Safety Science 106 (2018) 154–161

Contents lists available at ScienceDirect

Safety Science

journal homepage: www.elsevier.com/locate/safety

Generating synthetic aviation safety data to resample or establish new datasets

Andrej Lališ[a,*], Vladimír Socha[a], Petr Křemen[b], Peter Vittek[a], Luboš Socha[c], Jakub Kraus[a]

[a] Faculty of Transportation Sciences, Czech Technical University, Prague, Czech Republic
[b] Faculty of Electrical Engineering, Czech Technical University, Prague, Czech Republic
[c] Faculty of Aeronautics, Technical University of Košice, Košice, Slovakia

ARTICLE INFO

Keywords:
Aviation safety
Data resampling
Data simulation
Safety management system
Safety performance
Aerospace Performance Factor

ABSTRACT

Aviation safety data are limited in availability due to their confidential nature. Some aggregated overviews already exist but in order to effectively use the data, it is important to fill the gaps of their existing limitations. For some data, there are not enough data points in order to process them through advanced analysis. For other, only expert assumptions can be obtained. In both cases, these shortcomings can be addressed via proper data resampling or simulation where little effort can make the data suitable for various research and development initiatives. Examples of real aviation safety data made public are demonstrated together with key principles of how to perform their resampling. Then, for cases where only expert assumptions are available, general solution to the transformation of the assumptions into simulated data is introduced. The goal is to demonstrate how to transform accessible data or knowledge about aviation safety into data samples with sufficient granularity. The results provide general solution suitable not only for aviation safety data and knowledge, but also for similar transportation or high-risk industries related data issues, indicating that both the data resampling and simulation provide an option for generating datasets, which can be used for statistical inferential methods, linear regression modelling, recurrent analysis etc. Example of data resampling application is included in Aerospace Performance Factor calculation for years 2008 up to 2015.

1. Introduction

To date, aviation safety is subject of intensive research in terms of new information technology deployment. It is recognised, that further progress in this domain can be achieved by implementing technology, which collects, processes and analyses safety data in order to produce system-wide information of how the system performs on safety (ICAO, 2013). This information is to be used for safety-critical decision-making within safety management system as far as the aviation is concerned, but this principle is generally true for other high-risk industries as well (Niu and Song, 2013; Klein and Viard, 2013). One of the features of the system-wide information is that it cannot be reliably derived by individuals from the data available because aviation became very complex, i.e. hardly manageable for humans. The industry is distributed system of many types of stakeholders (airspace users, organisations, regulators, manufacturers, policy makers etc.) which use different technologies, different procedures and which overlap with each other to various extent. As a result, safety performance of one stakeholder may be severely affected by how safety is managed by other stakeholder and it can be difficult to identify this from either side.

Today's accidents only support this claim. They consist of long chain of events and contributing factors, which typically exceed responsibilities of one stakeholder and its safety management (Socha et al., 2014). From the perspective of managing safety, it is important to have some sort of full picture to be able to apply effective measures to prevent modern accidents. The distributed character of aviation, however, sets constraints for achieving such a full picture. Not only are the data and the full picture to be established distributed in parts among the stakeholders, but also the nature of both is often confidential and may have potential to damage someone's position on the market, if misused. Aviation authorities encourage organisations and other stakeholders to share safety data, experience and safety knowledge (ICAO, 2013; European Commission, 2010) but the degree of these activities is still not perceived to be satisfactory. This paper does not aim to resolve this issue but rather address its consequences, namely limited safety data availability.

Research and development initiatives in the domain of aviation safety are restricted by the safety data not being available. Whilst it may be possible to sign some bilateral confidential agreement between two parties, this is still rather difficult to achieve for multiple

* Corresponding author at: Czech Technical University in Prague, Faculty of Transportation Sciences, Department of Air Transport, Horská 3, 128 03 Prague, Czech Republic.
E-mail address: lalisand@fd.cvut.cz (A. Lališ).

https://doi.org/10.1016/j.ssci.2018.03.013
Received 22 September 2016; Received in revised form 20 February 2018; Accepted 8 March 2018
0925-7535/

stakeholders at the time. Fortunately, some of the data are regularly (annually) published by authorities in form of aggregated overview of key safety issues (such as Safety Regulation Commission, 2016; EASA, 2016 or Federal Aviation Administration, 2015), but this is true only for some segments of the industry, e.g. for air navigation services providers (ANSPs). These providers have a lot of advanced technology and data at their disposal and they are typically state-owned monopolies, which are not subject of market competition. The latter was likely the key factor for making some of their data publicly available.

To better understand the issue, it is important to note basic facts of data evolution in this domain. Safety was always measured indirectly, i.e. through its absence (Reason, 2000). It is quite hard to find any effective way to measure it directly as it is the case for conventional measurements related to more tangible issues (Hanakova et al., 2017). Overall safety is intangible system property and even where it is possible to measure it directly, it is often impractical because measuring the things which go right simply means a lot of effort to be spent in order to have meaningful records. Unlike safe state, unsafe outcomes are not only less frequent but they are much more tangible thus considerably easier to track (Hollnagel, 2014). Aviation accidents and incidents attract society from early days of its existence and for decades they were the best driver for safety improvements. As soon as they became rare, the focus just shifted to incidents and safety occurrences with their contributing factors, which, according to investigations, lead to the accidents.

Recently, a new type of data emerged in this domain. Tracking back the root causes of accidents led to the discovery of the so-called organisational factors denoting those contributing factors, which stem from how safety management and safety oversight work (ICAO, 2013). Until the discovery of the importance of how aviation organisations and regulatory bodies are set up as entities, no safety management system nor any sophisticated safety oversight were needed. Progressive requirements for gathering how organisations and regulatory bodies approach safety from management perspective appeared first around the year 2010 (European Commission, 2010; EASA). These requirements established datasets different in their very fundamentals; they assess activities which can hardly be associated with specific unsafe behaviour but which are capable of generating background on which unsafe behaviour emerges. Starting to collect this type of safety data was significant milestone for aviation safety as it brought the industry closer to generate the full picture.

Nowadays, we are closer to the full picture as the content of collected data evolved, but due to the insufficient data sharing and confidentiality restrictions, they are typically not available for research and development initiatives. This inhibits the progress of introducing new technology which could integrate and process the data so that all parties would benefit from industry-wide, open data based knowledge. So has the progress to be achieved the other way. Current research initiatives have to make the best use of public but restricted data samples to come with solutions that aviation organisations may trial and which would expedite establishing the full picture.

Data scarcity, however, is not a new issue. There are several studies available to date, which propose methodologies to overcome this issue in different applications. In fact, very few deal with this problem in scope of safety (such as Yu et al., 2017; El-Gheriani et al., 2017, which are only oriented to major accidents); much more frequent are studies oriented to system reliability, failure and risk assessment in terms of data uncertainty and its reduction. Both safety and reliability oriented studies are typically using Bayesian approach in some variations to produce a posterior distribution by combining data, expert knowledge or various simulation results. Among other methods, first order reliability method and Monte Carlo simulation (Awadallah et al., 2016), or grey system theory (Wen et al., 2011) are used in respective applications. Special attention in the literature is paid to expert elicitation, which was already formalised in several publications (such as Meyer, 2001; Keeney and von Winterfeldt, 1991 or Aven and Guikema, 2011). All the methodologies are, however, difficult to apply directly on the problem in this work as they require various inputs which are out of the scope of this paper. The problem here is of more generic nature, even though it can be complemented with the methods from other studies.

With respect to the afore-mentioned, this article describes the public aviation safety data in detail and provides solutions for how to overcome their limitations. It suggests generating either synthetic aviation safety data or resampling the data already available. The motivation to use data resampling is based on the need to decompose existing signals to increase their granularity for the purpose of further processing and analysis. Data simulation complements this approach by extending the possibility to generate entirely synthetic signals.[1] Synthetic data have their apparent limitations but the important aspect is that they can enable application of advanced analyses, even for experimental or learning purposes only, where real data do not allow it. Direct application of mathematical tools and methods, such as statistical inferential procedures, autoregression or recurrent analysis, to make inferences about safety performance (the full picture) would be otherwise impossible. To enable the tools and methods, it is important to resample the data, i.e. to transform annual figures into month, week or day distribution. For cases where no data are available, simulation based on expert assumptions can provide the solution.

Taking into account the goal, this paper deals with methodology of both data resampling and simulation. It describes data and identifies the gap for improvement. The methods are applied on selected figures from real datasets in the domain of aviation safety. At the end, aviation safety performance is computed using the resampled data to exemplify the contribution of the proposed solution.

2. Methods

This section details the proposed methodology to achieve the goal of this paper. At first, aviation safety data are specified, including their sources, relevant issues and examples. At second, data resampling follows with description of key principles of how to combine expert knowledge and real datasets to increase data granularity by the means of mathematical functions. Lastly, after the outline of data resampling principles, the methodology further specifies data simulation in order to extend the principles of generating synthetic data to situations where no real data are available.

2.1. Data characteristics

Aviation safety data comprise accidents, incidents and safety occurrences. The data are available in form of aggregated figures denoting number of observations of respective accident, incident or occurrence during given time interval. Additionally, new data types were recently introduced to aviation through the European Union-wide (EU-wide) safety key performance indicators (SKPIs) (European Commission, 2013), which are based on the so-called organisational factors. However, these are using artificial scores and due to their novelty, inherent bias and lack of relevant expert assumptions, they are not considered in the methods of this study.

Aviation accident records were gathered reliably till now and they are publicly available together with investigation reports, including conclusions and corrective measures. These data can be found on website of responsible body for respective investigation.[2] But because aviation accidents became rare, they solely cannot be used for safety management today. In terms of any research and development initiatives, much more valuable are data concerning incidents and safety

[1] For further reading on data resampling and simulation methods refer to Lahiri (2003) and Carsey (2014)).

[2] Such as Air Accidents Investigation Institute (2017) in the Czech Republic or Bundesstelle für Flugunfalluntersuchung (2017) in Germany.