ATM Safety Management: Reactive and Proactive Indicators

Forecasting and monitoring ATM overall safety performance

Giulio Di Gravio
Department of Mechanical and Aerospace Engineering
Sapienza University of Rome, Faculty of Engineering
Via Eudossiana, 18 – 00184 Rome, Italy
giulio.digravio@uniroma1.it

Maurizio Mancini
Operations Directorate
ENAV s.p.a.
Via Salaria, 716 – 00138 Rome, Italy

Riccardo Patriarca
Department of Mechanical and Aerospace Engineering
Sapienza University of Rome, Faculty of Engineering
Via Eudossiana, 18 – 00184 Rome, Italy

Francesco Costantino
Department of Mechanical and Aerospace Engineering
Sapienza University of Rome, Faculty of Engineering
Via Eudossiana, 18 – 00184 Rome, Italy

***Abstract* - Defining means to assess safety performance and delve into their causes is one of the current and future challenges of the ATM sector. Following the experiences of the Aerospace Performance Factor by FAA and EUROCONTROL, this research aims to apply the Analytic Hierarchy Process (AHP) in order to build synthetic and user-friendly safety reactive indicators. Therefore, it describe the process for evolve these indicators to a proactive perspective, in order to forecast future safety performance. This concept has been possible through the development of a statistical model of safety events, combining in a Monte Carlo simulation the results emerged from the literature analysis with the analytical models of historic data interpretation. Through the analysis and combination of the safety events over time (accidents, incidents and issues) and the relative control charts, this model will pinpoint critical situations and will address the interventions of the decision makers.**

Keywords – ATM; safety; APF; AHP; simulation; Monte Carlo.

I. INTRODUCTION

Based on international regulations, a rugged and proactive process of addressing current and emerging safety risks must be planned out in order to ensure that air traffic development is carefully managed and supported through strategic regulatory and infrastructure development [1].

Historically, ANSPs used basic metrics as traffic counts, number of accidents and incidents to gauge safety performances. Anyway, these standard indicators fail to effectively represent the overall safety perspective and do not constitute a system-wide performance measurement tool. In October 2009, the EUROCONTROL Performance Review Commission (PRC) and the US Federal Aviation Administration (FAA) identified common information and performance indicators [2] that could be used for monitoring safety in each region. EUROCONTROL Safety Regulatory Requirements (ESARRs) proposed a standard occurrence reporting and assessment scheme. In particular, ESARR 2 [3] Appendix A (and B) provides the minimum contextual/factual ATM related (and no-ATM related) to be collected and recorded for each safety occurrence.

II. ATM SAFETY EVENTS' DATABASE

The core idea of ESARR 2 is based upon Reason Swiss Cheese model, which relates a system's failure to an alignment of all the metaphoric barriers' weakness, permitting "a trajectory of accident opportunity", so that a hazard passes through all of the holes in all of the defences, leading to a failure [4]. According to Reason's interpretation, it is clear how the frequency of accidents is not sufficient to describe safety performances. The Performance Review Commission NLR [5] used the metaphor of an iceberg to picture that accidents constitute a small but visible subset of occurrences, while incidents and less serious events constitute a larger, often invisible, subset of the iceberg. Therefore, reporting also less serious events gains a primary role in safety analysis. According to this point of view, it appears strictly necessary to build a safety database, which contains both the more critical safety events and the less critical ones. This database includes the event types, as prescribed in ESARR 2 Appendix A and the number of monthly occurrences, classified for airport and route.

III. SAFETY REACTIVE INDICATORS' BUILDING PROCESS

In order to evaluate also less serious events' contribution, firstly, FAA and US Naval Safety Center [6] understood that new ways to measure and improve safety performance would be necessary. In early 2006, with the contribution of EasyJet

[7], they developed a new methodology, the Aerospace Performance Factor (APF), which helped to fit this gap, using AHP [8].

The APF aims to aggregate multiple operational safety risks, expressed as the weighted sum of incidents into one single value capable of showing macro changes in performance trends. Although this unique value gives the overall risk, according to the methodology, it can be broken down into its components to analyze specific causal factors. Since November 2008, the SAFety data REPorting and data flow Task Force (SAFREP TF), together with the FAA, has investigated the APF process and formulated a EUROCONTROL APF, based on the requirements of ESARR 2 and associated to Annual Summary Template (AST) data.

The linear combination of the weighted events, normalized by the traffic count, generates the Safety Indexes (1-2):

$$Event_i\ APF\ Safety\ Index = \frac{Event_i\ annual\ count}{Total\ traffic\ count} Event_i\ AHP\ weight \quad (1)$$

$$APF\ Safety\ Index = \sum_i^n Event_i\ APF\ Safety\ Index \quad (2)$$

Note that the normalization has permitted comparisons of results that do not depend on the specific monthly movements but are gradable in a general context. AHP, indeed, makes possible to integrate tangible events (data and quantitative measures) with intangibles (general indications, experiences, estimations, qualitative evaluations of experts) to create an effective safety monitoring system that could take into account both perceptions and events.

Di Gravio et al. [9] used this mathematical development in order to define several different Safety Indexes, replicating ESARR 2 requirement of differentiating the flight phase: Airport (APT) and En Route (ENR). They developed also a further partition, according to the ATM's role in the event (ATM contribution and no ATM contribution). Table 1 defines the different Safety Indexes, according to their features.

TABLE 1. SAFETY INDEXES' STRUCTURE

	Airport (APT)	**En Route (ENR)**
All events	Safety Index 1 APT	Safety Index 1 ENR
ATM contribution events	Safety Index 2 APT	Safety Index 2 ENR

By way of example, Fig. 1 shows the structure of Safety Index 1 ENR that collects all the events regardless the contribution of ATM, highlighting also its main clusters. ESARR 2 Appendix A describes all the analysed safety events, according to the HEIDI (Harmonisation of European Incident Definitions Initiative for ATM) tool. Table 2 just summarizes the acronyms used in Fig.1.

IV. SAFETY PROACTIVE INDICATORS' BUILDING PROCESS

Even though future events cannot be known with absolute certainty, it stands no reason that assessing the probabilities of alternative scenarios has an important role in safety analysis, considering that understanding future possibilities could help to make better decision today. Therefore, the aim of this research has been the evolution of the four reactive Safety Indexes described in Table 1 from their strengthened historic value to an innovative perspective one, in order to forecast ATM system performance. This operation has permitted to adopt the same four Safety Indexes in a proactive way, making them capable of forecasting and monitoring future performance, giving awareness of future scenarios and even trying to anticipate and prevent accidents. In particular, both probability distribution functions and Time Series analysis -based on historic data and factor analysis- have been developed in order to respectively obtain probability distributions capable of describing events' -historic and future- behavior, and a safety forecast database. This database, which had the same structure of the one cited in section II, contains the forecast monthly number of occurrence of each safety event.

This database, which has involved forecast number of occurrence of safety events, has revealed itself worthwhile to apply Time Series analysis and APF in order to evaluate the same Safety Indexes described in Table 1 in forecast future scenarios. This goal has been achieved by implementing three methodologies, which have been based on Monte Carlo simulation, specifically customized to describe methods able to propagate uncertainties in an input model into uncertainties in an output one.

TABLE 2. SAFETY EVENTS' ACRONYMS USED IN THE DEFINITION OF SAFETY INDEX 1 ENR

ACCIDENTS		
Accidents	ACD	Accidents
EVENTS		
Near collision	NCFIT	Near Collision Flight Into Terrain
	SMI	Separation Minima Infringement
Potential for collision or near collision	DATC	Deviation from ATC Instruction
	DATS	Deviation from ATS Procedure
	LBS	Level Bust
	PSMI	Potential Separation Minima Infringement
	UPA	Unauthorized Penetration of Airspace
System failure	AIS	AIS Occurrence
	ASP	ATM Specific
	MET	Radar
ISSUES		
Procedural	EME	Emergency
	PRI	Priority Requested
External	BS	Bird Strike
	TRA	TCAS Resolution False
	OTH	Other
Communication	SCS	Similar Call Sign
	CSC	Call Sign Confusion
	PLCC	Prolonged Loss of Communication

A. *Historic fit*

By using historic data it is possible to draw distributions capable of statistically describing a phenomenon (e.g. a safety event), by the so-called "fitting distribution to data" process. The principle behind this process allows finding the distribution type and its parameters that give the highest probability of reproducing the observed data. Estimating distribution parameters by Maximum Likelihood Estimators (MLE), it is possible to define probability distribution functions

that maximize the joint probability of obtaining the given data set.

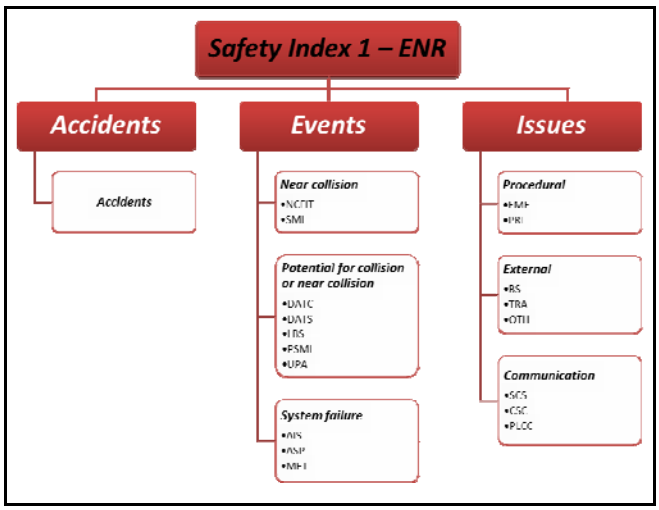

Figure 1. Safety Index 1 ENR structure.

By the way, it is quite rare that a variable could be represented by only one specific distribution type: distributions, even with subtle differences, can introduce significant variations in the resulting model. Thus, it is inescapable to fit several distributions to the data set comparing how well they fit the data with a quantitative criterion. Many goodness to fit indicators have been developed [10] over time but the Akaike Information Criterion (AIC) has been used to reach the aim of this research. Given a set of candidate distributions for the data, the preferred model is the one with the minimum AIC value, as described in (3).

$$AIC = 2k - 2\ln(L) \tag{3}$$

where k is the number of parameters in the statistical model and L is the MLE for the estimated model.

B. *Time series analysis*

Much of statistical methodology is concerned with models in which the observations are assumed to vary independently. A great deal in engineering occurs in the form of time series where observations are dependent and where the nature of this dependence is of interest in itself. The most common patterns are increasing or decreasing trend, cycle, seasonality and irregular fluctuations. The core idea of the analysis of such series is the use of available observations at time t in order to forecast their value at some future time $t+l$. Therefore, for this research, observations have been supposed to be available at discrete, equispaced time intervals (months).

In this research, the number of events z_t has been the Time Series process related to the safety database.

The number of events in the previous months has been indicated by z_{t-1}, z_{t-2}, z_{t-3},…, z_{t-m}, where m is the dimension, in months, of the safety database, (e.g. 36 months). These data might be used to forecast events for lead times $l=1, 2,\ldots, 12$ months ahead. According to Yule's formulation [11], Fig. 2 shows the implementation of the so-called linear filter, useful for obtaining the process z_t from a white noise at input.

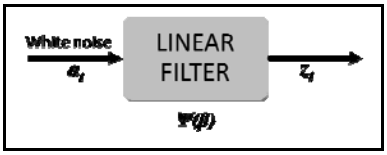

Figure 2. Schematic representation of a time series as the output from a linear filter, which receives in input white noise.

According to Box and Jenkins' theory and depending on z_t the best fitting model could has been either autoregressive or moving average or mixed autoregressive moving average models, just defining different filter's structures. For each event, the models have been compared according to their mean absolute error (MAE), in order to select the one which better fits.

C. *Causal fit*

Causal analysis painstakingly evaluates a number q of real variables smaller than the total number p of variables that are observable arising as linear combinations of the q factors just indicated. The core idea of this process is that it is possible to continuously compound a distribution by mixing various distributions.

The factor analysis has started by examining a specific event in order to focus on its causes [12]. Each cause has been represented by a standard defined distribution. These distributions, accurately weighted, have compounded the whole event by a mixing process [13].

Even though new tools should be developed to help the analyst in factor analysis [14] research evaluates three causal event sequence scenarios like technical, human and organisational factors, starting from Edwards' research [15]. In particular, this paper evaluates human factor, equipment and procedures as main responsible for safety events, arising from Safety Regulation Commission's review of accidents/incidents historical data [16].

In this approximate first step analysis, the weights assigned to the distribution of each safety event have been defined by a literature overview for each specific occurrence. Furthermore, investigation results on past events, as prescribed by ICAO [17], can help to assess the frequency of each cause to refine its weight. In order to enhance the accuracy of the research, a pool of ATCOs has been involved in the evaluations of the phenomena with less historic data. The accuracy of this approximation will be evaluated by the accuracy of the results, which pretends to represent each safety event behaviour. Then for each safety event type, each single cause's distribution has been modelled as it should be capable of completely describing the event. Basing on logic remarks, the process has started by the assignment of a qualitative distribution to each cause, at first ignoring its quantitative parameters. In order to determine them, it could be of service to reason about a borderline case. If the causal analysis determined that a particular event has dependence exclusively by one of the factors (e.g. Equipment),

then the distribution of this factor (Equipment) should have been able to perfectly reproduce the historic data of that event.

By repeating this reasoning for each cause, it has been possible to define the parameters of each distribution, capable of defining each cause. Afterwards, in order to characterize how much each cause was responsible for an event's happening, it has been necessary to evaluate each event type by a causal logic. Finally, these weighted distributions have been implemented in the mixture model in order to obtain an assembled distribution comparable to historic one.

A gap has been generally found by measuring the obtained distribution against the historic one. This value corresponded to the differences between what happened and what has been reported to be happened. This fact might reflect both the human behaviour in reporting [18] and the completely random events that can occur. In particular, in order to avoid under-reporting or wrong-reporting, a good reporting culture should be developed [19], considering the penalization of employees only when it is necessary (e.g. a willful act, the deliberate contravention of a correct procedure, several repeated mistakes that cannot be corrected by coaching, etc.).In order to fit this difference, it has been mandatory to add a fourth distribution, known in literature as noise, capable of representing these deficiencies of the causal model, as in (4).

$$f(x)=\lambda_{HF}\, f_{HF}(x)+\lambda_{EQUIP}\, f_{EQUIP}(x)+\lambda_{PROC}\, f_{PROC}(x)+\lambda_{NOISE}\, f_{NOISE}(x) \quad (4)$$

Thereafter by assigning a new weight to the noise and evaluating the compound distribution again, an iterative method has been implemented until the difference between the two distributions assumed a value near enough to zero, as sketched in Fig. 3.

The difference between the two distributions has been measured by evaluating the difference according to the AIC parameter, evaluated for the historic distribution and for the causal one. Note that another possible cause of the obtained gap might be due to the probability distribution parameters, which might not precisely represent the real situation. The mixture model, because of its merely mathematical nature of compounding other distributions is assumed to be correct.

- Human Factor

Human plays an important role in the operation process of most plants or systems and it can be often considered a root or significant contributing cause of system failure. This observation has led to the development of a range of methods under the general heading of human reliability analysis (HRA) to account the effect of human error in risk and reliability analysis [20]. Teperi and Leppänen [21] summarized the most relevant ones, highlighting the primary role of Human Factors in creating and managing safety in complex systems over years.

For this research, it has been useful to define a standard distribution capable of describing human errors during time. Thackray and Touchstone [22] have performed a monitoring task under relatively high task load for a time interval, observing a performance decay due to overload. Therefore, an exponential distribution could represent the waiting time for the first occurrence of a process that is continuous in time and of constant intensity. Note that, according to other studies it could have been selected also a Weibull [23] or log-normal distribution, without consequences on the experimentation. According to Human Factor qualitative logic, as summarized by Shorrock [24], an error should happen with a probability that increases over time, and it seems to arise from limitations in attentional resources due to stress, workload, organizational issues, handover, etc. Then, in the time domain human factor could have been represented by an increasing exponential curve, otherwise in the variable domain the same curve have become, indeed, a decreasing one.

- Equipment

It has been possible to include among the concept of Equipment all the items and the tools of the ATM system, according to a fully generic point of view. At this level, discerning each part of our system and studying its failure properties (e.g. through Weibull distribution) has not revealed itself particularly interesting. Looking at items' failures as a combination of many random independent factors on a set of independent complex systems [25] has been satisfactory enough for choosing both Normal and Lognormal distributions, which furthermore represent a constant failure rate.

Since Blanks [26], it is commonly assumed that a constant failure rate facilitates quantitative estimates of equipment failure probabilities and may introduce errors just in the region of early infant mortality or close to wear-out, conditions that go beyond the purpose of this research. In particular a Lognormal distribution has represented the best choice, because of its boundary values ($\mu>0$), its lower mean value and its capability to describe lifetime behaviour of components governed by fatigue processes and wear-out mechanism.

- Procedure

A longer term key risk area concerns airspace complexity and its impact on Air Traffic COntrollers (ATCOs)'s performance, traceable to some potential pre-disposing factors (short-sector and vertically changing traffic; traffic volume; communication congestion; training on the job). Via the FAA-EUROCONTROL Action Plan on Safety, EUROCONTROL, NATS, FAA and NASA have carried out a valuable work entailing extensive measures of procedural complexity [27]. A triangular distribution best models this core idea, because of its aptitude for expressing the mode of the distribution and its ease of application. A causal role has been assigned to procedures just to describe the effective complexity related to certain particular processes.

It is generally agreed that airspace complexity has both dynamic and static components, both of which can effect controller workload and influence the probability of occurrence of an ATC error. Dynamic complexity can include factors such as traffic volume, climbing/descending traffic, mix of aircraft type, military area activity and types of aircraft intersection. Static factors, meanwhile, encompass airspace structure, proximity of reporting points to sector boundaries and standing

agreements. All of the above complexity factors can potentially have an impact upon the safety of the ATM system.

For the purpose of this study, indeed, the mode is surely zero because it has been assumed that Procedures are almost lacking of internal errors. Therefore, in order to have a distribution capable of describing the safety events, it has been possible to use a simple set of standard parameters for minimum, maximum, mode (*0; 3μ; 0*).

- Noise

Safety events' reporting assumes a crucial role for monitoring and forecasting safety. Unfortunately, the majority of EUROCONTROL Member States deals with insufficient [28], inappropriate or incorrect reporting procedure.

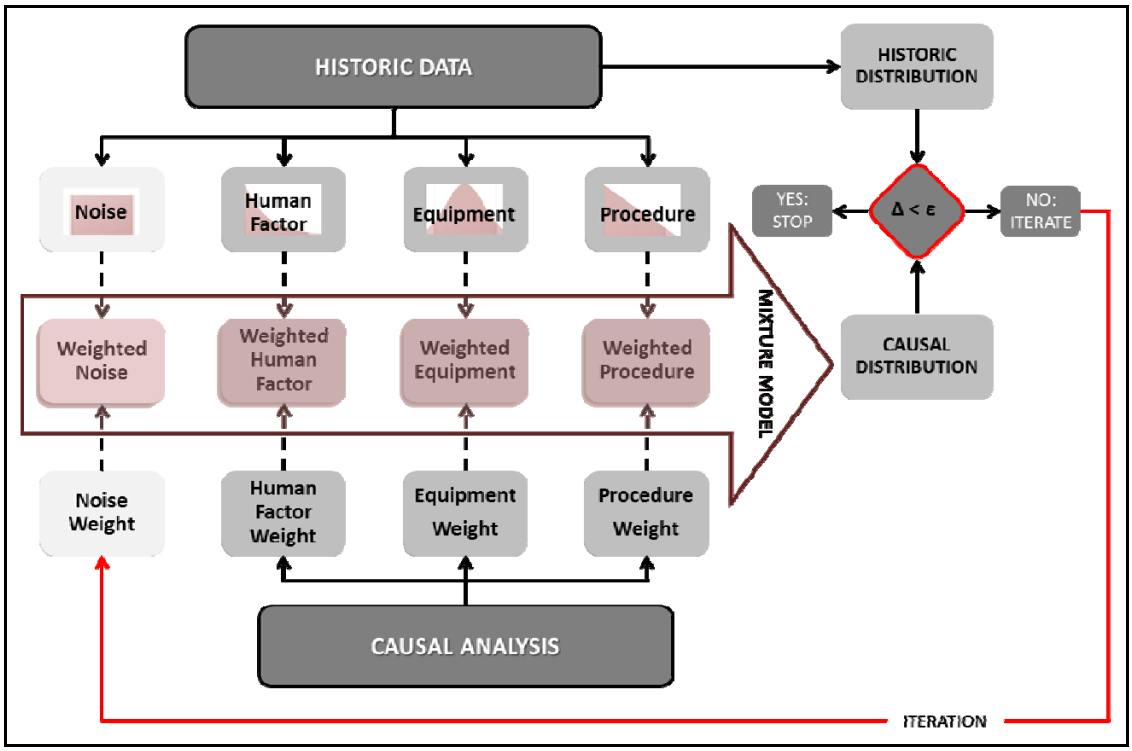

Figure 3. Causal fit application

A Uniform distribution best models the awareness that a reporting lack is equally probable over the considered time interval, making no further assumptions about the distribution structure. If the data showed any central tendency, the assumption of a Uniform distribution would lead to a higher standard deviation that might be appropriated. This could be seen as a conservative approach.

For certain, the quantitative weight of reporting lacks and truly random occurrences and so, in wider terms, of Noise distribution, has been determined by iteratively repeating the comparing process between the causal distribution and the historic one. Note that the Noise distribution includes also a potential error due to the distributions that represent the occurrences. The level of accuracy of the statistical model, indeed, is related to the Noise's weight. The iterative procedure sets this weight in order to allow the minimum value required to data fitting.

V. Case Study: Application to the Italian ATM System

In order to continuously improve operational safety, ENAV (the civilian Italian Air Navigation Service Provider) decided to adopt the above-described APF techniques. Based on Italian air traffic count, ENAV's APF succeeded in obtaining a well-structured database that has been the basis of current safety monitoring process and the starting point for the implementation of this evolving research. In order to reach practicality, this model has required a three years database (2008-2010), with the purpose of gaining one year forecast (2011), thus reaching a compromise between Box and Jenkins' formulation and available data. The time interval has been selected in order to compare the forecast data to the real ones and obtain in this way an evaluation about the model's reliability.

A. Preliminary assumptions

The Pareto principle has been adopted in order to increase the analysis' robustness and identify the top portion of causes that needed to be addressed in order to resolve the majority of

problems. In this research, it has been helpful to address and manage the most relevant safety events that have the most important impact on the system. The impact of each safety event has been evaluated through the risk consistent with its happening i.e. the likelihood of the risk being realised and the impact if the risk is realised. A risk/tolerability matrix has been built, and all the safety events have been evaluated to select, among them, the ones to analyse in order to address the top portion of risks. In order not to break ENAV intellectual property about safety criticality, the events to which the analysis of this paper has been applied, have been selected just on the basis of their frequency of occurrence, neglecting their risk impact Note that this assumption does not influence the testing results of the model and it is however in accordance with the iceberg of safety theory. Therefore, the safety events have been divided into three categories (respectively A :> 90%; B :> 80%; C :< 80%; quantifying each specific event's impact on the system as a whole) and the analysis has been accomplished only on the first two classes (A, B), neglecting the contribution of the less impacting class of events (C).

The proposed model has been entirely based on Monte Carlo method so the minimum number of iterations to run for a particular results' accuracy has been established by a probability-based approach [29], which relies upon the Central Limit Theorem and the pivotal method. Equation (8) specifically defines the required number of iterations' expression:

$$n > \left(\frac{\sigma \, \phi^{-1}\left(\frac{1+\alpha}{2}\right)}{\delta} \right)^2 \qquad (8)$$

This approach has been evaluated on the event that had less correlated historic data, which is furthermore one of the more frequent events, i.e. TRA (TCAS-related issues). In order to ensure a required level of accuracy ($\delta=0,5$) with a defined confidence ($\alpha=95\%$) for a standard deviation σ, the method has needed at least 609 iterations, increased to 1000 in a fully conservative approach.

Di Gravio et al. [9] present the implementation of the operations described in § IV.C to a specific phenomenon. In this paper, evaluating each element of the branches comprising a collection of the hierarchy and summing their score, the relative contribution of the element and the branch to the overall index have been "rolled-up" and a comprehensive system-wide performance measurement tool have been determined.

B. Reactive safety: global results

Fig. 5-7 represent the final step of APF implementation. Once the organizational factors that influence performance have been determined and analysed in a structured mind map and the relative importance of each factor has been evaluated by AHP, indeed, it is mandatory to display information for decision makers. Information need to provide a comprehensive and intuitive picture of organizational safety performance, graphically displaying the weighted mind map values and its changes over time. The pictured Safety Index is the time series of the reported and forecast events, weighted and summed up in a global index, in order to represent in a synthetic view the safety performance of the ATM system.

Considering this approach, safety performances can use the same principles of quality control analysis, following standard sequence Juran's framework (Fig. 4):

- choose the control subject (air transport safety performance);
- choose a unit of measure (weighted and normalized number of occurrences);
- set a goal for the control subject (EUROCONTROL's regulations);
- create a sensor to measure the control subject in term of the unit of measure (APF Safety Indexes: safety decreases when the Index increases);
- measure actual performance (APF monitoring);
- interpret the difference between actual performance and the goal (critical comparison);
- take action (if any) to fill the gap.

Because of the stochastic nature of the process, Shewhart control theory helps to monitor and control unexpected tendencies, irregular sequences and values out of the lower or upper control limits.

Figure 4. Quality control's sequence of steps

Given the characteristics of the safety process, it is possible to apply an individual/moving-range X-R control chart, according to a specific norm [30]. Fig. 5 shows an example of application over a specific ENAV's safety database for 2008-2011.

Any deviations from the normal behavior of the variable represent a warning device, capable of identifying unsafe conditions to intervene. In all these cases, it is possible to analyse which factors could have caused those increments and decide to adopt specific actions in order to reduce their contribution to the Safety Index's value and guarantee safety's enhancement. Fig. 5 shows the critical value of June 2009 that lies over the upper control limit and deserves a drill down analysis.

C. *Proactive safety: global results*

Fig. 6 therefore shows a sample path for time series process and two probability limits (5%; 95%) prescribed by historic and causal fit. Fig. 6 also highlights the three macro components of the Safety Index i.e ACD (accident), INCD (incident) and ATM Issues.

Fig. 7 shows the comparison between forecast index and the index based on real data, and two limits set by causal fit. The boundary set values define a probability of 95% to obtain the causal index within those values. The causal limits related to (e.g.) Safety Index 1 ENR have been largely satisfactory, globally displaying only one point out of control (as shown in Fig. 7, February 2011). The same analysis on the sub-components of the Index shows trends with the same level of accuracy, confirming the performance of the model.

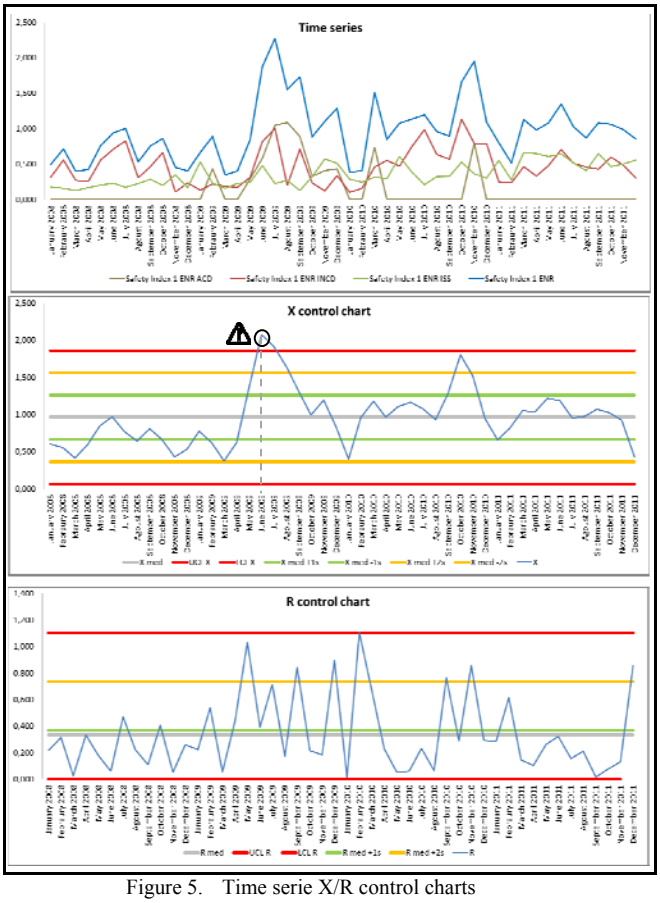

Figure 5. Time serie X/R control charts

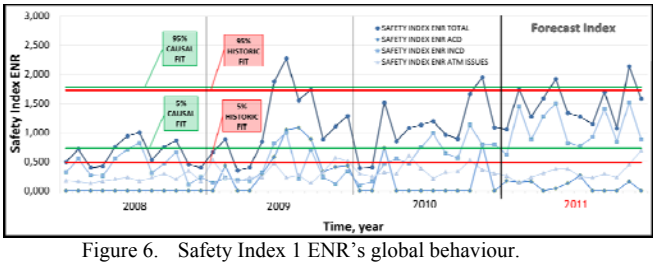

Figure 6. Safety Index 1 ENR's global behaviour.

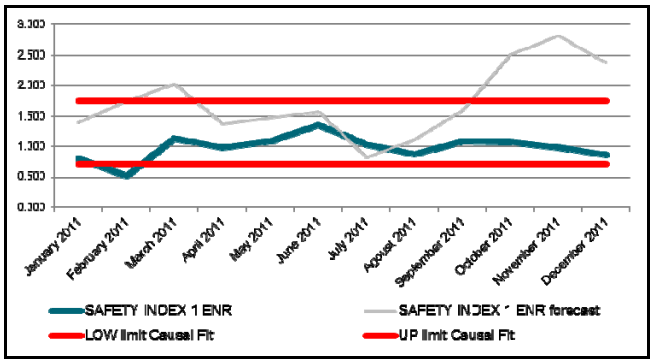

Figure 7. Safety Index 1 ENR's historic, time series and casual fit results' comparison for 2011.

VI. CONCLUSIONS

This entire research aimed to achieve an ambitious quality-related target about air transport sector: be able to evaluate, monitor and forecast ATM safety performance in order to give user-friendly results for ATC decision-makers, as prescribed by European regulations. By adopting X-R control charts, this paper has proved that the developed Safety Indexes are useful for monitoring and individuating critical situations in a reactive manner.

It is important to underline that the decision-makers have to complete the analysis, considering an in-depth analysis of the historic data for individuating the causes that leads to have that point out of control.

In order to evolve this point of view, the three methodologies implemented have given significant and remarkable enhancements for a proactive perspective.

In particular, Time Series analysis has been useful to describe future trends, based on historic results. Its mathematical construction has been however very laborious and it might require further study based on Fourier's series in order to detect the seasonal components of each event type's historic series and give more accurate results.

Historic fit has been instead able to accurately describe the historic events through a probability distribution function and it has proved itself able to satisfactorily summarize the historic ATM safety database.

The major innovation of this research has however surely been causal fit. This original methodology has allowed to obtain a probability distribution function, which would have been able to analyse past and forecast future safety events, with relevant effect on future air transport's quality control.

Causal fit, in effect, sets limits to Safety Indexes' behaviour according to an in-depth causal analysis of each safety occurrence. Therefore, if the weighting factors were impeccable and the database was flawless, then causal fit should represent what effectively would happen. In this case, comparing causal forecast indexes to real ones, any deviation from the set limits should be regarded very closely.

These deviations could be surely seen as a warning device, capable of identify those points which lie over of the upper causal limit, as unsafe conditions.

In this case, an in-depth analysis should be necessary in order to evaluate which factors could have caused those increments and decide to adopt precise actions in order to reduce the Safety Index' value and guarantee safety's enhancement. The specific index could be broken into its different branches in order to analyse each event type by specific tools. Once the necessary actions have been identified and performed, a new causal analysis should be developed, evaluating the effectiveness of the corrective undertaken actions and, if anything, iterate this procedure.

On the other hand, if the real data should lie beneath the forecast lower limit, there would be two possible eventualities. Firstly, it could be due to the fruitful consequences of eventual ATC improvement's actions. Anyway, if no actions have been performed a further investigation of the causes of that safety's enhancement could be required.

It might be possible that the Safety Index's decrement could represent only a reports' decrement. In this case, there would be mandatory a series of focusing actions on ATCOs in order to develop safety culture in the ATC structures [17] and avoid lacking reports.

Lastly, comparing the causal analysis to the historic one, it might be possible a database screening, neglecting those years which should suffer from lacking reporting. Therefore, according to the specific results presented in Fig. 6, 2008 database deserves a further study. In particular, its higher value of safety performance could be due to the fruitful consequences of safety improvement's actions or simply to a reports' decrement. Looking at this specific case study, because no safety actions have been developed in 2008 and even, 2009 has been the first year of new reporting procedures and Just Culture development, 2008 could be neglected because it should represent a dataset generated by different reporting process.

[1] ICAO, "2013-2028 Global Air Navigation Plan, 4th ed.," Montreal, 2013.

[2] EUROCONTROL & FAA, "U.S./Europe comparison of ATM-related Operational Performance," Bruxelles, 2012.

[3] EUROCONTROL, "Reporting and assessment of safety occurrences in ATM-ESARR 2", 2009.

[4] J. Reason, "The contribution of Latent Human failures to the Breakdown of Complex Systems", in Philosophical Transactions of the Royal Society of London, London, 1990.

[5] E. Hollnagel, "Industrial accident prevention: a scientific approach. Safer Complex Industrial Environments: A Human Factors Approach", McGraw-Hill CRC Press, 2009.

[6] Futron Corporation, "The Aerospace Performance Factor, How the APF is changing the way aviation views safety information", 2010.

[7] Lintner, T.M., Smith, S.D., Smurthwaite, S., "The aerospace performance factor: utilization of the analytical hierarchy process to develop a balanced performance and safety indicator of the national airspace system for the federal aviation administration", Proceedings of the International Symposium on the Analytic Hierarchy Process (ISAHP 2009), 29th July - 1st August, Pittsburgh, USA, 2009.

[8] Saaty, T. L., "Decision Making with the analytic hierarchy process", International Journal of Services Sciences, 1 (2008), 2008, pp. 83-98.

[9] Di Gravio, G., Mancini, M., Patriarca, R., Costantino F., "Overall safety performance of the Air Traffic Management system: indicators and analysis", Journal of Air Transport Management, forthcoming, 2014.

[10] Akaike, H., "A new look at the statistical model identification", IEEE Transaction on Automatic Control, vol. 19 (6), 1974, pp. 716-723.

[11] Yule, G.U., "On a method of investigating periodicities in disturbed series, with special reference to Wolfer's Sunspot Numbers", Philosophical Transactions of the Royal Society of London, ,1927, pp. 226- 267.

[12] Ale, B.J.M., Bellamy, L.J., Cooke, R.M., Goossens, L.H.J., Hale, A.R., Roelen, A.L.C., Smith, E., "Towards a causal model for air transport safety - an ongoing research project", Safety Science, vol. 44 (8), 2006, pp. 657-673.

[13] Shalizi, C.R., "Mixture model", advanced data analysis from an elementary point of view, University of Michigan, 2014. Accessed online in June, 2014, http://www.stat.cmu.edu/~cshalizi/ADAfaEPoV

[14] Roelen, A.L.C., Lin, P.H., Hale, A.R., "Accident models and organisational factors in air transport: The need for multi-method models", Safety Science, vol. 49 (8-9), 2011, pp. 1187-1199.

[15] Edwards, E., "Man and machine: systems for safety", Proceedings of British Airline Pilots Associations Technical Symposium, Airline Pilots Association, London 1972, pp. 21-36.

[16] EUROCONTROL, "ATM Contribution to aircraft accidents/incidents. Review and analysis of historical data", Safety Regulation Commision Document, SRC DOC 2, 2005.

[17] ICAO, "Aircraft Accident and Incident Investigation", Annex 13 to the Convention on International Civil Aviation, 10th ed, 2010.

[18] EUROCONTROL, "EVAIR Safety Bulletin No 8". Summer Seasons 2008-2011, 2012.

[19] INSAG, "Key practical issues in strengthening safety culture (No INSAG-15)". Vienna: International Nuclear Safety Advisory Group, 2002.

[20] French, S., Bedford, T., Pollard, S.J.T., Soane, E., "Human reliability analysis: A crititque and review for managers", Safety Science, vol. 49 (6), 2011, pp. 753-763.

[21] Teperi, A.M., Leppänen, A., "Managers' conceptions regarding human factors in air traffic management and in airport operations", Safety Science, vol. 49 (3), 2011, pp. 438-449.

[22] Thackray, R.I., Touchtone, R.M., "Effects of visual taskload on the behaviours involved in complex monitoring", Ergonomics, vol. 32 (1), 1989, pp. 27-38.

[23] Chen, L., Zhang, Y., Zhang, L., "The study of human reliability analysis based on dynamic bayesian netwroks theory", The 22nd International Offshore and Polar Engineering Conference, 17-22 June 2012, Rhodes, Greece.

[24] Shorrock, S.T., 2007. "Errors of perception in air traffic control", Safety Science, vol. 45 (8), 2007, pp. 890-904.

[25] Ahmad, R., Kamaruddin, S., Mokthar, M., Almanar, I.P., "Identifying the best fit distribution and the parameters of machine's component: a new approach", Proceedings of International Conference on Man-Machine Systems, (15-16), Malaysia, 2006.

[26] Blanks, H.S., "The generation and use of component failure rate data", Quality Assurance vol. 3 (85), 1977.

[27] Kirwan, B., Perrin, E., Gandolfi, C., and et al., "Preliminary Results from a European Safety R&D Program", EUROCONTROL Experimental centre, CEDEX, France, 2005.

[28] Wilke, S., Majumdar, A., Ochieng, W.Y., A framework for assessing the quality of aviation safety databases, Safety Science, vol. 63, 2014, pp. 133-145.

[29] Driels, M.R., Shin, Y. S., "Determining the number of iterations for Monte Carlo simulations of weapon effectiveness", Naval Postgraduate School, Monterey, CA, Technical report, 2004.

[30] ISO 7870-2:2013, "Control charts - Part 2: Shewhart control charts", 2013.