A Clustering Approach to a Major-Accident Data Set: Analysis of Key Interactions to Minimise Human Errors

Raphael Moura, Michael Beer
Institute for Risk and Uncertainty
University of Liverpool
Brodie Tower, Brownlow Street, Liverpool L69 3GQ, UK
Email: {rmoura,mbeer}@liverpool.ac.uk

Christoph Doell, Rudolf Kruse
Institute of Knowledge and Language Engineering
Otto von Guericke University Magdeburg
Universitätsplatz 2, 39106 Magdeburg, Germany
Email: {doell,kruse}@iws.cs.uni-magdeburg.de

***Abstract*—This work aims to scrutinise a proprietary dataset containing major accidents occurred in high-technology facilities, in order to disclose relevant features and indicate a path to the recognition of the genesis of human errors. The application of a tailored Hierarchical Agglomerative Clustering method, using the bray-curtis dissimilarity and two different linkage functions – complete and average – will provide means to understand data and identify key similarities among accidents. Significant interfaces between human factors, the organisational environment and the technology will be described. Main clustering results have shown that accidents featuring communication issues and interface problems were grouped together, and turned out to be the most deadly ones, considering the fatality rate. Another cluster highlighted relevant training shortcomings. Also, design failures, poor quality control and inadequate task allocation were identified as key contributing factors to major accidents. Conclusions to improve the human performance based on these clustering results are then discussed.**

I. Introduction

Major accidents appear to be an unfortunate side of the development of human activities, and the added complexity of high-technology systems seems to challenge the improvement of industrial safety records. Only in the last 5 years, a perplex society faced worldwide tragedies such as the Macondo Blowout in the Gulf of Mexico, the Fukushima Nuclear Plant disaster in Japan, the missing Malaysian airplane and the Korean Ferry which has capsized and killed 304 people. Similarly, Reason [1] listed several analogous events in past decades (e.g. Seveso, Three Mile Island, Bhopal, Chernobyl and Piper Alpha) referring to them as man-made organisational disasters.

Technical investigations generated a considerable amount of data, revealing an extremely intricate chain of events leading to disastrous consequences and highlighting the decisive part the humans involved have played (Cullen, 1990 [2]; National Commission on the BP Deepwater Horizon Oil Spill and Offshore Drilling, 2011 [3]; Kurokawa et al., 2012 [4]). Factors such as management problems, lack of information, communication failures, fear, poor working conditions and perceptive and reasoning shortcomings were combined with human erroneous actions during the operational stage, to explain the catastrophic failure of cutting-edge technologies and systems, initially developed to be highly reliable.

One of the methods used to minimise human errors is to apply a suitable Human Reliability Analysis (HRA) technique to predict the probability of human error while executing the tasks required during the operation of systems. Previous work [5] identified 72 different methods to assess human reliability, and some of these techniques are widely used by industry, such as the Human Error Assessment and Reduction Technique HEART [6], the Technique for Human Error Rate Prediction THERP [7], and A Technique for Human Error Analysis ATHEANA [8]. These techniques usually comprise a list of internal and external features called performance shaping factors, used to quantify the likelihood of human failure given the task and the existing contributing factors.

In spite of the adequateness of HRA techniques to estimate human error probabilities, it is clear that uncertainties related to human behaviour, which are highly associated with cultural issues, the organisational context, the work environment, work pressures and relationships, training and technological aspects, among others, turn the outcomes of this kind of study largely imprecise.

Therefore, in face of the decisive impact of human actions and decisions on the performance of engineering systems, associated with the uncertainties of current prediction methods, it is necessary to better understand the relationship among contributing factors. Therefore, the Multi-attribute Technological Dataset [9] will be scrutinised by a suitable data mining technique, in order to overcome barriers for dealing with complex data and reveal improvement opportunities. The dataset contains 216 major accidents from different industries, all classified under the same framework, build upon the taxonomy developed by Hollnagel [10] in the Cognitive Reliability and Error Analysis Method (CREAM).

In the following sections, a short description of the data is given. Then, data preprocessing steps are disclosed, hierarchical agglomeration clustering is reviewed and the distance similarity

978-1-4799-7560-0/15 $31.00 © 2015 IEEE
DOI 10.1109/SSCI.2015.256

IEEE computer society

criterion for the particular case is defined. Finally, conclusions are built upon the relationships among contributing factors, to reveal hints for the development of accident prevention schemes.

II. Data description

A dataset containing information on major accidents was analysed. For more than 200 major accidents, the documentation was examined. Thereby several comparable information of the accidents were gathered, like the number of casualties, the year, the location and the industry in which the accident occurred. In addition to these general information, the contributing factors and causes of the accidents are of special interest. Earlier work suggested that 53 hierarchically ordered attributes can define causes of accidents such that the known reasons can be adequately modelled. On the first hierarchical level, there are three entries: Man, Technology and Organisational Environment. The group Man generalises all human mistakes and errors where a human action or cognitive aspect were directly involved. Man can be split into the following four subgroups: Action, Specific Cognitive Functions, Temporary Person Related Functions and Permanent Person Related Functions. The technology part can be divided in the following four categories: Equipment, Procedures, Temporary Interface and Permanent Interface. Organisational Environment is split into Organisation, Training, Ambient Conditions and Working Conditions. Deeper levels of this hierarchy and more detailed information were described earlier by Moura et al [9].

The overall dataset contains 216 major accidents. To maintain the focus on the contributing factors, these will be the data points considered by the analysis. Each of the accidents is given as a vector in the 53 dimensional boolean space, stating which conditions were present or absent when the accident occurred. The dataset can be represented by a 216×53 boolean matrix. For these overall 11448 values, the dataset contains 1416 ones and 10032 zeros. Given these data, our goal is to find groups with common features, the so called data clusters.

Table I presents the factors list ordered by frequency in the MATA-D dataset. Design Failures was most frequent and appeared in 64.35% of the accidents, while 4 factors (Cognitive Style, Sound, Humidity and Other) were not identified.

III. Data Understanding and Preprocessing

In the early phase of data analysis, one aspect is data understanding [11, p. 33], applying simple statistics on the data to get a rough idea. Knowing that 1s are especially important, we look at rows and columns with few ones. There are 14 monocausal accidents (lines in the matrix), representing accidents which showed a single contributing factor or cause. For all of them, insurance companies have commissioned the reports from which the data was taken. This might imply that insurance companies do not invest further money on the causes of accidents when one strong reason, such as an equipment failure, has already been found. For the analysis, these monocausal accidents were removed, as they would probably be considered outliers by the cluster analysis.

TABLE I: Features and corresponding frequency

Feature	**Frequency**
Design failure	64.35%
Inadequate quality control	59.26%
Equipment failure	58.33%
Inadequate task allocation	58.33%
Inadequate procedure	43.98%
Insufficient skills	37.50%
Maintenance failure	35.19%
Insufficient knowledge	34.26%
Wrong Place	26.85%
Missing information	19.91%
Observation Missed	15.28%
Wrong Time	14.81%
Incomplete information	13.89%
Wrong Type	12.96%
Faulty diagnosis	12.96%
Wrong reasoning	12.04%
Communication failure	11.11%
Management problem	10.19%
Inadequate plan	9.72%
Decision error	8.80%
Cognitive bias	7.87%
Adverse ambient conditions	7.87%
Priority error	6.94%
Social pressure	6.94%
Distraction	6.48%
Excessive demand	5.56%
Delayed interpretation	5.09%
Irregular working hours	4.17%
Incorrect prediction	3.70%
Inadequate team support	3.70%
Fatigue	3.24%
Psychological stress	3.24%
Wrong Identification	2.78%
Software fault	2.78%
Ambiguous information	2.78%
Inadequate work place layout	2.78%
Wrong Object	2.31%
False Observation	2.31%
Fear	2.31%
Inattention	2.31%
Access problems	1.85%
Performance Variability	1.39%
Access limitations	1.39%
Mislabelling	1.39%
Temperature	1.39%
Memory failure	0.93%
Physiological stress	0.93%
Illumination	0.93%
Functional impairment	0.46%
Cognitive style	0.00%
Sound	0.00%
Humidity	0.00%
Other	0.00%

A look at the raw data of the 53 dimensions revealed that four dimensions did not appear at all: Cognitive Style (part of the group of Permanent Person Related Functions), as well as Sound, Humidity and Other (from the group of Ambient Conditions). Seventeen dimensions appeared less than 7 times each (less than 3%). Although this could be useful information for the accidents causation understanding, for statistical analysis these dimensions hardly contain much information.

In order to verify the amount of information contained, a Principal Component Analysis (PCA) was performed. The result, showing the explained variance of the dimensions, revealed that no further dimensions could be found, allowing its legitimate removal.

Given that the dimensions are hierarchically ordered, data was aggregated considering whether or not an attribute was present in at least one of the groups within the categories. This aggregated dataset (in contrast to the original unaggregated dataset) contained only 15 dimensions. After filtering mono-causal accidents and applying the aggregation, the boolean matrix had a size of 202×15. The PCA performed on these new data did not show any duplicate dimensions.

IV. Methodology

Having a boolean representation of the accidents, defining an adequate similarity measure to group them is the next step. Accidents were considered to be similar if they have similar contributing factors. For two given boolean vectors i and j, we counted the number of Ones for each event $S(i)$, $S(j)$ and further counted the dimensions in which both vectors show Ones simultaneously $C(i, j)$.

$$d(i, j) := 1 - 2C(i, j)/(S(i) + S(j)) \tag{1}$$

Equation 1 shows the bray-curtis dissimilarity of two accidents [12], which is not a distance, because the triangle inequality does not hold. Nevertheless, this measure ensures that accidents are similar if they have common causes, meaning common Ones in the boolean space. Dimensions with common Zeros do not increase the accidents' similarity. The measure is normalised by the division of the number of ones contained in both vectors so that high similarities occur if two vectors have many common and only few different causes.

As our goal is to find groups of common accidents, we apply a cluster analysis, e.g. described in [13, p. 390]. Hierarchical Agglomerative Clustering was applied. Every data point is considered as a cluster in the beginning, and then successively merged by similarity. We already clarified what similarity means for two given accidents, in our case. For given clusters, this can be extended by using an appropriate Linkage Method. For two clusters A and B with their contained points a and b, respectively, their distances can be defined in the following way for complete and average linkage:

$$d_{complete}(A, B) := \max_{a \in A, b \in B}(d(a, b)) \tag{2}$$

$$d_{average}(A, B) := \frac{1}{|A||B|} \sum_{(a,b) \in A \times B} (d(a, b)) \tag{3}$$

Clusterings can be evaluated with silhouette score [14]. This is an internal clustering evaluation measure returning values in $[-1, 1]$. Values are close to one when the resulting clustering return relatively compact clusters, also having fairly high inter-cluster dissimilarities.

V. Results

The following analysis details the dendrogram of the complete-linkage distances. Reading bottom to top, it shows the data points, which are iteratively merged into bigger clusters. This process ends when the cut value is reached. Above that value, the connecting edges are blue, and beyond they are differently coloured, depending on the cluster. The cutting was performed at the 0.65 distance, where a reasonable jump in the clustering levels can be observed, as shown in Fig. 1. The resulting clustering shows a moderate silhouette score of 0.228, which is slightly better than the score for average linkage 0.221

The clusters are enumerated from left to right. First single cluster comprises the only two events which contained just human factors as significant contributors to the accidents. This chunk is substantially different from the distribution in the remaining groupings, as the vast majority of the technological accidents of the MATA-D encompasses at least one organisational or technological issue. Therefore, these two events can be considered outliers.

All the remaining events involved organisational aspects (organisation, training, ambient conditions or working conditions) to generate the undesirable outcome. Clusters 2 to 5 were highly associated with technological issues (equipment, procedures or interface), while clusters 6 to 9 showed the manifestation of humans factors (execution errors, specific cognitive functions and person-related functions). It is important to notice that the latter groupings (from 6 to 9) form a single cluster containing almost the same amount of elements (102 of 106 events), when a different clustering criterion, i.e. the average-linkage method, is applied to quantify dissimilarities between clusters (Fig. 2). Further, the same points build the leftmost cluster of outliers.

The main results of the individual analysis of the complete linkage clusters are summarised in Table II and will be detailed as follows.

Cluster 2 (20 elements) was dominated by wrong procedures (100%) and organisational issues (95%). Training was also significant (70%). No erroneous actions were observed.

Third Cluster (29 elements) contained organisational issues (100%) with training problems (100%), also with a high incidence of equipment failures (75.9%). A marginal incidence (a single case) of erroneous actions was shown.

Cluster 4 (6 elements) main feature is the combination of organisational issues with communication problems (100% of the cases). No erroneous actions were observed, and a marginal incidence of temporary person-related functions, i.e. psychological stress, was shown.

Cluster 5 (39 elements) contained organisational issues (100%) with equipment failure (82%). No human issues were observed.

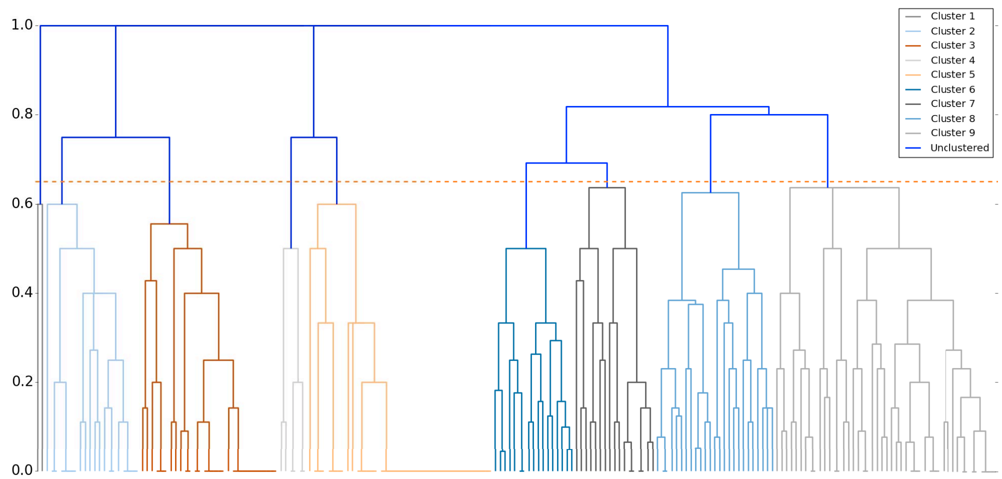

Fig. 1: Dendrogram for complete linkage clustering

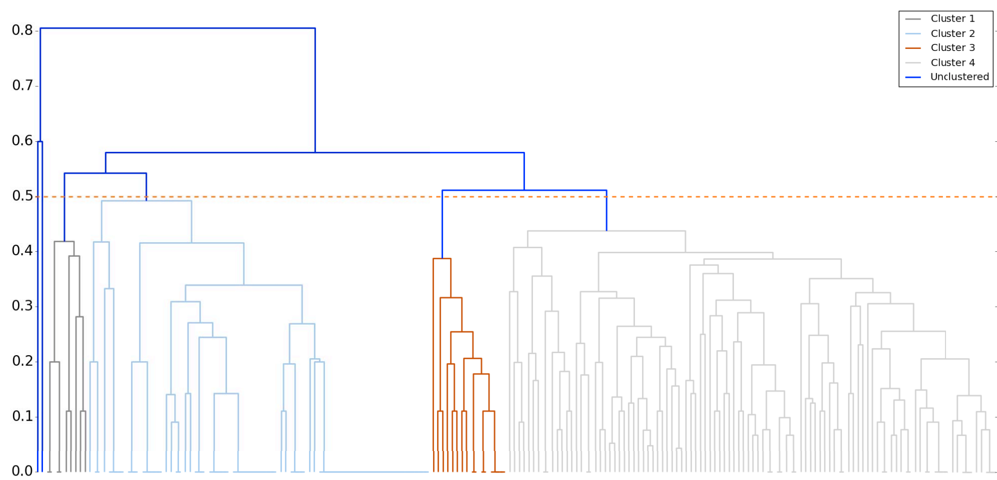

Fig. 2: Dendrogram for average linkage clustering

TABLE II: Aggregated Data, Complete Linkage

	C1	C2	C3	C4	C5	C6	C7	C8	C9
Accident#	2	20	29	6	39	17	17	25	47
Median	2.50	6.00	6.00	4.00	2.00	13.00	5.00	10.00	8.00
Average	2.50	5.20	5.66	4.33	2.77	13.65	7.24	10.00	8.30
Mode	N/A	7	5	4	2	13	5	6	9
Fatalities	5	159	67	2	248	471	6	132	388
Death Rate	2.50	7.95	2.31	0.33	6.36	27.71	0.35	5.28	8.26

The following 4 clusters contained direct human errors joining organisational issues to generate accidents. It is worth mentioning that Cluster 6 was the deadliest combination of factors, reflecting 471 fatalities and a death rate of 27.71 per event, followed by Cluster 9, which encompassed 388 fatalities and attained a death rate of 8.26, as can be seen in Table II.

All accidents in the sixth cluster (17 elements) contained organisational issues plus erroneous actions. These actions were explained by all three levels of specific cognitive functions, i.e. interpretation (94.1%), Observation (88.6%) and Planning (70.6%). Temporary person related functions were also significant (88.6%). An important feature of cluster 6 is that an interface problem (information provided by the systems, a technology problem) or communication issues (exchange of messages/information within the organisation) were shown in 100% of the cases. Training problems (82.4%) and poor working conditions (58.8%) as contributing factors were also above the overall average.

The accidents in Cluster 7 (17 elements) showed organisational issues with erroneous actions (100%) again, but mainly combined with one specific cognitive function level (observation, with 70.6%). Equipment failures and communication issues also showed 70.6% of incidence.

Cluster 8 (25 elements) contained the combination of organisational issues (100%) with erroneous actions (96%), but with an intermediate level of cognitive error (interpretation, with 84% of incidence). Wrong procedures (92%) were very significant, and equipment problems were also shown (72%).

Cluster 9 grouped 47 cases (100%) containing erroneous actions with organisational (97.9%) and training (91.5%) issues. 95.7% of the erroneous actions were accompanied by intermediate to advanced specific cognitive functions, i.e. interpretation (61.7%) and inadequate mental planning (44.7%).

VI. Discussion

A suitable cluster analysis algorithm was applied to a collection of 202 major accidents from the MATA-D dataset to disclose relevant relationships among contributing factors. All clusters were largely dominated by the Organisational Environment group, meaning that design failures, poor quality control and inadequate task allocation are key contributing factors to major accidents.

First cluster, which contained only two elements (less than 1% of the sampling), confirmed previous studies (Moura et al, 2015 [9]) indicating that the possibility of having a single-failure point leading to a major accident is low. The two cases represented human erroneous actions following a serious violation of a faultless work procedure, such as smoking during a fuel tanker offloading or deviating from the recommended navigating route without any apparent reason. Although violations are not uncommon, it appears to occur in specific cases, such as under an uncertain environment (e.g. system is in an unrecognised, non-routine status), associated with unclear rules or procedures, or when operators distinguish some kind of tangible benefit (e.g. save money, time or effort through an easier way of performing a task) from non-compliance. These associations would require further contributing factors (e.g. inadequate procedure, poor communication or training), thus this was not the case in the two analysed accidents.

The algorithm application set apart a very significant group, i.e. 106 accidents (From cluster 6 to cluster 9) in which organisational factors were accompanied by human erroneous actions. The exposed differences between these 4 clusters lie on the mental functions or disturbances which triggered the action error.

In spite of having only 17 events, cluster 6 is an extremely important chunk, as it is by far the deadliest grouping, with a rate of 27.71 fatalities per event. Also, the accidents within this cluster showed the largest amount of simultaneous contributing factors to generate an accident, with an average of 13.65 and mode of 13 significant features. The justification is that a very complex chain of events or simultaneous failures took place in sophisticated systems (e.g. oil & gas, chemical or aviation industry), which would have required the whole range of specific cognitive functions (observation, interpretation and planning) to recover the system to a regular state and minimise the effects of the accident sequence.

Although the path to recover the system to a normal state was apparently clear in most of the cluster's cases, the operator was unable to make progress, due to two main reasons. First, a technology problem related to the man-machine interface failed to provide an accurate information (an indistinct or incomplete error message, for instance), and the communication channels of the organisation failed to deliver a complete information or feedback. Second, inadequate working conditions (e.g. excessive demand and irregular working hours) resulted in temporary person-related functions (e.g. distraction, fatigue and psychological stress) to disturb the mental processing. Deprived from an accurate input from the system and from the organisation (information and training) to support the decision-making process, and frequently affected by an undermining working condition, the operator was unable to respond appropriately.

Cluster 7 erroneous actions can be explained by a simpler mental modelling, where a wrong observation was the triggering mechanism. Examples are failures to observe an indication/warning, a mistaken or partial identification of a status or an incorrect recognition of a signal. This cognitive failure to observe something can be directly associated with the substantial rate of concurrent equipment failures and communication issues, which prevented the operator from entering in a deeper state of cognition to build a more complex problem-solving mental plan.

The 25 major accidents contained in cluster 8 shared similar characteristics with the previous grouping, but with erroneous actions largely commanded by a faulty reasoning (induction or deduction error) or an imperfect diagnosis of the system state. The additional contributors interacting with these human-related issues suggest a reasonable explanation for this: inadequate procedures, which can be specified by incorrect/outdated or incomplete written instructions, directly affected the operator's capacity to construct a mental plan and understand the situation or system state, as the written and trained information, typically assumed to be precise, was not representative of reality.

Cluster 9 grouped the largest number of major accidents, with 47 events, and contained a very consistent amount of contributing factors per accident (average of 8.3). Accordingly, the causes for these accidents are quite well dissected, meaning that a reliable understanding of the grouping behaviour can be reached. The human erroneous actions were strongly influenced by advanced cognitive functions, supposing that a complex mental modelling was required to maintain the system under a regular operational state, i.e. a seamless interpretation and the construction of an accurate mental planning was necessary. However, the high incidence of training flaws, including not only the lack of suitable working instructions to improve the human performance and manoeuvring capacity, but also the dispossession of the necessary knowledge to fully understand the system behaviour, seem to have undermined the human capacity to act properly.

VII. Conclusions

A. Insights to improve human performance and minimise accidents

The examination of complex accidents under an analysis method, i.e. a Hierarchical Agglomerative Clustering using distance measures tailored to the dataset characteristics, can offer wide-ranging insights into the data, which could be advantageous for the development of accident prevention schemes. It is now clear that different mechanisms are able to trigger specific cognitive failures, leading to human erroneous actions and subsequently disastrous consequences, and the decision-making process at any stage of high-technology facilities' lifecycle can take advantage of the findings discussed earlier in this work. From the analysis of the technological accidents in cluster 6, which resulted in 471 fatalities (27.71 per event), it can be concluded that efforts towards the improvement of the communication fluidity from the organisation (to make sure operators receive and understood messages) and from the interface (clear warnings and error messages from the system) would have enhanced the operator's ability to recognise the system status. An evaluation of the whole range of information reaching the operators is indispensable to ensure a proper construction of a right cognitive line.

The study has also revealed that the largest cluster (C9, with 47 accidents) was intensely associated with training aspects. Current industry's response to training problems is usually short-duration courses and on-the-job evaluations. In spite of being valuable to some extent (primarily to improve practical skills and develop work experience), this approach is unlikely to improve the awareness level when operating complex systems. This is a robust indication that the knowledge level required to operate in high-technology environments is currently not sufficient.

As a result, the application of a multidisciplinary and more advanced recruiting and knowledge development programme is strongly recommended. This should be fully tailored to the system and industry in which the operator will be allocated, in order to ensure that the capabilities required to understand and operate the system consciously, including the recovery from complex and abnormal situations, are in place.

B. Future Developments

The pronounced incidence of organisational problems in all clusters is an issue which deserves further investigation. This category includes failures in maintenance, quality control, management, design and task allocation. Some of them (such as a design problem) can be embedded in the system for many years. The analysed accidents have shown that there was enough time to spot symptoms of some future flaws before being exposed by an accidental sequence. Also, a deeper study of certain flaws (e.g. training) may provide cues to improve organisations and technologies. Testing other data mining methods such as Frequent Itemset or Association Rule might reveal some extra factors to be addressed by future work.

C. Acknowledgements

Raphael Moura's contribution to this work has been partially funded by CAPES (Proc. no. 5959/13-6).

References

[1] J. Reason, *A life in error: from little slips to big disasters*. Ashgate Publishing, Ltd., 2013.

[2] L. W. D. Cullen, "The public inquiry into the piper alpha disaster," *Drilling Contractor;(United States)*, vol. 49, no. 4, 1993.

[3] D. Water, "The gulf oil disaster and the future of offshore drilling," *Report to the President [of the USA]*, 2011. [Online]. Available: http://www.gpo.gov/fdsys/pkg/GPO-OILCOMMISSION/pdf/GPO-OILCOMMISSION.pdf

[4] F. N. A. I. I. Commission *et al.*, *The official report of the fukushima nuclear accident independent investigation commission: Executive summary*. National Diet of Japan, 2012, (Accessed: 4 July 2015). [Online]. Available: https://www.nirs.org/fukushima/naiic_report.pdf

[5] J. Bell and J. Holroyd, "Review of human reliability assessment methods," *Health and Safety Laboratory, United Kingdom*, 2009.

[6] J. Williams, "Heart–a proposed method for assessing and reducing human error," in *9th Advances in Reliability Technology Symposium, University of Bradford*, 1986.

[7] A. Swein and H. Guttmann, "Handbook of human reliability analysis with emphasis on nuclear power plant application," NUREG/CR-1278, Tech. Rep., 1983.

[8] M. Barriere, D. Bley, S. Cooper, J. Forester, A. Kolaczkowski, W. Luckas, G. Parry, A. Ramey-Smith, C. Thompson, D. Whitehead *et al.*, "Technical basis and implementation guidelines for a technique for human event analysis (atheana)," *NUREG-1624, Rev*, vol. 1, 2000.

[9] R. Moura, M. Beer, E. Pattelli, J. Lewis, and F. Knoll, "Human error analysis: Review of past accidents and implications for improving robustness of system design," in *European Safety and Reliability Conference: ESREL2014, Wroclaw, Poland*, ser. Proceedings of the 24th European Safety and Reliability Conference, T. e. a. Nowakowski, Ed., September 2014, pp. 1037–1046.

[10] E. Hollnagel, "Cognitive reliability and error analysis method," *UK Oxford: Elsevier Science Ltd*, 1998.

[11] M. R. Berthold, C. Borgelt, F. Höppner, and F. Klawonn, *Guide to Intelligent Data Analysis: How to Intelligently Make Sense of Real Data*. Springer Science & Business Media, 2010.

[12] J. R. Bray and J. T. Curtis, "An ordination of the upland forest communities of southern wisconsin," *Ecological monographs*, vol. 27, no. 4, pp. 325–349, 1957.

[13] R. Kruse, C. Borgelt, F. Klawonn, C. Moewes, M. Steinbrecher, and P. Held, *Computational intelligence - A Methodological Introduction*. London: Springer, 2013.

[14] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of computational and applied mathematics*, vol. 20, pp. 53–65, 1987.