

Learning from accidents: interactions between human factors, technology and organisations as a central element to validate risk studies

R. Moura^{a,d}, M. Beer^{b,a,c}, E. Patelli^a, J. Lewis^a & F. Knoll^e

^a*Institute for Risk and Uncertainty, University of Liverpool, United Kingdom*

^b*Institute for Computer Science in Civil Engineering, Leibniz University Hannover, Germany*

^c*Tongji University, Shanghai, China*

^d*National Agency for Petroleum, Natural Gas and Biofuels (ANP), Brazil*

^e*NCK Inc., Canada*

ABSTRACT: Many industries are subjected to major hazards, which are of great concern to stakeholders groups. Accordingly, efforts to control these hazards and manage risks are increasingly made, supported by improved computational capabilities and the application of sophisticated safety and reliability models. Recent events, however, have revealed that apparently rare or seemingly unforeseen scenarios, involving complex interactions between human factors, technologies and organisations, are capable of triggering major catastrophes. The purpose of this work is to enhance stakeholders' trust in risk management by developing a framework to verify if tendencies and patterns observed in major accidents were appropriately contemplated by risk studies. This paper first discusses the main accident theories underpinning major catastrophes. Then, an accident dataset containing contributing factors from major events occurred in high-technology industrial domains serves as basis for the application of a clustering and data mining technique (self-organising maps – SOM), allowing the exploration of accident information gathered from in-depth investigations. Results enabled the disclosure of common patterns in major accidents, leading to the development of an attribute list to validate risk assessment studies to ensure that the influence of human factors, technological issues and organisational aspects was properly taken into account.

1. Introduction

1.1 Accident causation models and implications to validate risk assessments

Accident causation models lie beneath all efforts related with safety engineering, as they serve as basis for accident investigation and analysis, to prevent future accidents in new designs and for the development of risk assessment techniques (Leveson, 2012). The rising interest in understanding the genesis of major accidents and the growing importance of technological issues to societies directed many schools of thought to approach the accident causation problem from different perspectives, leading, to a certain extent, to conflicting ideas on how (and if) hazards can be appropriately addressed and controlled.

According to Perrow (1984), failures in complex, tightly coupled systems are inevitable, and thus the occurrence of accidents with catastrophic potential in some high-technology facilities (e.g. nuclear power and nuclear weapons) is unavoidable, constituting an expected or *normal accident*. His theory was developed after the Three Mile Island accident, a partial core meltdown that occurred in a USA nuclear power plant in 1979 which was his base case. To cut a long story short, he simply suggests

the discontinuation of technologies such as nuclear plants and weapons (which he deems hopeless) as he understands that the inevitable risks outweigh the perceived benefits. Operator errors are frequent elements of the scrutinised case studies, highlighting how complex interactions of a series of failures can lead to flawed mental models. Perrow alludes to a sole possible managerial style to safely run these facilities: a military-shaped organisation, authoritarian and rigidly disciplined. However, he claimed that this administration structure would be socially intolerable and unsustainable during peacetime, for industrial civil activities.

The Normal Accidents Theory was preceded by Cohen's Garbage-Can Model (Cohen et al., 1972, Davis et al., 1988), which presented an earlier recognition that organisations have high degrees of uncertainty, leading to ill-defined or competing preferences, ambiguous goals, unclear technology and fluid patterns of stakeholders' involvement in the decision-making process. While the Garbage Can theory indicates that major accidents will happen because organisational behaviour is extremely complex and unpredictable, the Normal Accidents Theory limits the inevitability of disasters to systems where complexity and tight coupling are observed. Though both theories share an unenthusiastic view of the human capacity to predict and control hazards, some distinct (and useful) elements can be extracted from them: the former clearly points towards organisational matters as the root-cause of catastrophes, while the latter blames technological aspects, albeit assuming that it could be somehow mitigated by a particular type of military organisation.

Taleb's book *The Black Swan – The Impact of the Highly Improbable* (2007) coined a popular and wide-reaching concept (Aven, 2015, Aven 2013, Paté-Cornell, 2012) to explain the occurrence of major accidents. He refers to events with extreme impacts as *Black Swans*, considering them as highly improbable events (or outliers) which are not prospectively foreseeable. His celebrated analogy was based on the fact that people in the "old world" only knew white-feathered swans before the English arrival in Australia, where the sight of a black swan came as a surprise. He concludes that predictions based on historical data cannot anticipate outliers, claiming that the usual focus on standard operations disregards the extreme or uncertain. According to his views, the dynamics in high-technology domains are far more complicated than can be anticipated, and conducting laborious pre-analysis and validation based on probabilistic modelling should be ruled out, as it has little effect in terms of major hazards control (or black swans prevention!).

It is worth noticing that many widespread accident causation theories appear to consider the understanding of all complex interactions leading to major accidents during the operation of high-risk industrial facilities as a significant challenge still to overcome. According to this approach, objectives and preferences are being randomly defined, technologies are not fully understood by managers and workers, complex interactions leading to major accidents are not predictable and stakeholders' groups are fluctuating during the facility's lifecycle.

Conversely, researchers on High Reliability Organisations (Roberts, 1990, Grabowski & Roberts, 1997, La Porte & Consolini, 1998) address cases where organisations managing operations with high potential for disasters achieved excellent levels of reliability for long periods of time, appearing to function better than others. Based on the observation of success cases, they believe that it is possible to recognise scientific methods to sustain a nearly error-free operation, even in very hazardous environments. It is worth noticing that the examples used to ratify the High Reliability

Organisations principles include nuclear power stations, putting it in sharp contrast with the Normal Accidents Theory. According to Perrow (1984), these are precisely the sort of facility susceptible to unavoidable failures, and thus society should consider abandoning it at once.

Sagan (1993) conducted an in-depth analysis of the Normal Accidents and the High Reliability Organisations theories, presenting some of the competing viewpoints below.

Table 1 – Competing Perspectives on Safety with Hazardous Technologies (Sagan, 1993)

<i>High Reliability Theory</i>	<i>Normal Accidents Theory</i>
Accidents can be prevented through good organisational design and management.	Accidents are inevitable in complex and tightly coupled systems.
Safety is the priority organizational objective.	Safety is one of a number competing objectives.
Redundancy enhances safety: duplication and overlap can make “a reliable system out of unreliable parts”.	Redundancy often causes accidents: it increases interactive complexity and opaqueness, and encourages risk-taking.
Decentralized decision-making is needed to permit prompt and flexible field-level responses to surprises.	Organisational contradiction: decentralisation is needed for complexity, but centralisation is needed for tight-coupled systems.
A “culture of reliability” will enhance safety by encouraging uniform and appropriate responses by field-level operators.	A military model of intense discipline, socialisation and isolation is incompatible with democratic values.
Continuous operations, training and simulations can create and maintain high-reliability operations	Organisations cannot train for unimagined, highly dangerous or politically unpalatable operations.
Trial and error learning from accidents can be effective, and can be supplemented by anticipation and simulations.	Denial of responsibility, faulty reporting and reconstruction of history cripples learning efforts.

Despite the evident disparity between these schools of thoughts, especially regarding the possibility of preventing a major accident, Sagan perceived some common ground regarding the frequencies of these events. While the normal accidents theory states that major accidents are inevitable, but *extremely rare*, high-reliability organisations theory postulates a *nearly* error-free operation by an enhanced safety management. Implicitly, there is a mutual recognition of the low probabilities of catastrophic events. After assessing several study cases on safety events involving U.S. nuclear weapon systems, Sagan (1993) concluded that the collected evidences provided stronger support to the Normal Accidents Theory. His observations indicated that factors such as excessive discipline (he identified evidences of extreme loyalty, secrecy, cover-ups, distain for external expertise and other self-protecting mechanisms), conflicting interests and constraints on learning have limited nuclear facilities’ organisational safety and could have resulted in major catastrophes if circumstances were slightly different.

Therefore, Sagan’s resulting analysis of the theories can be considered even more pessimistic than the Normal Accidents Theory. Despite the claim that accidents are inevitable, Perrow left the door open for a social incompatible but safety-efficient managerial style: a military-shaped organisation with rigid discipline. However, his allegations were challenged by Sagan’s nuclear weapons handling sample, which included an alarming number of close calls.

Other researchers recognise the difficulties in preventing major accidents, but focus on the development of strategies to reduce their likelihood. Following this principle, James Reason developed an acclaimed and widely-known accident causation approach, which evolved from Heinrich's et al. (1980) Domino Theory. Reason (1990) firstly developed the idea of having a combination of active failures and latent conditions to explain how complex systems can fail, later expanding it to a multi-barrier concept known as the Swiss Cheese Accident Model (Reason, 1997), which is widely used by academics and practitioners to describe the dynamics of accident causation. Successive cheese slices represent layers of defences, barriers and safeguards, all containing holes symbolising breaches caused by active failures and latent conditions. In the rare occasions when holes are perfectly aligned and all protective layers are overcome, an organisational accident will occur, usually having devastating consequences. A vital distinction between individual accidents and organisational accidents was highlighted by the theory, especially the risk that organisations will be tempted to rely on LTI (lost-time injury) or Bird's pyramid-type methodologies to demonstrate safety performance, overlooking latent conditions that degrade barriers and lead to major accidents. Many risk management approaches derive from the multi-barrier concept developed by Reason, believing that the underlying mechanisms causing organisational accidents can be correctly identified and properly managed. Human reliability approaches such as Human Factors Analysis and Classification System – HFACS (Shappell et al., 2007), Systematic Occurrence Analysis Methodology – SOAM (Licu et al., 2007) and the Sequentially Outlining and Follow-up Integrated Analysis – SOFIA (Blajev, 2002), and accident causation analysis methods such as Bow-Tie (Zuijderduijn, 2000) and Cause-Consequence Diagrams (Nielsen, 1971) are examples, to name but a few, of risk assessment techniques deeply aligned with Reason's approach.

Contemporary approaches on accidents causality models try to apply systems theory and system thinking (e.g. Leveson, 2011) to disclose deeper factors contributing to accidents, by adding higher hierarchical levels beyond immediate events and analysing the interactions among factors and broader circumstances. Examples are how public opinion and governments' movements influence the safety culture of an industrial sector. If the interaction among some of the constituent elements violates a set of constraints that guarantees the system safety integrity, an accident may occur. The focus of this systemic approach to accident causation is on understanding why the enforcement of constraints was unsuccessful.

A comparable perspective was previously conceived by Rasmussen's (1997) thoughts on system performance control. Instead of continually constraining individual elements to fit a pre-defined operational standard or limit, he focused on two features of system control theory: firstly, the need for adaptation of the system operation boundaries, i.e. increasing the margin from normal operation to loss-of-control; and secondly, increasing the awareness level of operational limits by making these boundaries visible to stakeholders. Rasmussen also noted that the pace of technology change is much faster than the modification time for management structures, and an even longer change lag is observed in higher hierarchical levels such as governments, regulations and society. This asynchrony defies risk modelling and challenges the rationale of using detailed methods and tools for analysing individual components or sub-systems, as satisfactory results in parts of a system might not reflect the safety status of the overall system.

Current research (Hopkins, 2002; Hollnagel et al., 2011; Arstad and Aven, 2017) highlights the need for taking into account the complexity and the dynamic nature of high-technology systems to prevent major accidents. In summary, these works discuss how it is crucial to challenge existing risk management assumptions, in order to identify oversimplifications and cope with the intricate interactions leading to the genesis of major events.

When the utmost objective is the validation of risk assessments for hazardous industrial process plants in a dynamic and fast-changing environment, the complexity of the interactions among system elements must be recognised, along with the unpredictability of organisational behaviour and the inherent difficulties to foresee extremely rare, low-probability events, as highlighted by accident causation theorists. Additionally, designed safety barriers are not static and tend to degenerate through time. Factors such as ageing, maintenance shortcomings, budget constraints, personnel fluctuation and pressure towards cost-effectiveness, to name but a few, can contribute to defeat barriers and thus defence-in-depth concepts, which largely serve as basis for risk assessment studies. Hence, how the confidence of wider stakeholder groups, such as the general public, investors and governments, which are particularly concerned with major accidents and might lack knowledge, interest or appropriate time to go into too much technicalities of risk assessment methods, can be enhanced?

The hypothesis underpinning the current work is that mapping patterns and common tendencies in major accidents, and demonstrating that these accident lessons can be fully understood and applied to new assessments in a logical and structured way, might reveal a realistic path to develop stakeholders' trust and to contribute to validation schemes.

1.2 Identifying common patterns and developing a risk assessment validation framework based on major accidents

The fact that accidents causation theories disagree whether a truthful representation of the multidimensional interactions in major events is achievable or not turns risk assessment validation and trust in risk management into a challenging research topic. Although any model will imply the reduction of the complexity of operational reality, some attributes can be extracted from accident causation models in order to establish an acceptable framework to verify the applicability and accurateness of risk management strategies.

It is disputed if the study of successful cases, as argued by high-reliability organisations theorists, will give some insight into the unusual, rare interfaces observed in major accidents. In contrast, the identification of common patterns arising from interactions between human factors, technological aspects and organisations during catastrophic events seems to be a reasonable approach to support a verification strategy for risk analysis, at least to certify that lessons learned from previous accidents were contemplated in current studies. This novel approach might help in reducing the gap pointed out by Skogdalen and Vinnem (2012) when analysing a number of quantitative risk analysis from the Norwegian Oil & Gas industry. They identified that human and organisational factors (HOFs) were not taken into account during the estimation of the probabilities of a blowout. In contrast, the Deepwater Horizon blowout was deeply associated with HOFs such as work practice, training, communication, procedures, quality control and management. Previous analysis of 238

major accidents (Moura et al., 2016) also indicated that 95% of these events presented some sort of organisational contribution to the undesired outcome, and 57% were directly associated with human factors, highlighting the importance of considering these significant features to develop realistic safety studies.

Barrier and defences-in-depth concepts will rely on the integrity and availability of the designed barriers to hold hazards or to minimise their consequences. Addressing common organisational and technological shortcomings contributing to the degradation of critical safety barriers can reveal tendencies which make them fail upon demand. The pattern identification process would also support the application of a safety check against recurrent damage mechanisms, reducing latent failures and providing useful data to endorse the expected positive effect of the barrier during a real event.

The disclosure of common patterns leading to major accidents will make operational boundaries visible to stakeholders, improving confidence in the decisions made and justifying the application of additional safety measures. The fact that the output will be directly associated with real events will facilitate the learning process and highlight the significance of addressing the identified concerns.

Therefore, this research will focus on the development of a risk assessment validation scheme, based on the interactions between human factors, technological aspects and organisations during major accidents. The collection of events constitutes the Multi-Attribute Technological Accidents Dataset (MATA-D) introduced by Moura et al. (2016), which captured major accidents occurred in high-technology industrial domains (e.g. aviation, oil & gas upstream, refineries and nuclear plants) and classified them under a common framework, the Contextual Control Model used as basis for Hollnagel's (1998) Cognitive Reliability and Error Analysis Method. This previous work presented one of the most complete statistical analyses of major accidents from different industrial sectors in the open literature.

The application of an artificial neural network approach, specifically Kohonen's (2001) Self-organising Maps (SOM), will result in the conversion of complex accident data into 2-D risk maps. Events will be clustered by similarity, allowing the combined treatment of accidents with similar interactions but from distinct industrial segments. The development of the data visualisation provided by the SOM application will give rise to the development of a set of properties, attributes and recommendations for the validation of systems, safety barriers, human-machine interfaces and risk studies, enhancing risk perception and stakeholders' trust.

2. Analysis Method

2.1 Using a major-accident dataset as a reliable data source

Previous works have applied past accidents data to produce insight into the genesis of adverse events, in order to support researchers and practitioners by offering valuable contributions to the development of risk management strategies and to disclose contributing causes to accidents. Most of the existing datasets arise from accident/incident data reporting systems, voluntarily developed by companies/associations (e.g. DNV-GL World Offshore Accident Database, International

Association of Gas Producers Process Safety Events Data) or enforced by states (e.g. UN International Civil Aviation Organization Accident Incident Data Reporting system – ADREP, UK HSE’s Reporting of Injuries, Diseases and Dangerous Occurrences Regulations - RIDDOR). These efforts to collect data are commonly limited to a single industrial sector (Baysari et al., 2008, Evans, 2011) or attempt to embrace from occupational accidents to process safety events (Bellamy, 2007, 2013). Generally, reporting systems also include a category called near-misses, which are hazardous occurrences that did not result in a loss/injury but had the potential to do so.

The events’ scrutiny level during the data acquisition stage will involve some expected variations, as it will mostly depend on the consequences of the event and secondly on the societal interest in the subject. Consequently, near-misses will be directly reported by companies, with the regulating body using this summarised data to develop performance indicators or to trigger further actions such as inspections. Regulators can investigate occupational accidents directly, or validate/rely on companies’ internal investigation procedures. Major accidents usually capture the media’s and societal attention, pushing governments and regulators to react accordingly. Due to the wide-ranging consequences observed, this type of event requires consistent investigation processes, usually undertaken by one or more regulators, independent investigation commissions or both. The European Safety, Reliability and Data Association (2015) has recently recognised that these events trigger comprehensive examinations concerning preventive and protective systems, along with a careful consideration of factors and surrounding conditions leading to accidents. An illustrative example would be the Transocean’s drilling rig Deepwater Horizon blowout and explosion occurred in the Gulf of Mexico in April 2010, which was investigated by the licensee (BP, 2010), regulators (USCG, 2010, BOMRE, 2011), an independent agency (US-CSB, 2016) and academic study groups (CCRM, 2011). Beyond doubt, catastrophic events lead to meticulous examinations and produce very detailed data about the conditions in which operations were inserted. Attributable to this extraordinary level of scrutiny, the data produced is indisputably more reliable and complete than any alternative source of information regarding accident causation.

2.2 The SOM data mining applied to the MATA-D

The current version of the MATA-D, containing 238 major accidents from different high-technology industries (e.g. aviation, hydrocarbons exploration and production, refining, chemical industry, nuclear) will be used as a data source for this research. The dataset was fed from detailed major accident investigation reports, obtained from reliable sources such as regulators, independent investigation commissions and insurance companies (Moura et al., 2016). The dataset framework comprises 53 factors distributed in three main categories: man, technology and organisation. The structured but comprehensive nature of the MATA-D framework allowed for the effective application of several data mining approaches in previous research (e.g. Doell et al., 2015, Moura et al., 2015a, 2015b), such as agglomerative clustering methods, association rule mining techniques and neural networks. Cross-industrial common patterns in major events as well as significant relationships among contributing factors were successfully disclosed.

In this work, key interfaces between human factors, technological aspects and organisations will be identified through the application of a suitable artificial neural network technique, namely self-organising maps - SOM (Kohonen, 2001). This data mining approach is especially effective when an

unsupervised method (i.e. the number of clusters or final categories in the output space are unknown) and the classification and visualisation of high-dimensional data are needed (Kohonen, 2013; Ultsch, 1993). Data mining efforts will result in the reduction from 53 dimensions (or contributing factors per accident) to two-dimensional maps. The 2-D SOM maps will be generated with the support of specialised software (Viscovery® SOMine expert version), to enhance the features' visualisation and facilitate the interpretation of the SOM output.

2.3 The SOM construction rationale and further data mining settings

The self-organising maps (SOM) is a widely used clustering and data classification approach developed by Kohonen (2001). It holds important properties to group data by similarity and regarding the preservation of the input data, enabling visual representation possibilities for multidimensional datasets. Proportionally to the amount of data, the computational complexity of the SOM can be considered low and it is an easy-to-implement algorithm (Cottrell et al., 2016), allowing its straightforward usage on practical applications, with minor adjustments.

In the current study, the objective of the data mining process is to provide an improved but simplified organisation and visualisation of the input data, which consists of a 238 x 53 matrix (i.e. 238 major accidents containing up to 53 contributing factors). The SOM output is a topographic two-dimensional map, where accidents with analogous interactions among contributing factors are mutually attracted and contiguously positioned in a general grid. In order to highlight some specific interactions, individual heat maps containing the presence or absence of single properties are also plotted. The clustering process by similarity, in conjunction with the examination of the individual properties' maps, enables the observation and interpretation of common tendencies associated with the major accidents contained in the MATA-D.

Basically, the neural network is trained according to Eq. 1 (Kohonen, 2001), which indicates that the position at the grid will be occupied by the model m that minimises the distance between the input vector $x(t)$ and an output node m_i .

$$v(t) = \arg \min_{i \in \Omega} \|x(t) - m_i(t)\| \quad (1)$$

In the output space, a neighbourhood function h_{ji} will define the influence zone around the winning output node (the one which is the closest to the input vector), which will be updated through the batch-learning rule (Eq. (2)) developed by Kohonen (2013) to eliminate possible convergence problems. The best matching node m_i^* will be located at the centroid of the influence region defined by the neighbourhood function h_{ji} ; the \bar{x}_j , which represents the mean value of a group of input vectors $x(t)$ previously defined by $v(t)$; and the number n_j of input space samples $x(t)$.

$$m_i^* = \frac{\sum_j n_j h_{ji} \bar{x}_j}{\sum_j n_j h_{ji}} \quad (2)$$

The sequence will be repeated until the map converges, i.e. until the models for the input vectors are accurately represented in the output space. Due to the neighbouring properties, the more contributing factors in common they have, the closer the vectors will be located in the output. If

more than one model in the output space contained exactly the same contributing factors, they would be positioned in the same grid node.

In the current study, the tension of the neighbourhood function (or the radius of the neighbourhood Gaussian function) was defined as the smallest value to produce meaningful clusters and maintain a close resemblance among adjacent neurons, and the map converged after 111 batches. Further details on the SOM algorithm, the network settings, the choice of the neighbourhood function and the clusters' quality criterion for the specific application can be found in Moura et al. (2017).

After the application of the SOM algorithm, the clusters where the highest incidence of interfaces was identified during major accidents will become apparent. Further examination of the intricate relationship among contributing factors within the clusters of interest will reveal common patterns and accident tendencies, highlighting principles that must be taken into account when developing risk assessment studies.

The conversion of relevant interfaces in a set of principles will support the validation of risk analysis and risk management documents, by applying the lessons learned from major accidents. Accordingly, a straightforward requirement list to be crosschecked against risk studies will be developed, and further implications to enhance stakeholders' trust will be then discussed.

3. Results

The application of the SOM algorithm to the MATA-D dataset resulted in four different accident clusters containing dissimilar influencing factors, as shown in Figure 1. The contributing factors label sizes are proportional to their effect within the grouping. For example, the Inadequate Task Allocation factor in Cluster 1 (blue) occupies 95% of the total cluster area, while Wrong Place occupies 52.5%, and the Incomplete Information frequency is 36.2%. This is one example of the usage of the visualisation power of the clustering method to interpret accident data. Figure 1 synthesizes information from a 238 x 53 Matrix (number of major accidents x possible contributing factor per event) in a single 2-D image.



Figure 1 – MATA-D SOM Clustering output labelled by most relevant contributing factors

The first cluster (blue) covered 35% of the SOM map area, containing the highest number of data points, with 34% of the accidents. Cluster 2 (red) has 25% of the total area and 24% of the dataset. The third grouping (yellow) occupies 20% of the total area and has the lowest event’s frequency, with 16%. Cluster 4 (green) also holds 20% of the map area, but embraces 26% of the dataset events. Figure 2 depicts the rate of contributing factors per event, discriminated by clusters.

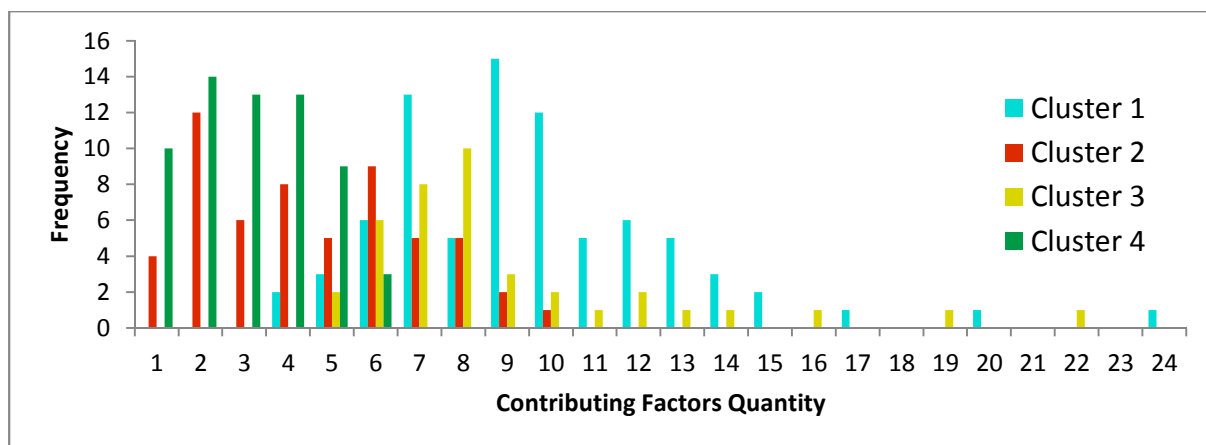


Figure 2 – Number of Contributing Factors Histogram

Figure 2 shows Cluster 1’s events with four to twenty-four contributing factors per accident and mode of nine, as it appeared in fifteen events. 86.2% of the accidents within this cluster have seven or more contributing factors, constituting a very rich grouping for further interpretation. Cluster 2 events were influenced by one to ten features with 72.2% of the grouping having six or less

contributing factors, while the totality of the events in Cluster 4 have six or less features. Both groupings show the same low mode of two factors, indicating a lower prospect for the identification of multiple interactions among contributing factors. For Cluster 3, the total number of contributing factors per accidents varied from five to twenty-two. 79.5% of the events contained seven or more contributing factors, with eight factors the mode value. This grouping also tends to provide good opportunities for enhanced interpretations of the genesis of major accidents.

Results show that the application of the SOM algorithm largely improved the visualisation of interfaces, by confining events with lower frequency of contributors in Clusters 2 and 4, as well as elevating the features' mode for Clusters 1 and 3, highlighting special structures within the dataset.

Table 2 details the results of the SOM clustering, indicating the outcome of the data mining process for selected contributing factors, in relation to the overall dataset. The variation columns compare the overall dataset statistics with the individual factors' influence in each cluster. Negative or very low variations are not indicated, as the preservation or reduction of the frequency of a contributing factor in a grouping (in relation to its overall incidence) means that the factor was not significant to the formation of the cluster. Twenty-seven features contributed to less than 10% of the individual clusters and will not be represented, due to their low significance to the groupings formation. Contributing factors with strong dominance (more than 50% of the individual cluster areas) are highlighted, as well as frequencies higher than 10% and with positive cluster effect in relation to the overall dataset.

Table 2. Dataset overall statistics vs. clustering distribution for significant features

Contributing Factor	Overall	C1	Effect	C2	Effect	C3	Effect	C4	Effect
Wrong Time	14.7%	13.8%	-	10.5%	-	41.0%	+178.8%	3.2%	-
Wrong Type	11.8%	11.3%	-	7.0%	-	30.8%	+161.8%	4.8%	-
Wrong Place	31.5%	52.5%	+66.6%	36.8%	+16.8%	12.8%	-	11.3%	-
Observation Missed	15.5%	20.0%	+28.6%	12.3%	-	23.1%	+48.6%	8.1%	-
Faulty diagnosis	13.0%	26.3%	+101.9%	8.8%	-	12.8%	-	0.0%	-
Wrong reasoning	11.3%	20.0%	+76.3%	1.8%	-	25.6%	+125.7%	0.0%	-
Decision error	9.2%	5.0%	-	17.5%	+89.3%	17.9%	+93.6%	1.6%	-
Inadequate plan	9.7%	10.0%	-	7.0%	-	25.6%	+164.9%	1.6%	-
Priority error	7.1%	6.3%	-	8.8%	+23.2%	15.4%	+115.6%	1.6%	-
Distraction	5.9%	11.3%	+92.1%	3.5%	-	7.7%	+30.9%	0.0%	-
Cognitive bias	7.1%	15.0%	+110.0%	1.8%	-	10.3%	+44.2%	0.0%	-
Equipment failure	55.0%	33.8%	-	22.8%	-	94.9%	+72.4%	87.1%	+58.2%
Inadequate procedure	44.1%	78.7%	+78.4%	42.1%	-	38.5%	-	4.8%	-
Incomplete information	17.6%	36.2%	+105.1%	7.0%	-	20.5%	+16.2%	1.6%	-
Communication failure	10.5%	16.3%	+55.2%	5.3%	-	20.5%	+95.2%	1.6%	-
Missing information	20.6%	37.5%	+82.1%	14.0%	-	15.4%	-	8.1%	-

Maintenance failure	34.9%	56.3%	+61.4%	14.0%	-	33.3%	-	27.4%	-
Inadequate quality control	60.9%	81.3%	+33.4%	24.6%	-	79.5%	+30.5%	56.5%	-
Management problem	9.2%	12.5%	+35.2%	5.3%	-	23.1%	+149.9%	0.0%	-
Design failure	66.0%	85.0%	+28.9%	50.9%	-	87.2%	+32.2%	41.9%	-
Inadequate task allocation	60.1%	95.0%	+58.1%	68.4%	+13.8%	48.7%	-	14.5%	-
Social pressure	7.1%	17.5%	+145.0%	3.5%	-	0.0%	-	1.6%	-
Insufficient skills	36.1%	56.3%	+55.8%	12.3%	-	76.9%	+112.8%	6.5%	-
Insufficient knowledge	35.3%	60.0%	+70.0%	17.5%	-	56.4%	+59.8%	6.5%	-
Adverse ambient conditions	7.1%	2.5%	-	14.0%	+96.0%	10.3%	+44.2%	4.8%	-
Irregular working hours	3.8%	10.0%	+164.4%	1.8%	-	0.0%	-	0.0%	-

Figure 3 summarises the most relevant contributing factors to the formation of the clusters, rearranged by categories according to the dataset framework.

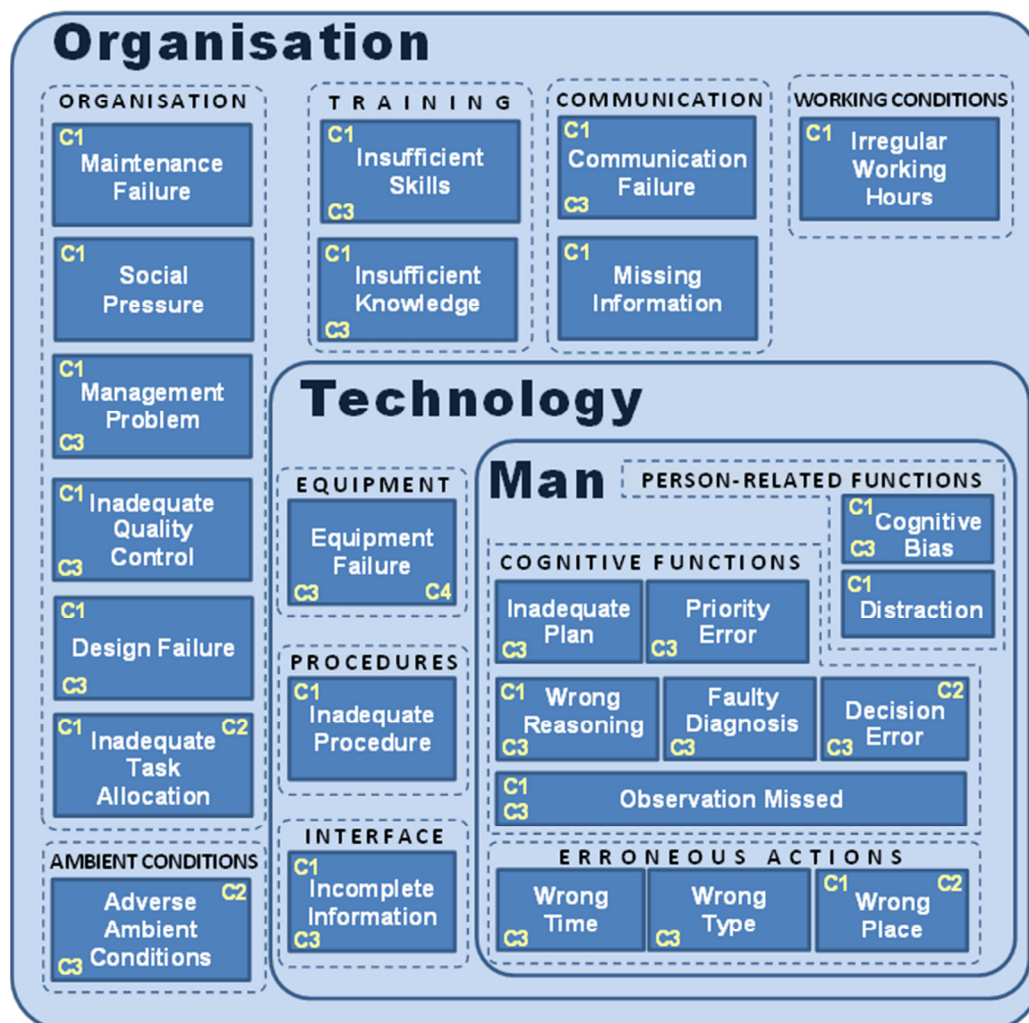


Figure 3 – Categories of the most significant contributing factors per cluster

From a human factors perspective, Cluster 1 accidents were dominated by the Wrong Place phenotype, when an action from an expected sequence is skipped, carried out in the incorrect order or substituted by an unrelated movement. Action errors interfaced with intermediate levels of human cognition, as operators were required to observe a signal or event (observation missed) and diagnose a situation or system state (faulty diagnosis). Inference or deduction errors (wrong reasoning) were also observed. This was the grouping where person-related features were more significant, as shifts in attention (distraction) or constraining the information search to confirm a pre-defined hypothesis, attributing events to specific factors or believing that actions have controlled the system state developments (cognitive bias) contributed to 11.3% and 15% of the cluster, respectively. Technology issues included procedure shortcomings (78.7% of the cluster) and situations where the information provided by the system interface was poor (incomplete information). Many organisational issues interacted within the cluster. Inadequate Task Allocation (95%), Design Failure (85%) and Inadequate Quality Control (81.3%) were the most significant ones, but training (Insufficient Skills and Insufficient knowledge) and communication issues (Communication Failure and Missing information) were considerable as well. Maintenance issues were visible in 56.3% of the cluster, and the effects of other organisational aspects such as social pressure (17.5%), management problem (12.5%) and irregular working hours (10%) were also majored by the application of the clustering technique.

Cluster 2 has Inadequate Task Allocation as the most relevant factor, covering 68.4% of the grouping, followed by an erroneous action (Wrong Place) associated with an inability to decide, a partial/incomplete decision or making the wrong decision among alternatives (decision error). Accidents where Adverse Ambient Conditions were significant are mostly grouped within this cluster.

As indicated by Figure 2 histogram, Cluster 3 shows several important interactions among contributing factors, being a rich grouping for further interpretation. Many action errors were captured during the investigation of these events, where movements were performed earlier or later than required (Wrong Time), or with insufficient force, wrong speed, direction or magnitude (Wrong Type). Erroneous actions were accompanied by all three levels of cognition (observation, interpretation and planning). The fact that complex cognitive functions such as Inadequate Plan (25.6%) and Priority Error (15.4%) contributed to the formation of the cluster, together with observation missed (23.1%), wrong reasoning (25.6%) and decision errors (17.9%), gives us an opportunity to understand how cognitive functions leading to erroneous actions interact with organisational and technological aspects. Equipment failures contributed to almost the totality of the grouping. As in Cluster 1, Design Failure, Inadequate Quality Control and training (Insufficient Skills and Insufficient Knowledge) records were very high, and other aspects such as incomplete information and communication failure were also significant for both groupings. Management problems were observable in 23.1% of Cluster 3.

Cluster 4 is largely dominated by Equipment Failures (87.1%), the only noteworthy factor to influence the formation of grouping.

Figures 4 to 22 represent the cluster results for individual features. They give further insight into the general map (Figure 1), in the sense that the interactions among individual contributing factors can be visualised. Figures are read as topographical or heat maps, where one can directly observe the regions of the general map affected by individual properties or dimensions. In the current work, blue tones indicate the absence of the contributing factor, while red tones represent its manifestation. Shaded regions highlight and exemplify special interactions among contributing factors.

Two graphical methods will be used to present individual maps and highlight the main results for further discussion:

- (i) disclosing multiple intersections (superposition of images) of the most frequent contributing factors, which represent strong interaction patterns between human factors, technology and organisations (e.g. Figures 4 to 10 and 18 to 22); and
- (ii) analysis of special features (e.g. communication issues in Figures 11 to 14, human-related factors in Figures 15 to 17).

In Cluster 1, three map regions (1A, 1B and 1C) represent the intersection between Inadequate Task Allocation, Design Failure, Inadequate Quality Control and Inadequate Procedure (Figures 4 to 7). Region 1A is deeply related to Insufficient Knowledge (Figure 8), while 1B is mostly associated with Insufficient Skills (Figure 9). Accidents represented in 1C tend to combine with Maintenance Failures (Figure 10).

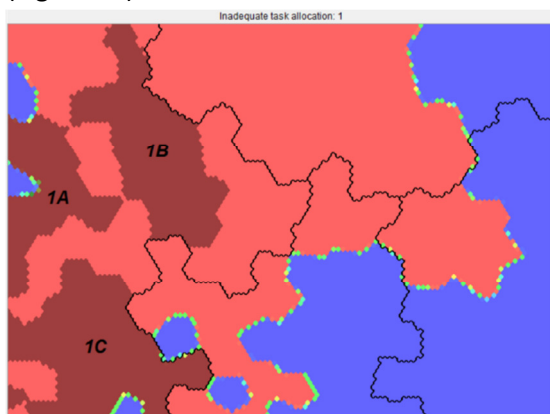


Figure 4 – Inadequate Task Allocation Map

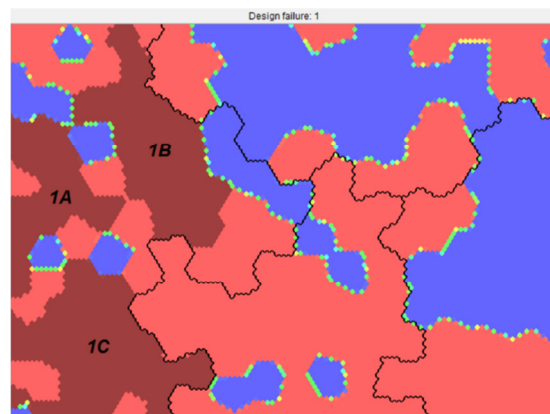


Figure 5 – Design Failure Map

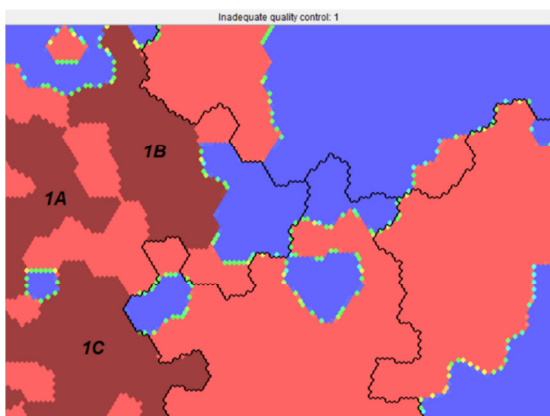


Figure 6 – Inadequate Quality Control Map

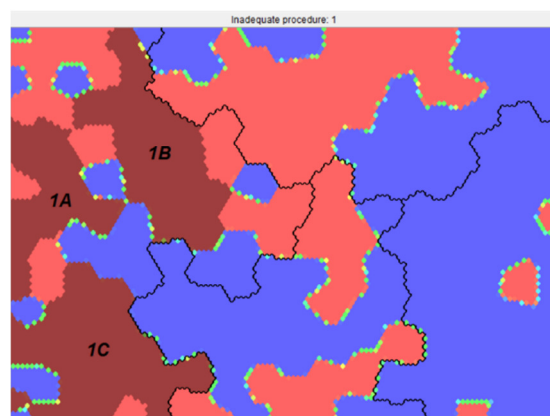


Figure 7 – Inadequate Procedure Map

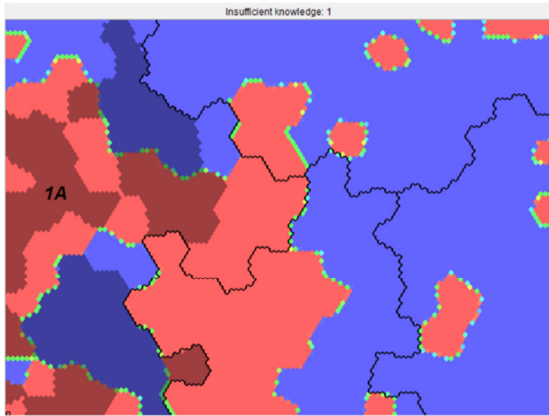


Figure 8 – Insufficient Knowledge Map

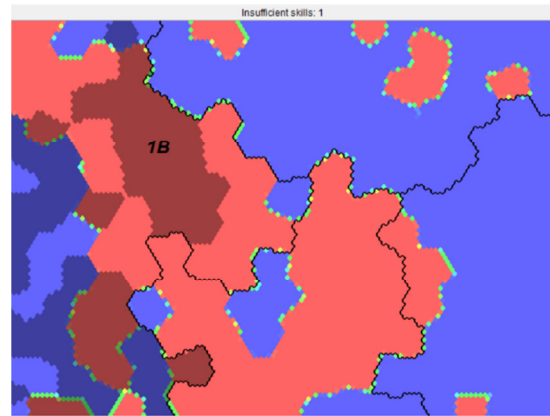


Figure 9 – Insufficient Skills Map

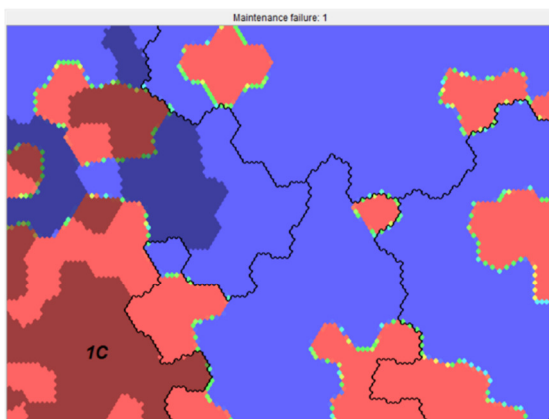


Figure 10 – Maintenance Failure Map

Figures 11 and 12 present the SOM maps for communication issues. These issues largely overlapped Inadequate Task Allocation in Cluster 1, as can be seen in the shadowed region in Figure 13. Exceptions are the two small circled areas, where task allocation issues were substituted by the person-related feature named Cognitive Bias (Figure 14).

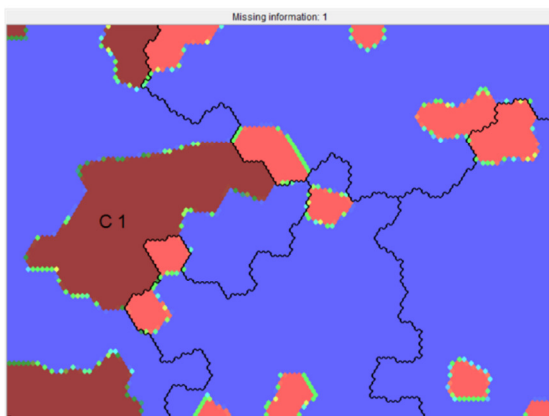


Figure 11 – Missing Information Map

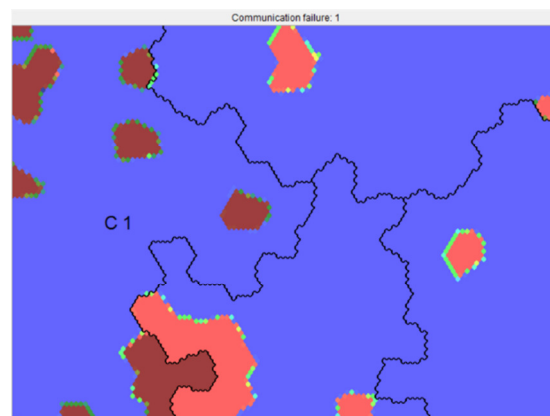


Figure 12 – Communication Failure Map

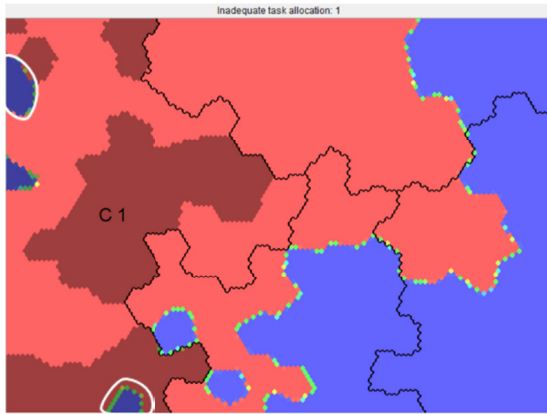


Figure 13 – Inadequate Task Allocation Map

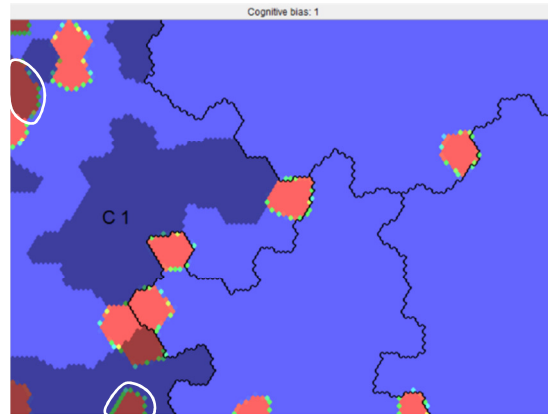


Figure 14 – Cognitive Bias Map

64.1% of Cluster 3's area contained two erroneous actions: Wrong Time (Figure 15) and Wrong Type (Figure 16). The faded region depicts the incidence of the three levels of specific cognitive factors within this grouping, showing the human-related contributing factors' representation. Consequently, a combination of observation (Observation Missed), interpretation (Wrong reasoning and Decision Error) and mental planning (Inadequate Plan and Priority Error) was expected to take place, suggesting that a profound judgement of the confronted situation was necessary to solve system deviations. It can be observed that a technological issue (Incomplete Information – Figure 17) interacted with erroneous actions related to timing in the regions where specific cognitive functions are not identified, suggesting that supervisory control system and data display limitations led to some of the Wrong Time occurrences. These areas are circled in Figure 15.

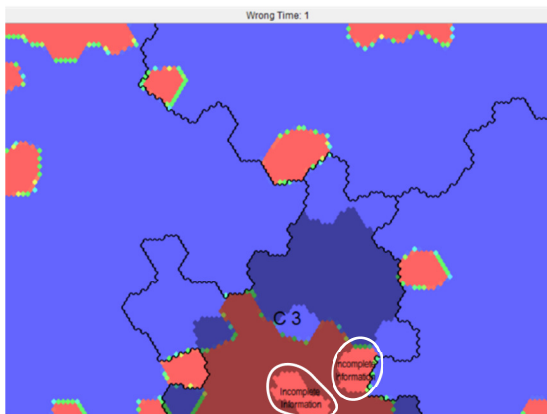


Figure 15 – Wrong Time Map

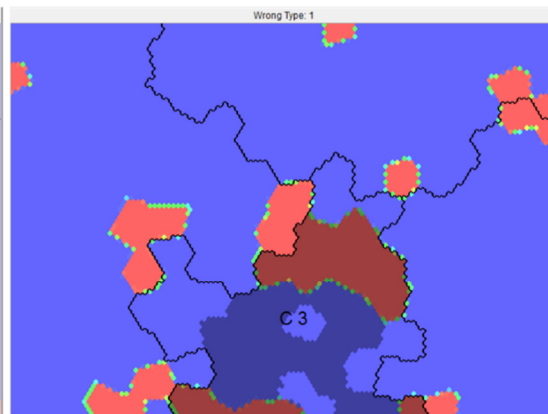


Figure 16 – Wrong Type Map

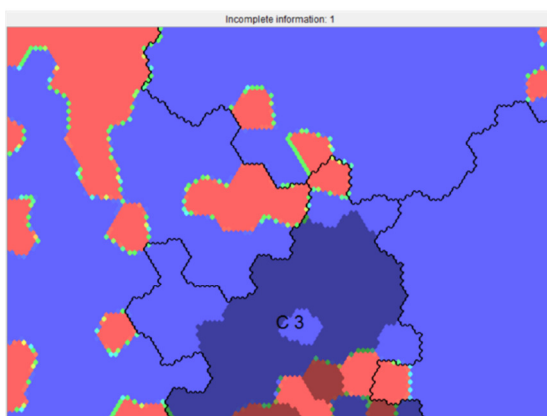


Figure 17 – Incomplete Information Map

Figures 18 to 22 show how the main technological (Equipment Failure) and organisational aspects (Quality Control, Design Failure and training) interacted among them and with human-related issues (shaded region) to result in system control problems within Cluster 3. The shaded region is 79.5% of the grouping area, representing the incidence of human erroneous actions, specific cognitive functions and person-related functions.

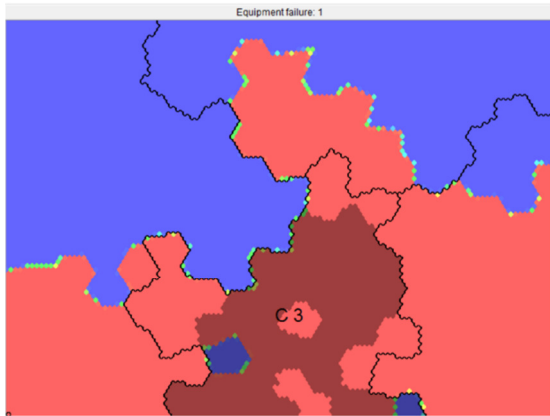


Figure 18 – Equipment Failure Map

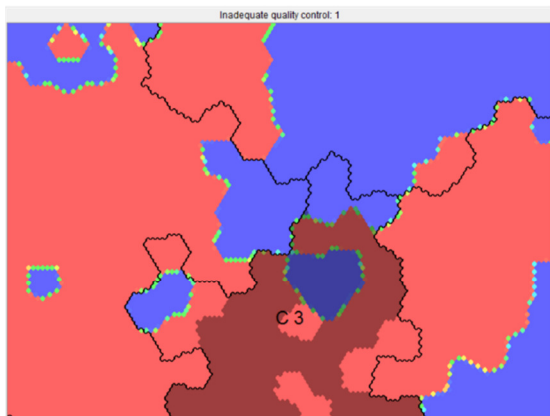


Figure 19 – Inadequate Quality Control Map

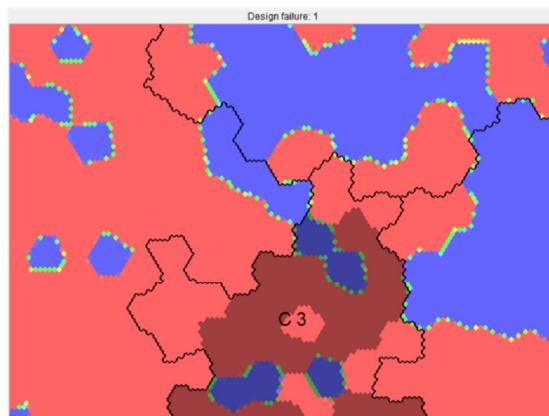


Figure 20 – Design Failure Map

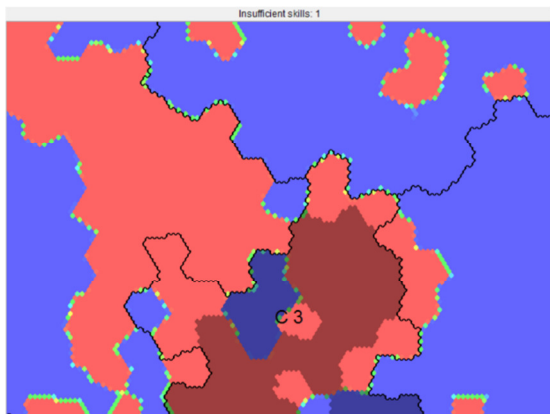


Figure 21 – Insufficient Skills Map

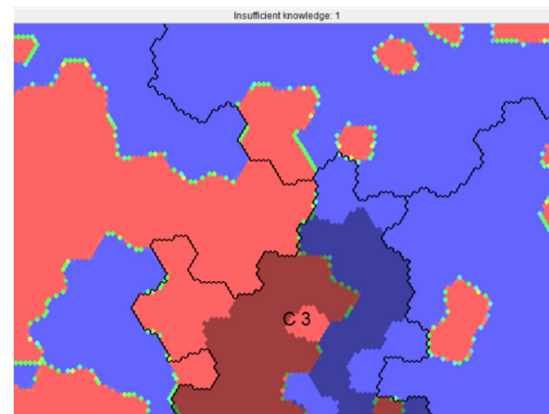


Figure 22 – Insufficient Knowledge Map

4. Discussion

4.1 Main Clusters Interpretation

The analysis of the maps indicate an intricate combination of factors contributing to the major accidents contained in the MATA-D database, including the significance of the human factors to the undesirable outcome. Previous studies (Graeber, 1999, McLaughlin et al., 2000, Leveson, 2004) using different industrial sectors as a data source also emphasised the importance of considering human issues when assessing risk, relating between 70% to 80% of accidents to some kind of operator error. Therefore, it seems to be clear that a satisfactory risk assessment study must take into account the relationship between humans, technology and organisations to convey realistic scenarios. Otherwise, the safety analysis will not offer a trustworthy and sufficiently broad view of the major hazards that industrial facilities are exposed to throughout their lifecycle.

So why has limited attention, especially if compared with traditional risk assessment approaches focused on engineering technical systems, been paid to the consideration of human factors in risk studies (Hollywell, 1996; Skogdalen & Vinnem, 2011, 2012)? When analysing occupational risk assessments, Cuny and Lejeune (2003) pointed out some problems to the consideration of the human influence, particularly the preparation of data for processing and the estimation of probabilities to feed deterministic approaches. The complexity of organisational interfaces and the variability of human behaviour also make a socio-technical system modelling a challenging task, maybe explaining the reason behind the disproportionate focus on purely technical aspects and discrete components in risk evaluation.

The interpretation of the maps enables the possibility of considering the whole range of contributors without previous assumptions of their conjectural importance, focusing on their interactions and on the disclosure of tendencies, instead of concentrating on individual factors. The application of the SOM algorithm and the joint analysis of maps highlighted topographical areas containing similar interfaces, allowing a targeted examination of the genesis of the MATA-D accidents and the development of an attribute checklist with the most frequent observations. Some of these interfaces will be illustrated with the accident narratives as positioned in the map, all accessible through the MATA-D database. Examples of specific items developed to raise the awareness for the discussed interactions will be given, and the complete list will be presented in the appendix.

An analysis of Cluster 1 accidents from area 1A (Figure 8) indicates that these events are related to situations where components were designed and implemented on an individual fashion, rather than as a holistic system. A deeper examination of the contributing factors description for the major accidents contained within this region, according to the MATA-D stored data, indicates that safety studies failed to adequately address risks related to the system interaction with the environment as well as possible interferences among individual components. The shortcomings in design, procedures, quality and task allocation joined the loss of situational awareness during operation, and insufficient theoretical knowledge led to the misperception of risks. A practical example of this tendency was the widely-known Varanus Island incident in June 2008 (Bills & Agostini, 2009), when a pipeline rupture and explosion caused a shortage in the gas supply for Western Australia, resulting in three billion Australian dollars in economic losses. In summary, the lack of an integrated approach to

design and risk management led to problems in the cathodic protection system, most likely due to electrical interferences from adjacent pipes and other structures, causing alternating current corrosion. The assumption that safeguards are always active and the sense that their failures are unconceivable are also patterns observed within the grouping. In order to ensure that these common major accident tendencies are addressed by risk studies, it is suggested that holistic validation questions (e.g. items 10, 11, 15 and 59 of the appendix checklist) should be developed to raise reviewers and designers' risk awareness level.

Accidents within area 1B (Figure 9) presented situations where process changes undermined the original recommendations from risk assessment studies. Equipment or system replacements, product modifications and procedures updates lacking a proper hazard evaluation (or management of change) enabled the deterioration of the system. The necessary training to operate under the new conditions was also insufficient, causing a human performance failure. The validation checklist items 4, 5, 23 and 33 are examples of checkpoints aimed at neutralising these accident patterns highlighted by the MATA-D data mining using the SOM.

The shadowed region 1C (Figure 10) contained many events where seemingly minor maintenance issues, i.e. keeping vessels and pipes free of deposits, consumable parts (e.g. filters) replacements, lubrication and calibration, drains obstruction and dust/particles accumulation, were combined with quality problems, task allocation issues, design shortcomings and inadequate procedures to generate a major failure. These common accident causation trends gave rise, for instance, to the items 26, 27, 28, 29 and 39 of the validation checklist.

Figures 11 and 12 highlighted the map regions where communication problems attained their highest incidence, mostly combined with task allocation issues (Figure 13). These events were prone to poor communication between workers, which was made worse by background noise (mainly alarms and usual process sounds) or by the low quality of the transmission. Deficiencies in reporting to supervisors some unusual situations observed in the process plant and to convey important information from hazard studies to the personnel were frequent within this grouping. In addition, data transfer from paper to computer-based systems, incorrect coding and poor communication between shifts were risk-increasing factors commonly observed.

Through the results shown in Figures 13 and 14, it is possible to scrutinise a few regions where inadequate task allocation was not as relevant as in the rest of the cluster. Nonetheless, communication issues tended to interact with person-related issues such as a Cognitive Bias, particularly when critical information was not communicated, supporting an illusion that actions taken were sufficient to control the situation, or when actions were constrained by a strong (and wrong) assumption of the current system status. An example extracted from these regions would be the 2011 helicopter crash in Missouri (NTSB, 2013) during a patient transfer from one hospital to another, which resulted in 4 fatalities. The Pilot knew that he has misinterpreted the fuel level to some extent (he reported 26% or forty-five minutes of fuel in the pre-flight check, but post-accident investigation indicated only 18%, or a thirty-minute supply), but his alternative refuelling plans were constrained by the hypothesis that he was able to reach a station 34 minutes away from the departure point. Having visual contact with the refuelling point (three-minute distance) when the gauge indication approached to zero, the pilot sustained his course (instead of landing immediately)

until fuel exhaustion. A communication with qualified land-based staff (available at the Operational Control Centre) would have exposed his plans as inadequate. Other interesting tendencies were also identified in the cluster region, such as having the attention caught by phone calls or texting using portable devices.

Examples on how to check if the tendencies identified by the interpretation of Figures 11 to 14 SOM maps were adequately tackled by safety studies are presented on items 19, 21, 34, 35, 36 and 65 of the validation list in the appendix.

Cluster's 3 erroneous actions and cognitive functions' frequencies are generally higher than in any other grouping, especially the most complex ones, involving the need for mental planning. These human-related factors merged with design shortcomings, equipment failures and quality control issues. A tendency to underperform under non-standard operations (e.g. start-up or partial plant operation) was also observed, repeatedly combined with training issues (Figures 21 and 22). Cases where an equipment failure caused a shutdown, and operators focused on fixing the equipment and restarting it without further consideration are recurrent in this grouping. Some of the common failure modes observed are: (i) catastrophic failures due to the hot flow of products into cold pipes and vessels (brittle fractures); (ii) valves and seals which were damaged or partially closed/opened during a pause in operations and were not inspected (a quality control problem) before the restart; and (iii) omitting to realign valves and restart control/signalling/alarm systems. After the identification of specific patterns in Cluster 3, some suggestions on how to check if safety studies considered these interactions are given by items 30, 32 and 33 of the risk validation checklist.

The grouping also contained some regions where insufficient information from supervisory control and data acquisition systems shaped human erroneous actions (Figures 15, 16 and 17). The growing dependence on information systems is a pattern to be considered when assessing hazards, thus validation schemes must verify if the risk growth due to inadequate/unsatisfactory human-machine interfaces is carefully addressed. The lack of direct indications of problems; panels not providing accurate process overviews; information that is not displayed in relevant places (e.g. in the control room and/or locally); general/critical alarms not taking precedence in relation to local, less important alarms; delays in the information presented, undermining operators' efforts to diagnose system status; and incorrect information display are some of the human-machine interface problems extracted from Cluster 3. These accident causation tendencies resulted in the development of the items 37, 38, 62, 63 and 64 of the proposed example of a safety study validation checklist.

4.2 An Application Example for Safety Studies Validation

A safety study generally comprises a planning process (describing the context, regulatory requirements, scope of the study, risk acceptance criteria etc), a hazard identification phase, a risk assessment (e.g. events frequencies, reliability, event modelling, consequences, level of risk estimation) and a final report (e.g. presentation of results, uncertainties appraisal, recommendations, study quality assurance), to generate input to the decision-making process.

Those stages involve different techniques, combining quantitative (e.g. probabilistic analysis), semi-quantitative (e.g. fault-trees) and qualitative (e.g. Failure Modes, Effects and Criticality Analysis –

FMECA; Hazards and Operability Studies – HAZOPs) approaches. The wide range of techniques and possibilities to assess risk turn the verification and validation of safety studies into a challenging task, especially if it is recognised that a non-integrated approach to validation may result in relevant gaps in the overall consideration of risks, leaving outside the scope important interactions that might contribute to major accidents.

The international standard for quality management systems (ISO 9001, 2015) defines verification as the process to ensure that input requirements are met by design and development outputs. It includes activities which aim to certify that the results of the safety study are aligned with the chosen assessment approach and accurately reflect the conceptual description of the system/facility. Correspondingly, validation (ISO 9001, 2015) includes a set of activities performed with the intent to ensure that the output meet the requirements for the projected usage. This process aims to confirm that the safety study contains a sound representation of risks, considering the intended application of the model.

Therefore, the current study contributes to validation activities by providing practical means to confirm if lessons from past major accidents have been contemplated by safety assessments. The ultimate objective is to ensure that patterns and tendencies leading to disasters are anticipated and mitigated, thus assisting the development of societal and stakeholders' trust on safety studies.

Once the application of the SOM algorithm successfully supported the disclosure of common patterns in major accidents, the interpretation of the most important interactions can be further used to check if safety studies present a consistent representation of risks, by testing its robustness against the complex and realistic scenarios extracted from the MATA-D.

An example of how to use the accident patterns to contribute to validation schemes is given in the appendix. Accident tendencies disclosed by the application of the SOM approach and the consequent analysis of the maps, as illustrated in section 4.1, are being used to construct a checklist comprising common hazards, major risks and shortcomings involving the interactions between humans, technology and organisations. The checklist aims to give some insight into how these lessons learned from past accidents in different industrial environments can be used as part of a validation scheme, to ensure that recurring accident causation patterns will not escape from scrutiny in new safety studies. A general, comprehensive language is being used, in order to facilitate the direct application of the list or the integration with existing verification and validation approaches.

It is suggested that a large number of positive answers would represent a safety study that intrinsically contains solutions for the accident causation patterns encountered in the MATA-D scenarios, which caused major disasters in high-technology systems. Negative answers would indicate possible weaknesses in the safety study, which should be addressed in order to improve trust. For items that are neither relevant nor related to the assessed installation or system, a neutral answer (non-applicable) can be given. After confirming that the major problems raised by the list were addressed, the safety study can be seen as robust, from a "lessons learned" perspective.

Although this work present a general list as an example, applicable to virtually any industrial sector, the tendencies drawn from the MATA-D with the SOM approach are perfectly adaptable to specific

industries such as oil & gas or aviation, and can be further developed for particular applications, e.g. design reviews. The accident causation patterns can be also used as input to the development of risk scenarios.

5. Conclusions

Validation schemes must analyse proposed risk reduction measures, taking into consideration that systems are dynamic. Assumptions such “as good as new” systems/equipment, perfect procedures and faultless operators are accurate only on paper, and should be challenged by verifiers. The appendix presented a sixty-six-item attribute list which enables this debate and exposes possible shortcomings, addresses major hazards and stimulates improvement. The objectives are to give impetus to broader considerations about risk in real projects and raise the discussion about the implementation or dismissal of recommendations and solutions, enabling the dialogue among stakeholders and bringing transparency to the whole process.

Also, the prime attribute of a project is its feasibility, which means cost. This attitude is absolutely normal and engrained in our social behaviour (Does anybody check safety records before booking a flight, or the price is the first – sometimes the only – attribute considered in the decision-making process?). Therefore, promoting the coexistence and balance between economic aspects (i.e. resources, budget) and safety performance is the ultimate goal pursued by risk managers. It is a permanent persuasion exercise to which the current research intends to contribute, by developing means to enlighten stakeholders to consider a wider picture of risk.

The problem of trust in risk management and risk validation is not surprising at all. Risk assessment is a complex and multidisciplinary matter, and there is no such thing as a definite standard reference on how to perform a safety study. Distinct techniques and approaches are not mutually exclusive and should be simultaneously used, making the development of a single validation method or procedure hardly possible. However, the most important outcome of a risk study is to support the decision-making process. Hence, it must be able to communicate risks to stakeholders, addressing potential problems and solutions in a clear way, and using visual aids such as maps can help tackling this challenge.

In this regard, the conversion of the MATA-D dataset into self-organised maps and their subsequent interpretation successfully converged into a comprehensive checklist containing items representing major accident tendencies, to be verified against risk studies and to help developing confidence that critical issues were taken into consideration. These concerns arose from shortcomings in many different industrial segments, also promoting an inter-industry exchange of valuable accident lessons. The questions can be easily traced back to regions in the maps, and practical examples of flawed interfaces between humans, technology and organisations can be extracted, in order to illustrate the possible adverse effects of not dealing with specific conditions. The 2-D SOM maps can be used to communicate and describe complex interfaces to a broader public in a simpler way, enhancing stakeholder’s confidence that genuine strategies to mitigate risks are in place and the study was adequately completed.

Acknowledging that there is not a single method to validate risk studies, the application of the widest possible range of approaches to stimulate the comparison of alternatives and different experts' opinion can give some insight into how to enhance trust in risk management. This work focused on ensuring that lessons from several past accidents are considered by new risk studies as good engineering practice and a sensible approach to reduce risk, by means of a straightforward risk study validation checklist.

Furthermore, the validation framework can be easily applied by a range of independent reviewers from industry and academia, which could use the checklist output to involve experienced people and develop innovative risk approaches, bringing new ideas and insights to safety studies in a structured way.

6. Acknowledgements

This study was partially funded by CAPES [Grant nº 5959/13-6].

7. References

Arstad, I and Aven, T., 2017. Managing major accident risk: Concerns about complacency and complexity in practice. *Safety Science* 91: 114-121.

Aven, T., 2013. On the meaning of the black swan concept in a risk context, *Safety Science* 57: 44–51.

Aven, T., 2015. Implications of black swans to the foundations and practice of risk assessment and management, *Reliability Engineering and System Safety* 134: 83–91

Baysari, M., McIntosh, A. and Wilson, J. 2008. Understanding the human factors contribution to railway accidents and incidents in Australia, *Accident Analysis and Prevention* 40: 1750-1757.

Bellamy, L.J. et al., 2007. *Storybuilder—A Tool for the Analysis of Accident Reports*. *Reliability Engineering and System Safety* 92: 735–744.

Bellamy, L.J. et al., 2013. Analysis of underlying causes of investigated loss of containment incidents in Dutch Seveso plants using the Storybuilder method, *Journal of Loss Prevention in the Process Industries* 26: 1039–1059.

Bills, K & Agostini, D., 2009. *Offshore petroleum safety regulation – Varanus Island Incident Investigation*. Government of West Australia. ISBN: 978-1-921602-56-6

Blajev, T. 2002. *SOFIA (Sequentially Outlining and Follow-up Integrated Analysis) Reference Manual*. Brussels: EATMP Infocentre.

British Petroleum., 2010. *Deepwater Horizon – Accident Investigation Report, 8 September 2010* [Online]. Available from: http://www.bp.com/content/dam/bp/pdf/sustainability/issue-reports/Deepwater_Horizon_Accident_Investigation_Report.pdf (Accessed 25 September 2016).

Bureau of Ocean Energy, Management, Regulation and Enforcement (BOMRE)., 2011. *Report regarding the causes of the April 20, 2010 Macondo well blowout* [Online]. Available at: <https://www.bsee.gov/sites/bsee.gov/files/reports/blowout-prevention/dwhfinaldoi-volumeii.pdf> (Accessed 25 September 2016).

Center for Catastrophic Risk Management (CCRM), 2011. *Final Report on the Investigation of the Macondo Well Blowout* [Online]. Available at: http://ccrm.berkeley.edu/pdfs_papers/bea_pdfs/dhsgfinalreport-march2011-tag.pdf (Accessed 25 September 2016).

Cohen, M., March, J. & Olsen, J., 1972. *A Garbage-Can Model of Organisational Choice*, *Administrative Science Quarterly* 17(1): 1–25.

Cottrell, M., Olteanu, M., Rossi, F. and Villa-Vialaneix, N. 2016. Theoretical and Applied Aspects of the Self-Organizing Maps, *Proceedings of the 11th International Workshop WSOM 2016*, Houston, Texas, USA, January 6–8, 2016.

Cuny, X. and Lejeune, M., 2003. Statistical modelling and risk assessment, *Safety Science* 41: 29–51.

Davis, G., Wanna, J., Warhurst, J. & Weller, P. 1998. *Public Policy in Australia*. 1st edn. Sydney: Allen & Unwin.

Doell, C., Held, P., Moura, R., Kruse, R., and Beer, M., 2015. Analysis of a major-accident dataset by Association Rule Mining to minimise unsafe interfaces, *Proceedings of the International Probabilistic Workshop (IPW2015)*, Liverpool, UK, November 4-6, 2015.

European Safety, Reliability and Data Association (ESReDA), 2015. *Barriers to learning from incidents and accidents* [Online]. Available from: <http://esreda.org/wp-content/uploads/2016/03/ESReDA-barriers-learning-accidents-1.pdf> (Accessed 25 September 2016).

Evans, A., 2011. Fatal train accidents on Europe's railways: 1980-2009, *Accident Analysis and Prevention* 43: 391-401.

Grabowski, M. & Roberts, K., 1997. Risk Mitigation in Large-Scale Systems: lessons from high reliability organisations. *California Management Review* 39(4): 152-162.

Graeber, C., 1999. *The Role of Human Factors in Aviation Safety in Aero Magazine QTR_04 1999* (p. 23-31). The Seattle: Boeing Commercial Airplanes Group.

Heinrich, H., Peterson, D. & Roos, N., 1980. *Industrial Accident Prevention*. 5th edn. New York: Mcgraw-Hill.

Hollnagel, E. 1998. *Cognitive Reliability and Error Analysis Method*. Oxford: Elsevier Science Ltd.

Hollnagel, E., Paries, J., Woods, D.D. & Weathall, J., 2011. *Resilience engineering in practice: A guidebook*. Ashgate.

Hopkins, A., 2003. *Working Paper 7 - Safety culture, mindfulness and safe behaviour: Converging ideas?* National Research Centre for OHS Regulation [Online]. Available at: http://regnet.anu.edu.au/sites/default/files/publications/attachments/2015-05/WorkingPaper_7_0.pdf (Accessed on 15 February 2017).

Hollywell, P.D., 1996. Incorporating human dependent failures in risk assessments to improve estimates of actual risk. *Safety Science* 22: 177–194.

Kohonen, T., 2001. *Self-Organizing Maps*. 3rd ed. Berlin: Springer.

Kohonen, T., 2013. Essentials of the self-organizing map, *Neural Networks* 37: 52–65.

La Porte, T., & Consolini, P. 1998. Theoretical and operational challenges of high reliability organisations: air traffic control and aircraft carriers. *International Journal of Public Administration*, 21 (6-8): 847-852

Leveson, N., 2004. A new accident model for engineering safer systems, *Safety Science Journal* 42: 237-270.

Leveson, N., 2011. Applying systems thinking to analyse and learn from events, *Safety Science Journal* 49, 55-64.

Leveson, N., 2012. *Engineering a safer world: systems thinking applied to safety*. Cambridge Massachusetts Institute: The MIT Press.

Licu, T. et al. 2007. Systemic Occurrence Analysis Methodology (SOAM) - A "Reason"-based organisational methodology for analysing incidents and accidents, *Reliability Engineering and System Safety* 92: 1162-1169.

McLaughlin, T., Monahan, S., Pruvost, N., Frolov, V., Ryazanov, B. & Sviridov, V., 2000. *A Review of Criticality Accidents*. New Mexico: Los Alamos National Laboratory

Moura, R., Beer, M., Patelli, E., Lewis, J. & Knoll, F., 2015a., Learning from Accidents: Analysis and Representation of Human Errors in Multi-attribute Events, *Proceedings of the 12th International Conference on Applications of Statistics and Probability in Civil Engineering, ICASP12*, Vancouver, Canada, July 12–15, 2015.

Moura, R., Beer, M., Doell, C., Kruse, R. 2015b., A Clustering Approach to a Major-Accident Data Set: Analysis of Key Interactions to Minimise Human Errors, *Proceedings of the 2015 IEEE Symposium Series on Computational Intelligence (SSCI2015)*, Cape Town, South Africa, December 8-10, 2015.

Moura, R., Beer, M., Patelli, E., Lewis, J. & Knoll, F., 2016. Learning from major accidents to improve system design, *Safety Science* 84: 37-45.

Moura R., Beer, M., Patelli, E. & Lewis, J., 2017. Learning from major accidents: graphical representation and analysis of multi-attribute events to enhance risk communication, *Safety Science* XX: XX-XX. [in press, DOI [10.1016/j.ssci.2017.03.005](https://doi.org/10.1016/j.ssci.2017.03.005).]

National Transportation Safety Board (NTSB), 2013. *Crash Following Loss of Engine Power Due to Fuel Exhaustion, Air Methods Corporation, Eurocopter AS350 B2, N352LN, Near Mosby, Missouri, August 26, 2011. Aircraft Accident Report AAR-13/02*. Washington, DC: NTSB.

Nielsen, DS. 1971. *The cause/consequence diagram method as a basis for quantitative accident analysis*. Risø-M 1374.

Paté-Cornell, M., 2012. On "Black Swans" and "Perfect Storms": risk analysis and management when statistics are not enough, *Risk Analysis* 32 (11): 1823-1833.

Perrow, C., 1984. *Normal Accidents: Living With High-Risk Technologies*. New York: Basic Books.

Rasmussen, J., 1997. Risk management in a dynamic society: a modelling problem, *Safety Science* 27: 183–213.

Reason, J., 1990. *Human Error*. Cambridge: Cambridge University Press.

Reason, J., 1997. *Managing the Risks of Organizational Accidents* 1st ed. Farnham: Ashgate Publishing Ltd.

Roberts, K. 1990. Some Characteristics of one type of high reliability organizations. *Organization Science* 1(2): 160-176.

Sagan, S., 1993. *The Limits of Safety: organisations, accidents and nuclear weapons*. New Jersey: Princeton University Press.

Shappell, S., et al. 2007. Human Error and Commercial Aviation Accidents: an analysis using the human factors analysis and classification system. *Human Factors* 49(2): 227-242.

Skogdalen, J. and Vinnem, JE., 2011. Quantitative risk analysis offshore - Human and organizational factors, *Reliability Engineering and System Safety* 96: 468–479

Skogdalen, J. and Vinnem, JE., 2012. Quantitative risk analysis of oil and gas drilling, using Deepwater Horizon as case study, *Reliability Engineering and System Safety* 100: 58–66.

Taleb, N., 2007. *The Black Swan: The Impact of the Highly Improbable*. 2nd Ed. York: Allen Lane.

Ultsch, A. 1993. Self-organizing neural networks for visualization and classification. In: *Opitz, O., Lausen, B., Klar, R. (eds.). Information and Classification*. Berlin: Springer: 307–313.

United States Chemical Safety Board (US-CSB)., 2016. *Investigation Report – explosion and fire at the Macondo well* [Online]. Available at: <http://www.csb.gov/macondo-blowout-and-explosion/> (Accessed 25 September 2016).

United States Coast Guard (USCG)., 2010. *Report of Investigation into the Circumstances Surrounding the Explosion, Fire, Sinking and Loss of Eleven Crew Members Aboard the Mobile Offshore Drilling Unit Deepwater Horizon* [Online]. Available at: <https://www.bsee.gov/sites/bsee.gov/files/reports/safety/2-deepwaterhorizon-roi-uscg-volume-i-20110707-redacted-final.pdf> (Accessed 25 September 2016).

Zuijderduijn, C., 2000. *Risk management by Shell Refinery/Chemicals at Pernis, The Netherlands*. EU Joint Research Centre Conference on Seveso II Safety Cases, Athens.

Appendix – Checklist for risk studies validation

No.	Item	Yes	No	n/a *
01	Were the premises, hypothesis and justifications for the chosen design concept clearly stated? Was a safer known alternative/approach to achieve the same objective discussed?			
02	Are the underlying basis and limitations of the method, the origin of the input data and further assumptions (e.g. duration of an event, flammable vapour clouds expected drifts, maximum spill size, release composition) that support probabilities, scenarios and results clearly stated? Are they consistent?			
03	Are events' frequencies used in probabilistic risk analysis reliable? Are they used exclusively when historical data is comparable (e.g. same operation type, facility or equipment)? Would alternative approaches (e.g. non-frequentist) be more suitable to estimate the events' likelihood in the study case (e.g. no sufficient past experience or previous operation data)?			
04	Although some regulations prescribe periodic reviews to risk studies, there is a tendency that assessments may fall into disuse due to people, process or environmental changes in between revision deadlines. Modifications usually lead to a management of change and some sort of risk analysis, but more complex, previous deeper safety studies are not revisited at this point. Are design verifications, as-built drawings, production checks, field data collection or other approaches required to confirm/maintain trust on the major/approved risk study throughout the facilities' lifecycle, instead of using a rigid deadline for review? Have the facility's critical factors / performance indicators that could indicate an up-to-date and trustworthy risk assessment been identified/listed?			
05	Were possible critical changes affecting the original studies (e.g. in the operational philosophy, control logic and process modernisations) acknowledged? Are the conditions with the potential to invalidate the current safety study clearly stated?			
06	The safety studies must contemplate a list of recommendations and safeguards, which can be rejected on a technical basis. Is the value of the implementation of risk reduction measures clearly stated? Are the justifications for favoured alternatives or rejections consistent with the best available knowledge? Do the underlying principles for rejections contemplate safety benefits over cost matters?			
07	Is the data extracted from databases and standards (as well as calculations made) logical, traceable and consistent with the operational reality?			
08	Were previous assessments in analogous installations used to give some insight into the hazard identification process?			
09	Were the recommendations and risk control measures previously applied to analogous facilities? Is there any feedback about their suitability from previous designers and operators?			
10	Safety studies have shown a tendency to fail to adequately address risks related to the system interaction with the environment as well as possible interferences among individual components and systems. Was a comprehensive and integrated approach to design and risk management achieved? Were components and systems designed and implemented in a holistic way rather than on an individual and secluded fashion? Are human factors analysis integrated with engineering studies?			
11	Some high-technology facilities are likely to start their operations before the whole system and all safeguards are in place. Offshore platforms may have to adapt their process while a pipeline is not operating or a pump/compressor is not commissioned. Refineries may be designed (or obliged) to operate without some processing modules, due to technical or economic reasons. Does the risk assessment contemplate all modes of operation (e.g. commissioning, start-up, partial operation, maintenance breaks) for the facility examined? Are transitory states (e.g. warm-up and cooling down times) also considered?			

12	Have the studies taken into consideration thermal properties, hydraulics and electrical/electronic parts of components, equipment and systems, not being overly focused on mechanical/structural aspects?			
13	Equipment and structural failures tended to arise from problems during the material selection stage and due to poor understanding and monitoring of well-known damage mechanisms. Has the material selected for construction, equipment fixation, pipelines and support structures identified and analysed by safety studies? Was a compatibility assessment (with loads, system and environment) conducted, including thermal, chemical and electrical properties?			
14	Are the specificities of the assessed facility or process clearly identified, in a way that specific risks will be identified and addressed? Where expert advice is required to assess risk, are the correspondent technical reports included in the safety studies (e.g. to assess the possibility of catastrophic failures due to stress corrosion cracking in stainless steels, or corrosion mechanisms emerging from the saturation of wet hydrocarbons with dissolved carbon dioxide and sour environments)?			
15	Are risks associated with the interaction of different materials addressed (e.g. with different temperature gradients leading to deformations and ruptures or with distinct electric potential resulting in galvanic corrosion)?			
16	Are major hazards, complex areas and critical operations clearly identified? Are the level of detail, the methodology to assess these problematic cases and the safeguards proposed by studies compatible with the magnitude of the risks identified?			
17	Are the steps taken to construct the risk scenarios developed in a logical way? Does the study sequence lead to a clear and rational understanding of the process and its possible outcomes?			
18	Does the criterion for setting accident scenarios, specially the worst-case one(s), consider common-cause, domino or cascading effects and simultaneous/multiple scenarios?			
19	Are the risks associated with third-party operations (material delivering, fuelling, electrical power, water supply) addressed by the safety studies? Are these risks considered in a holistic approach, occurring simultaneously and integrated with the facility's risks?			
20	Are risks associated with auxiliary systems (e.g. cooling and heating) contemplated?			
21	Is technology evolution naturally considered by safety studies? Is the increasing usage of operational and non-operational portable devices (e.g. mobile phones, tablets, cameras, smartwatches and fitness wristbands) considered, for instance, as potential ignition sources in explosive/flammable atmospheres? Does human reliability analysis and task allocation processes consider the new technologies potential to impact the performance of workers (e.g. attention shifters)?			
22	Have the studies evaluated the process plant safety when experiencing the effects of partial or total failures in critical elements (e.g. emergency shutdown valves fail in the safe position)?			
23	Are process changes that modify the risk level clearly identified when, for instance, safety critical equipment or systems are removed, deactivated or bypassed/inhibited for maintenance?			
24	Is the availability of safeguards and further risk control/mitigation measures addressed?			
25	Were critical equipment and components with limited life span properly identified? Were replacement operations affecting safeguards and/or increasing risk addressed?			
26	Is quality control an active element of the risk assessment? Is it compatible with operational requirements for systems and equipment?			
27	Are suitable quality indicators proposed to verify critical system elements status?			

	Is there an auditable failure log, to confirm that the expected performance of components and systems is maintained through time?			
28	Are chemical reactions and adverse events associated with housekeeping procedures (e.g. cleaning and painting substances, dust management), inertisation processes, equipment and pipeline deposits removal and necessary tests (e.g. hydrostatic tests) contemplated by the studies?			
29	Were the design and process reviewed aiming at their optimisation to avoid pocket/stagnant zones for dusts, gases, fumes and fluids (e.g. reducing elevated spaces and corners prone to dust/particles built-up or minimising lower pipeline sections subjected to particles/heavier fluids decantation)?			
30	Is the necessary information supporting non-routine tasks aiming at the risk reduction (e.g. pre-operational or restart inspections) sufficiently detailed, allowing the identification of process weak-points such as deposits accumulation, valve misalignments, damaged seals and rupture disks and equipment condition after, for instance, a process halt, or after maintenance works nearby and before resuming operations?			
31	Are permanent cues and signals (e.g. pipeline and equipment marking to indicate content, maximum pressure and direction of flow) proposed as risk reduction measures for standard and non-standard operations? If so, is the permanent marking wear through time a factor considered?			
32	“The operator” is an entity sometimes subjected to extreme variations. When human intervention is considered by safety studies, are the expected skills (e.g. practical experience, acceptable performance variability level) and knowledge (e.g. the situational awareness level and the academic level – technician, engineer, expert) clearly indicated?			
33	Underperforming when conducting non-standard operations (e.g. start-up, commissioning or partial plant operations) was also a noteworthy pattern. Were situations and conditions where an enhanced level of training (skills or knowledge) or even the support of specialised companies (e.g. to control an offshore blowout) are required to keep risks controlled or to reduce the consequences of undesirable events identified?			
34	Is the essential risk information and knowledge arising from safety studies, which should reach the involved personnel, identified? Are there any special provisions to ensure that critical information will be conveyed by proper means (e.g. awareness campaigns, training, written procedures, simulation exercises) and will be accessible where needed?			
35	Is operational reality such as process conditions (e.g. background noise, fumes, heat, wind from exhaustion systems or alarms) considered as a possible disturbance when some sort of communication is required to convey important information?			
36	Are administrative/management aspects affecting the seamless continuity of operations (e.g. loss of information due to shifts, personnel replacement or reduction) addressed during the identification of safety critical tasks hazards? Is the prospect that obvious unusual situations (e.g. seemingly small leakages, unfamiliar odours and a flange missing some screws) may not be reported to supervisors promptly, affecting the effectiveness of risk reducing measures such as process plant walkthroughs, considered?			
37	Do supervisory control and data acquisition systems produce a real-time operation overview, not being excessively focused on individual parameters?			
38	Were the accessibility and visibility of instruments and equipment identified as critical in the risk studies and been ensured by an examination of the design drawings? Were 3-D models and/or mock-ups used to facilitate the visualisation of complex areas and reduce the possibility of interferences/visualisation issues? Are the external critical indicators/gauges fitness to the operational environment verified (e.g. visual impairment or working issues due to snow, rain or sun radiation)?			

39	Was the possibility of obstruction of water intakes, air inlets, sensors and filters (e.g. by water impurities, air particles or formation of ice) assessed? Are mitigation measures in place?			
40	Have operators examined if the information supplied by indicators, panels and displays are sufficient, as active members of the safety assessment team? Do they have similar training level (skills and knowledge) as required for the operation of the system?			
41	Is there an assessment of the usefulness of the information provided by supervisory control and data acquisition systems? Are the functions and outputs clear, in particular to operators? Do they know when and how to use the information provided, or some of the signals are perceived as excessive/useless?			
42	Was the need to diagnose the system status and conduct special operations from alternative places (e.g. stop the operation from outside the control room) considered?			
43	Are supervisory control and data acquisition systems failure modes assessed as critical hazards? Is the possibility that spurious or ambiguous error messages or information insufficiency/delays triggering human or automatic actions that can jeopardise the stability or integrity of the system carefully analysed? Were adequate mitigating measures put in place?			
44	Is the damage to power and control cables, pipelines and hydraulic systems, their routing and its consequences to the supervisory control and data acquisition systems considered by the risk assessment?			
45	Are safety critical alarms clearly distinguishable from other operational alarms?			
46	Are process facilities and hazardous materials located within a safe distance from populations, accommodation modules, administrative offices and parking spaces? Is the storage volume of hazardous substances optimised to reduce risks? Is the transportation route for hazardous materials optimised in a way that the exposure of people to risks is reduced to the minimum practical?			
47	Are control rooms and survival/scape structures protected from damage and located within a safe distance from the process plants? Does the risk study consider a scenario of control room loss? Is there any redundancy in place for emergency controls (e.g. fire control systems, shutdown systems)?			
48	Are visual aids used as risk-reducing measures to increase the awareness level of operators? Are reactors, vessels and equipment arrangement and dimensions visually distinctive from each other (e.g. by position, size or colour) to minimise swap-overs or inadvertent manoeuvres?			
49	Is the possibility of inadvertent connections of similar electrical, mechanic and hydraulic connectors an assessed risk? Are measures in place (e.g. using different connector dimensions or distinct thread types) to minimise hazardous interchangeability among connectors, elbows and other parts from different systems or functions?			
50	Is the inadvertent operation of temporarily or permanently disabled components, equipment or systems considered as a risk-increasing factor? Are measures in place to enhance the visualisation of non-operational parts such as isolated valves? Are overpressure safeguards (e.g. safety valves and rupture disks) accessible and visible from the operational area of the equipment or system they are designed to protect?			
51	Are ignition sources (e.g. exhaust, electrical equipment) optimised in order to be located within a safe distance from significant inventories of flammable materials (including piping) or in a position in which ignition is minimised, in case of leakage? Was the position of flares and vents revised by safety studies? Are exhaust gases routed to and flares and vents located in areas where the risk of ignition is minimised?			
52	Are different scenarios (e.g. in distinct plant locations, with variable volumes) for			

	pipeline and vessels leakages considered by safety studies? Are there risk-reduction strategies to limit the released inventory in case of leakage (e.g. the installation of automatic emergency shutdown valves between sections)?			
53	Are safeguards prescribed by safety studies to minimise the possibility of creation of explosive atmospheres in enclosed compartments (e.g. deluge or inertisation (CO ₂ or N ₂) systems; exhaustion/vents)? Have the possibility of backflow in heating, refrigeration or ventilation systems been examined? Have the logic of automatic systems (e.g. automatic shutoff of air intakes after the detection of gases) and the reliability/availability of surrounding-dependent systems (e.g. positively pressurised rooms and escape routes) been assessed?			
54	Are fire systems, emergency equipment, escape routes and rescue services designed to withstand extreme conditions expected during an accident (e.g. blast, fumes and intense heat)? Are accident probable effects (e.g. impacts from fragments of explosions or the duration/intensity of a fire) considered in the evaluation of the effectiveness/survivability of these systems?			
55	Are alternative emergency power sources provided? Do the safety studies assess their functionality under distinct accident scenarios (e.g. main power cuts, flood, lightning storms and local fires)? Does the transition time from main to alternative power sources pose non-considered risks?			
56	Is there a main safe escape route and further alternatives designed, including load-bearing structures such as anti-blast and firewalls calculated to resist until the facility has been fully evacuated?			
57	Does the escape route contain clearance warnings by means of visual and audible cues? Are local alarm switches located in adequate positions to alert the remaining workers about the best available escape route? Are emergency lighting and alarms connected to the emergency power system (or have their own battery power source)?			
58	Have safety studies assessed the possibility of collisions (e.g. with cars, boats and airplanes) and external elements (e.g. projectiles from firearms) affecting equipment and the structure of the facility? Are measures in place (e.g. mechanical protection, administrative prohibitions, policing) to minimise these risks?			
59	Are distances among pipelines, equipment and modules optimised in order to consider the contents volatility, temperature, pressure and other risk-increasing factors? Is the separation among adjacent elements sufficient to avoid electromagnetic interferences, energy transfer or domino/cascading effects in case of failure? Were additional measures (e.g. physical separations and blast and fire protection walls) evaluated?			
60	When physical separation is not possible, does the safety study evaluated if the surrounding equipment endurance time is sufficient to withstand the consequence of possible failure modes (e.g. a release followed by a jet fire from a failed adjacent element, for the inventory depletion time)?			
61	Does the safety study consider multiple safety barriers prone to common cause failures as a single barrier? Are alarms and sensors subjected to the same failure modes (e.g. same power supply or same cable routing) considered as non-redundant systems? Were redundant safety barriers subjected to an independence evaluation by safety studies?			
62	Are the risk scenarios demanding automatized responses (e.g. fire alarm demanding the activation of deluge systems or gas detection demanding the neutralisation ignition sources) identified and assessed? Does the supervisory control and data acquisition system have the capability of interpreting multiple alarms and command automatized actions or present consistent diagnostics to operators though the interface? Is the harmonisation of automated functions and personnel actions assessed?			

63	Is the position and type of sensors representative of the category of information they intend to convey? Are failures in sensors and indicators auto-diagnosed and clearly indicated by the interface?			
64	Is there a consistent assessment of safety alarms? Is the alarm precedence logic based on its safety significance? Are they prioritised according to how quickly personnel should respond in order to avoid undesirable consequences?			
65	Is the number of simultaneous alarms considered as a risk-increasing factor capable of disturbing cognitive functions? Are less important signals and alarms reduced/supressed (to minimise mental overburden) when the supervisory control and data acquisition system diagnoses a critical situation demanding full attention from the personnel involved?			
66	Are reduction measures for the initiation and escalation of fires and explosions proposed (e.g. reduction of ignition sources, material selection based on flammability level, ability to spread flames, generate smoke or propagate heat and the toxicity level)? Is the likelihood of ignition assessed in susceptible sections of the installation, by consistent means?			
<i>Total</i>				

*non-applicable to the assessed study**