI has an ability to allow reporting on the safety performance of the whole European ATM system, but also on the level of its individual entities, e.g. Member States or even at the level of service providers. Moreover, scaling possibility allows measurement of CRI of individual types of safety occurrences as well.

The CRI however, should not be construed as an absolute measuring stick. It is only as good as the fidelity of the data that supports it. In general, specific probabilities of occurrence are not precisely known, and there is some subjectivity in the assessment of severity of the occurrence.

As mentioned before, besides the fact that CRI methodology can be customised to local environment, i.e. Weights can be re-modelled using local safety data, CRI methodology can be scaled up or down to satisfy monitoring of individual entities.

Based on individual local safety data availability, the CRI calculation can be improved by using higher granularity of safety-related data used to compute CRI. In other words, by

using safety data with higher granularity, so that Weights are computed separately for each different type of occurrence, (e.g. providing separate weights for different OTHER types of OPS occurrences).

Moreover, initial ranges of different Weights could be fine-tuned based on collective expert opinion. Adjustment of proposed Weights could be further improved via dedicated expert group, both locally and within aviation community. This would also help to better understand potential concept limitations and added value.

Finally, the CRI normalisation could also be done per different metrics, in order to allow inclusion of airspace size, capacity and/or complexity (for example, normalization per sector or number of flights). This could allow adding additional local specific operating conditions into equation.

References

1. Burns, M. (2016). Logistics and Transportation Security, A Strategic, Tactical, and Operational Guide to Resilience. CRC Press, Taylor and Francis Group.

2. Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., (2005). "Handbook on Constructing Composite Indicators: Methodology and User Guide". OECD Statistics Working Papers, STD/DOC (2005). Organisation for Economic Co-operation and Development.

3. Slovic, P. "Perception of Risk". (2016). Available via:
    https://www.researchgate.net/publication/325954197_The_perception_of_risk

4. EUROCONTROL, (2009). "Reporting and Assessment of Safety Occurrences in ATM," EUROCONTROL, Brussels.

5. EUROCONTROL, (1999). "Severity Classification Scheme for Safety Occurrences in ATM," EUROCONTROL, Brussels.

6. EUROCONTROL, (2013). "RISK ANALYSIS TOOL Guidance Material," EUROCONTROL, Brussels.

7. Jakobsen, J.C., Gluud, C., Wettersley, J., and Winkel, P. (2017). When and how should multiple imputation be used for handling missing data in randomised clinical trials – a practical guide with flowcharts. BMC Medical Research Methodology Volume 17, Article number: 162