Uncertainties in conditional probability tables of discrete Bayesian Belief Networks: A comprehensive review

Jeremy Rohmer

Uncertainties in conditional probability tables of discrete Bayesian Belief Networks: a comprehensive review

Jeremy Rohmer[1]

[1]{BRGM, 3 av. C. Guillemin - 45060 Orléans Cedex 2 - France}
Correspondence to: J. Rohmer (j.rohmer@brgm.fr)

Abstract

Discrete Bayesian Belief Network (BBN) has become a popular method for the analysis of complex systems in various domains of application. One of its pillar is the specification of the parameters of the probabilistic dependence model (i.e. the cause-effect relation) represented via a Conditional Probability Table (CPT). Depending on the available data (observations, prior knowledge, expert-based information, etc.), CPTs can be populated in different manners, i.e. different assumptions can be made and different methods are available, which might lead to uncertain BBN-based results. Through an extensive review study of the past ten years, we aim at addressing three questions related to the CPT uncertainties. First, we show how to constrain these uncertainties either using elicitation of expert inputs, or using a combination of scarce data and expert-derived information. Second, we show how to integrate these uncertainties in the BBN-based analysis through propagation procedures either using probabilities or imprecise probabilities within the setting of credal or evidential networks. Finally, we show how to test the robustness of the BBN-based results to these uncertainties via sensitivity analysis specifically dedicated to BBNs. A special care was paid to describe the best practices for the implementation of the reviewed methods and the remaining gaps.

Keywords: Bayesian Belief Network; Conditional Probability Table; Expert Elicitation; Sensitivity Analysis; Credal Network; Evidential Network.

1 Introduction

Bayesian Belief Network (BBN) has become an increasingly popular method for the analysis of complex systems in various domains of application, like ecosystems (Milns et al., 2010), genetics and biology (Scutari et al., 2014), agriculture (Drury et al., 2017), industry (Weber et al., 2012), finance forecasting (Malagrino et al., 2018), marine safety (Hänninen et al., 2014), human reliability assessment (Mkrtchyan et al., 2015), nuclear power plants (Kwag and Gupta, 2017), aviation risk analysis (Brooker, 2011), coastal systems (Jäger et al., 2018), structure reliability assessments (Langseth and Portinale, 2007), multi-hazard risk assessments (Gehl and D'Ayala, 2016), etc.

Its benefits are: (1) its high flexibility to model any causal relationships; (2) its capability to integrate information from any kind of sources, including experimental data, historical data, and prior expert opinion, and (3) its capability to answer probabilistic queries about them and to find out updated knowledge of the state of a subset of variables when other variables (i.e. the evidence variables) are observed.

Formally, a Bayesian belief Network (BBN) is a class of graphical model (see Jensen, 2001 for a complete and detailed introduction to BBNs), which allows to synthetically represent relations among random variables by means of a directed acyclic graph (DAG) composed of nodes (i.e. the states of the random variables) and arcs (i.e. dependency between nodes). The value of the nodes may be discrete or continuous, and we focus here on the former case, which is the most widely used. For instance, a Boolean node representing the state of a system component can be either "True" or "False". The nodes connected by an arc are called the parent nodes and child nodes respectively. One child node may have several parent nodes, meaning that this node is affected by several factors. Similarly, a parent node could have several child nodes, meaning that this factor may have influences on several other factors. Conditional probabilities are the probabilities that reflect the degree of influence of the parent nodes on the child node. For BBNs with discrete nodes, the probabilistic dependence (i.e. the cause-effect relation) is often represented via a table called a Conditional Probability Table (CPT).

As an illustration, Fig. 1 depicts the binary BBN adapted by van der Gaag et al. (2013) from Cooper (1984) in the field of oncology. The network is composed of 6 nodes and 6 arcs. Node MC refers to metastatic cancer, which may potentially lead to the development of a brain tumor (node B) and may give rise to an increased level of serum calcium (node ISC). The presence of a brain tumour can be established from a CT scan (node CT). Another indicator of the presence of a brain tumour can be related to severe headaches (node SH). A brain tumour or an increased level of serum calcium are both likely to cause a patient to fall into a coma (the node C is connected to node B and node ISC). The conditional probabilistic relationships between the nodes (CPT entries) are provided in Fig. 1 next to the corresponding nodes. For instance, the probability that a patient falls into coma given brain tumor and increased level of serum calcium corresponds to the first entry of the table (1st row, 1st column), namely P(C=True|C=True,ISC=True)=0.80.

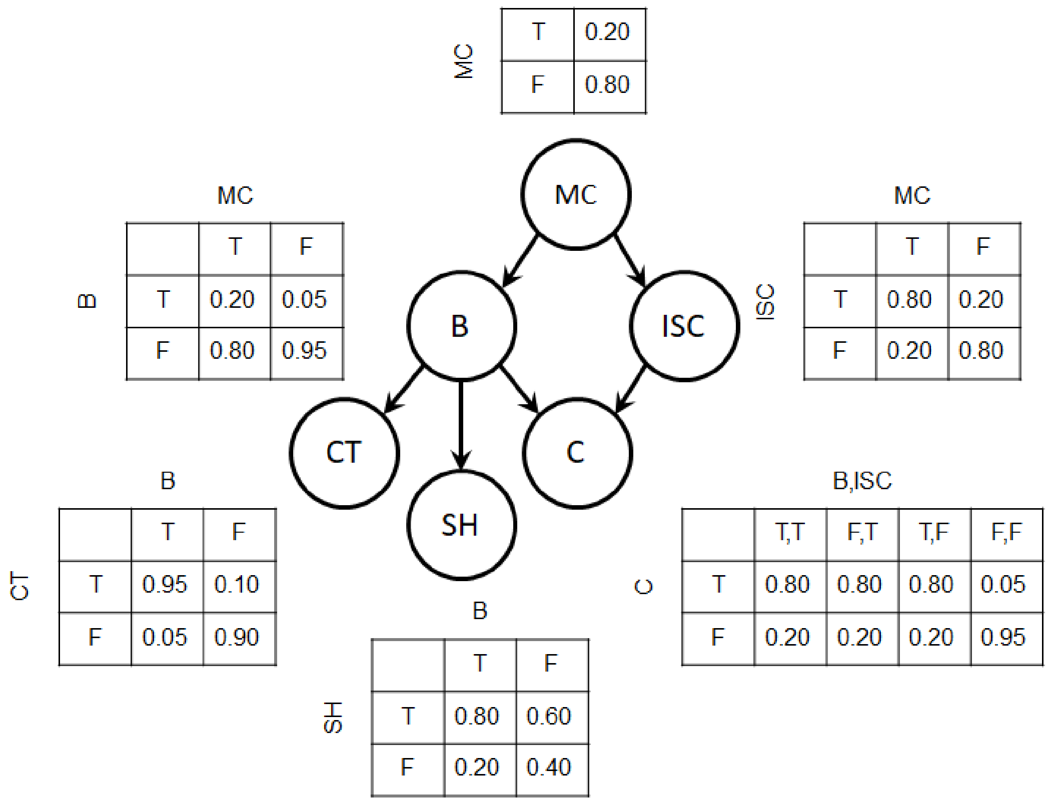

Figure 1. Binary BBN adapted by van der Gaag et al. (2013) from Cooper (1984) in the field of oncology. The tables (called CPT) next to the nodes provide the conditional probabilities values.

[*Figure 1 about here*]

Two key ingredients are necessary to build a BBN, namely (1) the graph structure with the direction of the arcs, i.e. the DAG; (2) the states of nodes and the strength of the relationships between nodes, i.e. the CPT. In the present study, we assume that the DAG model has already been determined and restrict the analysis to the quantification of the BBN relationships. The process of deriving the CPTs and its associated uncertainties is recognized in the literature as one of the most delicate part of the BBN development (e.g., Chen and Pollino, 2012; Druzdzel and van der Gaag, 2000; Marcot et al., 2006; Cain, 2001, etc.). It should, however, be noted that the process of DAG derivation (i.e. building the graph structure plus the directions; also known as causal structure learning) has its own challenges as well, in particular when the learning is based on data (see e.g., a comprehensive review by Heinze-Deml et al., 2018).

Depending on the available data (observations, prior knowledge, expert-based information, etc.), CPTs can be evaluated in different manners, i.e. different assumptions can be made and different methods are available leading to different BBN-based results, hence resulting in uncertain BBN-based results. This raises the following questions: (1) how to constrain the uncertainties related to CPT derivation, i.e. what are the methods that are available to minimize these uncertainties? (2) how to integrate these uncertainties in the BBN-based analysis, i.e. what are the methods for propagating these uncertainties? (3) how to test the robustness of the BBN-based results to these uncertainties, i.e. what are the methods for identifying the most influential uncertainties? These questions are addressed below through an extensive review of studies performed in the past ten years by focusing on discrete BBNs that can be used for modelling complex causal relationships, for merging different information sources, for prediction, and for belief/evidence propagation (i.e. probabilistic queries). Continuous BBNs (i.e. BBNs with continuous nodes) and dynamic BBNs (i.e. BBNs adapted to model systems evolving over time) are out of the scope of the review.

The paper is organized as follows. The first section describes more specifically the problem of populating the CPT parameters directly from data/observations. This first part highlights the necessity for overcoming the lack of data by complementing with additional sources of information. Sect. 3 explores an expert-based option for constraining the uncertainties related to data incompleteness, namely by completing with different expert-based sources of information. Sect. 4 provides an overview of the different approaches embedded in different uncertainty analysis settings for evaluating the impacts of CPT uncertainties, either using probabilities or imprecise probabilities. Sect. 5 further addresses the problem of screening these uncertainties by describing sensitivity analysis techniques. Finally, Sect. 6 summarizes the main findings and discusses the open questions.

2 Learning CPT from data

In this section, we address the issues related to deriving the CPT entries from data. Sect. 2.1 first discusses the problem of performing this task by using only data. Sect 2.2 and 2.3 further discuss two practical difficulties, namely: (1) the presence of missing values and (2) the problem of translating observations related to continuous variables into a limited number of discrete states. Finally, Sect. 2.4 describes methods that make the most out of scarce data while exploiting qualitative information provided by experts.

2.1 A pure data-driven approach

Let us consider a BBN composed of n discrete nodes $X_{i=1,\ldots,n}$. Let us denote r_i the cardinality of X_i and q_i the one of the parent set of X_i, denoted pa(X_i). The kth probability value of the conditional probability distribution is $\theta_{ijk} = P(X_i = k|pa(X_i) = j)$ where i=1,…,n; j=1,….,q_i; k=1,…., r_i.

In data rich contexts, CPT parameters can be evaluated by computing the appropriate frequencies from data. An example is provided by Chojnacki et al. (2019) for fire safety analysis where more than 1 million of numerical simulation results are used. This method corresponds to the maximum likelihood estimation (MLE), which is described below.

Let us consider a dataset D where a total number N_{ij} of data records are available for which pa(X_i) is in the state j and where N_{ijk} data records are available for which X_i is in the state k and pa(X_i) is in the state j. MLE aims at maximizing the log-likelihood function l(.) of θ given D as follows

$$l(\theta|D) = \log\big(P(D|\theta)\big) = \sum_{ijk} N_{ijk} log\,(\theta_{ijk}) \qquad \text{(Eq. 1)}$$

The solution is then $\frac{N_{ijk}}{N_{ij}}$.

The MLE method however fails to find good estimates due to data scarcity when $N_{ij}\approx 0$, i.e. when training data are not sufficient in number in some specific variable state configurations. Examples of such contexts are not rare in practice; see e.g. rare disease diagnostic (Seixas et al., 2014), accident prevention (e.g., Hänninen, 2014), reliability analysis (e.g., Musharraf et al., 2014), etc. This problem is even worsened when the number of nodes increases. Recall that the number of conditional probabilities is exponential with the number of its parent nodes, i.e. for a node with i states and k parent nodes and if each parent node has n states, (i-1)×n^k CPT entry values have to be specified. For instance, a binary node with 2 binary parent nodes imposes to specify 4 entries, whereas for a ternary node with 2 ternary nodes, this number reaches 18.

2.2 Dealing with missing values

The process for parameter learning of discrete BBNs may be complicated in the presence of missing values. This can be handled by means of different algorithms. The most popular ones are Expectation Maximization (Dempster et al., 1977) and Gibbs sampling (Geman and Geman, 1984). Yet, they both assume that the values are missing at random. This hypothesis may not always be true in practice. Alternative methods have been proposed to overcome this disadvantage, like AI&M procedure (Jaeger, 2006), the RBE algorithm (Ramoni and Sebastiani, 2001), and the maximum entropy method (Cowell, 1999). Other methods have also been developed to speed up the learning process, like generalized conjugate gradient algorithm by Thiesson (1995) or the online updating of rules (Bauer et al., 1997). To deal with both missing data and qualitative influences (as described in Sect. 2.4), some initiatives have been proposed like the one of Masegosa et al. (2016), who further improved the combined Isotonic Regression - EM approach.

2.3 Discretising continuous variables

A second practical difficulty for parameter learning of discrete BBNs is inherent to the main assumption introduced by discrete BBNs, namely that data should be represented by a limited numbers of outcomes. This imposes to discretize continuous variables. This process might, however, lead to a loss of information, and potentially to an increase of the associated computational effort, because the size of discrete BBNs increases approximately exponentially with the number of discrete states of its nodes. Nojavan et al. (2017) investigated the implications of several mathematical methods for constructing discrete distributions in an unsupervised manner. Using a simple 3-node BBN describing chlorophyll concentrations in Finnish lakes, the authors evaluated the impact on the developed BBNs of the number of intervals and of the choice of the type of discretization methods. Three techniques were investigated, namely in which the data are divided into groups: (1) of equal length; (2) of equal sample size; (3) for which the moments of the discretized distribution match with the moments of the continuous data. They showed that none of the models did uniformly well in all comparison criteria (sum of squared errors, accuracy, area under the receiving operating characteristic curve) for the considered case. They concluded that they cannot justify using one discretization method against others. Using a 4-node BBN from the domain of coastal erosion, Beuzen et al. (2018) extended the tests to other types of discretization methods, namely manual and supervised techniques. They showed, on their specific test case, that supervised methods led to a BBN of the highest average predictive skill, followed by the one with manual discretization. They also outlined the advantages of the different methods, namely that:

- Manual methods allow ensuring physical meaningful BBNs;
- Supervised methods can autonomously and optimally discretize variables and may be preferred when predictive skill is a modelling priority;
- Unsupervised methods are computationally simple and versatile.

Depending on the objective, some specific discretization algorithms have also been developed; for instance, Zwirglmaier and Straub (2016) developed specific methods to deal with rare events in reliability analysis; Neil et al. (2007) proposed a dynamic discretization method to perform inference in hybrid BBNs, i.e. both dealing with continuous and discrete variables.

2.4 Combining scarce data and expert judgements

When data are scarce, the parameter learning may be improved by incorporating additional information provided by experts. A popular approach relies on the Maximum a Posteriori (MAP) estimation using Dirichlet priors, which express experts' belief (e.g., Heckerman et al., 1995) about θ in the absence of data. Formally, the Dirichlet distribution for CPT column θ_{ij} is expressed as follows:

$$p(\theta_{ij}) = \frac{1}{Z_{ij}} \prod_{k=1}^{r_i} \theta_{ijk}{}^{(\alpha_{ijk}+1)-1} \qquad \text{(Eq. 2)}$$

with $\sum_k \theta_{ijk} = 1$, $\theta_{ijk} \geq 0$, Z_{ij} is a normalisation term $\int_{-\infty}^{+\infty} \prod_{k=1}^{r_i} \theta_{ijk}{}^{(\alpha_{ijk}+1)-1} d\theta_{ijk} = 1$, and α_{ijk} is the parameter of the Dirichlet distribution, which can be intuitively interpreted as "how many times the expert believes he/she will observe X_i=k in a sample of α_{ij} instances drawn independently at random from the distribution θ_{ij}" (Zhou et al., 2014). On this basis, MAP relies on the following equation:

$$p(\theta|D) \propto P(D|\theta)P(D) \propto \prod_{ijk} \theta_{ijk}{}^{(\alpha_{ijk}+N_{ijk})-1} \qquad \text{(Eq.3)}$$

This equation results in the estimate of θ_{ijk} as $\frac{N_{ijk}+\alpha_{ijk}-1}{N_{ij}+\alpha_{ij}-1}$, which combines information from the data and from the experts' prior guess. In their computer experiments using twelve publicly available BBNs (available at http://www.bnlearn.com/bnrepository/), Zhou et al. (2016a) showed that MAP achieves better performances than conventional MLE, which suffers from the absence of data in several state configurations in situations of limited sample size (typically 50).

Expert-based information can take several forms, and the one that corresponds to qualitative constraints have given rise to several developments. Instead of directly providing the exact value of the entries of binary BBN (denoted P_{1-2}), the expert may feel more conformable in providing an ordering like "$P_1>P_2$", "$P_1 \approx P_2$", "$P_1>0.80$", etc. Zhou et al. (2016a) showed that incorporating such expert knowledge about the monotonic influences between nodes (translated into probability constraints) further outperformed MAP and MLE and was also robust to errors in labelling the monotonic influences.

Different methods have been developped to incorporate qualitative constraints, namely:

- Convex Optimization (Niculescu et al., 2006; Zhou et al., 2016a; de Campos and Ji, 2008; Liao and Ji, 2009; Altendorf et al., 2005) is an extension of the MLE by incorporating constraints via penalty functions or by restricting parameter spaces;
- Constrained MAP approach has also been proposed by Yang et al. (2019) to learn BN parameters by incorporating convex constraints;
- Isotonic Regression (Feelders and van der Gaag, 2005; 2006) builds on qualitative information about the influences between the variables of a BBN. The most recent algorithm by Masegosa et al. (2016) also enables the analysist to learn the CPT parameters from incomplete data;
- Qualitative MAP (originally proposed by Chang and Wang (2010) and further improved by Guo et al. (2017)) constructs Dirichlet priors from Monte-Carlo random samples of the constrained parameter space, which are used by the MAP algorithm;
- Multinomial Parameter Learning with Constraints (Zhou et al., 2014; Hospedales et al., 2015) rely on auxiliary BBNs, which are hybrid BBNs, to infer the posterior distribution of BBN parameters.

2.5 Discussion

Following a pure statistical data-driven approach for populating the BBN conditional model requires a large amount of statistically significant data to cover all BBN relationships. To compensate the lack of data, a possible option is to complement the analysis with expert-based information. Sect. 2.4 shows that a broad range of different tools/methods are available to incorporate expert-based information either in the form of qualitative influences or constraints, namely constraints that should be almost linear and convex (i.e. concave constraints like $P_1 \neq 0.5$ cannot be accounted for). The improvement of the learning accuracy of the parameters in BBNs from a small data set has been shown using each of the described methods compared to conventional methods; for instance Guo et al. (2017) compared MLE, constrained MLE, maximum entropy and constrained maximum entropy estimator, MAP and their qualitatively MAP estimator. Yang et al. (2019) showed the higher performance of their constrained MAP estimator compared to conventional parameter learning algorithms, MLE and MAP, and to constrained maximum likelihood algorithm. Yet, to the author's best knowledge, no extensive benchmark exercise covering all the afore-mentioned estimators (as well as their pros and cons) is available yet; practical recommendations on how to implement them and their limitations is currently lacking in the literature.

Among the possible limitations, the problem of under-fitting related to the use of prior distributions (that are common ingredients of most of the methods of Sect. 2.4) is seldom tackled. As described by Gao et al. (2019), imposing certain a priori knowledge on the CPT parameters might decrease the likelihood of the parameters, hence a reduction of the fitness between parameters and data. Azzimonti et al. (2019) proposed a hierarchical procedure to improve the widely-used approach based on Dirichlet priors. Gao et al. (2019) proposed a Minimax Fitness algorithm combined with an improved constrained maximum entropy method to overcome this problem. They also concluded that there is a need for further investigation to develop learning methods that does not require specification of prior strength.

3 Learning from experts

In many situation, the primary source of information for learning the CPTs is not based on data, but on inputs from expert domain. For instance, for rare-event situations like reliability analysis, inputs from expert domain stem from questionnaires, interviews and panel discussions. Sect. 3 focuses on the process of deriving information from experts that is named "elicitation". The issues and methods related to this task were analysed by review articles in different domains of application, namely shipping accidents by Zhang and Thai (2016), human reliability by Mkrtchyan et al. (2015) and more broadly regarding dependence in probabilistic modelling by Werner et al. (2017). The objective is to focus the elicitation on specific pieces of information to efficiently populate the CPTs by ensuring quality and consistency of the elicited result and minimizing the workload on the experts owing to the large number of CPT entries. Elicitation for CPT generally relies on three (possibly combined) main approaches through: (1) the assessment of probabilities directly from an (or a panel of) expert (Sect. 3.1); (2) assumptions on the causal structure either by simplifying the network structure or by simplifying the causal dependence (Sect. 3.2); (3) filling-up methods (Sect. 3.3).

3.1 Direct elicitation

In a direct approach, experts are asked to give quantitative numbers (like frequencies or confidence intervals) using methods like probability wheel, probability scale and gambling analogy. An extensive discussion on the different types of biases are provided by Renooij (2001), and more specifically in the domain of ecology by Kuhnert et al. (2010). Overall, methods which map qualitative statements to numerical values like the probability scale (see an example in Fig. 2(A)) is preferred for its simplicity, which improves the consistency (as underlined by Wiegmann (2005), and as reported by Zhang and Thai (2016) for marine safety). Probability wheel is criticized for not being appropriate for the elicitation of small or large probabilities, and the gambling analogy is criticized for being too time-consuming.

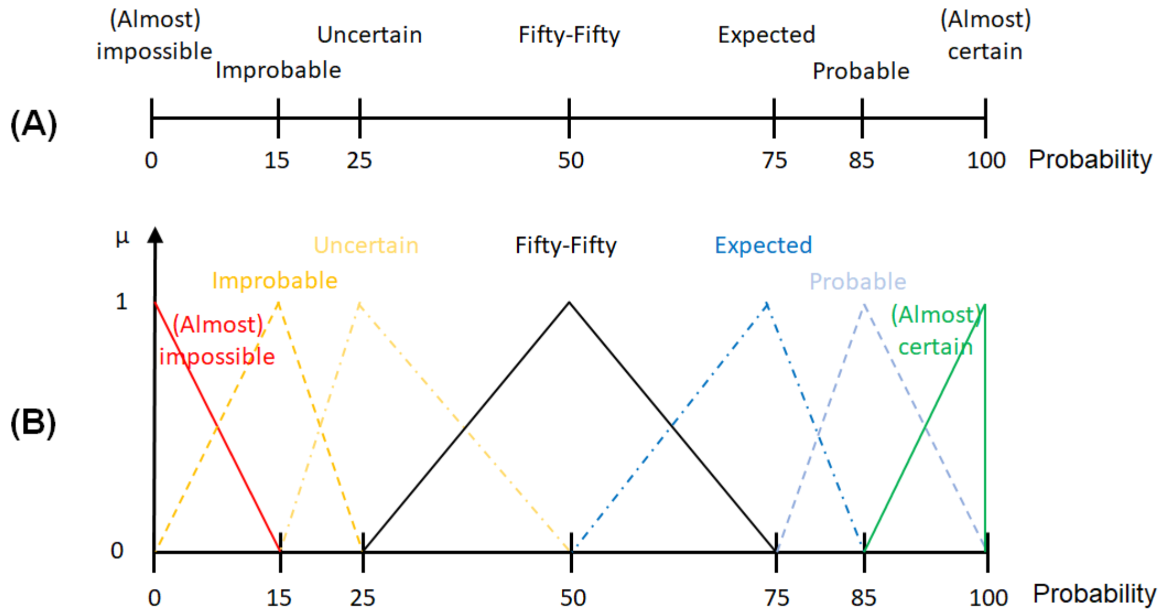

Figure 2. (A) Example of probability scale used to assist expert elicitation of CPTs (adapted from Knochenhauer et al. (2013)); (B) Translation of the probabilities qualified in (A) into Fuzzy sets (μ is the degree of membership).

[*Figure 2 about here*]

As an alternative, experts are preferably asked to give qualitative statements (like categorical or relative measure). To support this indirect approach, tools from the domain of multicriteria decision-making have been proposed. For instance, Chin et al. (2009) adapted the Analytical Hierarchy Process method for the task of probability elicitation and semi-automatic generation of the parameters of CPTs. The basic idea is to elicit paired comparisons about the relative likelihood of the possible events using predefined scores (equally possible, etc.) instead of directly asking the probability values. Yet, this procedure is at the expense of an increase in the number of comparisons as the number of conditional probabilities increases.

An alternative option proposes to directly process natural linguistic terms by mathematically modelling them using for instance a Fuzzy set (Zadeh, 1975). Let us consider the concept of membership function, which defines how each element x of the input space X (also named "universe of discourse") is mapped to a degree of membership (denoted μ). Under the classical theory of Boolean logic, the membership function of a set A is simply defined as a binary function that takes the value $\mu(x)$ =1 if the element belongs to A and the value $\mu(x)$=0, otherwise. The Fuzzy set theory of Zadeh (1965) introduces the concept of a set without a crisp (i.e. clearly defined) boundary. Such a set can contain elements with only a gradual (partial) degree of membership (μ is scaled between 0 and 1). The translation of the probability scale of Fig. 2(A) into Fussy sets is provided in Fig. 2(B). Some successful applications cover fault detection (D'Angelo et al., 2014), performance analysis of devices (Penz et al., 2012), safety risk analysis (Zhang et al., 2015), human reliability analysis (~~Peng-cheng et al., 2012;~~ Li et al., 2012), and offshore risk (Ren et al., 2009). Two viewpoints exist in the literature on Fuzzy BBNs. Fuzziness can be incorporated in the variables (nodes) or on the probabilities. For instance, Ren et al. (2009) carried out studies using fuzzy probability calculations in BBNs (as illustrated in Fig. 2(B)). Conversely, Tang and Liu (2007) used fuzzy events (i.e. Fuzzy node states) in BBNs for a machinery fault diagnosis problem. İçen and Ersel (2019) incorporated both aspects with application in medicine.

3.2 Making assumptions on the causal structure

To reduce the elicitation burden, the number of CPT entries to be elicited should be kept "reasonable". This can be performed by making assumptions regarding the causal structure. One option is by simplifying the structure through the introduction of "divorcing" nodes (Henderson et al., 2009). This involves aggregating a few of the nodes by adding a new node that summarizes them provided that the aggregations are logical and no interactions are lost in the procedure. Although this process adds nodes to the network, it reduces the combined size of CPTs in the network (Cain, 2001). Yet, divorcing might dilute the sensitivity of the final node(s) to the input nodes and might increase the uncertainty propagated through the network as underlined by Cain (2001).

A popular alternative aims at making some simplifications regarding the causal dependence based on the logical Noisy-OR gate (Pearl, 1988). In their typical implementation, Noisy-OR gates focus on binary BBN nodes and assume that the influence of the considered factor is independent from the presence of the other factors. This means that the probability of the outcome is the product of the probabilities of the outcome in presence of one factor at a time, with all other factors being absent. Formally, let us consider a binary variable Y with two states {False, True} and n binary parent variables $X_{i=1,..,n}$.

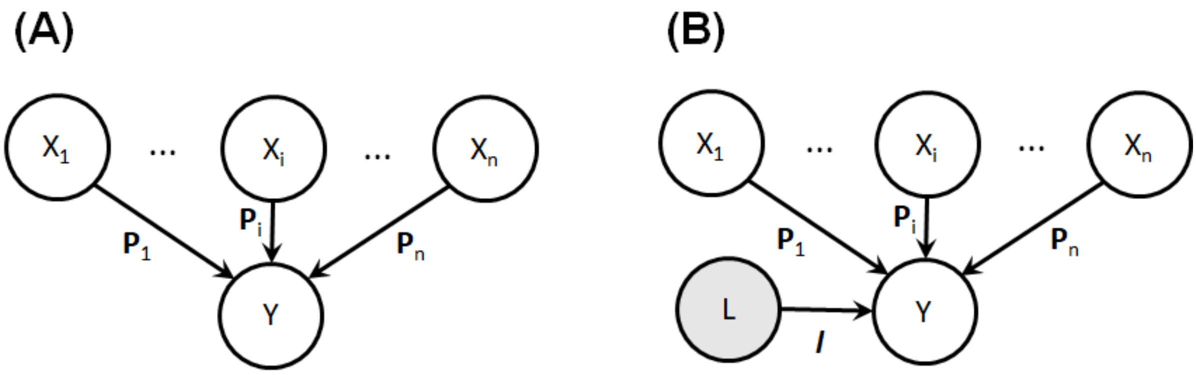

Figure 3. (A) Schematic representation of the Noisy-OR gate with $P_{i=1,..,n}$ the link probabilities; (B) Schematic representation of the Noisy-OR gate.

[*Figure 3 about here*]

The main principle of the Noisy-OR model is to define probabilities P_i (termed as link probability, Fig. 3(A)), which are defined as the probability that Y is False given that X_i is False and X_j is True for $i \neq j$. A Noisy-OR model is thus a disjunction "noisy" version of X_i (Pearl, 1988). This means that the distribution of Y conditional on X_1; X_2;…; X_n is $P(Y = F|X_1; \ldots; X_n) = 1 - \prod_{i:X_i \in \boldsymbol{X}_T}(1 - P_i)$ where $\boldsymbol{X}_T$ is the set of parent nodes whose states are True. The Noisy-OR model enables the analysist to specify fewer CPT parameters; the number of independent parameters being here reduced from 2^n to 2n. The extension of Noisy-OR gate to multi-valued variables is the Noisy-MAX gate model (Diez, 1993; Henrion, 1989). If the parent node X_i has n_{X_i} states, then the total number of parameters that have to be elicited using leaky Noisy-MAX gate is $N = \sum_{i=1}^{n}\left(n_{X_i} - 1\right)(n_Y - 1) + 1$ to be compared to the total number without Noisy-MAX gate, namely $N = (n_Y - 1).\prod_{i=1}^{n} n_{X_i}$.

Different empirical studies have been conducted to investigate the performance of the leaky Noisy-OR approach. Several authors (Oniśko et al., 2001; Anand and Downs, 2008; Bolt et al., 2010; among others) showed how this approach helped reducing the burden of elicitation in practical real-life applications without impacting too much the performance of the network. Besides, Zagorecki and Druzdzel (2012) explored to which extend the pattern of causal interaction induced by Noisy-OR(MAX) gates are common in real cases. Using three existing BBNs, they showed that the Noisy-MAX gate provides a good fit for as many as 50% of CPTs in two of these networks.

The Noisy-OR structure is based however on a strong assumption, i.e. that the node of interest is in the state False (considering the above illustrative case) with a probability equal to 1 if all its parent variables are in the state False. Yet, in many cases, it is often difficult to capture all the causes of the node of interest (e.g. for reliability purpose, it means to define all the failure modes of a component). To deal with this problem, Henrion (1989) proposed an extension called "leaky Noisy-OR" gate that includes a background probability that represents the influence of non-modelled causes as schematically depicted in Fig. 3(B). Zagorecki and Druzdzel (2004) proposed to elicit leaky and non-leaky Noisy-OR parameters as alternatives to conditional probabilities using statements like "What is the probability that Y is present when X_1 is present and all other causes of Y (including those not modelled explicitly) are absent?". They showed that the leaky Noisy-OR parameter was assessed as the most accurate (in terms of Euclidean distance to empirical distribution).

The leaky Noisy-OR method was further extended by relaxing the necessity to define a crisp precise leaky probability value, i.e. by introducing uncertainty on this parameter. This type of uncertainty has been addressed within different uncertainty treatment settings (which are introduced in more details in Sect. 4). Antonucci (2011) developed an imprecise leaky Noisy-OR gate model with uncertainty on the link probabilities modelled by intervals within the formalism of credal networks (see Sect. 4.2). Alternatively, Fallet-Fidry et al. (2012) (further extended by Zhou et al. (2016b)) proposed an imprecise extensions of the Noisy-OR within the formalism of evidential networks (see Sect. 4.3). Finally, Dubois et al. (2017) developed a version of noisy logical gates within the theory of possibility (Dubois and Prade, 1988) using possibilistic causal networks (as presented by Benferhat et al. (2002)) with illustration on an example taken from human geography.

3.3 Filling-up methods

Alternative methods to Noisy-OR(MAX) gate are based on filling-up techniques. These methods are typically based on extracting information on the factor effects from known relationships (named anchor conditional probability distributions, denoted CPD) and extrapolating to the whole CPTs. Considering two BBNs (of respectively 3 and 4 nodes) for a human reliability problem, Mkrtchyan et al. (2015) tested five popular methods for CPT derivation considering nodes with multiple states, namely:

- Method 1: the functional interpolation method (Podofillini et al., 2014) approximate CPDs elicited at the anchor positions by functions described by parameters (e.g., Normal functions); the parameters of the missing CPDs are then obtained by interpolating those corresponding to the anchor ones;
- Method 2: the Elicitation BBN method (Wisse et al., 2008) is based on piecewise linear functions interpolating among the elicited CPDs, and on state influencing factors and importance weights;
- Method 3: The Cain calculator (Cain, 2001) uses interpolation factors derived from CPDs to populate the missing relationships in CPTs;
- Method 4: The method presented by Røed et al. (2009) is also based on functional relationships between influencing factors and outcome nodes; the parameters of the function (exponential) are then determined based on the elicitation of selected CPDs;
- Method 5: the ranked node method by Fenton et al. (2007) (further improved by Laitila and Virtanen, 2016) is not based on interpolation of known CPDs. In this approach, all the nodes are defined on the interval [0–1]. For instance, let us consider a node with 5 states, namely ‘‘very low”, ‘‘low”, ‘‘average”, ‘‘high”, and ‘‘very high”; each of the state is assigned to an interval width of 0.2; for instance, the value ‘‘low” is assigned to the interval [0.2–0.4]. To generate CPTs, the experts are asked to provide the weight parameters and to choose one algorithm (the mean average, the Minimum, the Maximum and the MixMinMax). Using this method, if there are m ranked nodes and each node has n states, the expert will only need m +1 parameter values, while it requires $n \times m+1$ values for full elicitation.

Mkrtchyan et al. (2015) showed that:

- All methods allow representing the different importance of the various influencing factors;

- The representation of the interactions (combined effects of multiple factors) is problematic for methods eliciting information on the influence of factors taken one at a time (methods 2-4);
- Functional representation of the CPTs (methods 1, 5 and 4) can be traced more easily, because they allow an explicit representation of uncertainty in the factor relationships;
- But methods 4 and 5 have difficulties in representing the different degrees of uncertainty in the relationships;
- The method allowing the largest modelling flexibility is method 1 with respect to strong factor influences (single and multi-factor) and proper uncertainty characterization, but becomes too costly for large BBNs.

3.4 Discussion

The conclusions drawn by Mkrtchyan et al. (2015) serve as valuable recommendations regarding the use and applicability of the five most popular filling-up methods for reducing the expert burden of CPT elicitation. Despite the practical usefulness of this comparative exercise, it should be noted that they primarily focused, by construction, on the modelling aspects important for their application domain (here human reliability analysis), namely the representation of strong factor influences and interactions, and the characterization of different degrees of uncertainty in the relationships. Broader exercises are needed to cover a larger spectrum of methods (i.e. filling-up methods should be completed by Noisy- OR/MAX models, direct elicitation among others), of contexts (different network sizes, binary versus multivalued nodes, etc.), as well as of domains of application.

Despite the clear advantages of these methods for BBN engineering, they cannot be applied uncritically, because the probability values can only be considered approximations of the true probabilities and whatever the considered methods, they are all based on simplifications that may hamper the BBN performance. Initiatives like the one by Woudenberg and van der Gaag (2015) for the Noisy-OR model should be intensified. They identified the conditions under which ill-considered use of this method can result in large impact on output probabilities; in particular, when the yet-unobserved cause variables in the mechanism have relatively skewed probability distributions and/or the obtained parameter probabilities have small values. For this purpose, sensitivity methods as described in Sect. 5 can play an important role. Fenton et al. (2019) also dealt with the limitations of leaky Noisy-OR model for backward inference. When the binary node of interest Y of the example in Fig. 3 is observed to be in the state False, the normal “explaining away” behaviour fails, which means that after observing the state of any parent the remaining parents become independent, and the results may not result in what BBN practitioners expected. Fenton et al. (2019) described a simple extension of the model that requires the elicitation of only one extra parameter that can solve this problem for a large spectrum of cases in practice.

4 Propagating the uncertainties

Whatever the methods used to populate the CPTs, residual uncertainties may still prevail. This residual uncertainty should be reflected in BBN-based results. This means that the uncertainty on CPT entries should be propagated in order to evaluate their consequences on the BBN results. The propagation can either rely on probabilities (Sect. 4.1), or alternative mathematical representation tools like intervals (Sect. 4.2) or a generalization of a probability distribution (Sect. 4.3), i.e. within the theory of belief functions as introduced by Shafer (1976) and Dempster (1967). Fig. 4 summarizes the main principles of the different approaches by using a simple OR-gate model.

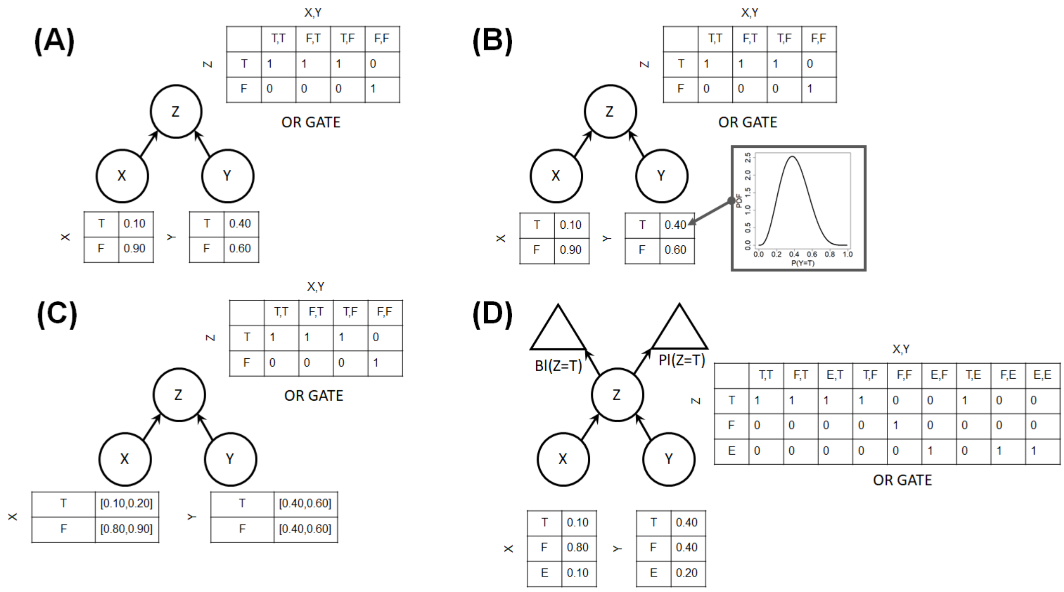

Figure 4. (A) Example of an OR gate model translated into a BBN with two binary parent nodes *X* and *Y* (with states corresponding to T=True or F=False). The truth table related to the OR gate corresponds to the table next to the child node *Z*. Illustration of a probability-based approach where uncertainties on CPT entries are represented by: (B) Beta probability distributions (with an example here for node *Y*); (C) Interval-valued probabilities (credal network approach); (D) Mass probability tables; here the truth table includes the epistemic state E={T,F}; Two nodes were added to the network to calculate the belief and plausibility functions (see Sect. 4.3 for more details).

[*Figure. 4 about here*]

4.1 Methods using probabilities

The problem of uncertainty propagation for BBN has originally been addressed using probabilities. This approach assumes that the uncertainty on CPT entries follows a Beta probability distribution (or for more generic cases, a Dirichlet probability distribution), as schematically depicted in Fig. 4(B). Kleiter (1996) originally described a Monte-Carlo-based random simulation procedure to carry out the approximation of the spread of the probability distribution for the considered query. The method requires, however, a large number of random samples to accurately characterize the true variance. Van Allen et al. (2008) proposed an improved method by avoiding Monte Carlo sampling through the combination of bucket elimination (Dechter, 1998) with the "delta rule" that linearizes the relationship between the query probabilities and the corresponding Dirichlet conditional probabilities connecting the query variable to its parents and children. They further proved that the Beta approximation (for binary BBNs) is asymptotically valid. The conditions of the exact Beta distribution has extensively been investigated by Hooper (2008). This problem has further been formalized within the setting of subjective logic (Jøsang, 2001; 2016) as proposed by Kaplan and Ivasnoska (2018), who developed an efficient belief propagation for inference in a binary Bayesian network with a singly-connected graph. To introduce any type of probability distribution on CPTs, Fenton (2018) proposed to extend the BBN with continuous nodes corresponding to the uncertain prior probability distributions, but at the expense of a potentially large increase of computational time cost when the number of nodes and of CPTs increases.

4.2 Methods using interval-valued probabilities

Instead of specifying a crisp single value of each CPT entry, the formal setting of credal network, denoted CN (Cozman, 2000; 2005), integrates BBNs with credal sets, i.e. set of probability measures. A CN can be viewed as the representation of a set of BBNs, which share the same graphical structure but are associated to different conditional probability parameters; the interest being to provide a richer representation of uncertainty. In Fig. 4(C), the uncertainty in the CPTs are presented by intervals.

Formally, given a variable X, we denote by Ω, the possibility space of X, x a generic element of Ω, $P(X)$ the probability mass function for X and $P(\mathrm{x})$ the probability of x. The credal set over X is $\mathrm{K}(X)$, which corresponds to a closed convex set of probability mass functions over X. For any $x \in \Omega$, the lower probability for x according to the credal set $\mathrm{K}(X)$ is $\underline{P}(x) =$

$min_{P(X)\epsilon K(X)} P(x)$. Similar expression can be given for the upper probability. Within Walley's theory of imprecise probabilities (Walley, 1991), credal sets can then be represented as polytopes, where each inner point has a valid probability mass, and can be obtained by computing the convex hull of a finite number of probabilities, called vertices (Cozman, 2000). A credal set for a random variable X_i is labelled $K(X_i)$, while the set comprising its extreme points is denoted by ext$[K(X_i)]$.

A credal network CN (Cozman, 2000) over a set of random variables is thus a DAG where dependencies among variables are defined by a set of conditional credal sets as $K(X_i|pa(X_i))$. By analogy with BBN, it is possible to define a joint credal set as follows:

$$K(X) = CH(P(X){:}\, P(x) = \prod_{i=1}^{n} P(x_i|pa(X_i)) \qquad \text{(Eq. 4)}$$

where $P(X_i|pa(X_i) \in ext[K(X_i|pa(X_i)]$, CH is the convex hull operator, applied to the probabilities computed for the combination of all the vertices of all the conditional credal sets.

In this setting, the task of inference aims at computing the probability bounds of the largest extension that satisfies the Markov condition (i.e., independence of each node of its non-descendant non-parents given its parents) under the assumption of strong independence (Cozman, 2000). This results in the convex hull of the set containing all joint distributions that factorize the overall joint probability of the network, where the conditional distributions $P(X_i|pa(X_i)=\pi_k)$ are selected from the local sets $K(X_i|pa(X_i)= \pi_k)$. This task is a NP-hard (de Campos and Cozman, 2005), for which a number of exact and approximate algorithms have been proposed (Antonucci et al., 2015; Mauá et al. 2012; Ide and Cozman, 2008; Cano et al., 2007), but only exact inference algorithms are suitable to polytree-shape binary networks.

Though CN allows quantifying and integrating the uncertainty on CPTs on the BBN inference results, this increase in expressiveness comes at the expense of higher computational costs. Some real case applications of CN exist in different domains (Table 1), but the number of them remain limited (with comparison to BBN), despite the availability of some open-source solutions like OpenCossan (Tolo et al., 2018), the linear programming algorithm[1] of Antonucci et al. (2015), the GL2U-II algorithm[2] of Antonucci et al. (2010).

	Domain of application	**Method**	**Reference**

[1] http://ipg.idsia.ch/software.php?id=135
[2] http://ipg.idsia.ch/software.php?id=142

1	Hazard assessment of debris flows	CN	Antonucci et al. (2007)
2	Military identification	CN	Antonucci et al. (2009)
3	Reliability analysis of a fire-detector system	EN	Simon et al. (2008; 2009; 2017)
4	Threat assessment	EN	Benavoli et al. (2009)
5	Convoy detection	EN	Pollard et al. (2010)
6	Reliability analysis of oil filter plug linked to aero engines	EN	Yang et al. (2012)
7	Railway dysfunction	EN	Aguirre et al. (2013)
8	Food processing	Dynamic CN	Baudrit et al. (2016)
9	Cyber attack analysis	EN	Friedberg et al. (2017)
10	Vulnerability analysis of Nuclear Power Plant subject to external hazards	CN	Tolo et al. (2017)
11	Reliability analysis of a safety instrumentation system for a pressurized vessel	EN	Zhang et al. (2017)
12	Medical prognostic and diagnostic	EN and Fuzzy sets	Janghorbani and Moradi (2017)
13	Fault diagnosis for railway	BBN vs EN	Verbert et al. (2017)
14	Maritime accidents	CN	Zhang and Thai (2018)
15	Terrorist attack analysis on a chemical storage plant	EN vs CN	Misuri et al. (2018)
16	Landslides	CN	He et al. (2018)
17	Risk assessment of an oscillating water column	CN	Estrada-Lugo et al. (2018)
18	Reliability analysis of a feeding control system	EN	Mi et al. (2018)
19	Human reliability for Nuclear Power Plant safety analysis	EN	Deng and Jiang (2018)
20	Safety assessment of a truss	EN	Khakzad (2019)

Table 1. Case studies of evidential networks (EN) and Credal Network (CN)

[*Table 1 about here*]

4.3 Methods based on Dempster-Shafer Theory

An alternative setting for representing imprecision is the theory of belief functions, also called Dempster-Shafer Theory, denoted DST (Shafer, 1976, Dempster, 1967). Let X be a variable taking values in the frame of discernment Θ composed of q mutually and exhaustive possible state of X. For instance, for a binary node, the frame of discernment is Θ={True, False}.

Formally, the theory introduces the concept of basic belief assignment (BBA) based on the belief mass function m: $2^{\Theta} \rightarrow [0,1]$ and satisfies $\sum_{A \subseteq \Theta} m(A) = 1$, and $m(\emptyset) = 0$ (which assumes that at least one element of Θ is true). Every $A \in 2^{\Theta}$ such that m(A)>0 is called a focal element.

In classical probabilities, a probability value can be assigned to the state True or False only. By defining the belief mass function based on the powerset of the frame of discernment 2^{Θ} (which corresponds in the binary example to {∅,True, False, {True,False}}) enables the analysist to allocate a quantity supporting an additional state termed as epistemic state E={True, False}. Due to uncertainty, the analysist may not always be able to determine the amount of masses to attribute to each state, and the variable X may then be in both states, True or False. This means that the method allows characterizing uncertainty about the state of a given node.

From a mass function m, two measures can be defined (instead of one for the probabilistic case) called the belief (Bl) and plausibility (Pl) measures. The latter are respectively defined, for any event A as follows:

$Bl(A) = \sum_{E \subseteq A} m(E)$, and $Pl(A) = \sum_{E \cap A \neq \emptyset} m(E)$ (Eq. 5)

where Bl measures how much event A is implied by the information (it sums masses that must be redistributed over elements of A), Pl measures how much event A is consistent with the information (it sums masses that could be redistributed over elements of A). These two measures can be associated to a (closed convex) set of bound probabilities $\{P \mid \forall A \subseteq \Theta$, (A) $\leq (A) \leq Pl(A)\}$. It is thus possible to associate an interval-valued probability to the event A, with minimum and maximum probabilities provided by Bel and Pl, respectively. This makes the formal link with CN. Conversely, it is also possible to reconstruct BBAs from Pl and Bel functions using a Möbius Transformation (Smets, 2002).

As an illustration, let us assume consider a binary node for which the expert only knows that the probability of the event {X=True} is at least 0.8. The corresponding BBA is m({True})=0.8, m(E={True, False})=0.2, m({False})=0. This means that Bl({True})= m({True}), and Pl({True})= m({True})+ m(E)=0.8+0.2=1.0. This also means that Bl({True})=m({True})=0.8, and Pl({True})=m({True})+ m(E)=0.8+0.2=1.0. Then 0.8≤P({True})≤1.0.

The evidence theory is the basis of evidential networks (EN), which is a DAG propagating belief masses. One of the first formulation by Xu and Smets (1996) is based on the Dempster's rule for combining and reasoning with the belief masses. Yet, one major limitation is that the inference algorithms in this formulation are less effective that the one for traditional BBNs as underlined for instance by Khakzad (2019). In the domain of system reliability analysis, Simon et al. (2008) proposed an alternative by mapping logical gates (like OR or AND typically used for fault tree analysis), as EN with the hypothesis described by Guth (1991). Despite its similarity with BBN, relations in EN between variables are not probabilities, but belief masses. The truth table of gates are replaced by conditional mass tables for AND and OR gates (see an example in Fig. 4(D)). To compute belief and plausibility measures in EN, specific nodes (as proposed by Simon and Weber 2009) are introduced. Three types of nodes (as represented in Fig. 4(D)) are thus defined (Simon and Bicking, 2017), namely:

- Root nodes to which BBA are assigned, correspond to components;
- Non-root nodes correspond to logical gates that encode its resulting states {True, False, {True,False}} given the states of its parents;
- Evaluation nodes correspond to nodes that aim at providing estimates of the belief and plausibility measures of the system state.

In the formulation by Simon and Weber (2009), the inference computation is based on the Bayes theorem, which is extended to DST by specifying a mass of 1 on one of the focal elements of the frame of discernment for a specific evidence (hard evidence). Non-specific evidence (soft evidence) corresponds to a mass distribution on the focal elements of the frame of discernment. This means that probability updating in such EN can be based on BBN inference algorithms.

Misuri et al. (2018) compared CN and EN with illustration on a terrorist attack analysis on a chemical storage plant. They highlighted that:

- When used for uncertainty propagation, EN and CN give the same results;
- In terms of implementation, EN is simplier to use, because they can be built using existing codes for BBN, whereas CN requires specific codes;
- In terms of interpretation, Misuri et al. (2018) concluded that EN is more intuitive, because experts directly assign some weight to the epistemic state (e.g. E={True,False} for a binary node), whereas they have to specify interval-valued probabilities for CN, which can become tricky for multivalued nodes.

Khakzad (2019) further filled the gap between CN and EN by proposing some heuristic rules to determine prior belief masses based on imprecise probabilities. They further modified Simon and co-authors' EN formulation to both improve the propagation and updating of the belief masses using BBNs. In order to deal with linguistic variables for the network node' states, the EN method can be combined either with Fuzzy sets (Zadeh, 1975) as applied by Janghorbani and Moradi (2017) for medical prognostic, or with a Naive Bayes classifier model as applied by Zhang et al. (2017) for safety analysis for nuclear power plant.

4.4 Discussion

Different settings are available to help the BBN analysist to deal with the problem of uncertainty propagation. A natural question is the justification for using approaches that are alternative to classical probabilities. A first argument often highlights the epistemic nature of the CPT uncertainties. Contrary to aleatory uncertainty (also referred to as randomness), which represents the variability of the physical environment or engineered system under study, epistemic uncertainty mainly stems from the incomplete/imprecise nature of available information (e.g. Hoffman and Hammonds, 1994). While tools from the probabilistic setting can appropriately handle aleatory uncertainties, it is the second type, which raises several problems in practice. In our situation, probability distribution cannot be inferred from data/observations, and should therefore be assumed; the procedure described in Sect. 4.1 is mainly based on the assumption that the uncertainties on the CPTs are described by a Beta (or for more generic cases, a Dirichlet probability distribution) probability distribution. Yet, this assumption may influence the final results of the BBN-based analysis (see. e.g., Ditlevsen, 1994 for an extensive discussion in reliability analysis); Relying only on probabilities masks this problem and might induce an appearance of more refined knowledge with respect to the existing uncertainty than is really present (Klir, 1989; 1994). Sect. 4.2 and 4.3 describe alternative non-probabilistic frameworks to represent uncertainty in situations characterized by limited available pieces of information, which are mainly restricted to expert judgements. Both approaches allow improving the expressiveness with respect to uncertainty representation (as shown by the few tens of application studies using these techniques, see Table 1), in particular by enabling the BBN analysist to translate his/her uncertainty on the node states or his/her imprecision on the CPT parameters by avoiding the need for specifying a probability model.

Yet, extra-probabilistic approaches (whatever the considered methods, CN, EN or networks combined with linguistic variables or based on alternative uncertainty theories like possibility theory, see Dubois et al. (2017)) might come at the expense of higher level of sophistication and of complexity of the inference algorithms (and potentially higher computational costs). The danger is to add more confusion than insights as discussed by Aven and Zio (2011) with the viewpoint of decision making for risk management. The question of selecting the most appropriate approaches for representing and characterizing the risk and uncertainties (in particular with application on BBNs) still remains open (see e.g., an extensive discussion by Flage et al. 2014).

5 Characterizing the uncertainties

Methods presented in Sect. 4 allows evaluating the impacts of CPT uncertainties on the BBN results. But, this tells nothing about the respective contribution of the different CPT entries on the total uncertainty, i.e. the influence of the different uncertainties. This is the purpose of sensitivity analysis (SA), which can be used, in the construction phase of the BBN model, to study how the output of a model varies with variation of the CPT parameters. Subsequently, the results from SA can be used as a basis for parameter tuning, as well as for studying the robustness of the model output to changes in the parameters (Coupé and van der Gaag, 2002; Laskey et al., 1995).

5.1 Description of the methods

For discrete BBNs, a widespread SA method relies on the use of sensitivity functions (Coupé and van der Gaag, 2002; Castillo et al., 1997), which describe how the considered output probability varies as one CPT entry value is changed. An example of application in the domain of marine safety is provided by Hänninen and Kujala (2012). Formally, consider the conditional probability P(Z=k|e), where e denotes the available evidence, and a CPT entry x=P(X=i|π) where i is a value of a variable X and π is a combination of values for the parents of X. The sensitivity function then corresponds to a quotient of two functions that are linear in x of the following form:

$$f(x) = \frac{c_1 x + c_2}{c_3 x + c_4} \quad \text{(Eq. 6)}$$

where the constants c_i are built from the values of the network's non-varied parameters. The numerator of Eq. 6 expresses the joint probability P(Z=k|e) as a function of x, and its denominator describes P(e) in terms of x. Using the example described in Fig. 1, we focus on the probability of having brain tumor given absence of coma but increased level of serum calcium, i.e. $P(B = T|C = F, ISC = T)$. Van der Gaag et al. (2013) estimated the sensitivity of this probability of interest to the probability of having coma given absence of brain tumor but increased level of serum calcium, $x = P(C = T|B = F, ISC = T)$. The sensitivity function was established as $\frac{-0.03}{x-1.03}$ (as depicted in Fig. 5(A)). This type of function shows that the probability of interest steeply increases when x exceeds 0.80, i.e. above the original parametrization given by Cooper (1984).

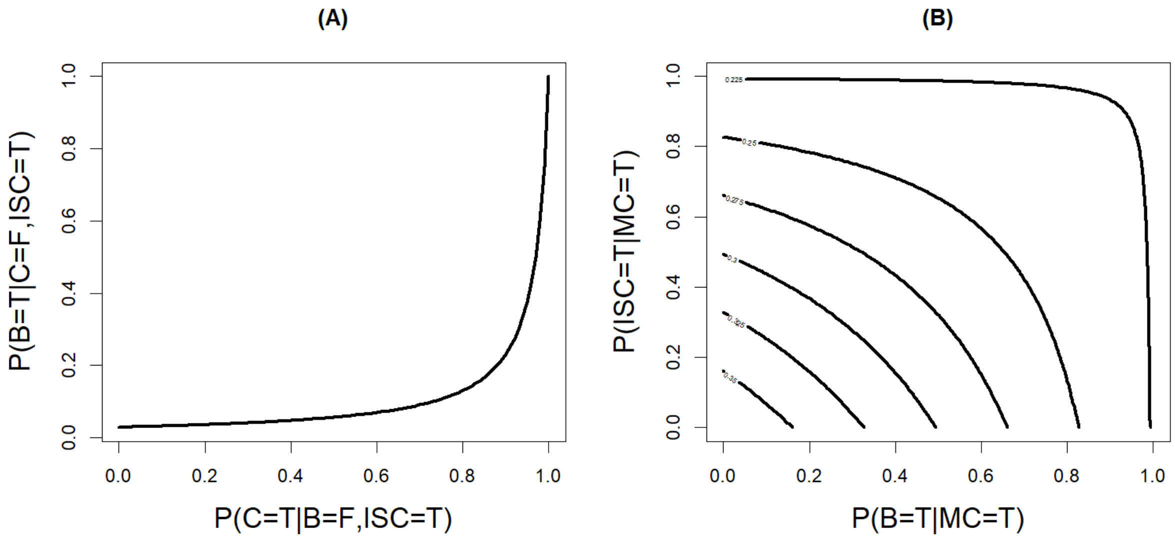

Figure 5. (A) One-way sensitivity function for the BBN described in Fig. 1; (B) Two-way sensitivity function for the BBN described in Fig. 1 considering the probability P(C=T) as the targeted probability.

[*Figure 5 about here*]

Two-way sensitivity functions can be expressed in a similar form as a quotient of two bi-linear functions. Consider the sensitivity function of $f(x, y)$ that expresses P(Z=k|e) as a function of the parameter probabilities $x = P(X = i|\pi_X)$ and $y = P(Y = j|\pi_D)$, where i and j are values of the variables X and Y, and π_X and π_Y are combinations of values for the parents of X and of Y. The function holds as follows:

$$f(x, y) = \frac{c_1 x.y + c_2 x + c_3 y + c_4}{c_5 x.y + c_6 x + c_7 y + c_8} \qquad \text{(Eq. 7)}$$

where the constants c_i are estimated from the values of the network's non-varied parameters. In the tumor BBN example (Fig. 1), we focus on the probability of having a cancer P(C=T) and its sensitivity to the simultaneous variation of the conditional probabilities x=P(B=T|MC=T) and y=P(ISC=T|MC=T), which was established by van der Gaag et al. (2013) as $0.374 +$

$0.15.x.y - 0.15.x - 0.15.y$ (Fig. 5(B)). This type of function shows that despite the large variation of x and y (from 0 to 1), the probability of interest varies over a moderate range of values of only ~15%.

A complementary approach for SA involves the study of the Chan–Darwiche (CD) distance (Chan and Darwiche, 2002; 2005), which is a measure for bounding probabilistic belief change. It is complementary in the sense that it gives insight in the effect of parameter changes on the global joint distribution, rather than on a specific (posterior) output probability of interest (as sensitivity functions do). In practices, the CD distance can be used to identify parameter changes, which lead the closer distance between the original and the varied BBN distributions (Chan and Darwiche, 2005). It should however be noted that the choice of the type of distance is rather arbitrary as outlined by Renooij (2014) and other distances like the KL-divergence (Kullback and Leibler, 1951) or the ϕ-divergences (Ali and Silvey, 1966) could also be of interest.

Recent studies have focused on the properties of the SA methods. Renooij (2014) thoroughly investigated the different schemes for varying a probability from a (conditional) distribution, while co-varying the remaining probabilities from the same distribution; the proportional co-variation scheme being the most popular one. Leonelli et al. (2017) further formalized the SA problem for discrete BBNs within the generic setting of multilinear models. They developed a unifying approach to sensitivity methods via the interpolating polynomial representation of discrete statistical models in the context of "BBNs single full CPT analyses", i.e. where one parameter from each CPT of one vertex of a BBN given each configuration of its parents is varied. This approach based on multilinear probabilistic models enabled them to address the problem of multi-way SA (with dimension ≥ 2). Furthermore, they proved the optimality of proportional covariation by showing that the CD distance is minimized when parameters are proportionally co-varied.

5.2 Discussion

Since the nineties, the BBN community has seen the developments of SA methods that are specifically dedicated to their respective needs regarding the BBN use and application. Though simple and efficient to implement, the approach based on sensitivity functions (combined with CD-distance analysis) remains local, because one parameter values are varied, while the other ones are kept constant. Multi-way SA methods have been proposed, but can rapidly become intractable. Interestingly, outside the BBN community, the problem of SA is commonly addressed with alternative tools; variance-based global SA techniques being the most popular one (Iooss and Lemaitre, 2015). Such techniques were adapted by Li and Mahadevan (2018) to bridge the gap between both communities. Their approach has the advantage to be global i.e. in the sense that all CPT parameters' values are changed all together. Besides, this approach can be applicable to any types of BBN (discrete, hybrid or continuous), i.e. it is model-free.

6 Concluding remarks

6.1 Summary

The current survey has investigated how to deal with uncertainties related to the specification of CPTs for discrete BBNs. Three questions were addressed, namely: (1) how to constrain the uncertainties related to CPT derivation; (2) how to integrate these uncertainties in the BBN-based analysis; (3) how to test the robustness of the BBN-based results to these uncertainties. Table 2 provides a summary of the main methods/approaches (together with their advantages and limits) to answer these questions.

Question	Approach	Advantages	Limits	Section
1	Learning CPT by combining data and expert prior knowledge via MAP estimation	- It improves the MLE-based fitting when the number of data is limited.	- The representation of expert belief is restricted to the use of Dirichlet priors; - There is a possible problem of "under-fitting" in sparse situations.	Sect. 2.4-2.5
1	Learning CPT by combining data and qualitative constraints	- The accuracy of the MLE/MAP-based fitting is largely improved when data are scarce; - The experts may feel more conformable in providing ordering than precise CPT values.	- Many new estimators are available, but many lack practical recommendations; - There is a possible problem of "under-fitting" depending on the chosen priors.	Sect. 2.4-2.5
1	Direct elicitation using qualitative statements	- The experts may feel more conformable in providing qualitative statements than quantitative estimates.	- Mathematical modelling of linguistic terms may lead to information loss or increased computation burden.	Sect. 3.1
1	Use of "divorcing" nodes	- The number of nodes is decreased through aggregation of nodes.	- Care should be paid to avoid the loss of interactions in the procedure; - It may dilute the sensitivity of the final node(s) to the input nodes;	Sect. 3.2

			- It might increase the uncertainty propagated through the BBN.	
1	Simplification of the causal structure using logical gates (e.g. Noisy-OR gate)	- The number of nodes to be elicited is largely decreased (e.g. from from 2^n to 2n for a binary node with n parents).	- The assumptions on the causal relationships might not always be valid in real life applications; - The simplifications may hamper the BBN performance.	Sect. 3.2
1	Extracting information on the factor effects from known relationships and extrapolating them	- A large variety of different “filling-up” methods exist to relieve the elicitation burden; - Some feedbacks on real case applications exist (e.g. for human reliability analysis).	- Simplifications are introduced and the derived probabilities can only be considered approximations of the true probabilities.	Sect. 3.3
2	Uncertainty propagation using probabilities	- The degree of confidence in the BBN-based results can be quantified.	- The uncertainty representation is restricted to the use of Beta/Dirichlet probability distributions; - It can become computationally intensive.	Sect. 4.2
2	Uncertainty propagation using intervals with credal networks (CN)	- It avoids selecting a probability model to represent the uncertainty; - The experts may feel more comfortable in assigning intervals than probabilities.	- The specification of interval-valued probabilities can become tricky for multivalued nodes; - It needs specific sophisticated inference algorithms and software solutions (with potential high computational costs).	Sect. 4.3
2	Uncertainty propagation within the Dempster-Shafer Theory by using evidential networks (EN)	- The expressiveness is improved like for CN; - EN is more intuitive than CN, because experts directly assign some weight to the epistemic state; - It can be implemented with existing BBN softwares.	- The translation of interval-valued probabilities within this setting can become difficult for multivalued nodes; - The inference algorithms for combining joint/disjoint belief masses are not so effective as those based on probability theory.	Sect. 4.4
3	Sensitivity Analysis (SA) using sensitivity functions	- The theory is well-established; - It is simple to implement;	- It focuses on the influence of one (or multiple) CPT parameters while the	Sect. 5.1

		- The graphical representation is straightforward to interpret.	other ones are kept constant; - It requires specific co-variations schemes; - Multi-way SA can rapidly become intractable.	
3	Sensitivity Analysis using Chan–Darwiche distance	- It complements the sensitivity functions by giving insight in the effect of parameter changes on the global joint distribution.	- It presents the same disadvantages than the sensitivity functions; - The choice of the distance can be rather arbitrary.	Sect. 5.1

Table 2. Summary of the advantages and limits of the main approaches

[*Table 2 about here*]

Considering the first question, we described methods for deriving CPT entries from different sources of information (observations, prior knowledge, expert-based information, etc.). Traditional estimators like MLE and MAP (or new ones) were proposed to make the best use of the data available even in scarce situations when completed by qualitative constraints like knowledge about the monotonic influences between nodes. For rare-event situations like reliability analysis, the main source of information relies on inputs from expert domain using different elicitation techniques; the main challenge being the minimization of the workload on the experts owing to the large number of CPT entries while preserving the quality and consistency of the elicited result. Elicitation for CPTs generally relies on three (possibly combined) main approaches: (1) through the assessment of probabilities directly from an expert or a panel of experts; (2) through a simplification of the causal structure using the popular Noisy-OR(MAX) model (and its improved versions like the leaky one); (3) through filling-up methods, which have in particular been thoroughly benchmarked on test cases in the domain of human reliability analysis.

The second question can be addressed using different approaches, either using probabilities, or imprecise probabilities either using interval-valued probabilities within the setting of credal networks or within the Dempster-Shafer theory within the setting of evidential networks. Though the latter approach enables an increase in expressiveness with respect to uncertainty representation (as shown by the few tens of application studies using these techniques, see Table 1), this might come however at the expense of higher complexity of the inference algorithms (and higher computational costs). Finally, the third question is investigated by methods specifically developed for sensitivity analysis of BBN; in particular through the use of one- or multi- way sensitivity functions.

6.2 Discussion and open questions

BBN is now viewed as a suitable tool for overcoming data gaps, estimating uncertainties, and visualizing complex causal relationships. Despite its clear advantages, it cannot be applied uncritically, and addressing the question of uncertainties in its construction, and more specifically in the CPT derivation, should become standard practices to increase the confidence in its use. The analysis of the literature (Table 2) shows that any analysist is now equipped with a handful of different tools/methods to address the question of uncertainties. This participates to the minimization of the concern of Neil et al. (2000): "In the literature much more attention is given to the algorithmic properties of BBNs than to the method of actually building them in practice".

Yet, the developments of these techniques is only one part of the problem, and effort should be intensified to bring them to an operative state. To this purpose, the implementation of these techniques within commonly-used BBN software packages (like Beuzen and Simmons 2019 for the widely-used Netica software, Norsys Software Corp., 2006 or Tolo et al. (2018) who proposed an open-source software package OpenCossan) should be strengthened. Second many methods lack practical recommendations. Therefore, more benchmark / comparative exercises are needed to cover broader situations and to serve as best practices for selecting the most appropriate tools depending the characteristics of the considered situation. For instance, the filling-up methods benchmarked by Mkrtchyan et al. (2015) should be completed by Noisy-OR/MAX models, direct elicitation among others, and applied in different contexts (different network sizes, binary versus multivalued nodes, etc.), as well as domains of application. Similarly, there is a need for comparing the pros and cons of using alternative frameworks for uncertainty representation and propagation in BBNs, i.e. comparing approaches using probabilities, or interval-valued or Dempster-Shafer structures or possibility distributions or Fuzzy sets, for instance by following the initiatives conducted for probabilistic risk analysis (e.g., Pedroni et al., 2013, Loschetter et al., 2016).

The current work has focused on CPT derivation for discrete BBN development. The second key ingredient of BBNs is the DAG specification, whose learning from data has been investigated in numerous studies (e.g., Heinze-Deml et al. (2018), Scutari et al. (2018), Beretta et al. (2018), etc.). To address the whole spectrum of uncertainties in BBN building, studies both covering DAG and CPT learning would be beneficial. To integrate both sources of uncertainty, possible lines of future research may either focus on the improvement of existing algorithm like the structural expectation-maximization algorithm (Benjumeda et al., 2019) to simultaneously learn the structure and parameters of a BN from incomplete data, or on the combination/aggregation of multiple BBNs, each of them being based on a different set of assumptions either regarding structure or CPT parametrisation (Kim and Cho, 2017; Feng et al., 2014).

Finally, it should be underlined that BBN modelling is a rapidly advancing field (see e.g., Marcot and Penman, 2019) that covers new applications and features (like the incorporation of the time and space dimension, the improvements in the treatment of discrete and continuous variables, its links with artificial intelligence, among others). The research on the uncertainty treatment for these new developments is active (see e.g., recent advances for sensitivity analysis of a wide array of graphical models by Leonelli (2019)), and the scope of the current work should be broadened in the future to include them.

Acknowledgements

This study has been carried out within the NARSIS project, which has received funding from the European Union’s H2020-Euratom Programme under grant agreement N° 755439.

References

Aguirre, F., Sallak, M., Vanderhaegen, F., Berdjag, D., 2013. An evidential network approach to support uncertain multiviewpoint abductive reasoning. Information Sciences, 253, 110-125.

Ali, S.M., Silvey, S.D., 1966. A general class of coefficients of divergence of one distribution from another J. R. Stat. Soc. Ser. B, 28, 131–142.

Altendorf, E.E., Restificar, A.C., Dietterich, T.G., 2005. Learning from sparse data by exploiting monotonicity constraints. In: Proceeding of the 21st Conf. Uncertainty in Artificial Intelligence, pp. 18-26.

Anand, V., Downs, S. M., 2008. Probabilistic asthma case finding: a noisy or reformulation. In: Annual AMIA Symposium, pp. 6–10.

Antonucci, A., 2011. The Imprecise Noisy-OR Gate. In: Proceedings of the 14th International Conference on Information Fusion, 709-715, Chicago.

Antonucci, A., Brühlmann, R., Piatti, A., Zaffalon, M., 2009. Credal networks for military identification problems. International Journal of Approximate Reasoning, 50(4), 666-679.

Antonucci, A., de Campos, C.P., Huber, D., Zaffalon, M., 2015. Approximate credal network updating by linear programming with applications to decision making. International Journal of Approximate Reasoning, 58, 25-38.

Antonucci, A., Salvetti, A., Zaffalon, M., 2007. Credal networks for hazard assessment of debris flows. In: Advanced Methods for Decision Making and Risk Management in Sustainability Science, pp. 237-256.

Antonucci, A., Sun, Y., de Campos, C. P., Zaffalon, M., 2010. Generalized loopy 2U: a new algorithm for approximate inference in credal networks. International Journal of Approximate Reasoning, 51(5), 474-484.

Aven, T., Zio, E., 2011. Some considerations on the treatment of uncertainties in risk assessment for practical decision-making. Reliability Engineering and System Safety, 96, 64–74.

Azzimonti, L., Corani, G., Zaffalon, M., 2019. Hierarchical estimation of parameters in Bayesian networks. Computational Statistics and Data Analysis, 137, 67–91.

Baudrit, C., Destercke, S., Wuillemin, P.H., 2016. Unifying parameter learning and modelling complex systems with epistemic uncertainty using probability interval. Information Sciences, 367, 630-647.

Bauer, E., Koller, D., Singer, Y., 1997. Update rules for parameter estimation in Bayesian networks. In: Proceedings of the 13th Conference on Uncertainty in Artificial Intelligence, pp. 3–13.

Benavoli, A., Ristic, B., Farina, A., Oxenham, M., Chisci, L., 2009. An application of evidential networks to threat assessment. IEEE Trans. Aerospace Electron Syst., 45(2), 620–639.

Benferhat, S., Dubois, D., Garcia, L., Prade, H., 2002. On the transformation between possibilistic logic bases and possibilistic causal networks. International Journal of Approximate Reasoning, 29(2), 135–173.

Benjumeda, M., Luengo-Sanchez, S., Larrañaga, P., Bielza, C. 2019. Tractable learning of Bayesian networks from partially observed data. Pattern Recognition, 91, 190-199.

Beretta, S., Castelli, M., Gonçalves, I., Henriques, R., Ramazzotti, D., 2018. Learning the structure of Bayesian Networks: A quantitative assessment of the effect of different algorithmic schemes. Complexity, 2018, Article ID 1591878, 12 pages.

Beuzen, T., Simmons, J., 2019. A variable selection package driving Netica with Python. Environmental modelling & software, 115, 1-5.

Brooker, P., 2011. Experts, Bayesian belief networks, rare events and aviation risk estimates. Safety Science, 49(8), 1142–1155.

Beuzen, T., Marshall, L., Splinter, K.D. 2018. A comparison of methods for discretizing continuous variables in Bayesian Networks. Environmental modelling & software, 108, 61-66.

Bolt, J., van der Gaag, L.C., 2010. An empirical study of the use of the noisy-OR model in a real-life Bayesian network. In: International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems. Springer, Berlin, Heidelberg, pp. 11-20.

Cain, J., 2001. Planning Improvements in Natural Resources Management. In: Guidelines for Using Bayesian Networks to Support the Planning and Management of Development Programmes in the Water Sector and Beyond. Centre for Ecology and Hydrology, Wallingford, UK.

de Campos, C.P., Cozman, F.G., 2005. The inferential complexity of Bayesian and credal networks. In: Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence, pp. 1313–1318.

de Campos, C.P., Ji, Q., 2008. Improving Bayesian network parameter learning using constraints, In: Proceedings of the 19th International Joint Conference on Artificial Intelligence, Edinburgh, pp. 1–4.

Cano, A., Gómez, M., Moral, S., Abellán, J., 2007. Hill-climbing and branch-and-bound algorithms for exact and approximate inference in credal networks. International Journal of Approximate Reasoning, 44(3), 261-280.

Castillo, E., Gutiérrez, J.M., Hadi, A.S., 1997. Sensitivity analysis in discrete Bayesian networks. IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans, 27(4), 412-423.

Chan, H., Darwiche, A., 2002. When do numbers really matter? J. Artif. Intell. Res., 17, 265-287.

Chan, H., Darwiche, A., 2005. A distance measure for bounding probabilistic belief change. International Journal of Approximate Reasoning, 38, 149-174.

Chang, R., Wang, W., 2010. Novel algorithm for Bayesian network parameter learning with informative prior constraints. In: Proceedings of the International Joint Conference on Neural Networks, 2010, pp.1–8.

Chen, S.H., Pollino, C.A., 2012. Good practice in Bayesian network modelling. Environmental modelling & software, 37, 134–45.

Chin, K.S., Tang, D.W., Yang, J.B., Wong, S.Y., Wang, H., 2009. Assessing new product development project risk by Bayesian network with a systematic probability generation methodology. Expert Systems with Applications, 36(6), 9879-9890.

Chojnacki, E., Plumecocq, W., Audouin, L., 2019. An expert system based on a Bayesian network for fire safety analysis in nuclear area. Fire Safety Journal, 105, 28-40.

Cooper, G.F., 1984. NESTOR: a Computer-based Medical Diagnostic Aid that Integrates Causal and Probabilistic Knowledge. Report HPP-84-48, Stanford University.

Coupé, V.M.H., van der Gaag, L.C., 2002. Properties of sensitivity analysis of Bayesian belief networks Ann. Math. Artif. Intell., 36, 323-356.

Cowell, R.G., 1999. Parameter learning from incomplete data using maximum entropy I: principles, Statistical Research Report, 21.

Cozman, F.G.. 2000. Credal networks. Artificial Intelligence, 120, 199–233.

Cozman, F.G., 2005. Graphical models for imprecise probabilities. International Journal of Approximate Reasoning 39(2–3), 167–184.

D'Angelo, M.F., Palhares, R.M., Cosme, L.B., Aguiar, L.A., Fonseca, F.S., Caminhas, W.M., 2014. Fault detection in dynamic systems by a fuzzy/Bayesian network formulation. Appl. Soft Comput., 21, 647–653.

Dempster, A.P., 1967. Upper and lower probabilities induced by a multivalued mapping. The annals of mathematical statistics, 325–339.

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. The Royal Statistical Society Series, B39, 1–38.

Dechter, R., 1998. Bucket elimination: A unifying framework for probabilistic inference. In: Learning and Inference in Graphical Models, Springer, Dordrecht, pp. 75-104.

Deng, X., Jiang, W., 2018. Dependence assessment in human reliability analysis using an evidential network approach extended by belief rules and uncertainty measures. Annals of Nuclear Energy, 117, 183-193.

Diez, F.J., 1993. Parameter adjustment in Bayes networks. The generalized noisy or-gate. In Proceedings of the ninth international conference on uncertainty. In: artificial intelligence, Morgan Kaufmann Publishers Inc., USA, pp. 99–105

Ditlevsen, O., 1994. Distribution arbitrariness in structural reliability. In: Proceedings of ICOSSAR'93. Structural Safety & Reliability, Rotterdam, The Netherlands, pp. 1241-1247.

Drury, B., Valverde-Rebaza, J., Moura, M. F., de Andrade Lopes, A., 2017. A survey of the applications of Bayesian networks in agriculture. Engineering Applications of Artificial Intelligence, 65, 29-42.

Druzdzel, M.J., van der Gaag, L., 2000. Building Probabilistic Networks: "where do the numbers come from?" IEEE Trans. Knowl. Data Eng., 12(4), 481–486.

Dubois, D. Prade, H., 1988. Possibility theory, New York: Plenum.

Dubois, D., Fusco, G., Prade, H., Tettamanzi, A.G., 2017. Uncertain logical gates in possibilistic networks: Theory and application to human geography. International Journal of Approximate Reasoning, 82, 101-118.

Estrada-Lugo, H.D., Patelli, E., de Angelis, M., Raj, D.D., 2018. Bayesian networks with imprecise datasets: application to oscillating water column. In: Safety and Reliability–Safe Societies in a Changing World. CRC Press, pp. 2611-2618.

Fallet-Fidry, G., Weber, P., Simon, C., Iung, B., Duval, C., 2012. Evidential network-based extension of Leaky Noisy-OR structure for supporting risks analyses. In: IFAC Proceedings 45(20), 672-677.

Feelders, A., van der Gaag, L.C., 2006. Learning Bayesian network parameters under order constraints, International Journal of Approximate Reasoning, 42(1–2), 37–53.

Feelders, A., van der Gaag, L.C., 2005. Learning Bayesian network parameters with prior knowledge about context-specific qualitative influences. In: Proceedings of the 31st Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI), pp. 193–200.

Feng, G., Zhang, J.D., Liao, S.S., 2014. A novel method for combining Bayesian networks, theoretical analysis, and its applications. Pattern Recognition, 47(5), 2057-2069.

Fenton, N.E., 2018. Handling Uncertain Priors in Basic Bayesian Reasoning. doi10.13140/RG.2.2.16066.89280

Fenton, N.E., Noguchi, T., Neil, M., 2019. An extension to the noisy OR function to resolve the ‘explaining away’ deficiency for practical Bayesian network problems. IEEE Transactions on Knowledge and Data Engineering, in press.

Fenton, N.E., Neil, M., Caballero, J.G., 2007. Using ranked nodes to model qualitative judgments in Bayesian networks IEEE Trans. Knowl. Data Eng., 19(10), 1420-1432.

Flage, R., Aven, T., Zio, E., Baraldi, P., 2014. Concerns, challenges, and directions of development for the issue of representing uncertainty in risk assessment. Risk Analysis, 34(7), 1196-1207.

Friedberg, I., Hong, X., McLaughlin, K., Smith, P., Miller, P.C. 2017. Evidential network modeling for cyber-physical system state inference. IEEE Access, 5, 17149-17164.

Gao, X.G., Guo, Z.G., Ren, H., Yang, Y., Chen, D.Q., He, C.C., 2019. Learning Bayesian network parameters via minimax algorithm. International Journal of Approximate Reasoning, 108, 62-75.

Gehl, P., D’Ayala, D., 2016. Development of Bayesian Networks for the multi-hazard fragility assessment of bridge systems. Structural Safety, 60, 37-46.

Geman, S., Geman, D., 1984. Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images, IEEE Transactions on Pattern Analysis and Machine Intelligence, 6, 721–741.

Guo, Z.G., Gao, X. G., Ren, H., Yang, Y., Di, R.H., Chen, D.Q., 2017. Learning Bayesian network parameters from small data sets: A further constrained qualitatively maximum a posteriori method. International Journal of Approximate Reasoning, 91, 22-35.

Guth, M., 1991. A probability foundation for vagueness and imprecision in fault tree analysis. IEEE Trans. Reliab., 40(5), 563–570.

Hänninen, M., 2014. Bayesian networks for maritime traffic accident prevention: benefits and challenges. Accident Analysis & Prevention, 73, 305-312.

Hänninen, M., Kujala, P., 2012. Influences of variables on ship collision probability in a Bayesian belief network model. Reliability Engineering & System Safety, 102, 27-40.

Hänninen, M., Banda, O. A. V., Kujala, P., 2014. Bayesian network model of maritime safety management. Expert Systems with Applications, 41(17), 7837-7846.

He, L., Topa Gomes, A., Broggi, M., Beer, M., 2018. Risk analysis of infinite slope failure using advanced Bayesian networks. In: 8th International Workshop on Reliable Engineering Computing, “Computing with Confidence”, 16–18 July 2018, University of Liverpool, Liverpool, UK

Heckerman, D., Geiger, D., Chickering, D.M., 1995. Learning Bayesian networks: the combination of knowledge and statistical data. Machine Learning, 20, 197–243.

Henrion, M., 1989. Some practical issues in constructing belief networks, Uncertainty in Artificial Intelligence 3, Elsevier Sciences Publisher, North Holland.

Henderson, C., Pollino, C.A., Hart, B.T., 2009. Appendix A: the review - workshop discussion document. In: Hart, B.T., Pollino, C.A. (Eds.), Bayesian Modelling for Risk-based Environmental Water Allocation, Waterlines Report Series No. 14, Canberra, pp. 33-83.

Heinze-Deml, C., Maathuis, M.H., Meinshausen, N., 2018. Causal structure learning. Annual Review of Statistics and Its Application, 5, 371-391.

Hooper, P.M., 2008. Exact distribution theory for belief net responses. Bayesian Analysis, 3(3), 615-624.

Hoffman, F.O., Hammonds, J.S., 1994. Propagation of uncertainty in risk assessments: the need to distinguish between uncertainty due to lack of knowledge and uncertainty due to variability. Risk analysis, 14(5), 707-712.

Hospedales, T., Zhou, Y., Fenton, N., Neil, M., 2015. Probabilistic graphical models parameter learning with transferred prior and constraints. In: Proceedings of the 31st Conference on Uncertainty in Artificial Intelligence, AUAI Press, pp. 972-981.

Ide, J.S., Cozman, F.G., 2008. Approximate algorithms for credal networks with binary variables. International Journal of Approximate Reasoning, 48(1), 275-296.

İçen, D., Ersel, D., 2019. A new approach for probability calculation of fuzzy events in Bayesian Networks. International Journal of Approximate Reasoning, 108, 76-88.

Iooss, B., Lemaître, P., 2015. A review on global sensitivity analysis methods. In: Uncertainty management in simulation-optimization of complex systems. Springer, Boston, MA, pp. 101-122.

Jaeger, M., 2006. The AI&M procedure for learning from incomplete data. In: Proceedings of the 22nd Conference on Uncertainty in Artificial Intelligence, pp. 225–232.

Jäger, W.S., Christie, E.K., Hanea, A.M., den Heijer, C., Spencer, T., 2018. A Bayesian network approach for coastal risk analysis and decision making. Coastal Engineering, 134, 48-61.

Janghorbani, A., Moradi, M.H., 2017. Fuzzy evidential network and its application as medical prognosis and diagnosis models. Journal of biomedical informatics, 72, 96-107.

Jensen, F., 2001. Bayesian Networks and Decision Graphs, Springer, New York.

Jøsang, A., 2001. A logic for uncertain probabilities, Int. J. Uncertain. Fuzziness Knowl. Based Syst. 9(3), 279–311.

Jøsang, A., 2016. Subjective Logic: A Formalism for Reasoning Under Uncertainty, Springer, Heidelberg.

Kaplan, L., Ivanovska, M., 2018. Efficient belief propagation in second-order Bayesian networks for singly-connected graphs. International Journal of Approximate Reasoning, 93, 132-152.

Khakzad, N., 2019. System safety assessment under epistemic uncertainty: Using imprecise probabilities in Bayesian network. Safety Science, 116, 149-160.

Kim, K.J., Cho, S.B., 2017. Ensemble bayesian networks evolved with speciation for high-performance prediction in data mining. Soft Computing, 21(4), 1065-1080.

Kleiter, G.D., 1996. Propagating imprecise probabilities in Bayesian networks. Artificial Intelligence, 88, 143–161.

Klir, G.J., 1989. Is theremore to uncertainty than some probability theorists might have us believe?. International Journal of General System, 15(4), 347–378.

Klir, G.J., 1994. On the alleged superiority of probabilistic representation of uncertainty. IEEE Transactions on Fuzzy Systems, 2(1), 27–31.

Knochenhauer, M., Swaling, V.H., Dedda, F.D., Hansson, F., Sjökvist, S., Sunnegaerd, K. 2013. Using Bayesian Belief Network (BBN) modelling for rapid source term prediction. Final report (No. NKS--293). Nordisk Kernesikkerhedsforskning, Roskilde, Denmark.

Kuhnert, P.M., Martin, T.G., Griffiths, S.P., 2010. A guide to eliciting and using expert knowledge in Bayesian ecological models. Ecol. Lett., 13, 900–914.

Kullback, S., Leibler, R.A., 1951. On information and sufficiency. The annals of mathematical statistics, 22(1), 79-86.

Kwag, S., Gupta, A., 2017. Probabilistic risk assessment framework for structural systems under multiple hazards using Bayesian statistics. Nuclear Engineering and Design, 315, 20-34.

Laitila, P., Virtanen, K., 2016. Improving construction of conditional probability tables for ranked nodes in Bayesian networks. IEEE Transactions on Knowledge and Data Engineering, 28(7), 1691-1705.

Langseth, H., Portinale, L., 2007. Bayesian networks in reliability. Reliability Engineering and System Safety, 92(1), 92–108.

Laskey, K.B., 1995. Sensitivity analysis for probability assessments in Bayesian networks IEEE Trans. Syst. Man Cybern., 25(6), 901-909.

Leonelli, M., 2019. Sensitivity analysis beyond linearity. International Journal of Approximate Reasoning, in press.

Leonelli, M., Goergen, C., Smith, J.Q., 2017. Sensitivity analysis in multilinear probabilistic models. Information Sciences, 411, 84–97.

Li, P.C., Chen, G.H., Dai, L.C., Zhang, L., 2012. A fuzzy Bayesian network approach to improve the quantification of organizational influences in HRA frameworks. Safety science, 50(7), 1569-1583.

Li, C., Mahadevan, S., 2018. Sensitivity analysis of a Bayesian network. ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering, 4(1), 011003.

Liao, W., Ji, Q., 2009. Learning Bayesian network parameters under incomplete data with domain knowledge, Pattern Recognition, 42(11), 3046–3056.

Loschetter, A., Rohmer, J., de Lary, L., Manceau, J., 2016. Dealing with uncertainty in risk assessments in early stages of a co2 geological storage project: comparison of pure-probabilistic and fuzzy-probabilistic frameworks. Stochastic Environmental Research and Risk Assessment, 30(3), 813-829.

Malagrino, L.S., Roman, N.T., Monteiro, A.M., 2018. Forecasting stock market index daily direction: a Bayesian network approach. Expert Systems with Applications, 105, 11-22.

Marcot, B.G., Steventon, J.D., Sutherland, G.D., McCann, R.K., 2006. Guidelines for developing and updating Bayesian belief networks applied to ecological modeling and conservation. Canadian Journal of Forest Research (Revue Canadienne De Recherche Forestiere), 36, 3063-3074.

Marcot, B.G., Penman, T.D., 2019. Advances in Bayesian network modelling: Integration of modelling technologies. Environmental modelling & software, 111, 386-393.

Masegosa, A.R., Feelders, A.J., van der Gaag, L.C., 2016. Learning from incomplete data in Bayesian networks with qualitative influences. International Journal of Approximate Reasoning, 69, 18-34.

Mauá, D.D., de Campos, C.P., Zaffalon, M., 2012. Updating credal networks is approximable in polynomial time. International Journal of Approximate Reasoning, 53(8), 1183-1199.

Mi, J., Li, Y. F., Peng, W., Huang, H.Z., 2018. Reliability analysis of complex multi-state system with common cause failure based on evidential networks. Reliability Engineering & System Safety, 174, 71-81.

Milns, I., Beale, C.M., Smith, V.A., 2010. Revealing ecological networks using Bayesian network inference algorithms. Ecology, 91(7), 1892-1899.

Misuri, A., Khakzad, N., Reniers, G., Cozzani, V., 2018. Tackling uncertainty in security assessment of critical infrastructures: Dempster-Shafer Theory vs. Credal Sets Theory. Safety science, 107, 62-76.

Mkrtchyan, L., Podofillini, L., Dang, V.N., 2015. Bayesian belief networks for human reliability analysis: A review of applications and gaps. Reliability engineering & system safety, 139, 1-16.

Musharraf, M., Bradbury-Squires, D., Khan, F., Veitch, B., MacKinnon, S., Imtiaz, S., 2014. A virtual experimental technique for data collection for a Bayesian network approach to human reliability analysis. Reliability Engineering & System Safety, 132, 1-8.

Neil, M., Tailor, M., Marquez, D., 2007. Inference in hybrid Bayesian networks using dynamic discretization. Statistics and Computing, 17(3), 219-233.

Neil, M., Fenton, N., Nielson, L., 2000. Building large-scale Bayesian networks. The Knowledge Engineering Review, 15(3), 257-284.

Norsys Software Corp., 2006. Sensitivity to Findings Netica Software Documentation available at www.norsys.com

Niculescu, R.S., Mitchell, T.M., Rao, R.B., 2006. Bayesian network learning with parameter constraints, J. Mach. Learn. Res., 7(3), 1357–1383.

Nojavan, F., Qian, S.S., Stow, C.A., 2017. Comparative analysis of discretization methods in Bayesian networks. Environmental modelling & software, 87, 64-71.

Oniśko, A., Druzdzel, M. J., Wasyluk, H., 2001. Learning Bayesian network parameters from small data sets: Application of Noisy-OR gates. International Journal of Approximate Reasoning, 27(2), 165-182.

Pearl, J., 1988. Probabilistic Reasoning in Intelligent Systems. Morgan Kaufmann, San Francisco, CA

Pedroni, N., Zio, E., Ferrario, E., Pasanisi, A., Couplet, M., 2013. Hierarchical propagation of probabilistic and non-probabilistic uncertainty in the parameters of a risk model. Computers & Structures, 126, 199–213.

Penz, C.A., Flesch, C.A., Nassar, S.M., Flesch, R.C.C., De Oliveira, M.A., 2012. Fuzzy – Bayesian network for refrigeration compressor performance prediction and test time reduction, Expert Systems with Applications, 39, 4268–4273.

Podofillini, L, Mkrtchyan, L., Dang, V.N., 2014. Aggregating expert-elicited error probabilities to build HRA models. In: Proceedings of ESREL 2014. Sept. 14–18, Wroclaw, Poland; 2014, pp. 1083–1091.

Pollard, E., Rombaut, M., Pannetier, B., 2010. Bayesian networks vs. evidential networks: An application to convoy detection. In: International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems. Springer, Berlin, Heidelberg, pp. 31-39.

Ramoni, M., Sebastiani, P., 2001. Robust learning with missing data. Machine Learning, 45(2), 147–170.

Ren, J., Jenkinson, I., Wang, J., Xu, D., Yang, J., 2009. An offshore risk analysis method using fuzzy Bayesian network. J. Offshore Mech. Arct. Eng., 131, 041101.

Renooij, S., 2001. Probability elicitation for belief networks: issues to consider. Knowled. Eng. Rev., 16, 255–269.

Renooij, S., 2014. Co-variation for sensitivity analysis in bayesian networks: properties, consequences and alternatives. International Journal of Approximate Reasoning, 55, 1022-1042.

Røed, W., Mosleh, A., Vinnem, J.E., Aven, T., 2009. On the use of the hybrid causal logic method in offshore risk analysis. Reliability Engineering and System Safety, 94(2), 445-455.

Scutari, M., Howell, P., Balding, D. J., Mackay, I., 2014. Multiple quantitative trait analysis using Bayesian networks. Genetics, 198(1), 129-137.

Scutari, M., Graafland, C.E., Gutierrez, J.M., 2018. Who Learns Better Bayesian Network Structures:Constraint-Based, Score-based or Hybrid Algorithms? In: Proceedings of Machine Learning Research, 72, pp. 1-12.

Seixas, F.L., Zadrozny, B., Laks, J., Conci, A., Saade, D.C.M., 2014. A Bayesian network decision model for supporting the diagnosis of dementia, Alzheimer's disease and mild cognitive impairment. Comput. Biol. Med., 51, 140–158.

Shafer, G., 1976. A Mathematical Theory of Evidence. Princeton University Press, Princeton, NJ.

Simon, C., Bicking, F., 2017. Hybrid computation of uncertainty in reliability analysis with p-box and evidential networks. Reliability Engineering & System Safety, 167, 629-638.

Simon, C., Weber, P., 2009. Evidential networks for reliability analysis and performance evaluation of systems with imprecise knowledge. IEEE Transactions on Reliability, 58(1), 69–87.

Simon, C., Weber, P., Evsukoff, A., 2008. Bayesian networks inference algorithm to implement Dempster Shafer theory in reliability analysis. Reliability Engineering and System Safety, 93(7), 950–963.

Smets, P., 2002. The application of the matrix calculus to belief functions. International Journal of Approximate Reasoning, 31, 1–30.

Tang, H., Liu, S., 2007. Basic theory of fuzzy Bayesian networks and its application in machinery fault diagnosis. In: Proceedings of Fourth International Conference on Fuzzy Systems and Knowledge Discovery, 24–27 Aug. 2007, Haikou, Hainan, China, pp.132–137.

Thiesson, B., 1995 Accelerated quantification of Bayesian networks with incomplete data, In: Proceedings of the First International Conference on Knowledge Discovery and Data Mining, pp. 306–311.

Tolo, S., Patelli, E., Beer, M., 2017. Robust vulnerability analysis of nuclear facilities subject to external hazards. Stochastic Environmental Research and Risk Assessment, 31(10), 2733-2756.

Tolo, S., Patelli, E., Beer, M., 2018. An open toolbox for the reduction, inference computation and sensitivity analysis of Credal Networks. Advances in Engineering Software, 115, 126-148.

Van Allen, T., Singh, A., Greiner, R.Hooper, P.M., 2008. Quantifying the uncertainty of a belief net response: Bayesian error-bars for belief net inference. Artificial Intelligence, 172, 483–513.

Van der Gaag, L.C., Kuijper, R., Van Geffen, Y.M., Vermeulen, J.L., 2013. Towards uncertainty analysis of Bayesian Networks. In 25th Benelux Conference on Artificial Intelligence, Delft, The Netherlands.

Verbert, K., Babuška, R., De Schutter, B., 2017. Bayesian and Dempster–Shafer reasoning for knowledge-based fault diagnosis–A comparative study. Engineering Applications of Artificial Intelligence, 60, 136-150.

Walley, P. 1991. Statistical reasoning with imprecise probabilities. Chapman and Hall, London.

Weber, P., Medina-Oliva, G., Simon, C., Iung, B., 2012. Overview on Bayesian networks applications for dependability, risk analysis and maintenance areas. Engineering Applications of Artificial Intelligence, 25(4), 671-682.

Werner, C., Bedford, T., Cooke, R.M., Hanea, A.M., Morales-Napoles, O., 2017. Expert judgement for dependence in probabilistic modelling: a systematic literature review and future research directions. European Journal of Operational Research, 258(3), 801-819.

Wiegmann, D.A., 2005. Developing a methodology for eliciting subjective probability estimates during expert evaluations of safety interventions: application for bayesian belief networks. Technical report from Aviation Human Factors Division AH FD-05-13/NASA-05-4. https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20050238445.pdf

Wisse, B.W., van Gosliga, S.P., van Elst, N.P., Barros, A.I., 2008. Relieving the elicitation burden of bayesian belief networks. In: Proceedings of the sixth Bayesian modelling applications workshop on UAI. Helsinki, Finland.

Woudenberg, S.P., van der Gaag, L.C., 2015. Propagation effects of model-calculated probability values in Bayesian networks. International Journal of Approximate Reasoning, 61, 1-15.

Xu, H., Smets, P., 1996. Reasoning in evidential networks with conditional belief functions. International Journal of Approximate Reasoning, 14(2-3), 155-185.

Yang, Y., Gao, X., Guo, Z., Chen, D., 2019. Learning Bayesian networks using the constrained maximum a posteriori probability method. Pattern Recognition, 91, 123-134.

Yang, J., Huang, H. Z., Liu, Y., Li, Y.F., 2012. Evidential networks for fault tree analysis with imprecise knowledge. Int. J. Turbo Jet Engines, 29(2), 111-122.

Zadeh, L.A., 1965. Fuzzy sets. Information and control, 8(3), 338–353.

Zadeh, L.A., 1975. The concept of a linguistic variable and its application to approximate reasoning. Information sciences, 8(3), 199–249.

Zagorecki, A., Druzdzel, M.J., 2004. An empirical study of probability elicitation under noisy-or assumption. In: Flairs conference, pp. 880–886.

Zagorecki, A., Druzdzel, M.J., 2012. Knowledge engineering for Bayesian networks: How common are noisy-MAX distributions in practice?. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 43(1), 186-195.

Zhang, X., Mahadevan, S., Deng, X., 2017. Reliability analysis with linguistic data: An evidential network approach. Reliability Engineering & System Safety, 162, 111-121.

Zhang, G., Thai, V.V., 2018. Addressing the epistemic uncertainty in maritime accidents modelling using Bayesian network with interval probabilities. Safety science, 102, 211-225.

Zhang, L., Wu, X., Qin, Y., Skibniewski, M.J., Liu, W., 2015. Towards a fuzzy Bayesian network based approach for safety risk analysis of tunnel-induced pipeline damage, Risk Anal. 36(2), 278–301.

Zhou, Y., Fenton, N., Neil, M., 2014. Bayesian network approach to multinomial param- eter learning using data and expert judgments. International Journal of Approximate Reasoning, 55(5), 1252–1268.

Zhou, Y., Fenton, N., Zhu, C., 2016a. An empirical study of Bayesian network parameter learning with monotonic influence constraints. Decis. Support Syst., 87(C), 69–79 .

Zhou, K., Martin, A., Pan, Q., 2016b. The belief noisy-or model applied to network reliability analysis. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 24(6), 937-960.

Zhang, G., Thai, V.V., 2016. Expert elicitation and Bayesian Network modeling for shipping accidents: A literature review. Safety science, 87, 53-62.

Zwirglmaier, K., Straub, D., 2016. A discretization procedure for rare events in Bayesian networks. Reliability Engineering & System Safety, 153, 96-109.