trajectory is not following the DMA boundary anymore after the passage $R_n$, the next section to overall trajectory is taken into account length $d(R_n, B)$. Section $d(R_n, B)$ is equal to 16.49NM. Total sum of path along sections $A-R_1; R_1-R_2; R_2-R_n; R_n-B$ (Fig. 2.) is equal to 34.2NM.

Summarizing the user preferred trajectory is 33NM. After solution with FDM, when to render points along DMA boundary, when to use the solution with intermediate points (on Fig. 2.), the trajectory increased to 35.86NM. By approximate solution (Fig. 6.) would to find application, where mesh density and DMA complexity of shape.

Proper mesh density and selection of shape function in order to most accurate and stable result for AAM and flight planning system elaborations, but described initial results indicate the expected monotonic relationships of SMS processes, operational activity and demographic data with safety outcomes. The researchers with the support and reviews from academic and regulatory institutions, subject matter experts and through pilot studies in aviation companies have developed a set of five new safety metrics. These metrics are based on the premise that the greater the gap between system design (Work-as-Imagined; Wal) and operation (Work-as-Done; Wad), the lower the safety/system performance. The metrics regard: (1) assessment of Safety Management Systems by addressing their design, operation, quality of delivery and internal and external interfaces and dependencies, (2) planning and implementation of safety culture prerequisites, (3) system complexity and coupling, (4) measurement of risk control effectiveness, and (5) utilization of resources. The metrics mentioned above operationalize and correspond to suggestions from academic literature, challenges in professional practice, depictions of system structures, and consideration of “soft” organizational aspects. The particular metrics meet the requirements for accuracy, construct, content and face validity and can be used jointly or separately to offer organizations insights into various aspects of their systems.

Keywords

- safety management
- safety metrics
- safety culture
- complexity
- coupling
- risk controls
- resource scarcity

1. Introduction

In September 2015, the Aviation Academy of the Amsterdam University of Applied Sciences initiated a research project entitled “Measuring Safety in Aviation – Developing Metrics for Safety Management Systems”. The project responds to specific needs of the aviation industry: Small and Medium Enterprises (SME) lack large amounts of safety-related data to measure and their safety performance practively, large companies might obtain abundant data, but they need safety metrics which are more leading than the current ones and of better quality; the transition from compliance-based to performance-based evaluations of safety is not yet backed with specific tools and techniques. Therefore, the aim of the research was to identify ways to measure safety proactively in scientifically rigorous, meaningful and practical ways without the benefit of large amounts of data and with an emphasis on performance rather than mere compliance (Aviation Academy, 2014). During the first phase of the project, the research concluded to the following findings:

- State-of-art academic literature, (aviation) industry practice, and documentation published by regulatory and international aviation bodies jointly suggest that (a) safety is widely seen as avoidance of failures and is managed through the typical risk management cycle, (b) safety metrics can be, conventionally, split in two groups: safety performance metrics which are more leading than the current ones and of better quality; the transition from compliance-based to performance-based evaluations of safety is not yet backed with specific tools and techniques. Therefore, the aim of the research was to identify ways to measure safety proactively in scientifically rigorous, meaningful and practical ways without the benefit of large amounts of data and with an emphasis on performance rather than mere compliance (Aviation Academy, 2014). During the first phase of the project, the research concluded to the following findings:
Taking into account the current situation, as explained in the previous section, the researchers initially reviewed recent literature on the prerequisites which are necessary to develop a safety culture (Karanikas et al., 2015), a framework was suggested for achieving the system objectives, because this requires deep knowledge of the context. This was out of the scope of the particular research.

To develop new safety metrics, the researchers initially reviewed relevant literature to identify how the Wal-WaD gaps could be described and quantified. This review described the safety concepts that were perceived as suitable to be operationalized through respective metrics (section 2.1) and the ones not addressed by the researchers (section 2.2). The primary criterion for the inclusion/exclusion of concepts in the research was their potential to be practical (i.e. not requiring vast amount of resources and operational data), scalable (i.e. applicable to SMS with various sizes and activity types) and comprehensible (i.e. not requiring in-depth knowledge of the underlying theoretical foundations). The degree to which the new metrics meet the research objectives and satisfy the quality criteria was thoroughly assessed during the reviewing process and pilot studies to companies (section 2.3).

### 2.1. Concepts included

#### 2.1.1. SMS assessment

The industry has recognized the need to move from a compliance-driven assessment of SMS to a performance-based evaluation scheme (ICAO, 2013; FAA, 2006). The Safety Management System Evaluation Tool developed by the Safety Management System Evaluation Group (SMICG, 2012), and the (FAA) Safety Operations Center (SOC) instrument, which was devised by the Eurocontrol (2012), have been introduced to support the aforesaid transition but they include highly subjective measurement scales and do not address the connections and dependencies of SMS processes (Karanikas, 2016). Therefore, although such tools introduce the transition from merely checking the existence of SMS elements and processes to considering the sufficiency of their output and indicating necessary improvements, the interlinks between SMS activities are not yet addressed. The researchers contemplated that filling the latter gap, would offer to the industry a more meaningful way to assess their SMS performance and detect its distance from optimum/deadline performance.

#### 2.1.2. Safety Culture

Safety culture has been for long a discussion topic in the academia and the industry. However, as Kaspers et al. (2016a) identified, there has been little consensus whether safety culture reflects the way an SMS is operated or the effects of SMS on safety performance. At the same time, safety culture is not consistently assessed within organizations (Kaspers et al., 2016b, 2016c). Recently, following research at a nuclear power plant (Karanikas et al., 2015), a framework was suggested for defining the prerequisites which allow a safety culture to be seen as a complex phenomenon. This framework is based on academic literature and industry standards, follows Reason’s (1995) conceptualization of the cycle and refers to the necessity to monitor the effectiveness levels.

### 2.2. Concepts excluded

Taking into account the objectives of the research as mentioned above (section 1), the project timeline and the research resources available it was concluded that the following safety-related concepts could not be examined, although they could be of interest for future research:

- **Dependency on initial conditions:** Dekker (2011) and Leveson (2011, 2015) discuss the importance of the decision of automating the equipment or not, the SMS processes and the effects of former on the performance of the latter over time. However, based on the experience of the researchers and informal discussions with partner organizations, such decisions and assumptions are not always documented and/or known by companies and are difficult to be obtained from system designers.
- **Decrementalism:** The drift of organizational performance due to small changes, which are typically judged against the success of the most recent change and not the distance from the original design, has been illustrated by Dekker (2011). However, as discussed for the dependency on initial conditions above, the research team were informed by the companies participating in the research that such information is not directly available to be collected.
- **Unruly technology:** Various authors (e.g., Sarter, Woods and Billsings, 1997; Dekker, 2011; Chow, Yortsos, & Meshkat, 2014) have pointed the fact that end-users do not always and fully obtain a sufficient understanding of how high-automated systems actually function (e.g., type of data collected and analyzed, algorithms used, connections of sub-systems and interdependencies). Thus, system operators are not able successfully to interact with automation surprises. Taking into account the variety of systems operated by aviation companies, the different levels of automation and systems operated and the diverse set of skills and knowledge across end-users, the researchers contemplated that the operationalization of a relevant metric could not be feasible within the boundaries of this research.
- **Resilience:** The introduction of resilience engineering into the field of safety (e.g., Nemeth and Herrera, 2015; Costella, Saunir, & de Macedo Guimaraes, 2009) has not yet been accompanied by research that quantifies the resilience abilities, namely anticipation, monitoring, response and learning. The Resilience Analysis Grid (Hollnagel, 2015) provides some guidance on the decompositions of the abilities above, but yet respective quantifiable metrics have not been developed. Based on the literature reviews, it seemed that the development of resilient metrics would require a separate research, a work that could not be undertaken during the current project.

### 2.3. Validation of metrics

The current research achieved accuracy, construct, content and face validity of the metrics mentioned in section 2.1 above were assessed as follows (adapted from Kaspers et al. (2016a) and addressing the limitations of current metrics presented in section 1 above):

- reflective of the respective theoretical framework;
3. Results

Following the design and validation phases, five safety metrics were developed. Due to paper length limitations, the following sections outline these metrics. Extended versions of each of the safety metrics will be published in future papers.

3.1. SMS assessment

To fill these missing areas of existing instruments, as mentioned in section 2.1 above, we developed an SMS assessment tool according to the Safety Management Manual of ICAO (2013) and based on the System Thoretic Process Analysis (STPA) technique (Leveson, 2011). The tool incorporates the view of SMS as a “system”, meaning that addresses its design and implementation along with time and internal/external process dependencies, as well as SMS effectiveness. Under this approach, the tool introduces two steps that examine (a) whether SMS is designed and implemented (WaD) according to standards (Wal), and (b) whether SMS is effectively implemented, where the actual effectiveness (WaD) can be assessed against a maximum/desired level (Wal). The former SMS dimension is assessed through documentation and log checks, and the evaluation of the former is achieved via a questionnaire survey.

Each of the two assessment areas leads to individual scores calculated based on the Euclidean distances, which in combination provide the total SMS performance score. Overall, the assessment tool aims at assuring organizations with the assessment of their SMS and the scores generated can enable the qualitative and numerical monitoring of SMS performance and in the quantification of the steps will be or implemented SMS changes. Apart from improving SMS, the tool can be used when SMS and safety performance targets are not met or are not within defined limits/thresholds.

3.2. Safety Culture Prerequisites

The tool designed can be used to gain insights into what safety culture prerequisites (SCP) are currently planned and the degree to which these prerequisites are operationalized. It is noted that the particular tool does not have the goal to measure safety culture which can be assessed with other relevant instruments that have been introduced in academia and the industry. The tool is based on three assessment areas: (a) documentation check to explore the plans the organization has in place to foster a positive safety culture, (b) a survey across managers as a means to examine the degree the safety culture prerequisites are realized, and (c) a survey across employees to assess their perceptions about each of the six culture areas mentioned in section 2.2 above.

It is clarified that the latter measurement regards a high-level approach to the perception about the functioning of each of the six areas within the organization and does not aim at assessing safety culture. Thus, the perception questions function as proxy for the quality of the implementation of the SCP. The tool is complemented with a scoring method that uses Euclidean distances to depict the gaps between design, implementation and perceptions in overall and across the six areas assessed. These gaps reflect in pairs the Wal and WaD differences. The results from the application of the tool and the scores generated can be used to assess organizations in prioritizing their intervention actions to improve safety culture.

3.3. System complexity and coupling

The research team developed a metric that can be applied to a defined system and encompasses two main factors: inherent system complexity and coupling, and perceived complexity. The latter factor can be assessed through a questionnaire survey across the system users, and the former is calculated with a formula. The parameters of the formula are (1) number of interacting elements within the system, (2) types and duration of interactions, and (3) time, human and technical buffers to cope with excessive internal and external disturbances. The integration and coupling are measured at the level of system design (Wal) and actual system configuration (WaD). The perception factor refers only to the actual conditions and moderates the gap between Wal and WaD.

3.4. Effectiveness of risk controls

Due to the lack of literature about the specific metric, the researchers employed engineering logic about the actual behavior of controls (e.g., frequency of failures of risk controls) and explored whether the functionality (Hollnagel, 2004) and hierarchy (e.g., Leveson, 2011) of controls could be exploited. We concluded to the following indicators for measuring risk control effectiveness:

- Number of successes of a risk control when challenged
- Number of occasions the risk control was challenged
- Number of successful test or inspection results of a risk control/Number of tests or inspections
- Number of unwanted events after a risk control was implemented
- Number of unwanted events before the risk control was implemented

4. Discussion

The metrics that were designed during the current research phase are based on a spectrum of suggestions from academia and industry that had not been consistently or entirely operationalized. Following two review rounds, the researchers developed the metrics by ensuring that those meet the quality criteria mentioned above and led to the final design of the metrics.
5. Conclusions
The five safety metrics presented in this paper meet
the resilience engineering perspective. Safety Science, 47, 1056-1067.


Dekker, S., 2016. Drift into failure: From hunting broken components to understanding complex systems. CRC Press.


Kaspers, S., Karankanis, N., Roelen, A.L.C., Piric, S., de Boer, R. J., 2016a. Review of Existing Aviation Safety Metrics, RAAM PRO Project: Measuring Safety in Aviation, Project SI0991, Aviation Academy, Amsterdam University of Applied Sciences, the Netherlands.

Kaspers, S., Karankanis, N., Roelen, A.L.C., Piric, S., van Aalst, R., de Boer, R. J. 2016b. Results from Surveys about Existing Aviation Safety Metrics, RAAM PRO Project: Measuring Safety in Aviation, Project SI0991, Aviation Academy, Amsterdam University of Applied Sciences, the Netherlands.


