Proceedings of the 35th European Safety and Reliability & the 33rd Society for Risk Analysis Europe Conference
Edited by Eirik Bjorheim Abrahamsen, Terje Aven, Frederic Bouder, Roger Flage, Marja Ylönen

doi: 10.3850/978-981-94-3281-3_ESREL-SRA-E2025-P9431-cd

A safety model for industrial environment based on the Bayesian network paradigm

Jacek Malinowski

Systems Research Institute, Polish Academy of Sciences, Poland. E-mail: jacek.malinowski@ibspan.waw.pl

A safety model for identifying hazards and predicting their possible consequences in an industrial facility is proposed. Complex cause-effect relations between safety-related events, starting from the initial ones, through the intermediate, to the final, are represented in the form of a Bayesian network. The input data originate from sensors and meters installed in safety-critical locations. Using the Bayesian network methodology, the impending hazards, accidents, or machinery breakdowns can be predicted from symptoms indicated by the monitoring devices. Also, the "reverse analysis" of the network can establish the root causes of these undesired events so that a preemptive maintenance can be carried out in order to avoid them. For illustration, a simplified safety model of a biogas plant is presented along with its basic analysis. Although there is abundant literature on Bayesian networks in the safety and reliability context, much of it is limited to theoretical considerations or provides only general guidelines for the construction of such networks. Thus, publications reporting specific applications of this methodology are rather rare. The current paper aims to contribute to filling this gap.

Keywords: Workplace safety, Bayesian network, biogas plant, hazards identification and prediction, root cause analysis

1. Introduction

This paper proposes a method of analyzing the safety hazards in a biogas plant or any industrial facility, using a Bayesian network (BN) paradigm. Biogas plants are becoming increasingly used for the purpose of combined electricity and heat production. The share of fossil fuels (oil, natural gas, coal) in energy production gradually decreases, mainly due to environmental issues, but also for economic reasons such as increasing costs of their extraction, transportation and processing, and, last but not least, taxation of CO2 emissions in a number of countries. This results in switching to alternative energy sources, one of which is biogas obtained in the process of anaerobic digestion (fermentation) of biomass. The biogas production facilities or biogas plants are mainly built in rural areas, where biomass in large quantities is easily available and doesn't need to be transported over long distances. They serve as distributed energy generation units increasing the reliability of energy supply on local level, especially in the event of an outage in the main power grid. When connected to the central grid, biogas plants contribute to the country-wide energy production and supply system. Although undoubtedly an important energy source, they also pose some specific security threats, both to the personnel and resources of the facility.

The issue of biogas plant safety has been widely addressed in the subject literature, which can roughly be divided into three categories. Publications in the first category define various hazards specific to biogas plants and formulate safety measures for their construction and operation (Marrazzoa and Mazzini 2024, Scarponi et al. 2015). Those in the second describe possible accident scenarios or report hazardous or destructive events that have occurred in biogas facilities and analyze their root and intermediate causes (Hegazy et al. 2024, Moreno et al. 2016, Stolecka and Rusin 2021). The third category includes descriptions and analyses of biogas plants safety models aimed at estimating various statistics related to dangerous incidents, as well as determining their causes and preventing future occurrences (Moreno et al. 2018, Trávníček et al. 2018, Torretta et al. 2015, Lu et al. 2020). The latter is not concerned with a biogas plant, but gives an example of applying BN-based risk analysis in a similar environment. Since this paper does not aim to be a comprehensive literature review, only a few example works from each category are referenced.

The current paper can be classified into the third category. Its author attempts to build a comprehensive safety model of a biogas facility, depicting cause-effect chains of events leading to hazardous situations and destructive or catastrophic consequences. The model has the form of a BN illustrating direct dependencies between individual events (the events are represented as nodes of a direct acyclic graph whose edges connect directly dependent events). It is divided into two sub-models, one for the biogas production facility and the other for the upgrading unit. These sub-models along with guidelines for their analysis are presented in sections 3 and 4 respectively. They are preceded by Section 2 that recalls BN basics, introduces the concept of generalized BN and provides guidelines for its probabilistic cause-effect analysis. Section 5 contains concluding remarks and prospects for future research and is followed by the list of references.

2. Generalized Bayesian Networks and their Analysis

BNs are graphical representations of causal dependencies between multiple inter-related events occurring within a complex system (Scutari and Denis 2021, Kjaerluff and Madsen 2013, Koller and Friedman 2009). A "classical" BN is a direct acyclic graph (DAG) whose nodes represent random variables or events and links represent direct dependence relations between parent and child variables or events. A BN node can have multiple parent nodes (unlike a gate in a fault tree) and child nodes. Each node is assigned the conditional probability table (CPT) whose single entry stores the probability that the random variable represented by this node takes one of its possible values, conditioned by a combination of possible values of the parent variables. Clearly, the number of entries in a CPT is large if the variables have many possible values (the number of entries is equal to the product of the numbers of the parent and child variables' values). However, it may happen that a child variable is a (logical) function of its parent variables, defined by an algebraic formula. In such a case the (possibly large) CPT can be replaced by a (technically equivalent) equality that defines the conditional probabilities using that formula. Networks in which the CPTs of selected (or all) nodes can be represented as algebraically defined functions will be referred to as generalized BNs. Also, the CPTs or the respective functions in such networks can be time-related, as explained in Section 4. For better understanding, let us consider a simple BN shown in Fig. 1.

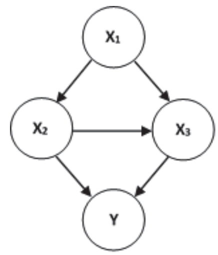

Fig. 1. A simple BN

Its nodes represent the following events: 1) fire, 2) activation of automatic sprinkler, 3) use of manual extinguisher and 4) fire quenched. X_1, X_2, X_3 and Y are the corresponding (binary) random variables. Table 1 is the CPT for the variable X_2.

Table 1. CPT for X_2

X_1	X_2	$\Pr(X_2=b \mid X_1=a)$
1	1	0.95
1	0	0.05
0	1	0
0	0	1

The above CPT converts to the following expression:

$$\Pr(X_2 = b \mid X_1 = a)$$
$$= a[0.95b + 0.05(1 - b)]$$
$$+(1 - a)(1 - b) \quad (1)$$

The CPTs of X_3 and Y are defined by the following formulas:

$$\Pr(X_3 = c \mid X_1 = a, X_2 = b)$$
$$= a(1 - b)[0.98c + 0.02(1 - c)]$$
$$+[1 - a(1 - b)](1 - c) \quad (2)$$

and

$$\Pr(Y = d \mid X_2 = b, X_3 = c)$$
$$= (b \vee c)d + [1 - (b \vee c)](1 - d) \quad (3)$$

where $a,b,c,d \in \{0,1\}$. As follows from Eq. (2), in the case of fire and failed sprinkler, manual extinguisher is used with probability 0.98, or is not used with probability 0.02. If there is no fire or the sprinkler is activated, the extinguisher is not used. Further, according to Eq. (3), Y depends deterministically on X_2 and X_3, i.e. the probabilities in its CPT are either zeroes or ones. Such CPTs can be called as deterministic and those with arbitrary probabilities - as stochastic.

Eqs. (1)-(3) can be generalized to multivalued variables and other scenarios. Considering Eqs. (1) and (2), we conclude that not only deterministic dependence between a node and its parents can be expressed algebraically, but also stochastic one. However, it should be pointed out that only CPTs exhibiting a regular pattern can be defined by simple algebraic formulas, hence not all CPTs can be thus represented.

CPTs or functions defining the conditional probabilities are necessary for quantitative probabilistic analysis of BNs, which involves computing probabilities of resulting events given the causing events or those of causing events given the resulting ones. These probabilities are essential for the cause-effect analysis of the events occurring in the system modeled by a BN.

For illustration, let us compute the probabilities $\Pr(Y=d|X_1=a)$ and $\Pr(X_1=a|Y=d)$. We have:

$$\Pr(Y = d|X_1 = a) = \frac{\Pr(Y=d, X_1=a)}{\Pr(X_1=a)}$$

$$= \sum_{b,c} \frac{\Pr(Y=d, X_2=b, X_3=c, X_1=a)}{\Pr(X_1=a)}$$

$$= \sum_{b,c} \begin{array}{c} \Pr(Y = d \mid X_3 = c,\ X_2 = b,\ X_1 = a) \times \\ \Pr(X_3 = c \mid X_2 = b,\ X_1 = a) \times \\ \Pr(X_2 = b \mid X_1 = a) \end{array}$$

$$= \sum_{b,c} \begin{array}{c} \Pr(Y = d|X_3 = c,\ X_2 = b) \times \\ \Pr(X_3 = c|X_2 = b,\ X_1 = a) \times \\ \Pr(X_2 = b|X_1 = a) \end{array} \quad (4)$$

where $\Sigma_{b,c}$ denotes the sum over all values of (X_2, X_3). The last equality follows from one of the basic properties of BNs, i.e. the probability that a variable takes a specific value is determined by the values of its parents. Thus, to compute $\Pr(Y=d|X_1=a)$, we need Eq. (4) and CPTs of X_2, X_3 and Y defined by Eqs. (1)-(3). Note that, proceeding as in Eq. (4), we can derive a recursive formula for the probability $\Pr(Y=b \mid X=a)$, where Y is a non-root variable of an arbitrary BN and X is one of root variables located above Y. This formula will involve all combinations of Y's parent's values and the respective entries from the Y's CPT (possibly algebraically defined). In general, Eq. (4) adjusted to a particular BN can be used to compute the probabilities $\Pr(Y_1=b_1,\dots,Y_k=b_k \mid X_1=a_1,\dots,X_j=a_j)$, where $Y_1,\dots,Y_k$ are non-root variables and $X_1,\dots,X_j$ are root ones. Such probabilities are useful for detailed analysis of a BN.

Now let us compute the "reverse" probability $\Pr(X_1=a|Y=d)$. We have:

$$\Pr(X_1 = a|Y = d) = \frac{\Pr(Y=d|X_1=a)\Pr(X_1=a)}{\Pr(Y=d)}$$

$$= \frac{\Pr(Y=d|X_1=a)\Pr(X_1=a)}{\sum_x \Pr(Y=d|X_1=x)\Pr(X_1=x)} \quad (5)$$

where Σ_x denotes the sum over all values of X_1. Thus, to compute $\Pr(X_1=a \mid Y=d)$, we need Eq. (4) and the probabilities $\Pr(X_1=a)$ for all values of X_1. These probabilities are assumed to be given.

Eqs. (4) and (5) demonstrate the computing technique to find the conditional "forward" and "reverse" probabilities in a generalized BN. The presented approach eliminates the necessity to store a large CPT if it admits an algebraic representation. This saves memory and accelerates the computations. Though there exists abundant literature on multi-aspect analysis of BNs (Fan et al. 2024, Kitson et al. 2023, Darwiche 2022, Yu et al. 2020, Ding 2010), the author could not find a literature source where an explicit algorithm for computing the above considered probabilities is presented. Such an algorithm, following the above demonstrated technique, will be a topic of the author's further research.

The nodes of a generalized BN bear similarity to the gates of a fault tree, but a BN node can have multiple outputs and its CPTs can be defined by a wide variety of functions, while a fault tree gate has one output and only a few logical functions are used to convert the inputs of a gate to its output (e.g. OR, AND, XOR, k-out-of-n).

Beside generalized BNs, there are also other extensions of classical BNs (e.g. dynamic or temporal BNs). They are described in Sucar 2001.

3. Safety analysis of a biogas production unit

The main component of a biogas production unit is the digester. It is a closed airtight tank where the biogas is obtained from biomass in the process of anaerobic digestion (fermentation), i.e. biomass is decomposed by anaerobic bacteria into gas mixture (methane, carbon dioxide and hydrogen sulfide) and digestate (liquid and solid residue that can be used as a fertilizer). There are four basic digester types – covered lagoon, complete mix, plug flow and fixed film (see Uddin & Wright, 2023). We will analyze here a digester of the complete mix type, since such a digester is used in a biogas plant whose supervising engineer has been consulted in the course of the II.PN.05 project, which provided support for this research.

consequences. This model may not be complete, but it has been designed so that the network structure does not need to be changed when the model is extended by adding new events and links.

As follows from the character of the specified events and their interdependencies, the respective variables are binary and their CPTs are defined using logical functions. For example, the CPT of Y↔{Serious injury or fatality} (i.e. Y=1 if the event in braces occurs, else Y=0) is defined using the function $f(a,b,c,d) = (a \vee b \vee c) \wedge d$, where a, b, c and d are binary values of the following variables:
X_1↔{External explosion or fire},
X_2↔{Digester explosion},
X_3↔{Digester fire} and
X_4↔{Personnel in the vicinity}.

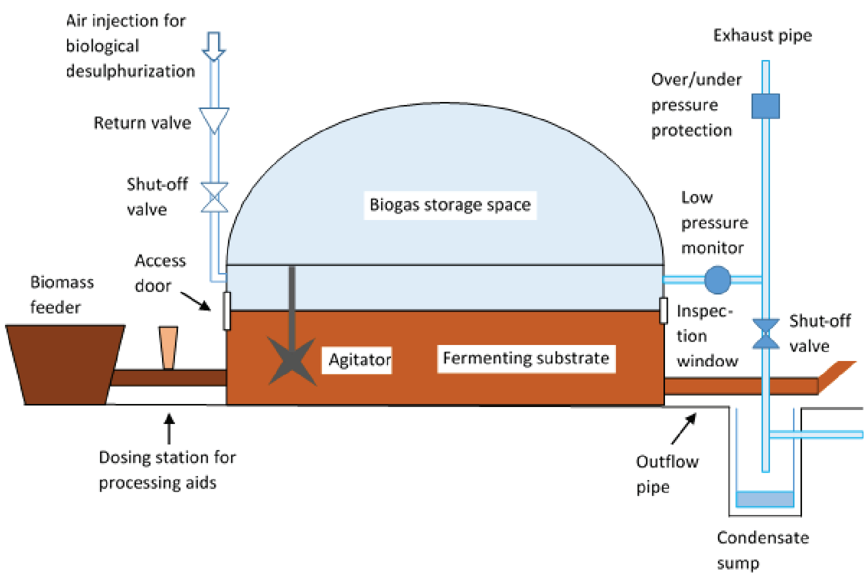

Fig. 2. Diagram of a biogas production unit

A schematic diagram of a biogas production system based on a complete mix digester is shown in Fig. 2. For lack of space, the heat exchanger providing the right temperature for the fermentation process, supplied from the combined heat and power unit located in the upgrading station, is not shown, but its presence should be taken into regard.

Fig. 3 shows the structure of a BN that illustrates the dependencies between various hazardous events that can occur in or around a digester and their damaging or destructive

If the CPT of Y is deterministic (see the definition in Section 2), then

$$\Pr(Y = e \mid X_1 = a, X_2 = b, X_3 = c, X_3 = d)$$

$$= f(a,b,c,d)e + [1 - f(a,b,c,d)](1 - e) \quad (6)$$

If all the CPTs are assumed deterministic (a simplified model), then for the probabilistic analysis of the network we only need probabilities of the initial events, i.e. those having no parents. A model with stochastic CPTs would require learning the respective probabilities from statistical data.

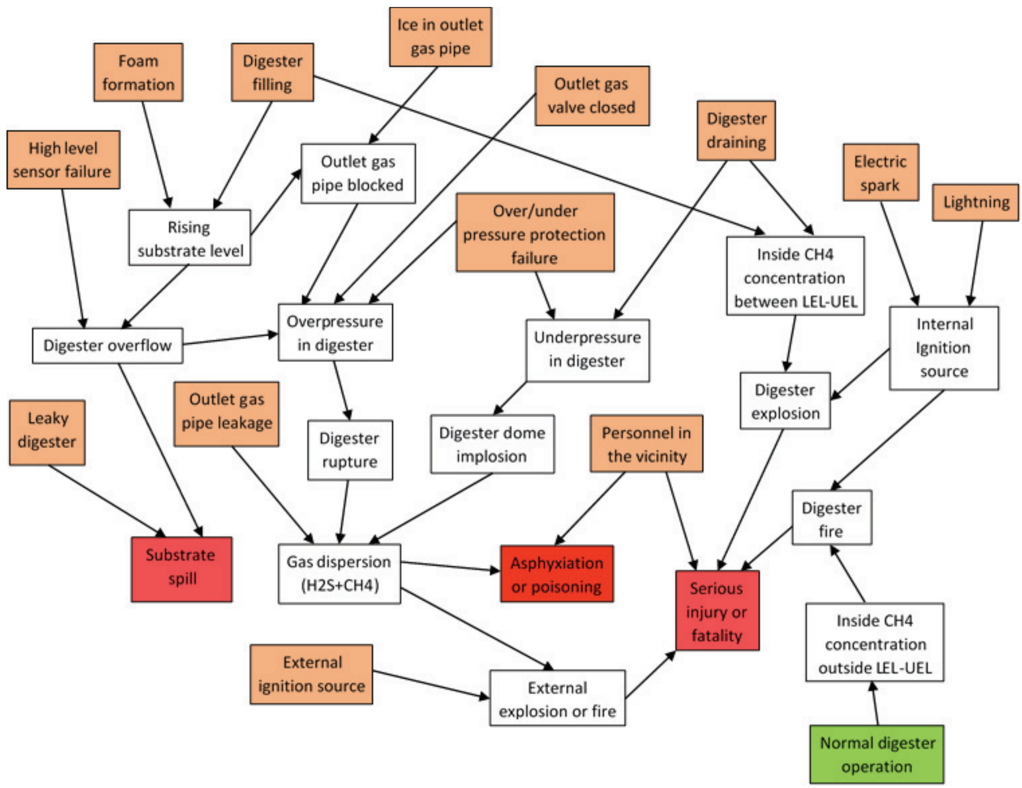

Fig. 3. Structure of a BN for the biogas production unit

A detailed network analysis should take into account that some events have definite duration (e.g. digester filling or draining), the duration of others is finite but indefinite (e.g. icy outlet pipe, valve closed, sensor failure), while others are instantaneous (e.g. electric spark or lightning).

Handling of such events, referred to as temporal or time-related, is the subject of (Mrowca et al. 2022) – its results can be helpful in the analysis of the considered system. In the next section it is shown how to define probabilities associated with time-related events.

4. Safety Analysis of the upgrading Station

The biogas produced in the digester is a mixture of methane (CH4), carbon dioxide (CO2), hydrogen sulfide (H2S), nitrogen (N2), hydrogen (H2) and oxygen (O2). It also contains some water vapor (H2O). The first two components make up approximately 95% of its composition.

Before being fed into a gas network, biogas undergoes a purification process to extract methane and remove other components. This process takes place in the upgrading station whose diagram is shown in Fig. 4. In the first processing stage, H2O and H2S (highly corrosive agents) are removed (H2O by chilling and H2S in the carbon filter). Next, the compressor pumps biogas to the upgrading unit whose task is to remove CO2 and other gases to obtain biomethane, i.e. almost pure CH4. In the conditioning unit the biomethane is odorized (for safety reasons) and its calorific value is increased (e.g. by addition of propane). The injection unit increases the gas pressure in order to feed it into the gas grid. Gas flare is necessary to burn the surplus biogas that cannot be accommodated by the grid (lower recipient demand or machinery failure). Flame arrester prevents flashbacks that can occur as a result of pressure fluctuations in the pipework (flame moving upstream may cause damage or explosion). For a detailed description of the biogas upgrading process see Galloni and Di Marcoberardino 2024.

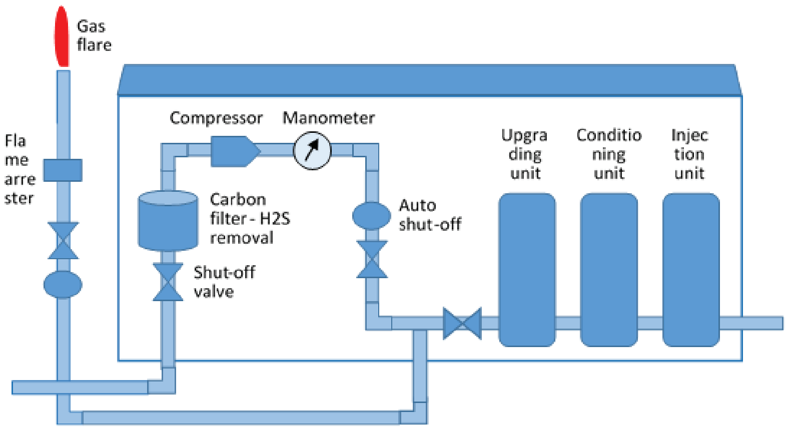

Fig. 4. Diagram of the upgrading station

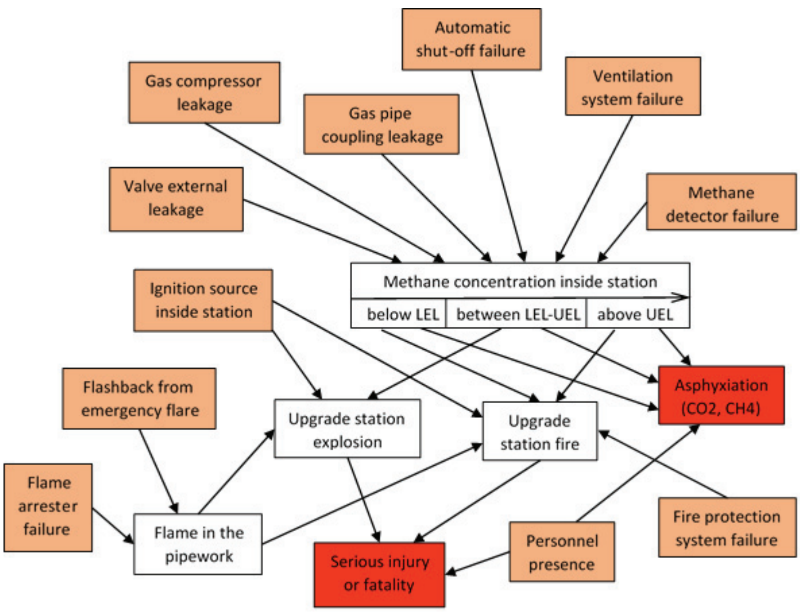

Fig. 5. Structure of a BN for the upgrading station

The structure of a Bayes network for the analysis of safety issues specific to a biogas upgrading station is presented in Fig. 5. The remarks concerning the CPTs and types of events are the same as in the previous section. For example, the CPT of Y↔{CH4 concentration inside station is between LEL-UEL} is defined using the function $f(a,b,c,d,e,h) = (a \vee b \vee c) \wedge d \wedge (e \vee h)$ where a, b, c, d, e and h are binary values of the following variables:

X_1↔{Valve leakage},
X_2↔{Compressor leakage},
X_3↔{Pipe coupling leakage},
X_4↔{Ventilation failure},
X_5↔{CH4 detector failure} and
X_6↔{Automatic shut-off failure}

Note that {CH4 concentration below LEL}, {CH4 concentration between LEL-UEL} and {CH4 concentration above LEL-UEL} are time-related events, i.e. if a leakage occurs, CH4 concentration is initially below LEL, then between LEL-UEL, and afterwards above UEL.

This is indicated by the time axis in the respective node. Since the time points at which the explosive limits are reached cannot be accurately predicted, only the chronological order of the above events is presented under the time axis (similarly as in SEQ gate in a dynamic fault tree). The time can be counted from the beginning of the earlier leakage event. The probabilities of the initial events should also be time-related and expressed as $p_i(t)$=Pr[X_i(t)=1] for the lasting events or as $p_i(s,t)$=Pr(event #*i* occurs in the time interval [s,t]) for the instantaneous events (flashback, spark). The event numbers are assigned according to the adopted numbering scheme. Some events can be merged to a single multiple-stage event with a common set of inputs and different sets of outputs for each stage (see the event "CH4 concentration inside station …" in Fig. 5). Each stage has its own CPT defined for selected inputs (not necessarily all). Such an event is in fact a random point process with a finite state space (state-to-state transitions occur at random points in time.

It should be noted that the probabilities of initial events and stochastic CPTs are not needed for the purely qualitative cause-effect analysis, since they are only necessary for the quantitative probabilistic one. This is because the qualitative analysis only identifies possible causes or effects of particular events without computing the associated (conditional) probabilities, as demonstrated by Eqs. (4)-(5).

5. Conclusion and Future Work

In this paper a method of safety analysis of an industrial facility, based on the Bayesian network paradigm, is proposed. The author introduces the concept of a generalized BN, whose selected or all CPTs are defined by algebraic formulas. The events represented by the nodes of such a network can be time-related, which means that the associated probabilities and CPTs are time-dependent, or, in general, they are be multiple-stage events viewed as random point processes. In Section 2 the guidelines for a generalized BN analysis, using Eqs. (4)-(5), are formulated. In the following sections two example networks are presented that illustrate the dependencies between various safety-related events occurring in two main parts of a biogas plant.

Safety models based on generalized BNs are expected to be more concise than those applying fault trees, Petri nets or "classical" BNs with extensive CPTs. Also, the quantitative probabilistic analysis of such BNs can be numerically less complex compared to the above mentioned models. The author's future work will include constructing more detailed BN safety models of industrial facilities (biogas plants in particular) and developing algorithms for their cause-effect analysis (both qualitative and quantitative).

Acknowledgement

This work was supported by the project II.PN.05 funded by the National Center for Research and Development, an executive agency of the Polish Ministry of Science and Higher Education.

References

Darwiche, A. (2022). *Tractable Boolean and Arithmetic Circuits*. arXiv: 2202.02942 https://arxiv.org/abs/2202.02942

Ding, J. (2010). *Probabilistic Inferences in Bayesian Networks*. arXiv: 1011.0935 https://doi.org/10.48550/arXiv.1011.0935

Fan, Z., L. Zhou, T.E. Komolafe, Z. Ren, Y. Tong and X. Feng (2024). Learning Bayesian network parameters from limited data by integrating entropy and monotonicity. Knowledge-Based Systems 291, 111568.

Galloni, M. and G. Di Marcoberardino (2024). Biogas Upgrading Technology: Conventional Processes and Emerging Solutions Analysis. *Energies 17*(*12*), 2907. https://doi.org/10.3390/en17122907

Hegazy, H., N.M.C. Saady, F. Khan, S. Zendehboudi, and T.M. Albayati (2024). Biogas plants accidents: Analyzing occurrence, severity, and associations between 1990 and 2023. *Safety Science 177*, 106597. https://doi.org/10.1016/j.ssci.2024.106597.

Kitson, N.K., A.C. Constantinou, Z. Guo, Y. Liu and K. Chobtham (2023). A survey of Bayesian Network structure learning. *Artificial Intelligence Review 56*, 8721-8814.

Kjaerluff, U.B. and A.L. Madsen (2013). *Bayesian Networks and Influence Diagrams: A Guide to Construction and Analysis. Second Edition*. Springer.

Koller, D. and N. Friedman (2009). *Probabilistic Graphical Models. Principles and Techniques*. The MIT Press.

Lu, Y., T. Wang, and T. Liu (2020). Bayesian Network-Based Risk Analysis of Chemical Plant Explosion Accidents. *International Journal of Environmental Research and Public Health 17*, 5364.

Marrazzoa, R. and C. Mazzini (2024). Safety Considerations and Major Accidents Prevention in the Biogas Production. *Chemical Engineering Transactions 109*, 451-456.

Moreno, V.C., D. Guglielmi, V. Cozzani (2018). Identification of critical safety barriers in biogas facilities. *Reliability Engineering and System Safety 169*, 81-94.

Moreno, V.C., S. Papasidero, G.E. Scarponi, D. Guglielmi, and V. Cozzani (2016). Analysis of accidents in biogas production and upgrading. *Renewable Energy 36, Part B*, 1127-1134.

Mrowca, A., F. Gyrock, and S. Günnemann (2022). Temporal state change Bayesian networks for modeling of evolving multivariate state sequences: model, structure discovery and parameter estimation. *Data Mining and Knowledge Discovery 36*, 240–294. https://doi.org/10.1007/s10618-021-00807-y

Scarponi, G.E., D. *Guglielmi, V.C. Moreno, and V. Cozzani (2015).* Risk Assessment of a Biogas Production and Upgrading Plant. *Chemical Engineering transactions 43*, 1921-1926.

Scutari, M. and J.-M. Denis (2021). *Bayesian Networks with Examples in R*. Chapman and Hall/CRC.

Stolecka, K. and A. Rusin (2021). Potential hazards posed by biogas plants. *Renewable and Sustainable Energy Reviews 135(C)*, 110225.

Sucar, L.E. (2021). *Probabilistic Graphical Models. Principles and Applications*. Springer Cham.

Torretta, V., S. Copelli, S. Contini, E.C. Rada (2015). Safety and Reliability in biogas plants. *WIT Transactions on the Built Environment. 151*, 227-238.

Trávníček, P., L. Kotek, P. Junga, T. Vítěz, K. Drápela, and J. Chovane (2018). Quantitative analyses of biogas plant accidents in Europe. *Renewable Energy 122*, 89-97.

Uddin, M.M and M.M. Wright (2023). Anaerobic digestion fundamentals, challenges, and technological advances. *Physical Sciences Reviews 8*, 2819-2837. https://doi.org/10.1515/psr-2021-0068

Yu, H., J. Moharil and R.H. Blair (2020). BayesNetBP: An R Package for Probabilistic Reasoning in Bayesian Networks. Journal of Statistical Software, 94, 1–31. https://doi.org/10.18637/jss.v094.i03