Accepted Manuscript

Understanding complex blasting operations: A structural equation model combining Bayesian networks and latent class clustering

S. Gerassis , M.T.D. Albuquerque , J.F. García , C. Boente , E. Giráldez , J. Taboada , J.E. Martín

Please cite this article as: S. Gerassis , M.T.D. Albuquerque , J.F. García , C. Boente , E. Giráldez , J. Taboada , J.E. Martín , Understanding complex blasting operations: A structural equation model combining Bayesian networks and latent class clustering, *Reliability Engineering and System Safety* (2019), doi: https://doi.org/10.1016/j.ress.2019.03.032

HIGHLIGHTS

- Workplace scenarios using explosives are analyzed with a Bayesian approach based on artificial intelligence (AI)

- Workers at different corporate hierarchies can engage with Bayesian modeling

- Blasting design and corporate accountability are the main drivers of accidents

- Uncertainty reduction is computed for accident risk groups

- Latent class analysis can avoid cognitive biases in safety policy formulation

Understanding complex blasting operations: A structural equation model combining Bayesian networks and latent class clustering

S. Gerassis[1], M.T.D. Albuquerque[2,3], J.F. García[1], C. Boente[4], E. Giráldez[1], J. Taboada[1], J.E. Martín[1]

[1] Department of Natural Resources and Environmental Engineering, University of Vigo, Lagoas Marcosende, 36310 Vigo, Spain
[2] Instituto Politécnico de Castelo Branco, QRural, 6001-909 Castelo Branco, Portugal
[3] ICT and CERNAS Research Centers, Portugal
[4] INDUROT and Environmental Technology, Biotechnology and Geochemistry Group, University of Oviedo, Campus de Mieres, 33600 Mieres, Spain

Abstract

A probabilistic Structural Equation Model (SEM) based on a Bayesian network construction is introduced to perform effective safety assessments for technicians and managers working on-site. Using novel AI software, the introduced methodology aims to show how to deal with complex scenarios in blasting operations, where typologically different variables are involved. Sequential Bayesian networks, learned from the data, were developed while variables were grouped into different clusters, representing related risks. From each cluster, a latent variable is induced giving rise to a final Bayesian network where cause and effect relationships maximize the prediction of the accident type. This hierarchical structure allows to evaluate different operational strategies, as well as analyze using information theory the weight of the different risk groups. The results obtained unveil hidden patterns in the occurrence of accidents due to flyrock phenomena regarding the explosive employed or the work characteristics. The integration of latent class clustering in the process proves to be an effective safeguard to categorize the variable of interest outside of personal cognitive biases. Finally, the model design and the software applied to show a flexible workflow, where workers at different corporate levels can feel engaged to try their beliefs to design safety interventions.

Key Words: Decision making; Bayesian learning; complex systems; risk analysis, structural design; blasting accidents

1. Introduction

Project execution requires meticulous management of the numerous risks present at the different operational stages. Recent works have underlined the need to evaluate risks from a holistic point of opinion, where technical aspects are connected with human and organizational factors [1,2]. Therefore, the challenge has been centered on the formalization of models that merge these three realities. To this effect, the industry has attempted to increase the process efficiency, improving accessibility, quality and

reducing the performance time under a framework committed to safety and the environment. However, one of the main problems encountered is how to model the propagation of uncertainty to the extent that the complexity of systems grows [3].

Complexity in the workplace has been furthered, in many cases, by abstraction and lack of cohesion in the integration of corporate hierarchies [4]. Risk analysis results obtained in the strategic levels are frequently hard to put through in practice, with operational levels regarded as the chief culprits when the targets are not accomplished. It has been demonstrated that labor responsibility affects decision making under risk [5]. In work scenarios where technical complexity and high competitiveness prevail, highly pressured managers and technicians prefer to take risk-averse decisions [6].

A particularity in complex systems is the impossibility to satisfactorily predict an accident from the study of its components [7]. The whole system should be considered in a deep evaluation of uncertainty and its network's propagation. While these methods deepen the understanding of the system, from the perspective of risk analysis the focus has always been on how to treat uncertainty [8]. Many researchers, including Levenson [9], Hollnagel [10] or Aven and Zio [11] have discussed the advantages and limitations of the application of probability-based approaches in models such as the System-Theoretic Model and Processes (STAMP) or the Functional Resonance Analysis Method (FRAM). Indeed, the lack of available information and the scarcity of resources to create accurate models was the major drawback for the probability approach acceptance [12].

Occupations regarded as dangerous increased notably this challenge. An example of this is the mining industry, which is one of the industries with the highest scores of occupational accidents worldwide [13,14]. Mining engineers are needed to assure the success of operations deployed with the right decisions for working conditions continuously changing. In the interim, occupational risk assessment in construction

3

sites was experiencing important advances during the last years, both in the techniques [15-17] and in the prediction capabilities [18-20]. Particularly, a better understanding was reached regarding the safe use of explosives to excavate the rock mass. Additionally, the emergence of new models allowed to successfully predict the environmental impacts associated with ground vibration or flyrock phenomena [21-22].

However, the results obtained thought theoretical models' fitting, usually take a long validation time, and thus, imposing a delay in the implementation of a workday routine [23-24]. This situation leads to the necessity of contrasting decisions, accordingly to the changing business conditions. The problem is that the technical staff is still resistant to the incorporation of digital tools based on AI techniques as the human creativity is not required in the same degree to face the work challenges [25].

To tackle this issue, a probabilistic SEM based on Bayesian networks is developed using the last AI software technology available. The aim is to provide an innovative digital approach that also considers expert knowledge as criteria to analyze different scenarios and define strategies in a practical manner. This resolution is anticipated to offer flexibility in the workplace and, at the same time, bring down the traditional preoccupation of engineering complex systems.

The great evolution of AI and the implementation of its techniques in many computer applications span a new horizon, making feasible to perform tasks that were simply impossible some years ago. This is the case of Bayesian networks, which have been during decades an indisputable analytical element of human reasoning under uncertainty. Yet, it is their current implementation in decision-support systems with increasingly improved software estimation capacities [26] what makes possible to learn them from data and thereby becoming an ideal tool for on-site workplace analytics. Their precision together with their powerful graphical outputs allows orientation to the analysis and evaluation of possible scenarios and strategies, avoiding carrying out arduous programming tasks [27].

4

This access is used to study the reasons why blasting accidents occur in complex mining and civil works. The model built aims to unveil relevant contributions that help-out technicians to design custom safety interventions. Additionally, this research is also intended to promote AI in engineering applications, developing a renewed probabilistic model that can obtain reliable results reflecting the difficult nature of the work context. To that end, expert criteria are used as a solution and complement when there is scarce information, including a new safeguard against the cognitive biases that employees may induce. It is important to emphasize that the promotion of these characteristics could situate AI as a disruptive element in decision-making not only in strategic phases but also in the operative parts of a project. Digital transformation in the workplace is expected to make professionals adapt their job tasks to AI as information specialists [28]. This could have a great impact in areas such as the mining sector where the incorporation of technology is slower in comparison to other industries [29].

The manuscript continues as follows: Section 2 explains the methodology employed to build a probabilistic SEM based on the latent class clustering of a target variable. Section 3 applies this methodology to the study of accident risks in the execution of blasting operations using an innovative AI software tool. Section 4 includes a discussion of the results obtained and how they can contribute to the reduction of accidents. Finally, concluding remarks and future prospects are made in Section 5.

2. Methodology for combining latent class clustering and Bayesian networks in a probabilistic SEM

The methodology proposed is based on the principles exposed by Conrady and Jouffe [30] about Probabilistic Structural Equation Models (PSEMs), which represent a conceptual evolution from traditional SEMs. The origin of SEM dates back to 1918 when geneticist Sewall Wright developed path analysis [31]. From that moment on, years of development were conditioned to the growing needs of both researchers and

social science practitioners to understand latent phenomena [32]. In 1973, Jöreskog [33] exposed the maximum likelihood for estimating SEM with computer intensive implementations. Since then, SEM has become very popular, continuing its development during the 21st with the adoption of multi-level and Bayesian approaches [34], and the introduction of algorithms from AI.

In this research, PSEMs are based on a Bayesian network structure, partially or fully machine learned from data, where all relationships are probabilistic and nonparametric, facilitating the incorporation of categorical variables. However, when modeling complex scenarios, it becomes extremely difficult to accurately define representative subgroups of analysis. For this reason, latent class modeling is introduced as a powerful first step to obtain meaningful segments of a specific attribute of interest to further PSEM development and so, tailoring interventions accordingly to specific subgroups.

3.1 Latent Class analysis

Latent Class Clustering (LCC) is an unsupervised data mining task that involves the identification of unobserved (latent) segments within a population. The latent classes are established by using the responses of the cases on a set of observed variables (indicators). Cases in a specific latent class are homogeneous in their responses to these indicators, while cases within different latent classes present important dissimilarities. In a formal approach, a latent class cluster model (LCM) is represented by K distinct classes of a nominal latent variable X. The latent variable X is measured with a set of observed variables, $Y_1,\ldots, Y_n$ where one observation can only be a member of one k class. Noting that P represents probability and P_{Xk} denotes the probability of a certain observation being in a latent class (k=1, 2, ...,K),   the expression of an LCM is given by

$$P_{Yi} = \sum_{k=1}^{K} P_{Xk}\, P_{(Yi|Xk)} \qquad\qquad (1)$$

where $P_{(Yi|Xt)}$ is the conditional probability of obtaining a case with a response pattern $Y_i = (y_1,...,y_n)$, given its belonging to the k class of a latent variable X. The model parameter estimation is normally carried out with iterative numerical methods. In this study, the criterion of maximum likelihood is used, although recent research in the area is starting to implement Markov Chain Monte Carlo (MCMC) and Gibbs sampling [35]. When the LCM is computed every case is classified to their most likely latent class. This is directly achieved using Bayes' theorem to obtain the posterior probability of a case membership in a certain k class

$$P_{(Xt|Yi)} = \frac{P_{(Yi|Xk)}\, P_{(Xk)}}{P_{(Yi)}} \qquad\qquad (2)$$

The goal of LCC is to obtain the number of clusters that best represent reality. There exist different possibilities to assess the model fit. The rise of information theory during the last years has displaced traditional methods sometimes questionable because of their strong assumptions [36]. For this reason, different information criteria were calculated to determine the optimum number of discrete and non-overlapping latent classes. Concretely, three representative parsimony indices widely found in the literature were used: (i) Akaike information criterion (AIC) [37]; (ii) Consistent Akaike information criterion (CAIC) [38]; and (iii) Bayesian information criterion (BIC) [39].

3.2 Unsupervised Bayesian learning

The PSEM starts with the creation of a global Bayesian network. For that purpose, unsupervised learning is an extended first step in machine learning applications when a human expert is insufficient to build a model that accurately represents reality. To carry

7

out this, the variable of interest, previously segmented with LCC, is excluded from the rest of the covariates. The temporary excluded variable constitutes the target node of the probabilistic SEM and it is undesired that becomes part of the Bayesian network structure that it will be used for discovering hidden patterns. Granted the importance of getting a robust network from which later perform posterior analyses, divers set of unsupervised learning algorithms are proven to create different network structures. The network whose structure offers the best performance is selected, increasing the probabilities of obtaining the optimal resolution for the topic under study.

3.3 Network variable clustering

The purpose of this step is the identification of relevant groups within the variables used in the unsupervised Bayesian network computation. For this purpose, hierarchical agglomerative clustering was performed using a minimum arc force value, below which clusters are not merged. For computing the arc force, the Kullback-Leibler Divergence (D_{KL}) is used to measure the strength of a direct relationship between two nodes. In the context of machine learning, the D_{KL} is often called the information gain [40], which compares two joint probability distributions (JPDs) P and Q. For probability distributions of a discrete random variable X, the D_{KL} is defined as

$$D_{KL}(P_x||Q_x) = \sum_{x} P_x \log_2 \frac{P_x}{Q_x} \qquad (3)$$

which constitutes the expectation of the logarithmic difference between P_x and Q_x. The stop threshold and the maximum cluster size depending on the purpose of variable clustering. If the goal is exclusively dimensionality reduction, cluster size should have a high value to obtain a few groups of interest. Nonetheless, when building PSEMs, a value between 5 and 7 is advisable [30] to obtain latent variables with a meaningful and

8

manageable number of manifest variables. This approach is highly dependent on policy makers who have here the opportunity to evaluate different clustering strategies.

3.4 Model completion

The target node, excluded in the second step, was introduced in the Bayesian network together with the manifest variables and their respective latent factors. This action completes the PSEM generating a final Bayesian network where the latent factors and the target node conform an overall representation of the study domain. To the accomplishment of an authentic SEM-type network structure, the constraint of all the links to the target node occurs directly with the latent factors, was introduced. In this manner, the manifest variables keep connected exclusively to their latent factors depicting two differentiated levels of complexity.

3. Engineering application

3.1. Data description

In this case study a total of 163 records of accidents from blasting operations, that took place in complex mining and civil engineering activities between 2009 and 2014, was used. The dataset was gathered from accident reports supplied by the companies. However, further information was collected through personal interviews and questionnaires conducted to prevention technicians and relevant employees. This allowed to adequately register the effects of adverse weather conditions, poor blasting designs, and deficient communications as key elements for the complexity of blasting operations.

All the data available was used to specify a set of variables that illustrate the broad range of conditions existing when an accident happens. In effect, this approach opens the possibility to the internalization of the "accident risk", which stands for the probability of a certain accident's typology to occur. In total, 28 variables were outlined, having each variable two or three possible outcomes representing the characteristics of

the accident. The variables can be grouped in two categories: (i) General variables: these include risk factors widely discussed in the literature [41,42] as common causes of accidents at work (e.g. Operator training, machinery age or order and cleanliness) (ii) Specific variables: these represent factors that are associated exclusively with the activity developed (e.g. explosive conditions or state of haul roads). See in Appendix I a description of all the variables employed in this study.

3.2. Segmentation of blasting accidents

As the final variable and target node of the PSEM, the accident risk is introduced describing the cause of why workers suffered an accident. A total of 11 different accident typologies were identified (Table 1). To obtain representative groups of study, the latent class analysis was carried out using XLSTAT-LG 2018 excel add-in. To find the optimal number of clusters AIC, CAIC and BIC information criteria were estimated through the generation of 10 evaluation models (1 to 10 clusters).

These information criteria favor a model that produces a high log-likelihood value using relatively few parameters, where a lower value represents a better fit. The growth in the number of clusters diminishes the criteria values, but a high number of clusters also provoke more complexity. Thus, it is necessary to determine the best tradeoff between complexity and statistical fit. In Fig. 1. goodness of fit is illustrated according to the clustering structure, showing the three criteria its lowest value when the number of clusters is 5. This is consistent with Depaire et al. [43] who adopted the model for which CAIC and BIC stabilize their improvement. In this regard, De Oña et al. [44] stressed the importance of ignoring a marginal improvement in a statistical fit if this implies a notable increase in the complexity.
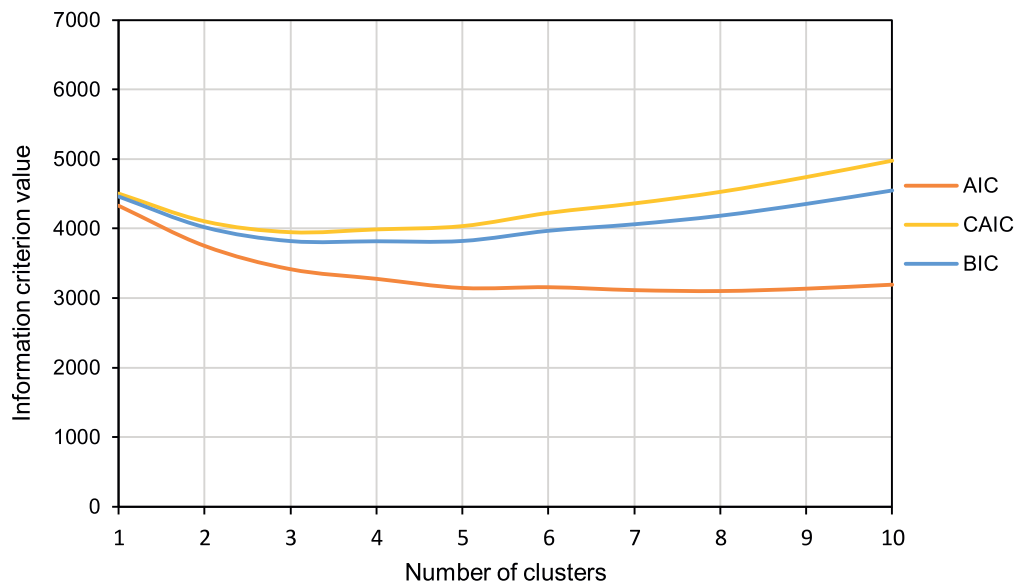
10

Fig. 1. Estimation of AIC, CAIC, and BIC increasing the model number of clusters.

Once the number of clusters is ascertained, it is required the correspondent characterization. Indeed, accidents involving objects are almost entirely found in cluster 1, representing 92% of accidents for this group. An exception is an entrapment between objects which constitutes 41% of cluster 4, been linked to accidents related to overexertion (38% of cluster 4). For cluster 2, it was found that 96,5% of the accidents occur due to falls from the same or different height. Projection of fragments and particles constitute 87% of the accidents in cluster 3. Lastly, thermal, electrical or chemical exposures are primarily found in cluster 5 and corresponding to 79% of the cases in this group. A worldwide overview of the newly obtained clusters is represented in Table 1.

Table 1

Accident typology and cluster definition for latent class analysis.

Accident typology	Cluster	Characterization	Size %
Detachment or falls of loose objects			
Stepping on objects			
Collisions with mobile or stationary objects	C1	Collisions or falls involving objects	67.82%
Blows from objects or equipment			
Falls from the same/different height	C2	Falls from the same/different height	12.07%
Projection of fragments and particles	C3	Projection of fragments and particles	8.62%
Entrapment between objects	C4	Entrapment and overexertion	5.75%
Overexertion			
Thermal contact			
Exposure to electrical contacts	C5	Exposure to physical or chemical agents	5.75%
Exposure to caustic and corrosive substances			

3.3. Initial Bayesian design

The 28 variables defined, excluding the accident risk, are used to create a primary unsupervised Bayesian model. For network modeling, BayesiaLab software version 8.0 was used. This AI platform provides a modern environment for machine learning through a wide range of structural learning. All unsupervised learning algorithms in BayesiaLab are based on the Minimum Description Length (MDL) [30], which is a two-component score widely used in AI applications that must be minimized to obtain the best structure in terms of bits required for representing the model, $DL(B),$ and its underlying data, $DL(D|B)$. Formally, the MDL is expressed as

$$MLD(B, D) = \alpha DL(B) + DL(D|B)$$

where α represents the network structural coefficient. Therefore, for each possible Bayesian network the MDL score is computed, where minimizing the score implies assessing the quality of each candidate network with respect to the available data. By

comparing the MDL score of the resulting unsupervised networks, the one built with Maximum Weight Spanning Tree (MWST) algorithm (Fig. 2) produces an MDL score of 2,742.277, slightly lower than the other algorithms performed, such as Taboo, EQ or SopLEQ whose values range from 2,791.554 to 2,981.634.
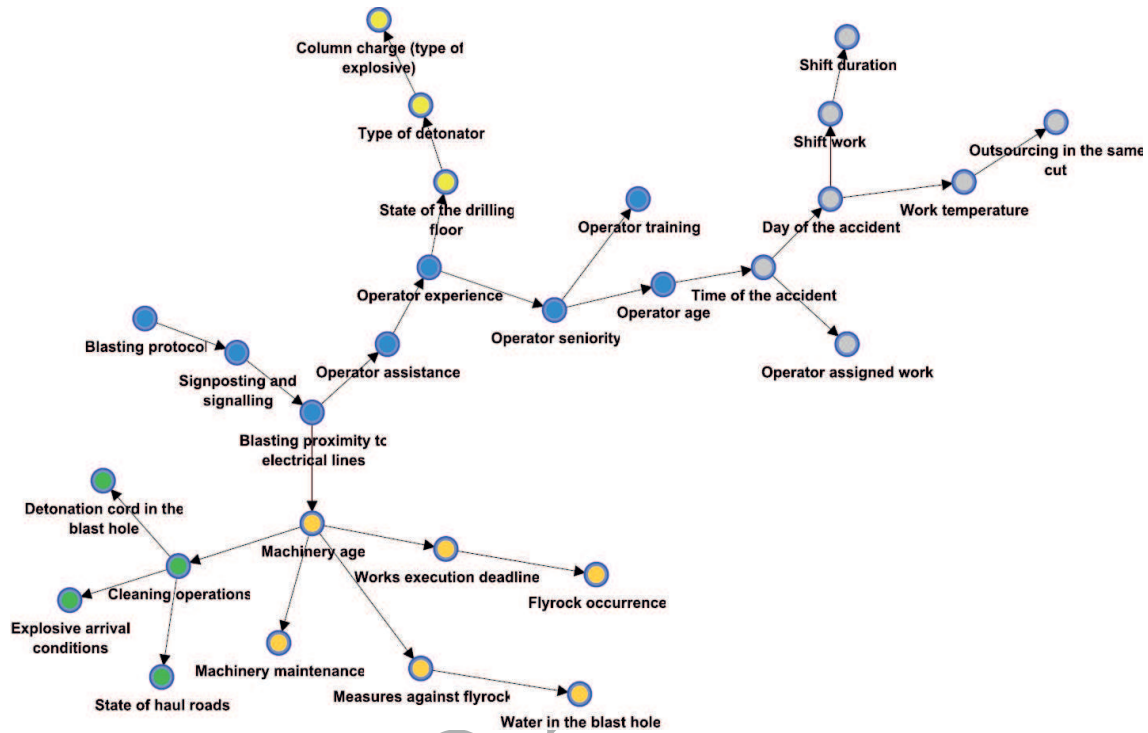

Fig. 2. Unsupervised Bayesian network using MWST algorithm with variable clusters coded in colors

The network established is evaluated in terms of its probabilistic relationships between the nodes. The BayesiaLab's variable clustering algorithm identifies clusters of variables by using arc force as a probabilistic measure based on the D_{KL} (section 3.3). Initially, each manifest variable is taken as a distinct cluster where the DKL merges gradually the closest variables into different groups. For building a comprehensible PSEM, 5 clusters are selected with groups between 3 and 7 variables (Fig. 2). However, this process offers the possibility of editing the proposed clusters at any time according to expert criteria.

3.4. Conception of multiple networks

All the variable clusters embedded in the initial Bayesian model are divided into independent Bayesian sub-networks with the formalization of a new latent factor variable for each structure. This procedure represents a key step in the PSEM due to the possibility of disaggregating information for modeling specific conceptual frames that are unknown beforehand. The only condition that is necessary to define is the number of states in the future latent factor. In this case, this parameter has been set between 2 and 5, allowing data clustering algorithm in BayesiaLab to determine the optimal number of states for representing the JPD.

Upon completion of this process, the PSEM layout shows 5 Bayesian sub-networks with the induced latent factors surrounded by their manifest variables (Fig. 3). These new Bayesian models follow the same color code set for the clusters defined in Fig. 2. Each latent factor (white node) was characterized by a name representative of the conceptual frame that embodies. For this task, mutual information (I) between the manifest variables and the latent factor was computed in order to identify the most influential variables regarding the knowledge of the latent factor.
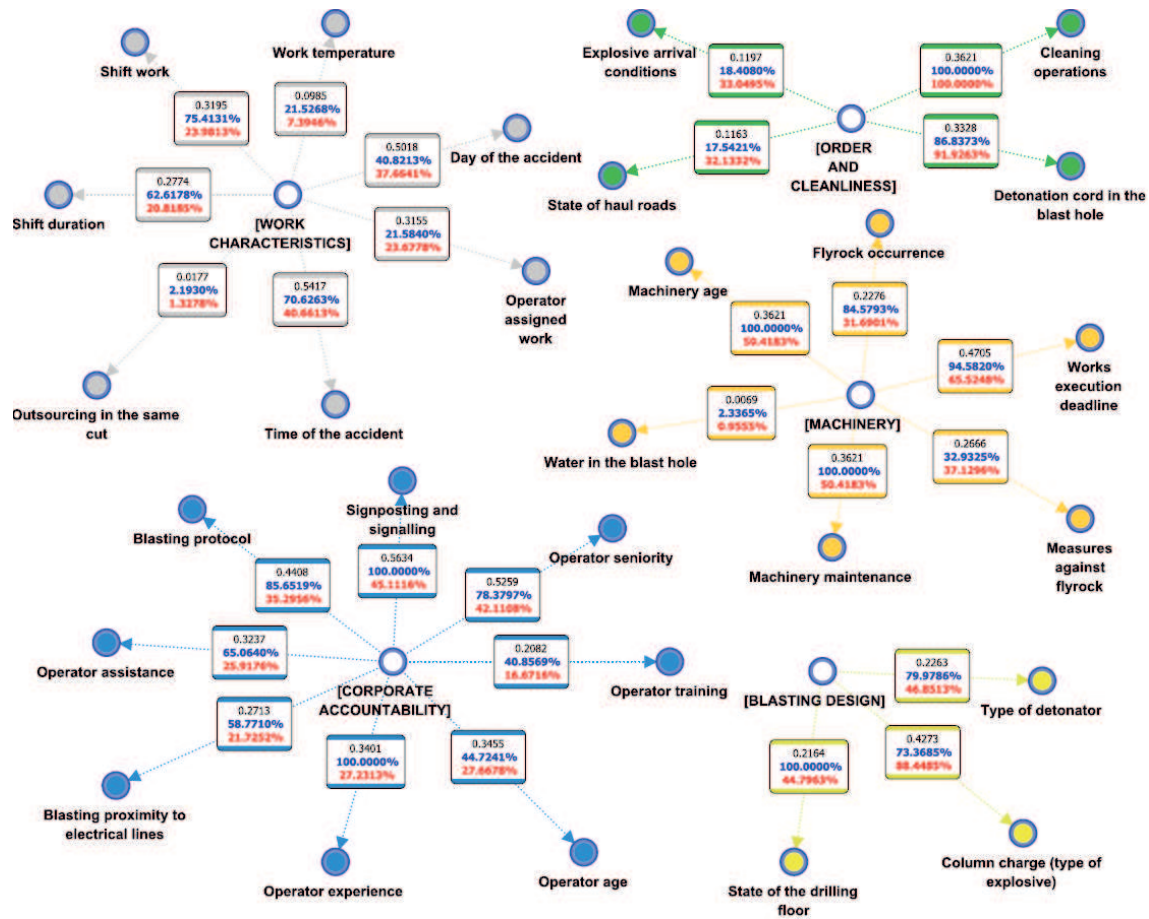
Fig. 3. Multiple Bayesian networks generated from variable clustering and arcs' mutual information analysis

In Fig. 3 the mutual information analysis is displayed in a box over the networks' arcs, where the number in the top reflects the I value. Nevertheless, to provide clarity in the interpretation of the relative mutual information (RMI) is likewise presented. For example, in the gray sub-network (Fig. 3) the variable *time of the accident* shows the highest MI value (0.5417 bits). The blue number shows the RMI regards the child node, whereas the red number refers to the parent node. Thus, the knowledge of *the time of the accident* contributes to reducing the uncertainty regarding the latent factor by 70.63%. Conversely, knowing the latent factor the uncertainty is reduced for the *time of the accident* by 40.66%. Other variables having strong predictive importance of this sub-network are *the day of the accident*, *shift work,* and *shift duration*. Therefore, it is

concluded that this sub-model could be denominated "Work characteristics". The same reasoning was given for the continuing sub-nets.

3.5. Final Bayesian structure

The last step in the PSEM building process incorporates the variable accident risk, which becomes at this point available for machine learning. Given the objective of obtaining a probabilistic network with an SEM typical composition, it is necessary to guarantee the creation of a final Bayesian model with a hierarchical structure. Taboo learning is an unsupervised algorithm in BayesiaLab that ensures this constraint. This algorithm has the capability to ascertain a new structure on top of an existing network and at the same time forbid new relations between the target node (accident risk) and the manifest variables. This is possible due to the adaptive memory of this metaheuristic, which conducts the search method to escape the trap of local optimality [45]. Taboo search employs a local or neighborhood search procedure to move from one potential outcome x to an improved solution x' until the stopping criterion is fulfilled. The solutions accepted to the new neighborhood $N(x')$ are selected using memory structures, typically known as taboo lists.

The emerging model delivered after completing the learning process is shown in Fig. 4. As it turns out, in a first layer the manifest variables are connected to their respective latent factors, and, in a second layer all the latent factors are linked to the accident risk according to the new relations discovered.
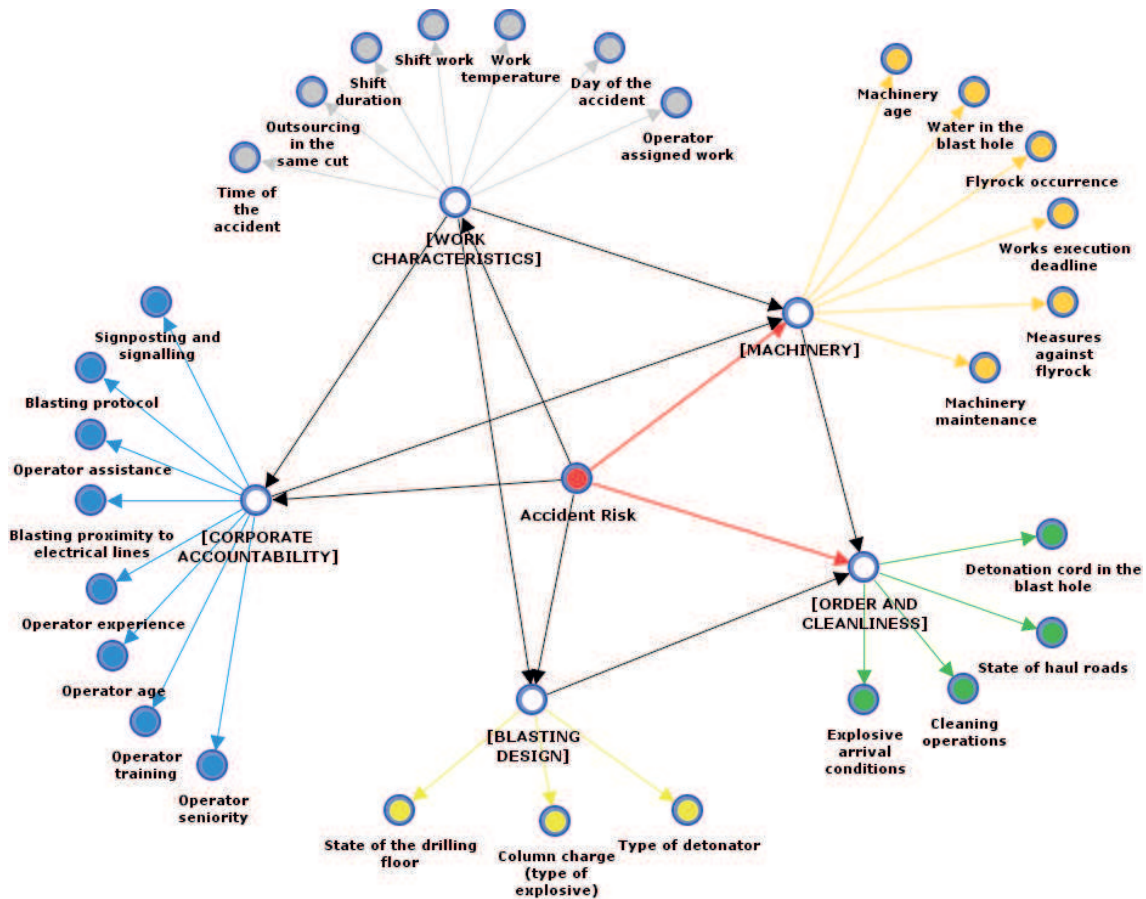
Fig. 4. Final PSEM structure with two distinct levels of complexity built using Taboo algorithm

The final structure elucidates that the accident risk is directly linked to work characteristics, blasting design and corporate accountability. However, in order to analyse the influence of all the latent factors in the occurrence of accidents, it was assigned manually the probabilistic relation between accident risk with machinery and, order and cleanliness (red lines in Figs 4, 5). In Fig. 5. arcs' mutual information was computed exclusively between the latent factors and the accident risk. The results obtained show the multicausal nature of accidents. Blasting design is the factor that provides a higher reduction of uncertainty regarding the accident risk by 2.52% on average, which is like corporate accountability and work characteristics, with 1.82% and 1.42% respectively. Finally, machinery together with order and cleanliness have the lowest influence, which justifies that these arcs were not added to the network built by the Taboo algorithm due to its weak probabilistic relationships.
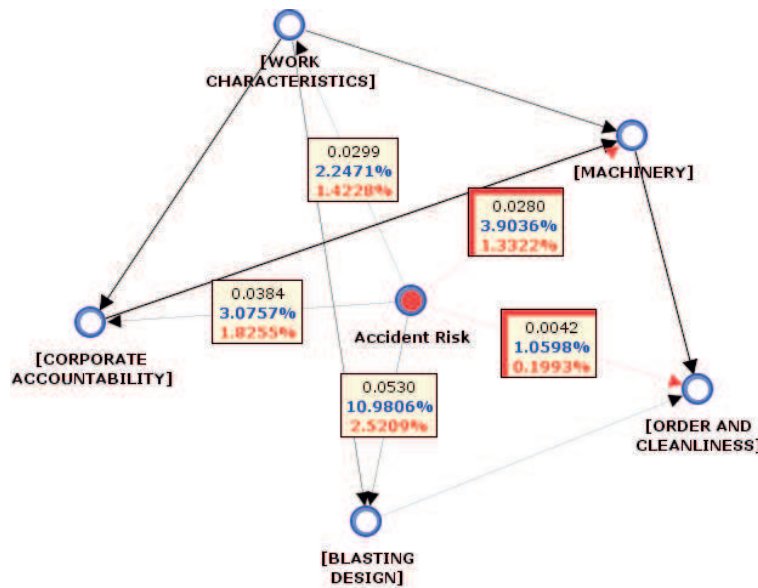
Fig. 5. Second, the complexity level of the PSEM with arc's mutual information visualization

3.6. Model validation

Validation of the built PSEM is of capital importance to sustain the matching between the obtained results and reality. In this study, two types of validation tests are considered to assess the reliability of the model: data perturbation and contingency table fit (CTF). The data perturbation algorithm implemented in BayesiaLab consists of of performing cross-validation, adding random noise to the weight of each observation of the original dataset. This disturbance is generated from a normal distribution with a mean of 0 and an initial standard deviation of 1. To examine the robustness of the models, the networks' MDL score is analyzed. Upon realization of this algorithm, it is found that the score has changed from 55,986.204 to 54,828.156, although no variations are perceived for the probabilistic relationships of the model (Fig. 4). Hence, this validation process has allowed escaping to a local optimum, obtaining a final model with a lower MDL that offers reasonable confidence.

The CTF is a useful metric to measure the quality of the induced factors. The main benefit of using CTF in BayesiaLab as a quality measure is that offers normalized values ranging from 0% to 100%. A higher CTF value implies a good representation of

18

the JPD. A common alert threshold below which the factor should be relearned is 70%, but in those cases, including more than 5 manifest variables per sub-network, this level should be reduced [30]. In Table 2 are shown the CTF values obtained for the latent factors induced at the conception of multiple networks stage. Only work characteristics are below the alert threshold but given that this group holds 7 manifest variables this result seems totally acceptable.

Table 2

Performance indices of induced factors during multiple clustering

Induced Latent Factor	CTF	Mean Purity	Hypercube cells per state
Work Characteristics	65.44%	98.61%	106.02
Machinery	96.26%	99.88%	23.10
Corporate accountability	79.97%	100.00%	172.74
Order and cleanliness	96.49%	100.00%	5.14
Blasting design	100.00%	100.00%	6.00
Mean	87.63%	99.69%	62.60

4. Results and discussion

The completion of the PSEM implies that the integrative modeling process that conceptualizes the accident risk is available to carry out causal inference. A major advantage of creating a probabilistic SEM using Bayesian networks is the possibility of computing the posterior probabilities of all nodes in the network omni-directionally, regardless of arc direction. This potentially allows exploiting to the informative content of the model for policy analysis.

4.1. Design of safety interventions

The design of safety interventions that minimize future accident risk in blasting operations is a fundamental pillar of this study. As an application example, in Fig. 6 is shown the latent factor blasting design together with its corresponding manifest variables. The factor induced is labeled with 3 cluster state representatives of the JPD generated during the multiple cluster algorithm application (section 4.4). Given the marginal distribution of the states, Cluster 2 constitutes the most frequent scenario across the manifest variables. Electric detonators, emulsion as column charge and difficult terrain with soft rock constitute a common pattern (91.95%) in blasting accidents. This becomes obvious when hard evidence is set for Cluster 2 (Fig. 6). Cluster 1 and 3 represent scenarios much less likely to occur.
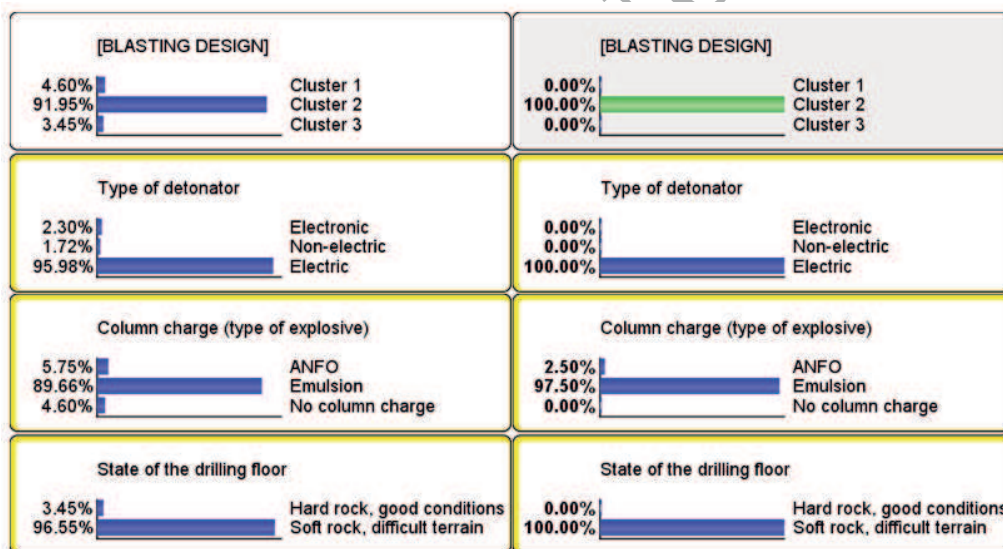

Fig. 6. Probability distribution results for "blasting design" sub-network

The PSEM group assessment offers valuable insight into the necessary actions to implement aiming the reduction of blasting accidents. Nevertheless, future prevention programs are thought to put a premium on priorities [46]. By using mutual information blasting design was identified as the most influential factor regarding risk accident (Fig. 5). This can be extremely useful when companies must sacrifice resource allocation in priority rankings due to budget constraints. The PSEM developers can contribute to

going through strategic planning that promotes efficient interventions by maximizing the yield on investment (ROI).

4.2. Training and accident analysis

A classic challenge in blasting operations is flyrock phenomena. The fierce and uncontrolled projection of rock fragments due to the energetic effects of the blast is one of the main accident risks (Table 1). To contain the outcomes of these issues has always been a major topic for engineers. In order to address this issue, Bayesian inference is carried out to analyze the causes behind accidents due to the projection of fragments and particles, which are a direct consequence of flyrock. The posterior probability computed is shown in Fig. 7 for several nodes in the model that represent a special interest in flyrock occurrence. It was found that signposting and signaling are insufficient in 84% of these accidents, whereas a blasting protocol with a defined blast area security is followed just by 13.91% of the accidents. This lack of involvement of operators and contractors with the protocols can be reflected when assessing the measures undertaken against flyrock. Special protection systems were only adopted in 8.42% of the accidents, being a common unsafe practice the use of machinery (82.36%) as a protective element.

Regarding geology and rock structure, a total of 88% of accidents occurred in soft rock, where difficult terrain with the existence of joints and fissures favors the release of energy contributing to flyrock generation. In this regard, blasthole overloading is a risk factor to prevent and emulsion correct charge must be brought into consideration (Fig. 6). Lastly, the work execution deadline is also found to be linked with the projection of fragments in blasting operations. The highest ratio (88.80%) of delayed works is a precursor to get around restrictions on occupational health and safety.
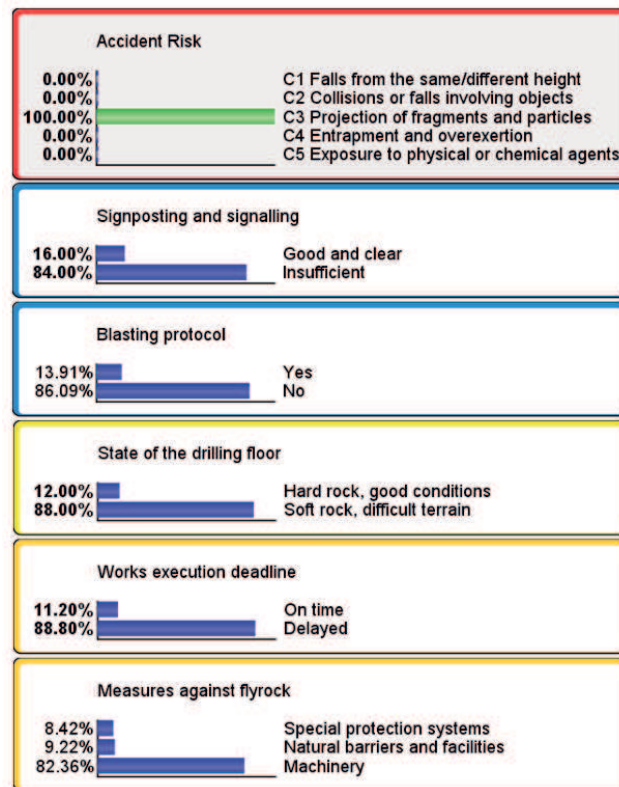
Fig. 7. Inference results in some variables when hard evidence is set for "Projection of fragments and particles" in accident risk

These results are a powerful source of information about training courses about the safe use of explosives. Having a trained workforce with knowledge about occupational risks is critical to reduce blasting accidents. However, there are differences in how workers perceive risks [47]. The reasons why this is happening can be establish in the failure of training programs and communication. Many of these programs are based on regulations which deal with general aspects of the design and execution of blasting operations. Traditionally, the purpose of this approach is the diminution of the accident's risk by ensuring workers' commitment to safety regulations. Inevitably, over the years, this has failed to owe to the lack of adaptation and capacity required to explain how the different risks interact in a complex system like this.

In the herein case study, the introduction of AI solutions at the formative process stage constitutes an example of how disruptive technologies can be used to reduce blasting accidents. In Fig. 8 is shown the inference simulation for collaboration between workers

and its impact on work execution deadline and, signposting and signaling. When an operator has assisted, the works improve their execution time by 17 percentage points having a good signposting and signaling now over half of the time (52.63%). Indeed, the possibility of using this kind of tools can provide the responsible entities with a better understanding about safe practices and uncertainty propagation.
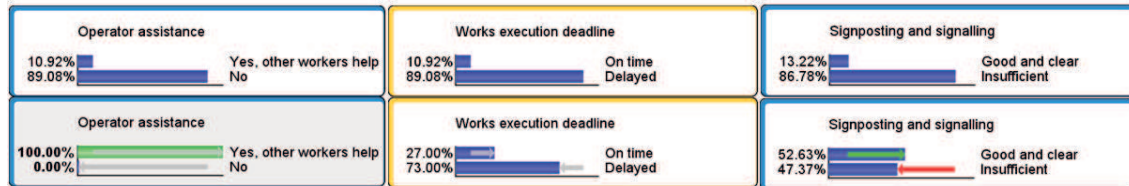

Fig. 8. Intercausal reasoning for the analysis of "operator assistance" impact in blasting operations

4.3. The power of clustering networks

Despite the great advances in machine learning during the last years, cluster analysis still represents a versatile technique that incorporated in new AI software systems allows unifying the best of the old with the best of the new. In this study, two different clustering techniques were applied in the development of the PSEM. First, LCC offered a meaningful segmentation of blasting accidents in statistical terms, where 5 major groups were identified (Table 1). This procedure avoids cognitive biases that expert criteria can include providing a solid foundation for the construction of the PSEM.

In the second place, a hierarchical agglomerative clustering using the Kullback-Leibler Divergence was created to find out relevant groups of interest between the variables in the model. For this study, 5 variable clusters were selected from which 5 Bayesian sub-models were generated (Fig. 3). The validation metric CTF with a mean of 87.63% (Table 2) shows that little information was lost by inducing the latent variables. Therefore, this represents a good balance between conceptual gain and information loss.

Network variable clustering has the advantage of involving engineers and analyst teams so they can assess their expert judgement. Consequently, a broad range of stakeholders, regardless of their quantitative skills, can engage in an easy way with Bayesian network modeling and contribute with their expertise to produce value-based decisions. This may include policy-related considerations of risk and safety that are not covered by the PSEM.

5. Conclusions

This article describes the construction of a probabilistic SEM by using Bayesian networks in order to analyze the risk of accident in complex blasting operations. The methodology proposed counts as an innovating element with the incorporation of LCC as an initial step in the modeling process to reduce the cognitive biases that engineers intrinsically make when designing the strategic planning of safety interventions. The model built combines two hierarchical structures where a total of 28 manifest variables representative of accident occurrences is grouped into 5 sub-networks with a latent factor representing its conceptual frame. Each latent factor induced is machine learnt to the accident risk where information theory is used to assist the analyst about which factors have a higher impact on the accident occurrence.

The functioning of the model exhibits its potential to quantify the uncertainty associated with the causes of accidents, offering policy-makers the possibility to infer multiple risk scenarios that shed light about the strategies that are more likely to succeed. The AI software employed to set up the model shows a smooth workflow with highly sophisticated graphical results. This implies the evolution of Bayesian networks and many AI platforms, which have improved their existing visualization tools decreasing the challenge that supposed in the past to explore complex systems and their unmanageable structures for non-statisticians. Finally, the introduction of PSEM into safety policy design can boost the level of resilience for an organization responding in a

more flexible way to threats, anticipating better potential risks. Future research requires the implementation of solutions like the one proposed to solve complex problems in different domains that seemed to be intractable until very recently. This scenario suggests the necessity to keep bringing unstructured data into play creating AI solutions based on real experiences that have response capacity.

Acknowledgments

References

[1] Okun, A.H., Guerin, R.J., Schulte, P.A., 2016. Foundational workplace safety and health competencies for the emerging workforce. J. Saf. Res. 59, 43-51.

[2] Mullen, J., Kelloway, E.K., Teed, M., 2017. Employer safety obligations, transformational leadership and their interactive effects on employee safety performance. Saf. Sci. 91, 405-412.

[3] Lee, J., Hong, Y.S., 2018. Data-driven prediction of change propagation using dependency network. Eng. Appl. Artif. Intell. 70, 149-158.

[4] Floricel, S., Michela J.L., Piperca S., 2016. Complexity, uncertainty-reduction strategies, and project performance. Int. J. Proj. Manag. 34, 1360-1383.

[5] Pahlke, J., Strasser, S., Vieider, F.M., 2015. Responsibility effects in decision making under risk. J. Risk Uncertain. 51,125-146.

[6] Wang, Z., Kuang, Y., Tang, H., Gao, C., Chen, A., Chan, K.Q., 2017. Are decisions made by group representatives more risk averse? The effect of sense of responsibility. J. Behav. Decis. Mak. 31, 311-323.

[7] Ottino, J.M., 2004. Engineering complex systems. Nature. 427, 339. doi:10.1038/427399a

[8] Paté-Cornell, M.E., 1996. Uncertainties in risk analysis: Six levels of treatment. Reliab. Eng. Syst. Saf. 54, 95-111.

[9] Levenson, N., 2012. Engineering a safer world: systems thinking applied to safety. Cambridge, The MIT press, 2012. ISBN: 978-0-26-253369-0

[10] Hollnagel, E., 2013. An application of the Functional Resonance Analysis Method (FRAM) to risk assessment of organizational change. Swedish Radiation Safety Authority. Report number, 2013:09. ISSN: 2000-0456.

[11] Aven, T., Zio, E., 2011. Some considerations on the treatment of uncertainties in risk assessment for practical decision making. Reliab. Eng. Syst. Saf. 96, 64-74.

[12] Bjerga, T., Aven, T., Zio, E., 2016. Uncertainty treatment in risk analysis of complex systems: The cases of STAMP and FRAM. Reliab. Eng. Syst. Saf. 156, 203-209.

[13] Chen, J., Zorigt, D., 2013. Managing occupational health and safety in the mining industry. J. Bus. Res. 66, 2321-2331.

[14] Breuer, J., Höffer, E.M., Hummitzsch, W., 2002. Rate of occupational accidents in the mining industry since 1950--a successful approach to prevention policy. J. Saf. Res. 33, 129-141.

[15] Rivas, T., Paz, M., Martín, J.E., Matías, J.M., García J.F., Taboada, J., 2011. Explaining and predicting workplace accidents using data-mining techniques. Reliab. Eng. Syst. Saf. 96, 739-747.

[16] Seo, J., Han, S., Lee, S., Kim, H., 2015. Computer vision techniques for construction safety and health monitoring. Adv. Eng. Inform. 29, 239-251.

[17] Golizadeh, H., Hon, C., Drogemuller, R., Reza, M., 2018. Digital engineering potential in addressing causes of construction accidents. Autom. Constr. 95, 284-295.

[18] Guo, B., Wing, T., González, V.A., 2016. Predicting safety behavior in the construction industry: Development and test of an integrative model. Saf. Sci. 84, 1-11.

[19] Sarkar, S., Verma, A., Maiti, J., 2018. Prediction of occupational incidents using proactive and reactive data: A data mining Approach. In: Maiti, J., Ray, P., (eds) Industrial Safety Management. Managing the Asian Century. Springer, Singapore.

[20] Debnath, J., Biswas, A., Sivan, P., Nirmalya, K., Sahu, S., 2016. Fuzzy inference model for assessing occupational risks in construction sites. Int. J. Ind. Ergon. 55, 114-128.

[21] Armaghani, D.J., Hasanipanah, M., Bakhshandeh, H., Tonnizam, E., 2018. Feasibility of ICA in approximating ground vibration resulting from mine blasting. Neural Comput. Appl. 29, 457-465.

[22] Saghatforoush, A., Monjezi, M., Faradonbeh, R.S., Armaghani, D.J., 2016. Combination of neural network and ant colony optimization algorithms for prediction and optimization of flyrock and back-break induced by blasting. Eng. Comput. 32, 255-266.

[23] Luo, L., He, Q., Xie, J., Yang, D., 2017. Investigating the Relationship between Project Complexity and Success in Complex Construction Projects. J. Manage. Eng. 63. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000471

[24] Antunes, R., Gonzalez, V., 2015. A Production Model for Construction: A Theoretical Framework. 5, 209-228. Buildings, 9, 22-39.

[25] Brynjolfsson, E., Mitchell, T., 2017. What can machine learning do? Workforce implications. Science. 358, 1530-1534.

[26] Zare, F., Zare, H.K., Fallahnezhad, M.S., 2016. Software effort estimation based on the optimal Bayesian belief network. Appl. Soft. Comput. 49, 968-980.

[27] McCann, R.K., Marcot, B.G., Ellis, R., 2006. Bayesian belief networks: applications in ecology and natural resource management. Can. J. For. Res. 36, 3053-3062.

[28] Makridakis, S., 2017. The forthcoming Artificial Intelligence (AI) revolution: Its impact on society and firms. Futures. 90, 46-60.

[29] Perrons, R.K., McAuley, D., 2015. The case for "n«all": Why the Big Data revolution will probably happen differently in the mining sector. Resour. Policy. 46, 234-238.

[30] Conrady, S., Jouffe, L., 2015. Bayesian Networks & BayesiaLab – A Practical Introduction for Researches. Franklin, TN, Bayesia USA (ISBN: 987-0-9965333-0-0)

[31] Westland, J.C., 2015. Structural equation models. From paths to networks. In: Studies in Systems, Decision and Control Vol. 22. Springer International Publishing.

[32] Tarka, P., 2018. An overview of structural equation modeling: its beginnings, historical development, usefulness and controversies in the social sciences. Qual. Quant. 52, 313-354.

[33] Jöreskog, K.G., 1973. A general method for estimating a linear structural equation system. In: Goldberger, A.S., Duncan, O.D. (eds.) Structural Equation Models in the Social Sciences, pp. 83–112. Academic Press, New York.

[34] Depaoli, S., Clifton, J.P., 2015. A Bayesian Approach to Multilevel Structural Equation Modeling with continuous and dichotomous outcomes. Struct. Equ. Modeling. 22, 327-351.

[35] White, A., Wyse, J., Bredan, T., 2016. Bayesian variable selection for latent class analysis using a collapsed Gibbs sampler. Stat. Comput. 26, 511-527.

[36] Tein, J., Coxe, S., Cham, H., 2013. Statistical power to detect the correct number of classes in latent profile analysis. Struc. Equ. Modeling. 20, 640-657.

[37] Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Autom. Control. 19, 716-723.

[38] Bozdogan, H., 1987. Model selection and Akaike's informnation criterion (AIC): The general theory and its analytical extensions. Psychometrika. 52, 345-370.

[39] Schwarz, G., 1978. Estimating the dimension of a model. Ann. Statist. 6, 461-464.

[40] Van Erven, T., Harremos, P., 2014. Rényi divergence and Kullback-Leibler divergence. IEEE Trans. Inf. Theory. 60, 3797-3820.

[41] Danna, K., Griffin, R.W., 1999. Health and well-being in the workplace: a review and synthesis of the literature. J. Manag. 25, 357-384.

[42] Haslam, R.A., Hide, S.A., Gibb, A.G.F., Gyi, D.E., Pavitt, T., Atkinson, S., Duff, A.R., 2005. Contributing factors in construction accidents. Appl. Ergon. 36, 401-415.

[43] Depaire, B., Wets, G., Vanhoof, K., 2008. Traffic accident segmentation by means of latent class clustering. Accid. Anal. Prev. 40, 1257-1266.

[44] De Oña, J., López, G., Mujalli, R., Calo, F.J., 2013. Analysis of traffic accidents on rural highways using Latent Class Clustering and Bayesian Networks. Accid. Anal. Prev. 51, 1-10.

[45] Glover, F., Laguna, M., 1998. Tabu search. In: Du DZ., Pardalos P.M. (eds) Handbook of Combinatorial Optimization. Springer, Boston, MA.

[46] Longoni, A., Cagliano, R., 2015. Environmental and social sustainability priorities: Their integration in operations strategies. Int. J. Oper. Prod. Manage. 35, 216-245.

[47] Aven, T., 2016. Risk assessment and risk management: review of recent advances on their foundation. Eur. J. Oper. Res. 253, 1-13.