

Safety Barriers Failures in Cold Wave-Triggered Events: a Data-Driven Approach

Giulia Collina^a, Federica Ricci^b, Alessia Bortoluzzi^{a,b}, Dimitrios Tzioutzios^a, Nicola Paltrinieri^{a*}

^aDepartment of Mechanical and Industrial Engineering, Norwegian University of Science and Technology NTNU, Richard Birkelands vei 2B, 7034 Trondheim, Norway

^bLISES - Dipartimento di Ingegneria Civile, Chimica, Ambientale e dei materiali, Università degli studi di Bologna, via Terracini 28, 40131 Bologna
giulia.collina@ntnu.no

Natural hazard-triggered technological accidents (Natech) add a layer of complexity to disaster risk management. They involve the combination of both natural and anthropogenic hazards. Noticeably, there is an uneven distribution in the extent of research dedicated to the different natural hazards as causes of accidents. Extreme weather events have been overshadowed, despite low temperatures being the third cause of Natech accidents in Europe, ranking only behind lightning and floods.

The present study aims to comprehensively analyse the role of safety barriers within the context of Natech events triggered by cold waves. Data selected from established databases are investigated through Unsupervised Machine Learning and Bayesian Networks. This approach facilitates identifying and quantifying the relationships between undesired event features and safety barrier failures. The model obtained is designed to predict the behaviour of safety barriers, providing a dynamic and updatable structure adaptable to future developments and integrable with additional available information.

This analysis provides valuable lessons to prevent the recurrence of similar events in the future and a robust foundation for the proposition of targeted safety barrier protection programs.

1. Introduction

Natech, which stands for natural hazard-triggered technological accidents, describes incidents where natural hazards lead to the release of hazardous materials in industrial settings (Krausmann et al., 2017).

The increased frequency of meteorological events, intensified by climate change, has brought wider attention to Natech incidents. However, a noticeable imbalance remains in addressing various natural hazard categories: extreme temperature events, for instance, are relatively overlooked (Suarez-Paba et al., 2019).

In January 2024, Northern Europe experienced a severe cold wave. In Sweden, temperatures dropped to -43.6 °C, marking the coldest January temperature in 25 years, and Oslo recorded temperatures as low as -30 °C (Earth Observatory NASA, 2024). The immediate and tangible impact of extremely cold weather on critical infrastructure became evident, making electricity supply limitations a significant influencing factor for Natech events. These occurrences align with studies by Krausmann and Baranzini (2012), highlighting low-temperature events as the third leading cause of Natech accidents in Europe.

Understanding how cold waves can influence safety barriers is essential for enhancing preparedness and response to Natech incidents in extreme climatic conditions. The main objective of this paper is to analyse incidents caused by cold waves, focusing on safety barriers. The expected impact of this study is to enhance preparedness and response strategies for Natech incidents by understanding how cold waves affect safety barriers in industrial settings. This paper is structured as follows. The “Database description” section describes the cold wave database used for this investigation. The “Methodology” section explains the procedure adopted in this work: the database is analysed to understand the relationships between incident features and the behavior of the involved safety barriers. Firstly, clustering based on unsupervised machine learning is used to

group the events; subsequently, the behavior of safety barriers is modeled using Bayesian Networks. This model is designed to provide a dynamic and updatable structure that is adaptable to future developments and integrable with additional available information. The “Results and Discussion” section shows and comments on the outcomes of the analysis, focusing on some quantitative results. Finally, the “Conclusion” section summarises the main findings of this study.

2. Database description

The Cold Wave Database (CW Database), established in 2023 by Ricci et al. (2023), aims to comprehensively analyse past accidents involving hazardous substances triggered by cold waves impacting industrial infrastructure. The database counts over 740 events and comprises 25 columns categorised as follows: Original Source (tracking the incident’s database origin), Time (specifically the year), Geographical Area (continent, country, USA state), Industrial Sector, Equipment Item, Direct Cause, Technological Scenario, Substance, Safety Barrier (indicating effectiveness or not), Damage to Humans, Economic Loss, and Summary (providing an incident overview).

The primary source of data for the CW Database is the National Response Center (NRC) Database (U.S. Coast Guard, 2023), which contributes approximately 75% of the total records. The ARIA Database (The French Bureau for Analysis of Industrial Risks and Pollution, 2023) follows, providing 19.3% of the total records. In addition, eMARS (European Major Accident Hazards Bureau, 2023) and TAD IChemE (Institution of Chemical Engineers, 2000) offer records with a higher level of detail.

3. Methodology

The work involves several steps, as depicted in Figure 1: machine learning algorithms such as k-means clustering are applied to group of incidents in the aforementioned database based on common characteristics. The information obtained is then analysed using Pearson’s Chi-square test and Information Gain to identify the relationships between incident features. Once the relationships are identified, the Bayesian network structure is constructed, and finally, the probabilities associated with each node are calculated.

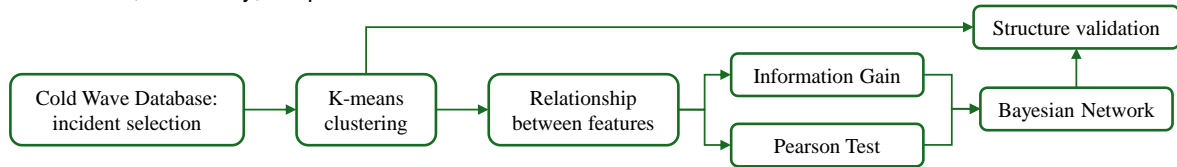


Figure 1: Flowchart showing the methodology of this study.

3.1 Unsupervised machine learning

Unsupervised machine learning allows grouping events based on their similarity to uncover hidden relationships and identify the combination of features that most significantly contributes to their occurrence. In this study k-means and clustering algorithms are applied as machine learning techniques to the incidents with at least one safety barrier. Before applying any algorithms, the dataset requires a pre-process: cleaning, handling missing values, removing outliers, and standardising data. Term Frequency – Inverse Document Frequency (TF-IDF) vectorisation is then performed, and dimensionality reduction is adopted to visualise a 2D representation (Liu et al., 2021).

The k-means clustering divides the dataset into a predetermined number of clusters. Each cluster contains similar patterns, while dissimilar ones are allocated to other clusters (Chokor et al. 2016).

A key element is the centroid, representing the center of mass of observations within the cluster, serving as a representative prototype.

Considering a set of N points denoted by $X = \{x_1, \dots, x_n\}$, the k-means algorithm seeks a set of k vectors μ_k minimising the Within-Cluster Sum of Squares (WCSS):

$$WCSS = \sum_{h=1}^k \sum_{x_i \in C_h} d(x_i, \mu_h) \quad (1)$$

where C_h is the h -th cluster, and μ_h is the centroid of the h -th cluster.

The algorithm begins by randomly defining k centroids in the feature space, each associated with a cluster. Patterns are then assigned to the nearest centroid using the Euclidean distance $d(x_i, \mu_h)$. Subsequently, each

centroid is recalculated as the mean of observations in the cluster. These steps repeat until no new assignments occur, meaning that WCSS converges to a local minimum.

The number k of clusters is a hyperparameter to be estimated. An iterative approach is adopted to define the optimal number of clusters (k): different values of k are assigned, ensuring through a visual analysis well-defined, non-overlapping clusters with sufficiently separated centroids (Chokor et al. 2016).

3.2 Bayesian Network

A Bayesian Network serves as a probabilistic graphical model that visually represents causal relationships among variables through a Directed Acyclic Graph (DAG) (Cui et al., 2024). Within this network, nodes are categorised into parent nodes and child nodes, each serving distinct roles. The term 'parents' is attributed to the nodes from which the arcs are directed, whereas the nodes with arcs directed into are referred to as 'children'. Eventually, nodes with no parents are called 'root nodes', while 'leaf nodes' is used to indicate nodes with no children. The edges connecting nodes illustrate the causal dependencies between them. The application of Bayesian Networks consists of two phases: establishing the network structure and calculating probabilities for each node.

In this study, a multi-technique approach is adopted to determine the structure of the Bayesian Network due to the complexity of the data and the need to mitigate the limitations of individual analytical techniques. The dataset presents various challenges, such as the presence of textual variables and the complex relationships between them. Specifically, Pearson's Chi-Squared Test (Turhan et al., 2020) and Information Gain (Qu et al., 2023) are employed for feature selection and creating the Bayesian Network structure. Pearson's chi-square test examines relationships in data that can be into a contingency table that displays the frequency distribution for two categorical variables. Then, the test compares the observed frequencies in the table to the frequencies that would be expected under the assumption of independence between the variables. The Information Gain technique consists of constructing a graph with weighted edges where the graph nodes correspond to the incident features of the dataset. The stronger the value shown in the edges, the stronger the node relationships. This approach facilitated the establishment of a reliable Bayesian network structure and minimised the risk of identifying relationships between columns that do not reflect reality.

Determining the probability of parent nodes in Bayesian networks relies on calculating the frequency of appearance of the word corresponding to the state in the column under consideration, divided by the total number of words in the same column. This approach is reasonable when dealing with variables that are independent of other factors within the system.

Finally, the conditioned probabilities of the states of child nodes are calculated using Conditional Probability Tables (CPTs), which are used in Bayesian networks to represent the probabilistic relationships between a child node and its parent nodes (Chen et al., 2012). CPTs specify the degree of belief (expressed as probabilities) that the child node will be in a particular state given the states of the parent nodes.

4. Results and discussion

The selection of the incidents with at least one safety barrier resulted in a significant reduction in the total number of records from 746 to 126. Then, k-means clustering facilitated the grouping of incidents based on common characteristics; the optimal number of clusters was three, as shown in Figure 2a.

Figure 2b shows, as an example, the Word Cloud created for Cluster 0. The Word Clouds enabled the identification of the most recurrent and distinctive incident features for each group based on the frequency of word appearances. In detail, Word Clouds showed that cluster 0 is predominantly associated with toxic releases of chlorine and ammonia involving shutdown procedures. Cluster 1 focuses on flammable compounds, particularly liquid hydrocarbons, with the involvement of secondary containment. Finally, Cluster 2 pertains to relief valves in the Chemical and Petrochemical sector.

The centroids of each cluster were also characterised by analysing the word frequency relative to each incident feature. In this way, one fictitious incident for each cluster was defined, representing the combination of the most frequent incident characteristics within the cluster.

The incident built through the characterisation of the centroid of Cluster 0 is a release of toxic gas, specifically ammonia and ammonium-based substances, occurring in the chemical and petrochemical sector. This incident involves an effective shutdown procedure, with the implication of one or more valves as crucial equipment.

In the case of Cluster 1, the identified incident is a near miss involving flammable hydrocarbon liquid, namely fuel oil. The safety barrier involved is secondary containment, which has proven effective. The equipment involved in this case includes a tank and a pipework.

The incident identified by Cluster 2 is a release of methane involving an effective relief valve. The equipment involved in this case includes a tank and a valve.

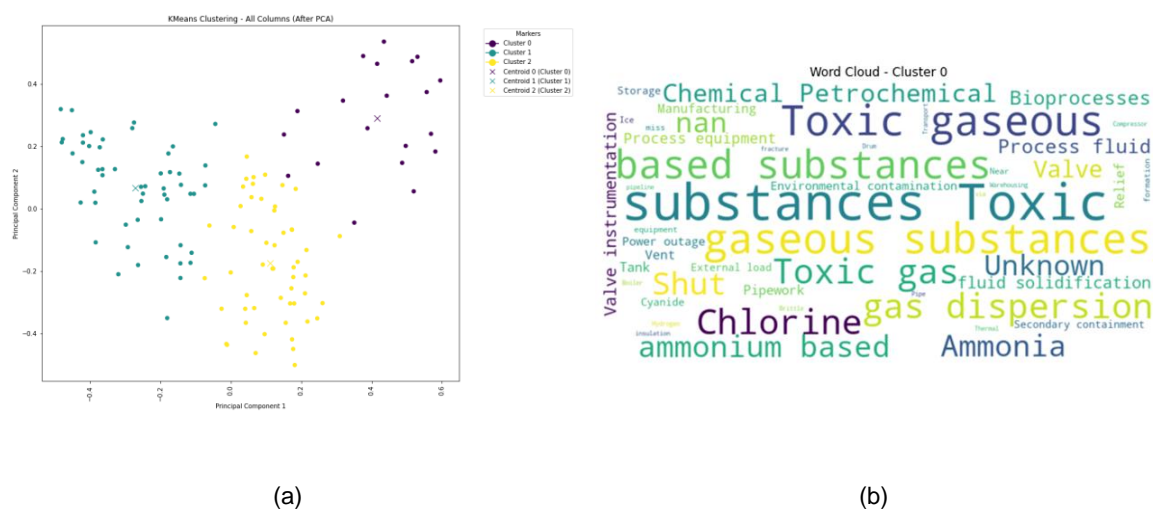


Figure 2: (a) Visual analysis determining 3 as the optimal k value. (b) Word cloud representing Cluster 0.

Table 1 shows the results of the characterisation of the centroid of Cluster 1 as an example.

Table 1: Results of the characterisation of the centroid of Cluster 1. The final scenario is a Near miss.

Column	Most frequent cell	Second most frequent cell
Direct cause	Process fluid solidification	Ice formation
Compound	Fuel oil	Crude oil
Hazard	Flammable liquid hydrocarbons	NC
Type of final scenario	Near miss	Near miss
Type of item involved	Pipework	Storage equipment
Macro sector	Chemical & Petrochemical	Transport via pipeline
Safety barrier 1	Secondary containment	Relief valve
Safety Barrier 1 failure	Effective	Failed
Safety Barrier 2	NO	Secondary containment
Summary	Tank	Tank dike

It is interesting to note that the final incident scenario occurred in all three cases despite the effectiveness of safety barriers. The first possible cause is that the safety barrier may be specifically designed to prevent a final incident scenario different from the one that occurred. Alternatively, it is likely that the safety barrier may indeed have been designed to withstand cold temperatures and would be effective in such case, but the plant section involved in the incident was not. The entire plant must be designed to withstand extreme temperatures, not only safety barriers.

After analysing the incidents, the Bayesian Networks were constructed. The intersection of Pearson's Chi-Square Test and Information Gain ensured a reliable structure.

Two root nodes and four leaf nodes were identified. The probabilities of the root nodes were calculated through a simplified procedure, as described in section 3.2. However, the frequency of word appearance may not accurately reflect the actual probability of the event, especially if limited available data or other factors influence the distribution of words in the column. Additionally, the reliability associated with compiling source databases must be considered. Non-expert personnel often report incidents that may not accurately describe the sequence of events leading to the final incident scenario or may incorrectly fill in the database entries. Therefore, relying only on the frequency of word appearance in the column related to the node under examination carries the risk of error propagation.

In this case, the simplified procedure was justified by the low number of root nodes (only two) and their nature. The compound node, which corresponds to 'Hazard' node in Figure 3 as a reference, expects a certain specificity of the entered textual values, and in case of missing information about the substance, 'Unknown' was entered. Therefore, the risk of error was considered very low, if not negligible.

The probabilities of the child nodes were calculated using Conditional Probability Tables (CPT). Once the probabilities for all states of all nodes were calculated, they were entered into the GeNIe software (BayesFusion LCC, 2024), which allows for a graphical representation of the network.

From the statistical analysis conducted, the effectiveness of safety barriers is not related to any node. Therefore, an analysis of safety barrier failure was performed by selecting a type of final incident scenario as a reference in GeNIe, allowing the visualisation of which safety barriers are involved in the occurrence of the specific incident scenario. Then, the probability of failure for these specific safety barriers was calculated using the data available in the CW Database.

The same type of analysis (identifying the most involved characteristics of a feature for a specific type of final incident scenario) was used to assess the involved equipment. The three identified incident scenarios during clustering, namely Toxic cloud, Near miss, and Release, were analysed.

For instance, Figure 3 shows the association of the following safety barriers to the Near miss scenario: secondary containment (28%), relief valve (18%), and flare (17%). The equipment primarily involved included pipes (44%), tanks (22%), and valves (18%).

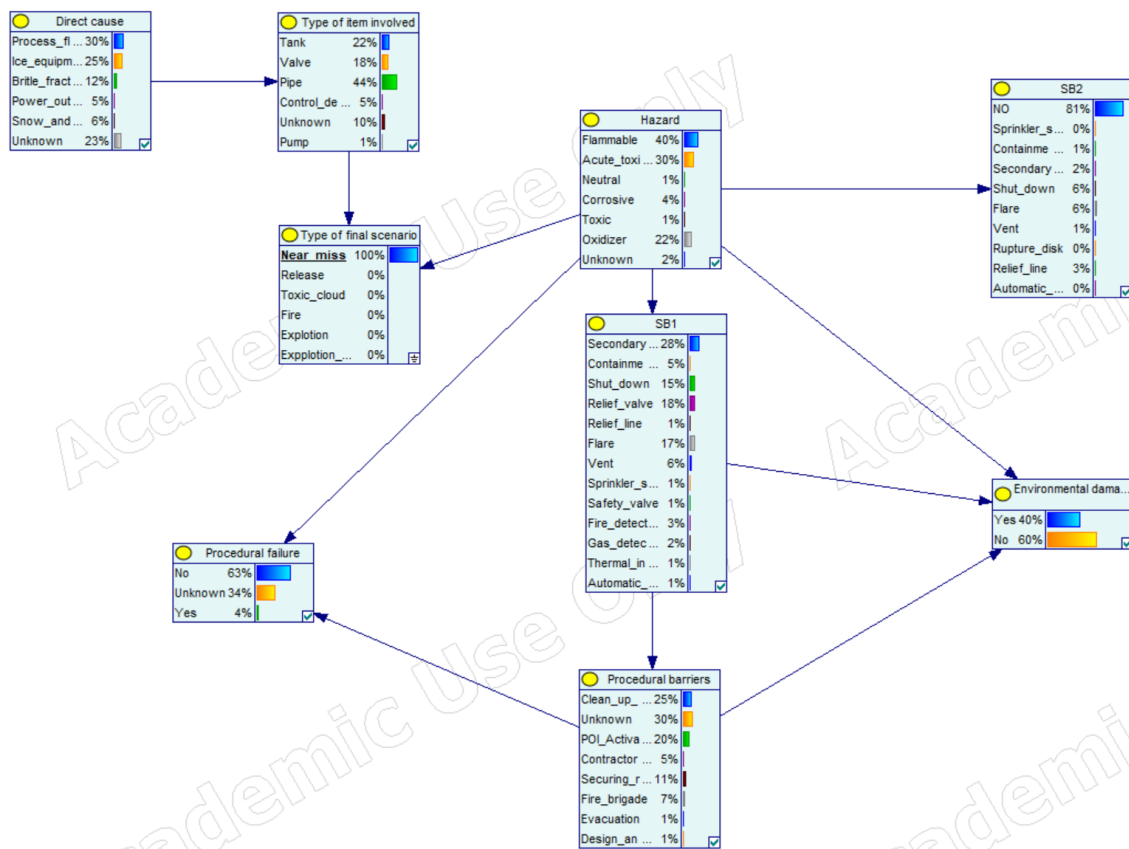


Figure 3: Bayesian Network with focus on Near miss as type of final scenario. It is possible to see the probability values of other node's states.

Moreover, the failure frequencies of safety barriers were analysed in the database, resulting in 20% for secondary containment, 0% for shutdown procedures, 72.7% for relief valves, and 28.6% for flares. The significant aspect is the high incidence of relief valve failures, particularly sensitive to cold waves.

By setting the final incident scenario type on GeNIe as one identified through the characterisation of cluster centroids, it is possible to analyse the prevalent characteristics for each incident feature related to that scenario. These characteristics were then compared with the incident features obtained from the three centroids of the clusters, yielding positive results. In fact, the Bayesian network results reflected the clustering analysis outcomes, suggesting that the created model is reliable and promising for further development.

5. Conclusions

This study proposed an approach based on Unsupervised Machine Learning and Bayesian Network for learning from past undesired Natech events. The analysis of the Cold Wave database identified three clusters representing the main characteristics of Natech events triggered by Cold Waves. On the other hand, the statistical analysis led to the creation of the Bayesian Network consisting of two root nodes and four leaf nodes. This study is a starting point for analysing past incidents through an innovative methodology. However, it is essential to consider that the use of Unsupervised Machine Learning requires a critical interpretation of the results. Additionally, the limited attention given to Cold Wave-induced incidents in the literature makes the available data constrained, representing a significant limitation in applying techniques such as Machine Learning, which relies on a large amount of data. A substantial increase in samples would lead to a considerable improvement in results, making them more reliable. Moreover, a much larger number of incidents would notably alter the proposed structure of the Bayesian Network.

References

- Chen S. H., Pollino C.A., 2012, Good practice in Bayesian network modelling, *Environmental Modelling & Software*, 37, 134-145.
- Chokor A., Naganathan H., Chong W.K., El Asmar M., 2016, Analysing Arizona OSHA Injury Reports Using Unsupervised Machine Learning, *Procedia Engineering*, 145, 1588-1593.
- Brussels: European Commission. <unisdr.org/files/2631_FinalNatechStateofthe20Artcorrected.pdf>. Accessed 10.02.2024.
- Cui J., Kong Y., Liu C., Cai B., Khan F., Li Y., 2024, Failure probability analysis of hydrogen doped pipelines based on the Bayesian network, *Engineering Failure Analysis*, 156, 1350-6307.
- European Major Accident Hazard Bureau (MAHB), The eMARS (Major Accident Reporting System) Database. <emars.jrc.ec.europa.eu/en/emars/accident/search> accessed 08.09.2023.
- BayesFusion LCC, 2024, GeNle Modeler. <bayesfusion.com/genle/> accessed 15.09.2023
- Institution of Chemical Engineers (IChemE), The Accident Database. <icheme.org/knowledge/safety-centre/resources/accident-data/> accessed 10.11.2024.
- Krausmann E., Cruz A. M., Salzano, E., 2017, Natech Risk Assessment and Management: Reducing the Risk of Natural-Hazard Impact on Hazardous Installations, Elsevier.
- Krausmann E., Baranzini, D., 2012, Natech risk reduction in the European Union. *Journal of Risk Research*, 15, 1027–1047.
- Liu G., Boyd M., Yu M., Halim S. Z., Quddus N., 2021, Identifying causality and contributory factors of pipeline incidents by employing natural language processing and text mining techniques, *Process Safety and Environmental Protection*, 152, 37-46.
- NASA Earth Observatory, 2024, Extreme Nordic Cold, <earthobservatory.nasa.gov/images/152288/extreme-nordic-cold> accessed 20.02.2024.
- Qu K., Xu J., Hou Q., Qu K., Sun Y., 2023, Feature selection using Information Gain and decision information in neighbourhood decision system, *Applied Soft Computing*, 136.
- Ricci F., Casson Moreno V., and Cozzani V., 2023, Natech Accidents Triggered by Cold Waves, *Process Safety and Environmental Protection*, 173, 106–19.
- Suarez Paba M. C., Perreux M., Munoz F., Cruz A. M., 2019, Systematic Literature Review and Qualitative Meta-Analysis of Natech Research in the Past Four Decades, *Safety Science*, 116, 58–77.
- The French Bureau for Analysis of Industrial Risks and Pollutions (BARPI), The ARIA (Analysis, Research and Information on Accidents) Database. <aria.developpement-durable.gouv.fr/the-barpi/the-aria-database/?lang=en> accessed 12.09.2023.
- Turhan N. S., 2020, Karl Pearson's Chi-Square Tests Educational Research and Reviews, 16.9: 575-580.
- U.S. Coast Guard, 2023. The NRC (National Response Center) Database. <epa.gov/emergency-response/national-response-center> accessed 10.09.2023.