PRC TECHNICAL NOTE

COMPOSITE RISK INDEX METHODOLOGY

Technical note prepared by the EUROCONTROL Performance Review Unit (PRU) and commissioned by the Performance Review Commission (PRC)

December 2020

Background

This Technical Note, commissioned by the Performance Review Commission (PRC) has been prepared by the EUROCONTROL Performance Review Unit (PRU).

The PRC was established in 1998 by the Permanent Commission of EUROCONTROL, in accordance with the ECAC Institutional Strategy (1997). One objective of this strategy is "to introduce a strong, transparent and independent performance review and target setting system to facilitate more effective management of the European ATM system, encourage mutual accountability for system performance…"

The PRC conducts independent measurement, assessment and review of the performance of the Pan-European Air Navigation Services (ANS) system, including its contribution to the efficiency of Pan-European aviation. The PRC strives to identify future improvements and makes recommendations as appropriate.

The PRC maintains open and transparent dialogue with relevant parties, including but not limited to States, Air Navigation Service Providers, Airspace Users, Airports, social dialogue partners, civil-military organisations, international and national organisations, etc.

The PRC conducts research into the development of performance measurement. This includes, inter alia, investigating how performance could best be described/measured in the long-term, developing and testing proposals for future indicators and metrics and contributing to future improvements in performance.

The PRC disseminates the results of its analysis to relevant parties, provided that no sensitive data are involved, in order to demonstrate the PRC's commitment to transparency and to promote the application of PRC analysis.

The PRC produces independent ad-hoc studies, either on its own initiative and/or at the request of relevant parties.

The PRC's website address is: https://www.eurocontrol.int/air-navigation-services-performance-review

PRC publications are also available on the website: www.ansperformance.eu

Notice

The PRU has made every effort to ensure that the information and analysis contained in this document are as accurate and complete as possible. Only information from quoted sources has been used and information relating to named parties has been checked with the parties concerned. Despite these precautions, should you find any errors or inconsistencies we would be grateful if you could please bring them to the PRU's attention by sending an email to: PRU-support@eurocontrol.int.

Printed by EUROCONTROL, 96, rue de la Fusée, B-1130 Brussels, Belgium.

To contact the PRC, please send an email to pru-support@eurocontrol.int.

DOCUMENT IDENTIFICATION SHEET

DOCUMENT DESCRIPTION

Document Title

Composite Risk Index Methodology

PROGRAMME REFERENCE INDEX	EDITION	EDITION DATE
Report commissioned by the PRC	Final	December 2020

Summary

In 2019 the PRC published (in PRR2018) its preliminary concept of Composite Risk Index (CRI) and corresponding methodology that could be used to measure the performance of the European ATM system, either as a whole or of its individual entities (service providers, Member States). Since then, the PRC has refined its CRI methodology to take account of local specific operating conditions and has published its methodology in this Technical Note.

This document will be of interest to ATM key decision makers, policymakers and operations staff as well as to Safety specialists, because the PRC's methodology by nature is scalable and therefore can provide high level information for decision makers as well as more in-depth knowledge about risk exposure to safety experts.

The PRC's CRI methodology contains four components: Safety data, Traffic/exposure data, Complexity and Reporting practices. This Technical Note explains in detail several statistical methods used to model index weights, overall computation, logic behind it, its use and limitations, and lastly areas of further improvement and expansion.

Keywords

Safety Performance Indicator, Risk Index, Composite Risk Index, Safety Performance EUROCONTROL Performance Review Commission, ATM Performance

CONTACT:	Performance Review Unit, EUROCONTROL, 96 Rue de la Fusée, B-1130 Brussels, Belgium. E-mail: pru-support@eurocontrol.int Website: www.ansperformance.eu

DOCUMENT STATUS & DISTRIBUTION

STATUS		DISTRIBUTION	
Draft	☐	General Public	☑
Proposed Issue	☐	EUROCONTROL Organisation	☐
Released Issue	☑	Restricted	☐

Table of Contents

List of Figures

List of Tables

1 TARGET READERSHIP

This PRC Technical Note will be of interest to Air Traffic management (ATM) key decision makers, policymakers and operations staff as well as to Safety specialists.

It contains the PRC's Composite Risk Index (CRI) methodology that could be used to measure ATM system performance at European level, or at local level (service providers, States) taking into account local conditions. CRI should be seen as a tool for an easy overview of risk exposure, to monitor and possibly compare the safety risks, either for one provider, or for State to make an overview of annual developments. Ultimately, based on data availability, it can be used as a benchmark between ANSPs, States or regions at one moment in time or over several years.

In 2019, the PRC first presented (in PRR2018) its preliminary concept of CRI and corresponding methodology. Since then, the PRC has refined its work to take account of local specific operating conditions. The PRC's updated CRI Methodology is contained in this Technical Note.

The methodology has four distinct components:

Component 1:	Safety data (with the following parameters: number and type of safety occurrences, severity classification, and human perception of risk),
Component 2:	Traffic / exposure data (consisting of the following parameters: flight hours and airport movements),
Component 3:	Complexity (namely adjusted density, structural index and Complexity index), and
Component 4:	Reporting practices (described through parameters of reporting rate and reporting culture).

The Technical Note explains in detail several statistical methods used to model index weights, overall computation, logic behind it, its intended use, benefits and limitations, and lastly areas of further improvement and expansion.

2 INTRODUCTION

Safety in aviation is paramount. Although safety procedures and policies have become well established, implemented and enforced, there is always a risk, however remote, of complacency creeping in. To ensure that this never happens, a CRI methodology can assist the key parties concerned to ensure that all aspects of safety management systems, for example policy, safety oversight, safety data reporting, pro-active actions to prevent safety threats, etc., are given top priority at all times.

A standard safety report with typical incident or accident trends may highlight some aspects. However, such safety reports do not always give tan insight and overall impression of how safety is performing over time or enable comparisons with different States or stakeholders.

It is from this s perspective that the PRC has developed a CRI Methodology. It is not designed to replace reporting on specific aspects of safety. Its purpose is to assist key decision makers, policymakers, and operations staff as well as Safety specialists by providing an easy overview to monitor and possibly compare the safety risks, either for one provider or for State to make an overview of annual developments. Ultimately, based on data availability, it can be used as a kind of benchmark between ANSPs, States or regions at one moment in time or over several years.

2.1 Problem: How can total safety risk be measured?

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 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.

2.2 PRC solution to measure total safety risk

In PRR 2018, the PRC presented the preliminary concept of Composite Risk Index (CRI) and corresponding methodology that could be used to measure the performance of the European ATM system as a whole or also its individual entities (service providers or Member States). This initial calculation of the CRI was mainly based on reported safety occurrences. More specifically, the CRI was represented as a cumulative risk value calculated aggregating all reported, assessed and severity classified key safety-related incidents to form an index.

The measure of risk exposure was based on probability and severity that considers the human perception of equivalent risk. The overall idea behind the CRI was that the performance of the safety system can be analysed in three important broad categories: the quality of the reporting system with reporting entity, measured risks within the system, and the human perception of risk (https://ansperformance.eu/methodology/cri-pi/).

At that time, the PRC also has highlighted that there could be possibilities to further improve the CRI by considering specific local operating conditions, airspace size, capacity and/or complexity.

This technical note represents, therefore, further development of CRI methodology, considering several proposed improvements, its computation, logic behind it, its use and limitations, and lastly areas of further improvement and expansion.

2.3 List of acronyms

Acronym	Definition
AST	Annual Summary Template
ATM	Air Traffic Management
CA	Cluster Analysis
CRI / CRInorm	Composite Risk Index / Normalised CRI
ICI	Intermediate Composite Index
FA	Factor Analysis
OPS	Operational
PC	Principal Components
PCA	Principal Component Analysis
PRC	Performance Review Commission
RAT	Risk Assessment Methodology
TECH	Technical

Table 1: List of acronyms used in the report

3 COMPOSITE RISK INDEX METHODOLOGY

In order to improve initial CRI methodology, further consideration of local operating conditions, airspace size and complexity of the airspace were taken into consideration. As a result, three additional components, besides the original one (Safety Data), were considered in the new methodology. The new CRI methodology includes hence four distinct components, or Intermediate Composite Indexes (ICIs):

1. Safety data (with the following parameters: number and type of safety occurrences, severity classification, and human perception of risk),
2. Traffic / exposure data (consisting of the following parameters: flight hours and airport movements),
3. Complexity (namely adjusted density, structural index and Complexity index), and
4. Reporting practices (described through parameters of reporting rate and reporting culture).

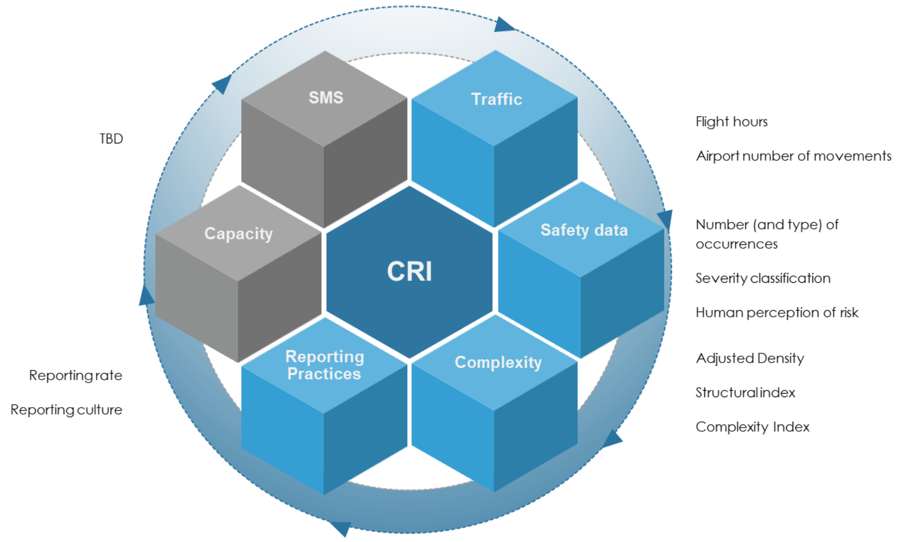


Figure 1. CRI Methodology Components

3.1 Intermediate Composite Indexes

In total, there were 38 different variables (parameters), together in four components, available for development of the new CRI methodology. These included both raw input information (such as number of specific incidents, or number of flight) to calculated values (such as reporting ratios, or reporting clusters).

Components were chosen based on the criteria that each of them represents an aspect of safety performance that could impact an entity's exposure to risk.

Based on expert opinion, "importance" of each ICI / component was assigned as in Figure 1. As an example, this means that *Safety Data* intermediate index has higher weight (importance) than amount of traffic within analysed airspace.

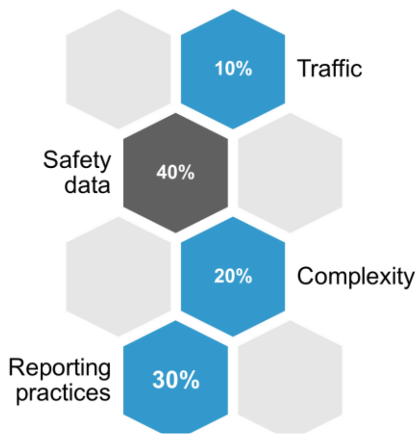


Figure 2. CRI Intermediate Composite Indexes

At the end, once the ICIs have been constructed, they are aggregated by allocating a weight (importance) to each of them to form the overall CRI.

The next sections explain how further CRI development looked into different statistical ways to assign weights to all selected parameters (variables) within each component.

3.2 Weights

The effect of weights used in aggregating indicators is very complex. One important step is the aggregation of indicators, where typically the variables are combined in a weighted average to give the resulting value of the composite indicator. Apart from the decision of which kind of weighted average to use (e.g. arithmetic, geometric), it is necessary to select values of weights to apply to each variable. The values of these weights can have a large impact on the subsequent rankings. Therefore, understanding the impact of weights on the variation of the composite indicator scores is important.

For this reason, various uncertainty and sensitivity analysis on composite indicator assumptions had to be performed. Nevertheless, it is important to notice, that even though considerable attention was given to the development of the Weights and aggregation methods used in development of the CRI methodology, many subjective choices are made in their construction. Additional detailed analysis of the importance of weights assigned to the index components can be at later stages also introduced to further investigate effects of weights. However, at this stage of development the importance was mainly given to the selection of components and variables to be included as well highlighting of different possibilities by introduction of the CRI concept.

3.3 Safety Data

Intermediate Index – *Safety Data* calculation kept its original form, as previously explained in the original methodology *(https://ansperformance.eu/library/Composite_Risk_Index_methodology.pdf)*.

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 EUROCONTOL 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. By definition, these ATM occurrences include:

- accidents;
- (serious) incidents:
    - Near collision (encompassing specific situations where one aircraft and another aircraft/the ground/a vehicle/person or object is perceived to be too close to each other):
    - Potential for collision or near collision (encompassing specific situations having the potential to be an accident or near collision, if another aircraft is in the vicinity).
- altitude deviations reported within the EUR RVSM airspace (above FL285);
- ATM-specific occurrences (encompassing those situations where the ability to provide safe ATM services is affected, including situations where, by chance, the safe operations of aircraft has not been jeopardised);
- Other defects or malfunctioning of an aircraft, its equipment and any element of the Air Navigation System which is used, or intended to be used, for the purpose or in connection with the operation of an aircraft, or with the provision of an air traffic management service or navigational aid to an aircraft.

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).

3.3.1 Severity definitions

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).

3.3.2 Frequency definitions

The RAT classification scheme [6] specifies five qualitative frequency categories (repeatability). However, as 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.

3.3.3 Methodology

As a proxy of safety risk within certain airspace or a State, the concept of exposure to risk, based on reported / historical safety information was proposed. This *risk index* presents a cumulative risk value calculated aggregating all reported, assessed and severity classified safety-related incidents to form an index.

The Safety Data ICI calculations were based on the following input (Figure 3):

- Accidents,
- Operation occurrences (high/medium risk incidents with Severity A to C):
    - runway incursions,
    - separation minima infringements,
    - unauthorized penetrations of airspace,
- Other operational occurrences;
- Technical occurrences (high/medium risk incidents with Severity AA/A to C).

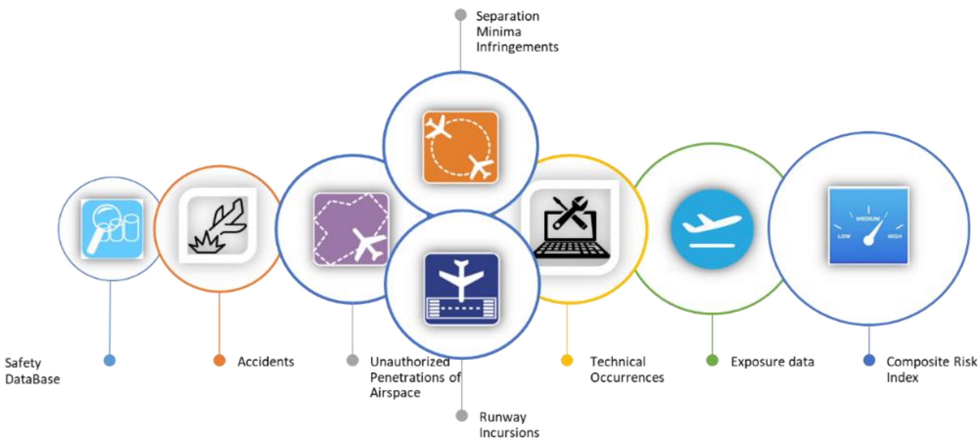


Figure 3. Safety Data Intermediate Index elements

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).

3.3.4 Occurrences with no severity classification

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 (Figure 4), based on historical distribution (determined using the last four 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.

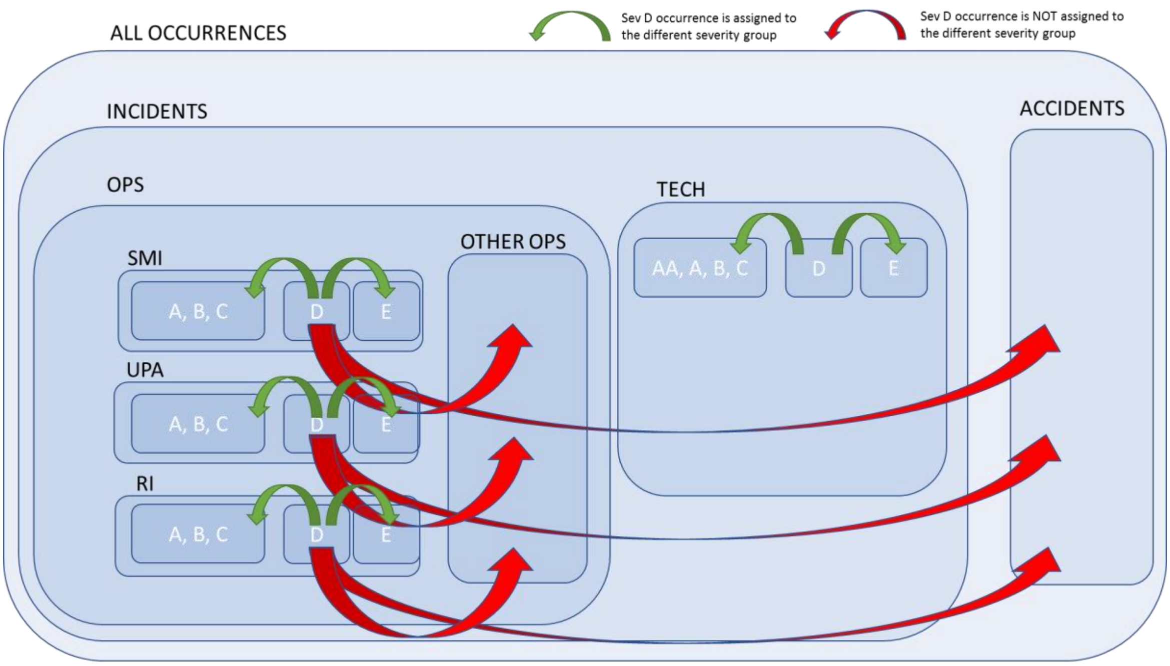


Figure 4: Severity D distribution into other severity groups

3.3.5 Estimated numbers of occurrences

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

$$No\_occ\_est_i = \left((No\_D_i * P_i) + No\_occ_i\right)$$

$No\_occ\_est_i$ *– 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:

$$P_i = \frac{No_{occ_i}}{Total\ No_{occ_i}}$$

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*

3.3.6 CRI calculation using 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, find which combination of weights 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*[1] was done in several stages:

1. Selection of Weights for accident, all OPS occurrences and TECH occurrences (Base Weights);

[1] Note that for definition of the *Weights*, which should explain the human perception of risk for each type of safety occurrence, 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 occurrences 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.).

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 techniques 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 were as in Table 2:

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 2: Safety Data component Weights

Using estimated number of occurrences and adding human perception of their risk (expressed as Weights) ICI for each entity was calculated separately. The simple formula to calculate ICI - Safety Data is presented below:

$$ICIdata_i = No_{occ\_est\_i} * w_i$$

$$ICIdata = \sum_i \frac{ICIdata_i}{Total\ N_{occ}}$$

$ICIdata_i$ – ICI 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 ICIdata 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 normalised ICIdata was calculated based on the following formula:

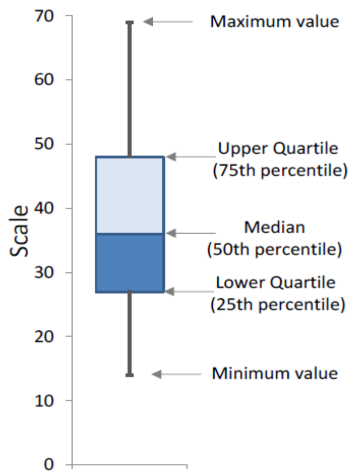
$$ICIdatanorm = \frac{ICIdata}{Number\ of\ Flight\ Hours}$$

The values of risk exposure for ICI Safety Data of all EUROCONTROL Member States in period 2015 – 2018 are presented in Figure 5.

Values are presented using BoxPlot, which is a standardized way of displaying the distribution of data based on a five number summary ("minimum", lower quartile (Q1), median, upper quartile (Q3), and "maximum").

Boxplots are useful as they show the average score of a data set; the skewness of a data set, is the data set symmetrical and how tightly is it grouped. Lastly, they show the dispersion (variability) of a data set.

Figure 5 shows that average value of ICI Safety Data is more or less stable over the four year period, however the number of outliers is reducing. Most importantly it can be seen that majority of the States have low Safety Data component risk index.



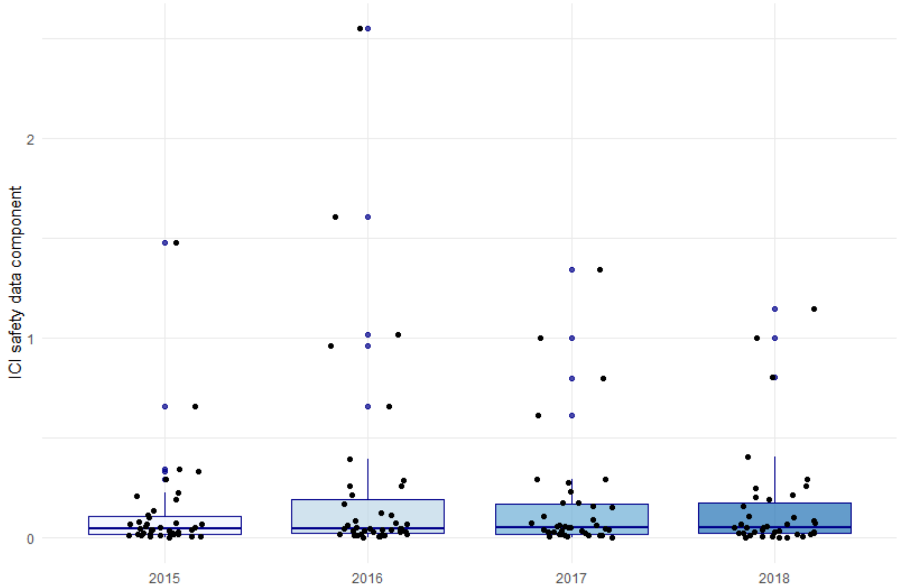


Figure 5. ICI of Safety Data

3.4 Traffic and complexity

As traffic volume and complexity are highly correlated and both describe local conditions, these two intermediate components were modelled together. Principal Component Analysis (PCA) was used to select parameters of both components that best describe Traffic and Complexity intermediate indexes. According to [8] PCA and Factor Analysis (FA) are the most frequently used multivariate statistical techniques used in the weighting of composite indices.

PCA analyses a data observations described by several dependent variables [9], which are, in general, inter-correlated. Its goal is to extract the important information from the data and to express this information as a set of new orthogonal variables called Principal Components (PCs). In simple words, the principle of PCA is to identify a small number of "averages" (PCs) that explain most of the variance observed in the data. Each PC represents a new variable computed as a linear combination of the original (standardized) variables.

The correlations between the original variables and the PCs are expressed as component loadings. Using the component loadings one can calculate component weights (for the different indicators) which are nothing but the proportion of component loadings to the sum of all the component loadings. These weights can then be used to compute component scores that are the desired composite indices.

Preliminary analysis showed that five (5) variables/parameters were highly correlated and should be kept for further modelling:

- two (2) traffic parameters, *flight hours* (total flight hours in a year) and number of movements (annual total number of airport movements), and
- three (3) complexity parameters:
    - *adjusted density*, parameter that assesses the potential interactions resulting from density, including uncertainty in the trajectories and time;
    - *structural index,* describing the potential number of interactions in specific situations classified as vertical, horizontal and mix of aircraft performances, that balances the density metrics according to the interaction geometry and aircraft performance differences, and;
    - *complexity score* (index describing the difficulty to manage the presence of several aircraft in the same area at the same time, particularly if those aircraft are in different flight phases, have different performances, and/or have different headings).

The standard method when applying PCA as a weighting technique is to use the factor/component loadings of the measuring indicators on the first component [10]. However, if the explanatory value of the first component is not sufficient to represent the data these methods do not lead to the construction of representative composite indices. To address this shortcoming, the method developed by [11] was used. This method considers the factor loadings of the first extracted component as well as the factor loadings of the consecutive extracted components to weight a composite index. The benefit of this method is that a higher proportion of the variance in the data set is explained.

Results of PCA using traffic and complexity parameters show that two (2) PCs are capable to describe ~ 85% of the sample (see Cumulative Var in Table 3).

To minimise the number of individual indicators that have a high loading on the same factor the varimax rotation is used. The idea behind transforming the PC axes is to obtain a “simpler structure” of the PC (ideally a structure in which each indicator is loaded exclusively on one of the retained PCs). Rotation changes the loadings and hence the interpretation of the PC, while leaving unchanged the analytical solutions obtained ex-ante and ex-post the rotation.

Results after varimax rotation	PC1	PC2	squared loadings (scaled to unity sum)	
flight_hours	-0.231	0.963	0.0229	**0.4707**
movements	-0.256	0.946	0.0281	**0.4543**
adj_density	-0.819	0.269	**0.2879**	0.0367
structural_index	-0.733	0.107	**0.2306**	0.0058
complexity_score	-0.952	0.253	**0.3890**	0.0325
	[,1]	[,2]		
Sum of squares (SS) loadings	2.233	1.970		
Proportion Var	0.447	0.394		
Cumulative Var	0.447	**0.841**		
Expl. Var.*	2.233	1.970		
Expl. Tot.**	**0.531287**	**0.468713**		

* is the variance explained by the PC
** is the explained variance divided by the total variance of the 2 factors

Table 3. Results of PCA for Traffic and Complexity

Individual measuring indicators with the highest loadings on a specific extracted component were grouped into intermediate composite indices. The weighting of each of the variables in the intermediate composite index is derived by squaring the loadings of the variables and scaling it to unity sum within each intermediate composite index. The squared loadings represent the proportion of the total variance of the indicator which is explained by the component.

Table 4 shows how *Traffic* and *Complexity* component Weights (normalized squared factor loadings) by using the coefficients of first two PCs, are calculated. For example, Weights for flight hours is calculated by dividing squared loading (of PC2 for flight hours) 0.4707 by total explained variance 0.468713 and lastly normalized.

Once the Weights of parameters have been defined intermediate indices can be constructed for each parameter. ICI is calculated by aggregating all intermediate indices for traffic and complexity.

	Not-normalized	Normalized
Flight Hours	0.220645	**0.240962**
Movements	0.212923	**0.232530**
Adjusted Density	0.152947	**0.167031**
Structural Index	0.122513	**0.133794**
Complexity Score	0.206656	**0.225685**

Table 4. Traffic and Complexity component Weights

Using *Weights* in Table 4, ICI for *Traffic* and *Complexity* for EUROCONTROL area in period 2015-2018 is shown in Figure 6 below.

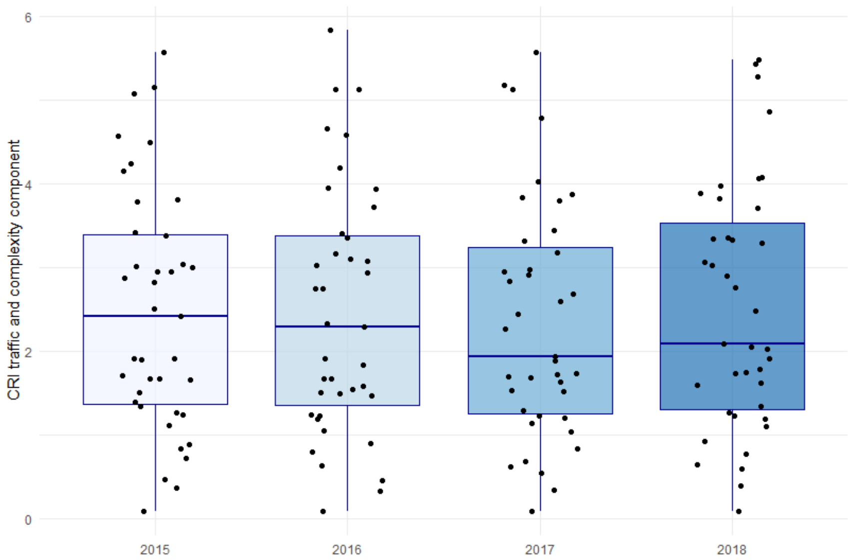


Figure 6. ICI for Traffic and Complexity

3.5 Reporting Practices

In order to take into account reporting culture within different entities, the third component considered in development of CRI had to, based on existing data available, extract information about both quantity and potentially quality of reporting. For that reason, three different variables are introduced:

- *Reporting Cluster*: determined using Cluster Analysis (CA) on total number of reported occurrences;

- *Reporting Ratio*: explained as share between High Severity Occurrences (A+B) and Low Severity Occurrences (C+E) and;
- *Reporting Rate*: calculated as total number of occurrences per flight hours.

Reporting Cluster parameter was estimated using CA to group information on EUROCONTROL Member States/entities based on their similarity in terms of total number of reported occurrences.

Cluster Analysis is one of the important data mining methods for discovering knowledge in multidimensional data. The goal of clustering is to identify pattern or groups of similar objects within a data of interest [12].

Prior to performing CA, the optimal number of clusters was determined. Most of the clustering algorithms depend on some assumptions in order to define the subgroups present in a data set. As a consequence, the resulting clustering scheme requires some sort of evaluation as regards its validity. To overcome this, NbClust R package was used, as it uses 30 indices which determine the number of clusters in a data set and it offers also the best clustering scheme from different results to the user obtained by varying all combinations of number of clusters, distance measures, and clustering methods [13].

Based on the analysis, selection of optimal number of clusters, was three (3). Using three different groups based on reporting "volume" (high, medium and low reporting), EUROCONTROL Member States Reporting Clusters are presented in dendogram in Figure 7. *Reporting Cluster* parameter, therefore, presents the first categorical variable used to calculate this intermediate index.

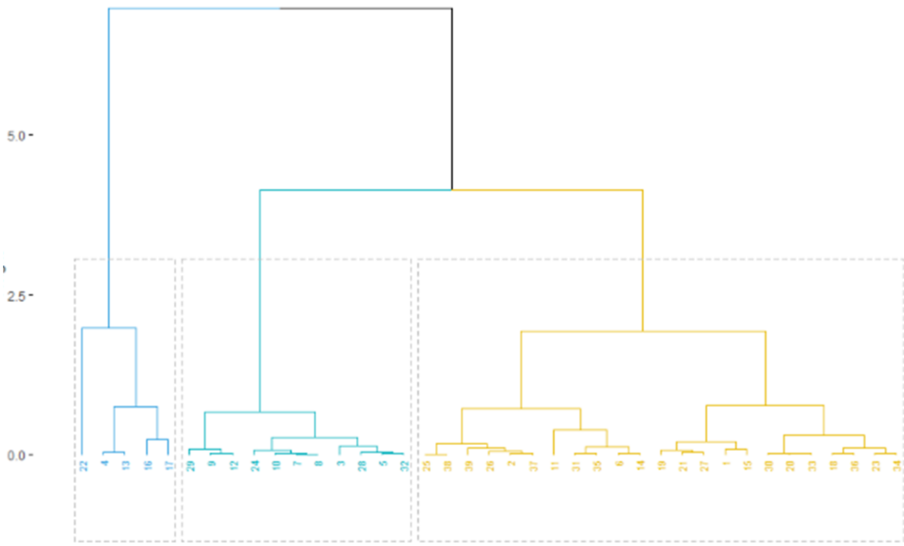


Figure 7. EUROCONTROL Member States per Reporting Cluster

The second and third parameter of *Reporting Practices* intermediate index were calculated directly from safety reported data available (reporting ratio and rate). *Reporting Ratio* was calculated as a share between High Severity Occurrences (A+B) and Low Severity Occurrences (C+E), whilst *Reporting Rate* was determined as total number of occurrences per flight hours.

Ultimately, the third intermediate index – ICI *Reporting Practices* was estimated using PCA of mixed data [14]. As not all parameters in this component are numerical, principal component analysis of mixed data was needed as a set of observations is described by a mixture of qualitative and quantitative variables (a mixture of numerical and categorical variables).

Results of PCA of mixed data (Table 5) show that three (3) PCs are capable to describe ~ 98% of the sample.

	Eigenvalue	Proportion	**Cumulative**			
Dim1	2.01997731	50.499433	50.50			
Dim2	1.02324984	25.581246	76.08			
Dim3	0.88968889	22.242222	**98.32**			
Dim4	0.06708396	1.677099	100.00			
Squared loadings after rotation						
	Dim1rot	Dim2rot	Dim3rot			
occ_per_flhr	0.95	0	0.02	0.47054	0.00000	0.00039
rep_ratio	0.01	0	0.99	0.00005	0.00000	0.96572
cluster	0.96	1	0.01	0.48050	0.99997	0.00010
Expl. Var.	1.91800	1.00003	1.01489			
Expl.Tot.	**0.48768**	**0.25427**	**0.25805**			

Table 5. Results of MCA for Reporting Practices

Reporting Practices component *Weights*, to be used to further compute this intermediate index are presented in Table 6.

	Weights	
	Not-normalized	Normalized
Reporting Cluster	3.93269	0.45518
Reporting Ratio	3.74237	0.43315
Reporting Rate	0.96486	0.11167

Table 6. Reporting Practices component Weights

Using selected weights intermediate index for Reporting Practices for EUROCONTROL area in the period 2015-2018 is shown in Figure 8 below.

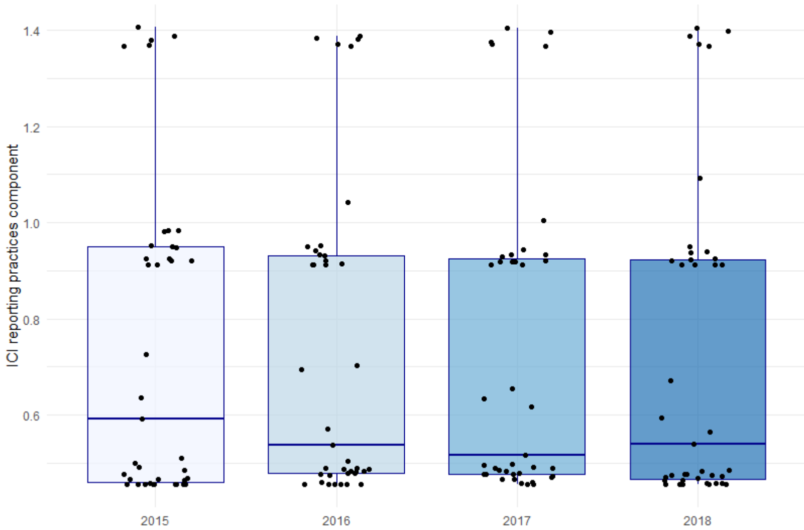


Figure 8. ICI for Reporting Practices

3.6 Composite Risk Index

The final CRI is calculated by aggregating all intermediate indexes weighted per their "importance" as per formula below:

$$CRI = ICI_{data} * W_{data} + ICI_{traffcompl} * W_{traffcompl} + ICI_{reppract} * W_{reppract}$$

where ICI_i *– intermediate index*

W_i – weight ("importance") assigned to specific CRI Component (see Figure 2)

CRI estimated values for all EUROCONTROL Member States in period 2015-2018 are presented in Figure 9. It can be see that composite risk value has been decreasing until 2017, however for some reason has increase in 2018 back to 2016 levels more or less.

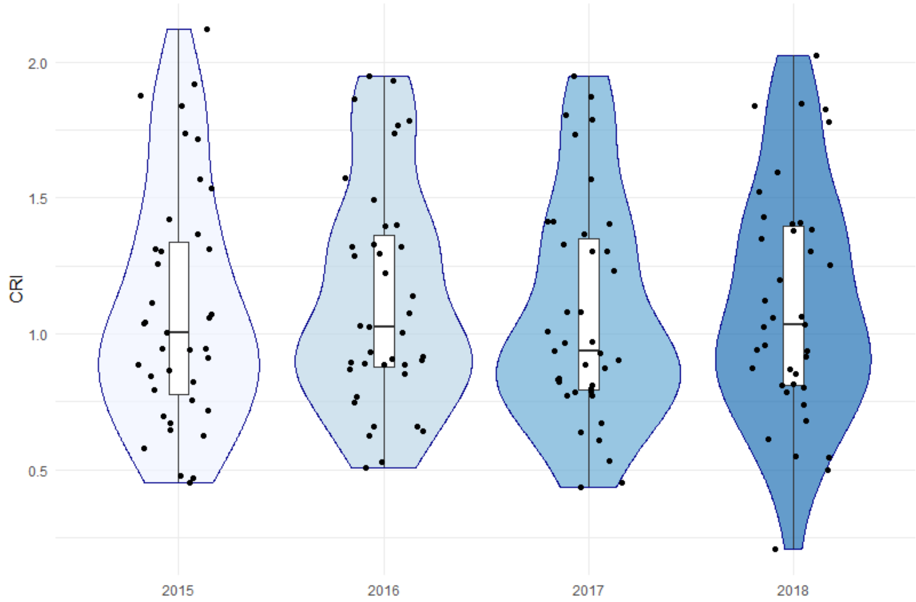


Figure 9. CRI for EUROCONTROL Member States (2015-2018)

This specific view is suitable for observations of CRI changes at network level between the years. However, if more information about variability between years is required per State, the next figure (Figure 10) provides more information. It can be seen that CRI for most of the EUROCONTROL Member States vary very little from year to year.

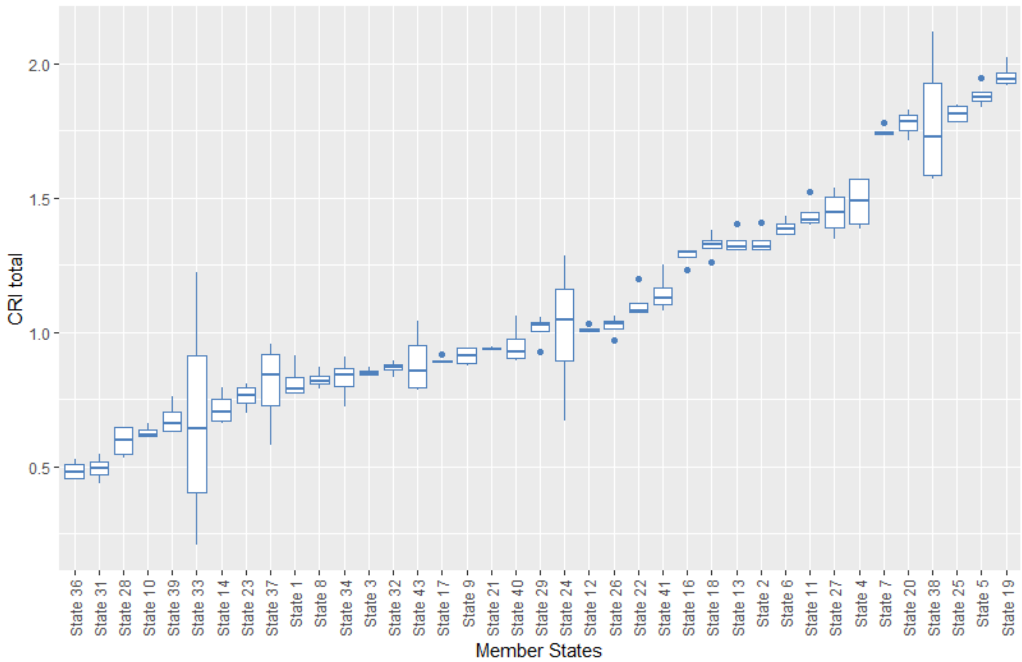


Figure 10. BoxPlot of CRI for all EUROCONTROL Member States (2015-2018)

Lastly, Figure 11 shows CRI in 2018 for all EUROCONTROL Member States in relation to the total number of reported occurrences. It can be seen that CRI not necessarily depends on the amount of reports and that inclusion of local conditions (such as reporting practices and airspace elements) prevent this potential distortion.

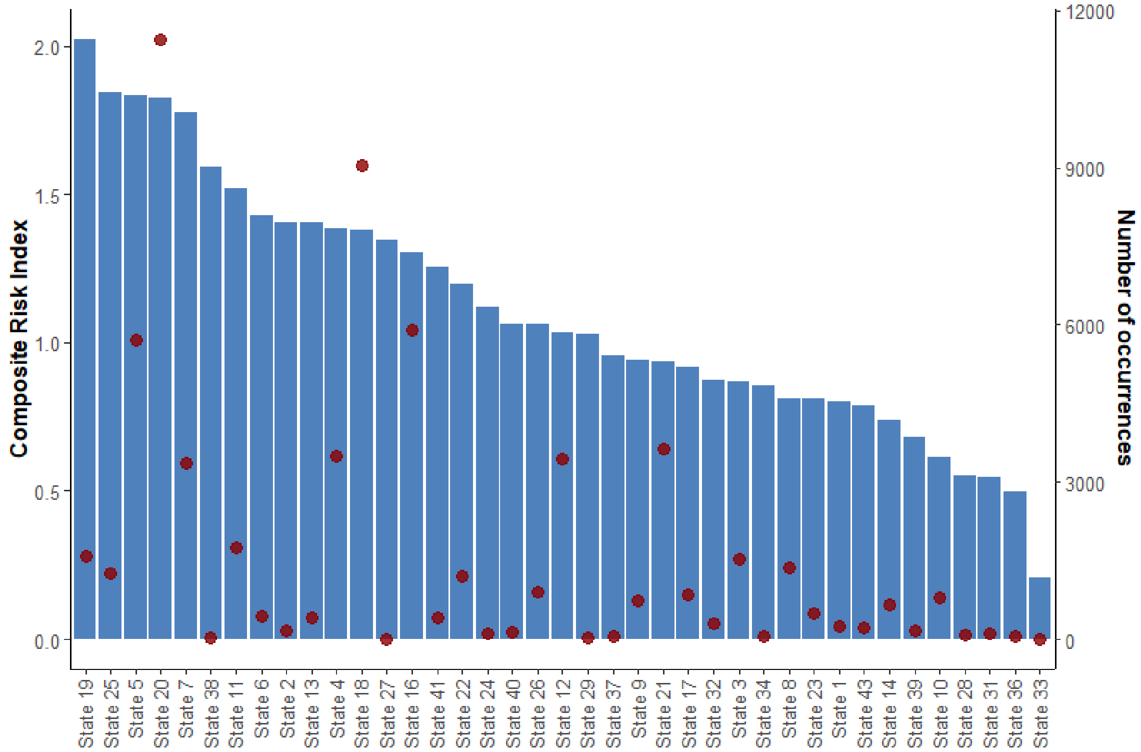


Figure 11. CRI for all EUROCONTROL Member States in 2018

3.7 Blind Benchmarking

Using CRI methodology it is possible to blind benchmark against other entities. CRI methodology also allows benchmarking on different levels and with scalable granularity.

Figure 13 shows how State 11 (blue dots) compares to other EUROCONTROL Member States using the overall CRI as benchmarking value in a period 2015 to 2018. CRI values of benchmarked Entity is presented as a blue dot. Figure 12 however, shows how State 11 (blue bar) compares to other states in a single year.

Benchmarking is also possible on a single CRI component, example of benchmarking of State 11 (blue dot) with other entities, for ICI *Reporting Practices* component, is shown on Figure 14. Moreover, benchmarking is possible on a single parameter (variable) of ICI. Figure 15 shows how State 11 (red bar) compares to other EUROCONTROL Member States in 2018, using risk index value of separation minima infringements (SMIs).

In conclusion, the CRI methodology has potential to be used for blind benchmarking of entities using different risk index values, either on the CRI component or ICI variable level.

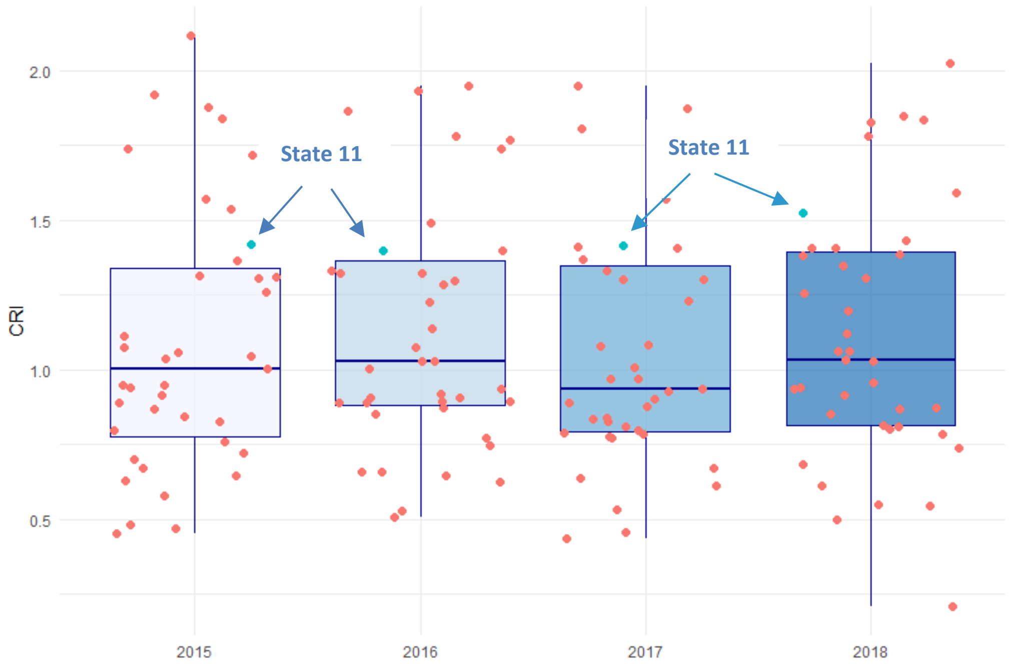


Figure 12. Entity blind benchmarking (trend)

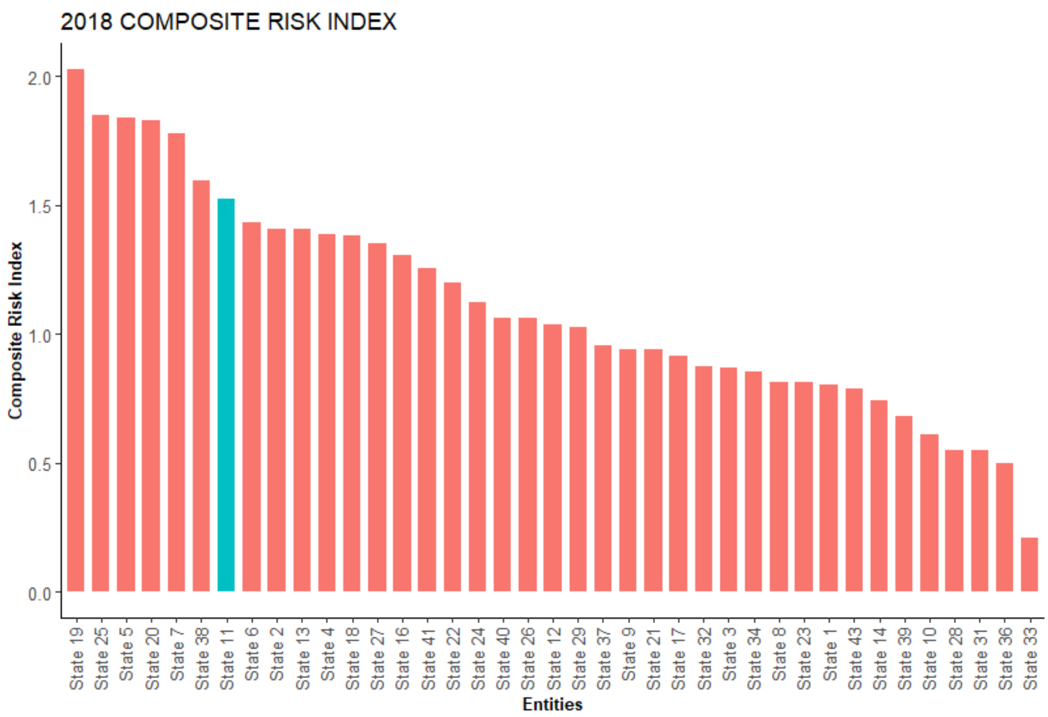


Figure 13. Entity blind benchmarking (single year)

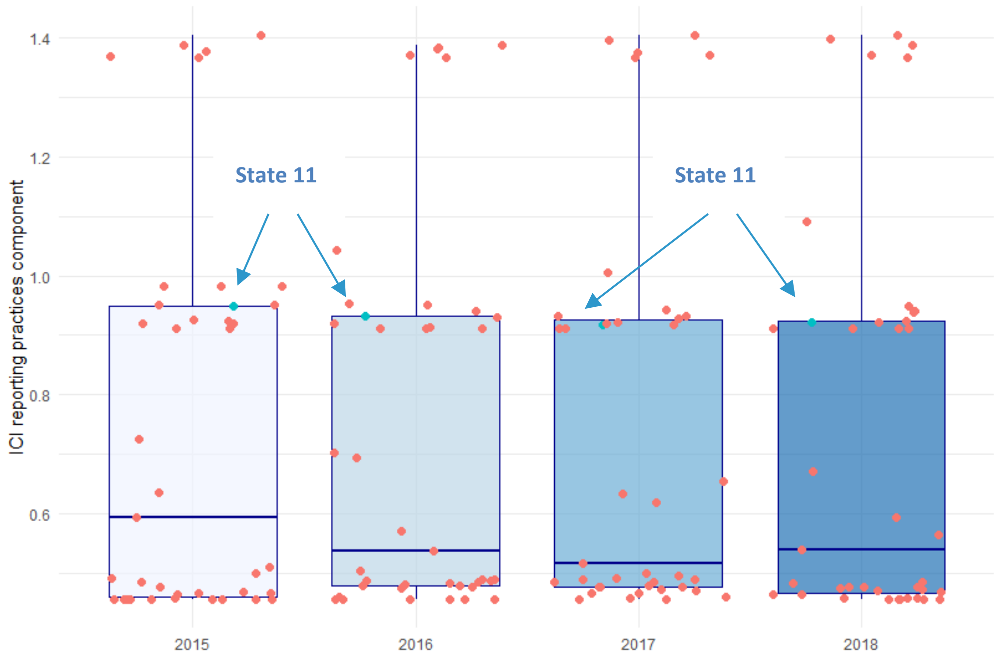


Figure 14. Entity blind benchmarking (single component /ICI)

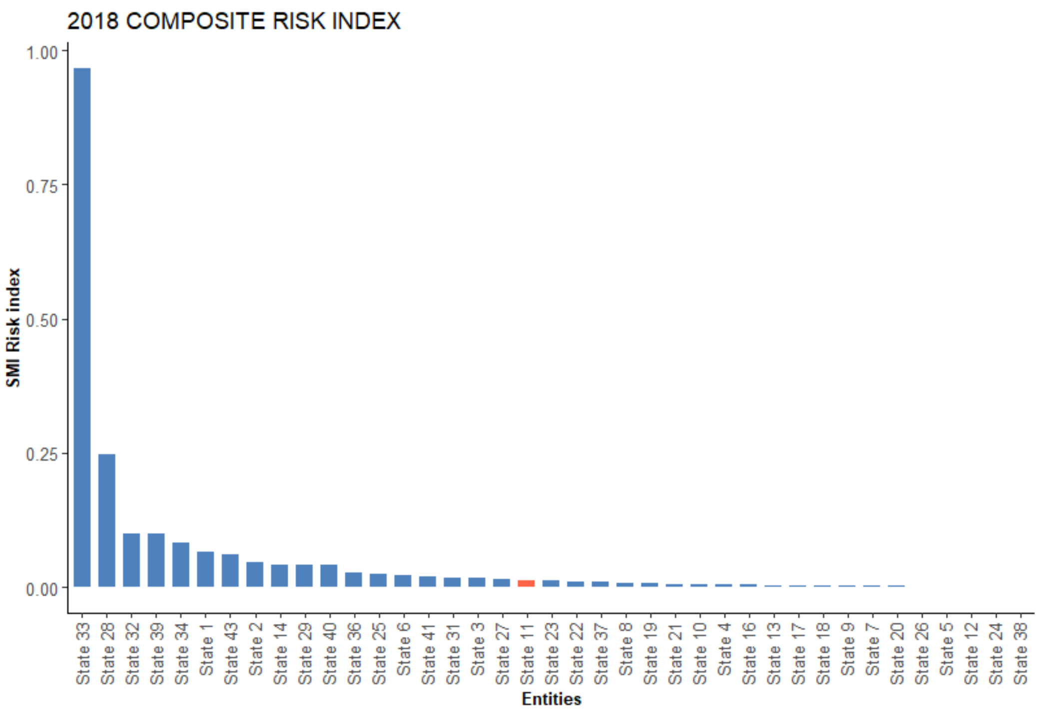


Figure 15. Entity blind benchmarking (single ICI variable)

4 SUMMARY

Composite index methods are increasingly recognized as a useful tool in policy analysis and public communication. They provide simple comparisons of units that can be used to illustrate complex and sometimes elusive issues. These indicators often seem easier to interpret by the general public finding a common trend in many separate indicators and have proven useful in benchmarking country performance.

In PRR 2018, the PRC presented the preliminary concept of CRI and corresponding methodology that could be used to measure the performance of the European ATM system as a whole or also its individual entities (service providers or Member States). The measure of risk exposure was based on probability and severity that considers the human perception of equivalent risk. The overall idea behind the CRI was that the performance of the safety system can be analysed in three important broad categories: the quality of the reporting system with reporting entity, measured risks within the system, and the human perception of risk. At that time, the PRC also has highlighted that there could be possibilities to further improve the CRI by considering specific local operating conditions, airspace size, capacity and/or complexity. As a result, three additional components, besides the original one (Safety Data), were included in the new methodology: Traffic, Complexity and Reporting Practices.

The concept of a CRI, as a cumulative risk value calculated using different components, has potential to become a proxy of exposure to risk within certain airspace for top management information and decision making.

The idea behind the CRI is to provide States, ANSPs, organisations with a composite index, as a tool for an easy overview to monitor and possibly compare the safety risks, either for one provider or for State to make an overview of annual developments. Ultimately, based on data availability, it can be used as a kind of benchmark between ANSPs, States or regions at one moment in time or over several years.

CRI is a relatively simple index figure that can be used as a thermometer for different underlying safety developments, by making overviews per single Entity or region, over single or multiple years. When differences become apparent, it will stimulate analysis of contributing aspects and their importance.

Overall, it is possible for States or ANSP to use the CRI locally by themselves, to monitor and compare safety risk, as the data for calculation is available either internally (such as occurrence reports and exposure data) or externally via PRC Dashboard (such as complexity). Nevertheless, using historical information available, the PRC can also offer to EUROCONTROL Member States confidential blind benchmarking up until 2019.

The nature of CRI methodology allows scalability of measurement and benchmarking. Preliminary analysis shows that it could be used to measure the performance of European ATM systems as a whole and also its individual entities. Moreover, this scaling possibility allows to measure CRI of individual components separately (e.g. only measure/benchmark ICI for reporting practices) or even more, benchmark entities comparing only individual types of safety occurrences. Overall, CRI could be used in any local environment to allow States or service providers to perform a blind benchmarking.

Due to its improved design and structure the new CRI methodology allows better ability to take into account local conditions, introducing the elements of traffic demand, airspace complexity, and reporting practices.

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.

4.1 CRI Customisation

One of the main benefits of proposed methodology is that CRI methodology is customisable to local environment (e.g. Weights can be re-modelled using local data) and it 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. In addition, as mentioned before additional analysis of the importance of weights assigned to the index components can be introduced to further investigate effects of weights on the overall CRI.

Lastly, the CRI normalisation could also be done per different metrics, in order to allow inclusion of airspace size and/or capacity (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

8. Booysen, F. (2002). An overview and evaluation of composite indices of development. Social Indicators Research,59 , 115-151.

9. Tabachnick, B. G. & Fidell, L. S. (2007). Using multivariate statistics. 5ed. Boston: Pearson education.

10. Klasen, S. (2000). Measuring Poverty and Deprivation in South Africa. Review of Income and Wealth, Series 46, Number 1(March), 33-58.

11. Nicoletti, G., Scarpetta, S. & Boylaud, O. (2000). Summary indicators of product and market regulation with an extension to employment protection legislation. Retrieved from www.oecd.org/eco/eco

12. Kaufman, L., Rousseeuw, P. J. (2005) Finding Groups in Data: An Introduction to Cluster Analysis. Wiley.

13. Charrad, M., Ghazzali, N., Boiteau, V., Niknafs, A. (2014) NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. Journal of Statistical Software, October 2014, Volume 61, Issue 6.

14. Chavent, M., Kuentz-Simonet, V., Labenne, A., Saracco, J. (2017). Multivariate analysis of mixed data: The PCAmixdata R package, arXiv preprint.