Bayesian networks with imprecise datasets: Application to oscillating water column

H.D. Estrada-Lugo, E. Patelli & M. de Angelis
Institute for Risk and Uncertainty, University of Liverpool, Liverpool, UK

Daniel D. Raj
Department of Ocean Engineering, Indian Institute of Technology Madras, Chennai, India

ABSTRACT: The Bayesian Network approach is a probabilistic method with an increasing use in the risk assessment of complex systems. It has proven to be a reliable and powerful tool with the flexibility to include different types of data (from experimental data to expert judgement). The incorporation of system reliability methods allows traditional Bayesian networks to work with random variables with discrete and continuous distributions. On the other hand, probabilistic uncertainty comes from the complexity of reality that scientists try to reproduce by setting a controlled experiment, while imprecision is related to the quality of the specific instrument making the measurements. This imprecision or lack of data can be taken into account by the use of intervals and probability boxes as random variables in the network. The resolution of the system reliability problems to deal with these kinds of uncertainties has been carried out adopting Monte Carlo simulations. However, the latter method is computationally expensive preventing from producing a real-time analysis of the system represented by the network. In this work, the line sampling algorithm is used as an effective method to improve the efficiency of the reduction process from enhanced to traditional Bayesian networks. This allows to preserve all the advantages without increasing excessively the computational cost of the analysis. As an application example, a risk assessment of an oscillating water column is carried out using data obtained in the laboratory. The proposed method is run using the multipurpose software OpenCossan.

1 INTRODUCTION

Nowadays, experimental research in engineering deals with systems with high complexity due to the number of components used in the procedures. Occasionally, the limited capacity of the experimental arrangements to test different configurations and obtain a number of relevant measurements, hinders the impact of the study. In addition, the effects of epistemic uncertainty derived from the procedure and, the uncontrollable conditions under which the study is carried out can provide results with limited applicability or lack of meaning. There are different methods developed for modeling the dependability and evaluation of large engineering systems allowing to take into consideration both qualitative and quantitative information. Among the most used methods, the reliability block diagrams, fault trees and, event trees, can be identified as the techniques with reliable results providing a robust mathematical background (Hamada et al. 2008). However, several assumptions are made with these techniques.

The Bayesian Network (BN), is a probabilistic method to study and analyze the genuine dependencies or independences of variables that make up a system. This concept was proposed by Judea Pearl

in 1988, originally for the artificial intelligence area (Pearl 1991). Currently, the BNs have many more applications ranging from system dependability (Castillo et al. 1997) and risk analysis (Hudson et al. 2002), to system maintenance (Kang and Golay 1999). It is worth noticing that this method has attracted an increasing interest, reaching 800% according to (Weber et al. 2012), during the last 20 years. The success of Bayesian networks rests on the graphical representation of the system, which renders them intuitive and easy to understand even by for non-experts. In addition, this method can be used to provide a diagnostic or predictive reasoning, a combination of both (Korb and Nicholson 2004) and also they accept new evidence that can be used to update the network and to adapt the model to the new parameters. Moreover, information of different types (e.g. expert judgment, experimental data, historical records, feedback experience, theoretical models, etc.) can be merged in the same network, inside structures called probability tables (or conditional probability tables in the case of children nodes). These tables are filled with crisp probability values, providing a global dependability estimation (Jensen and Nielsen 2007).

On the other hand, the high acceptance of the traditional Bayesian networks for uncertainty

reasoning is limited to the use of only crisp probabilities (Spiegelhalter 1987). This type of probabilities leads to discretization methods or hard assumptions, impoverishing the quality of the analysis (Tolo et al. 2016a). In order to work with continuous probabilities that can take into account the uncertainty of the variables in the network and avoid discretization of the input information, Daniel Straub and Armen Der Kiureghian (Straub and Der Kiureghian 2010) proposed to enhance BNs with structural reliability methods since these techniques support the use of continuous random variables. This approach embraces all the advantages of BNs and furthermore allows working with discrete, continuous, as well as small probability variables (ideal for low-probability high-impact events). Some applications have been done with this method, focusing on the risk assessment of technological facilities considering climate change (Tolo et al. 2016a).

Although the enhanced Bayesian networks consider a broader spectrum of variables, usually the information available rarely meets the requirements of the method. For example, often expert beliefs do not agree in an exact same probability value or the information is scarce, in which cases data have to be averaged or the analyst makes assumptions to perform the study. In engineering, it is common to perform measurements during an experiment, the results obtained will have attached an epistemic uncertainty that cannot be eliminated and underestimated. Consequently, the incorporation of imprecise probabilities becomes an imperative need that can improve the employability of BNs.

The proposed method attempts a naive implementation of Credal sets and p-boxes as a way to characterize imprecise probabilities in discrete and continuous variables, respectively. This approach is expected to overcome two main problems when dealing with uncertainty in Bayesian networks. The method is implemented on the multipurpose software OpenCossan, (Patelli et al. 2018). On one hand, the use of all the advantages of parametric and non-parametric p-boxes to work with continuous imprecise random variables and obtain the quantile bounds of the final distribution representing the system under analysis so they can be used after with the structural reliability methods. The network reduction process can be done with the *advanced line sampling* method, since it is capable of approximating the upper and lower bounds of the failure probability under the assumption of a monotone system. Once the network is reduced and the uncertainty of continuous variables propagated to the reduced network through the bounded probability of failure, the result will be a Credal network with only discrete but bounded variables. The method used, allows to compute the exact bounds of the query probability in the absence of evidence. In the case of evidence introduced in the network

the method will provide the intervals, such that the true bounds of the query probability are located.

2 THEORETICAL BACKGROUND

2.1 *Bayesian network*

A Bayesian network, as established before, is presented in the form of a directed acyclic graph (DAG) made of *nodes* and arrows (called *links*) connecting those nodes. Each node represents a random variable with information about observable quantities or hypothesis of the system, whilst the links show the dependency among the nodes. Nodes can be differentiated as *parent* and *child*. A child node depends directly on another node, called the parent node and graphically they are connected by a link starting in the parent and ending in the child. Nodes with no parents are called *roots*. The dependence of nodes is ruled the d-separation concept. Two variables, namely, A and B, are d-separated if there is an intermediate variable, C, different from A and B, such that in a serial or diverging connection C is instantiated (with a specific probability value, i.e. evidence). In the case of a converging connection, if neither C nor any of its descendants are instantiated they are d-separated (Jensen and Nielsen 2007).

The arrangement of parents and children nodes connected by links allows performing diagnostic and predictive reasoning. The first approach can be done to know the causes of an event by querying a parent node given the information in children nodes. The predictive reasoning follows the direction of the links and uses the causes (information in the parents) to predict the effects (children). The probability values denoting the dependency between a child node with its parents is stored in a conditional probability table. In this arrangement, the probability of each state of the child is provided given each of the states of the parents. The total dependability of the network is quantified by the *joint probability distribution* which is defined as the product of all the conditional and unconditional probabilities specified in the network. This is governed by the chain rule for Bayesian networks (Jensen and Nielsen 2007) and, it is given as follows,

$$P(X_i) = \prod_{i=1}^{n} P[X_i \mid pa(X_i)] \tag{1}$$

where X_i represents each of the random variables of the network and, pa(X_i) is the probability of the parents of X_i. As an example, the joint probability of the Bayesian network presented in Figure 1 is given by the expression,

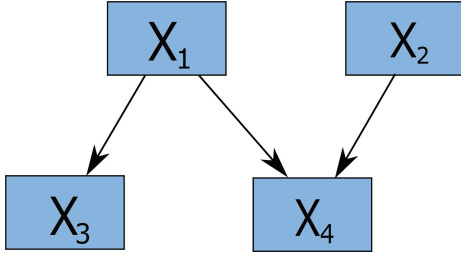$$P(X_1 X_2, X_3, X_4) = P(X_1) P(X_2) \\ \cdot \ P(X_3 \mid X_1, X_2) P(X_4 \mid X_1) \tag{2}$$

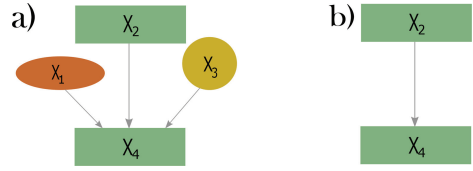Figure 1. Example of a traditional Bayesian network.

Figure 2. a) Simple enhanced Bayesian network with discrete nodes (rectangular shaped), probability function nodes (circle shaped) and, interval discrete nodes (elliptical shaped). b) Reduced Bayesian network with crisp probabilities.

The joint probability distribution function can be used to calculate the probability of any individual variable (Hosseini and Barker 2016). For instance, to calculate the probability only of the node X_3, the rest of the nodes are marginalized out from Eq.(2). So, $P(X_3)$ will be given as follows,

$$P(X_3) = \sum_{X_1, X_2, X_4} P(X_1)P(X_2) \times P(X_3 \mid X_1, X_2)P(X_4 \mid X_1) \tag{3}$$

Marginalization is a distributive process to calculate the total probability of a variable of interest by the summation of the products of all the possible combinations of local joint probabilities. This process allows to isolate the probability of the parameter of interest and to remove the rest of the variables from the joint probability distribution (Jensen and Nielsen 2007).

It is possible to remove variables that are not relevant to know the probability of the variable of interest, namely A. This process is called *variable elimination*. It consists of simply removing from the joint probability, variables that are outside the Markov blanket of A (i.e. variables that are parents or children of A, or sharing a children with A). The eliminated variables (those out of the Markov blanket), do not influence the probability measures of the variable of interest.

The type of information that can be adopted in this method involves real probability values (discrete nodes) or Gaussian distribution functions. The latter works with crisp value probabilities (Tolo et al. 2016b). However, this characteristic turns into the main drawback of this technique when the data comes in the way of continuous distributions. Nevertheless, this disadvantage is overcome with the use of structural reliability methods.

2.2 *Enhanced Bayesian networks*

Structural Reliability methods (SRMs) are used to work out the conditional probability tables of a BN containing both discrete and continuous random variables (enhanced Bayesian network) resulting in the reduction of the network to a traditional BN. Suppose the nodes in the network on Figure 2

a) correspond to independent random variables. Here X_1 is an interval node (with $f(X_1)$ its cumulative distribution function, CDF), X_2 and X_1, are discrete nodes (representing probability mass functions, $P(X_2)$, $P(X_4)$, respectively) and X_3 a continuous node (representing a CDF $f(X_3)$). According to (Straub and Der Kiureghian 2010) the enhance Bayesian network joint probability can be computed by approximating the equation below,

$$P(X_i) = \int_{X_1, X_3} f(X_1)f(X_3)P(X_2) \times P(X_4 \mid X_1, X_2, X_3)dX_1dX_3 \tag{4}$$

if the Markov properties are considered, node X_3 and X_1 are d-separated from X_2 since X_4 has not received any evidence yet. So, the joint probability of X_4 given X_2, from the equation above, can be written as,

$$P(X_4 \mid X_2) = \int_{X_1, X_3} f(X_1)f(X_3) \times P(X_4 \mid X_1, X_2, X_3)dX_1dX_3 \tag{5}$$

It has to be noticed that each entry in the conditional probability table of X_4 is defined by the domain $\Omega_{X_4, x_2}^{x_4}$ in the continuous space of X_1 and X_3 for a given value x_2 of the variable X_2. So the Eq. 5 can be further reduced as,

$$P(X_4 \mid X_2) = \int_{\Omega_{X_4, x_2}^{x_4}} f(X_1)f(X_3)dX_1dX_3 \tag{6}$$

The integral shown in Eq. 6 is equivalent to that of a reliability problem (Tolo et al. 2016b). Solving a structural reliability problem, i.e. approximating the system failure probability, a reduction of a network with continuous nodes to one with only discrete probability values is obtained.

There are certainly several approaches for the solution of problems like that shown in Eq. 6. These methods range from numerical approximations like Monte Carlo simulations, to the well-known and widely used *first-order* and *second-order reliability methods* (Hasofer and Lind 1974). Moreover, some advanced sampling techniques like Importance Sampling, Stratified Sampling or Advanced Line

Sampling (Hasofer and Lind 1974), among others, have been used as an alternative to the computationally expensive numerical approximations.

2.3 Credal networks

Credal networks can be referred as an extension of Bayesian networks to manage intervals of discrete probability values representing the lack of information and uncertainty about the variables involved. A Bayesian network constructed exclusively with discrete nodes such that, only one probability value is associated with the state of the variable. Such a state can belong to the variable itself, in the case of roots, or can be conditioned on the parents of that node, in the case of children. However, in a Credal network, probabilities are presented in the form of intervals that are associated with probabilistic inequalities. In this manner, a Credal network will represent the different variable states, each of them associated with one specific probability value inside the interval, of the same Bayesian network (Tolo et al. 2018). The graphical structure of such a network is the same as the Bayesian case, as well as the Markov blanket concept and d-separation of nodes. Nevertheless, the probability of a variable x is indicated in the form of the so-called credal set $K(X)$, whilst the set of joint probability measurements $P(X_i \mid pa(X_i))$ is named a joint credal set, given by $K(X_i \mid pa(X_i))$ (Cozman 2000). So, two different interval probabilities of variables X and Y (where Y is the complement of X) can be characterized by their upper and lower bounds, $[\bar{p}(X), \underline{p}(X)]$ and $[\underline{Y} = 1 - \bar{p}(X), \bar{Y} = 1 - \underline{p}(X)] \, \epsilon [0\,1]$, respectively.

2.4 Probability boxes

Probability boxes (or p-boxes) allow making fewer assumptions about the values used in the study when correlations of the variables employed in the study are ignored due to the effect of aleatory and epistemic uncertainties. A p-box is specified for a random variable X by the interval bounds $\left[\underline{F}, \bar{F}\right]$ on a cumulative distribution function F with values between 0 and 1, such that $\underline{F}(X) \le F(X) \le \bar{F}(X)$ (Ferson et al. 2003). If a probability measure $\underline{p}$ (since it is the lower bound of that measure) for the random variable X_1 is given, the lower, $\underline{F}(X_1)$, and upper $\bar{F}(X_1)$, bounds of the p-box can be computed as follows (Walley 1991),

$$\underline{F}(X_1) = \underline{p}(X_i \le X_1), \bar{F}(X_1) = 1 - \underline{p}(X_i > X_1) \quad (7)$$

The p-box has a dual interpretation, i.e. $\bar{F}$, represents the probability (CDF axis) upper bound and quantile (x-values axis) lower bound. The opposite happens with $\underline{F}$. Therefore, this concept is applicable in cases of imprecise continuous probabilistic distributions and two types can be differentiated, parametric non-parametric p-boxes. A parametric p-box is defined when the shape of the probability distribution is known, but there is no precise information about the parameters of that distribution. The non-parametric case is rarer but can exist especially when an experiment has been performed and a set of measurements was obtained. It occurs when parameters regarding the probabilistic distribution, e.g. mean and variance, of a variable, are known but no information about the type of distribution is available (Ferson et al. 2003).

2.5 Computational tool

The open source software OpenCossan exploits the Object-Oriented programming paradigm that Matlab offers with the use of *classes* (entities containing the data and the functions or methods that the members of the same class have in common) and *objects* (instances of a class). This basis is employed in order to efficiently provide solutions to problems regarding uncertainty quantification, sensitivity and reliability analysis, robust design, among others (Patelli 2015). The object-oriented methodology allows reutilizing parts of code to create more complex objects in a systematic and condensed way. OpenCossan offers a wide flexibility to integrate new methods that enhance, improve or complement the current tools available in this software. This opens the gate for new developments that enrich the software robustness to provide solutions.

Within the framework of OpenCossan, three main toolboxes can be used for Bayesian networks. These are, *BayesianNetwork*, *EnhancedBayesianNetwork* and, *CredalNetwork*. The first one can be used in cases were variables only correspond to crisp probability values whilst the second one, also considers continuous probability distributions and bounded variables. The third toolbox is the one chosen for this study, since it allows working with continuous probabilities and interval variables representing imprecise probabilities. However, the graphical display process is done with the same method, *makeGraph* which is a class of *EnhancedBayesianNetwork*.

3 CASE STUDY: OSCILLATING WATER COLUMN

An Oscillating Water Column (OWC) is a type of the so-called wave energy converters that capture the energy that sea waves deposit once reaching the named structure. The popularity of this type of energy converters has increased over the last few decades, since it is an alternative for the clean energy production (Falcão 2010). The structure of an OWC consists of a chamber, partially
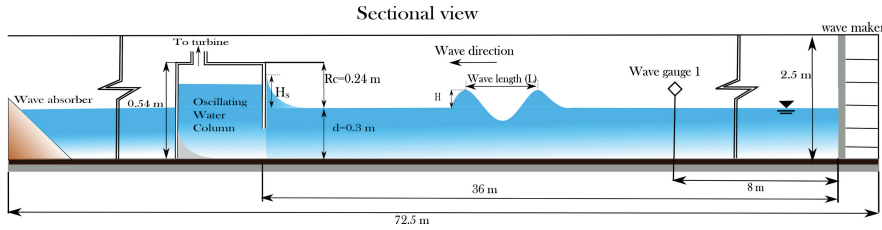
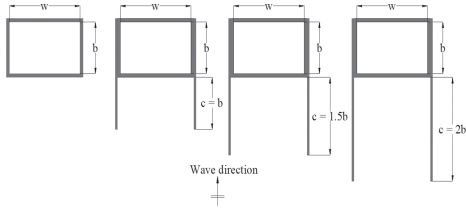Figure 3. Sectional view wave flume with OWC on the left-hand side.

Figure 4. Schematic layout from a top view of harbour walls with different lengths (Daniel Raj et al. 2016a).

submerged in the sea, generally with two orifices. One is typically at the top of the chamber inside which turbine is placed, and the other one is below the water line facing the coming sea waves, see Figure 3. An entombed mass of water column formed inside the chamber oscillates as a result of the wave inside in the structure. This, in turn, drives an air flow through the orifice coupled to a turbine, thus generating electricity (Cruz 2008).

Since the system is always surrounded by sea waves (especially when it is built away from the shoreline), it is susceptible to overtopping events that can cause serious damages to the structure, particularly in the operational mode. Horizontal sliding, overturning, scouring and collapsing of the structure may be possible during extreme sea conditions (Cruz 2008). It is preferred to study such overtopping possibilities to be more on the conservative side. Generally, a conventional-OWC device is constructed with adequate height such that the overtopping event is quite not possible. As the addition of harbor walls increases the wave amplification, the possibility of an overtopping event should be carefully addressed.

A very simple Bayesian network was built to test the inclusion of imprecise probability nodes in the computational toolbox based on the OpenCossan software. The network represents the components involved in an experimental work carried out by Daniel Raj and his team (Daniel Raj, D. et al, 2016) at the Indian Institute of Technology Madras in India, to study the influence of harbour walls of an OWC on its energy efficiency characteristics. The present network is used to provide an assessment of the risk of structure overtopping triggered by the waves generated in the laboratory.

3.1 *Description of the experiment*

The experimental arrangement was 72.5 m long, 2 m wide, and with a deep wave flume of 2.5 m, please refer to Figure 3. The scaled OWC model was 0.540 *m* height, in a 1:20 ratio from a real prototype. To generate random water waves (it is able to reproduce either shallow or deep water waves) the flume is equipped with a wave maker system to generate waves with steepness characteristics within the limits of the operational range of the system. The generated waves for this experiment covered a range of relative water depths, d/L, from 0.074 to 0.23 and a wave steepness, H/L, from 0.0074 to 0.065. Here, d denotes water depth, H corresponds to the wave height (in metres) registered from the first wave gauge 1 (situated at 8 *m* from the water generator), and L denotes wavelength (in metres). The crest periods, T_p, adopted were from 1 to 2.5 *s* in 0.25 *s* intervals. For more information about this experiment please refer to (Daniel Raj et al. 2016a), (Daniel Raj et al. 2016b).

The experiment was carried out in two stages; the first stage involved the identification of efficient resonating length of harbour wall, as seen in Figure 4, which enhances the energy conversion capacity of the system. Four testing criteria have been chosen in the first stage, so as to investigate the effect of projecting sidewalls length on the efficiency of the OWC. Among them, one is without the sidewalls (conventional) and rest with the projecting sidewalls in three criteria such as c/b of 1, 1.5 and 2. In the second stage, the effect of the harbour walls on each side of the OWC was studied by varying the angle of the harbour walls within the range of $[4\pi/8, 7\pi/8]$ at intervals of $\pi/8$ with respect to the front lip wall of OWC, as seen in Figure 5. This angle variation is called *wall inclination* from now on. The wall length, *c*, was maintained constant in this stage at its optimal configuration identified from the first stage of experiment.

3.2 *The Bayesian network*

The goal of the experiment was to analyse the influence of different configurations of harbour walls in order to modify the resonance frequency of the water wave coming towards the structure so
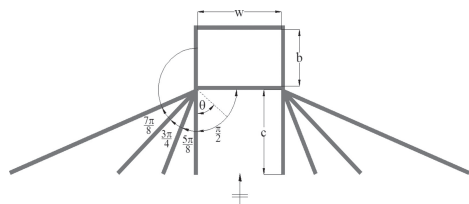
Figure 5. Schematic layout different harbour wall inclinations (top view) (Daniel Raj et al. 2016b).
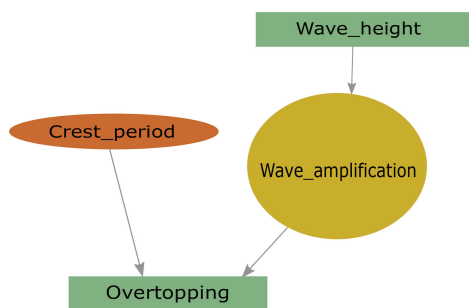

Figure 6. Enhanced Bayesian network for OWC experiment.

that the wave amplification was maximized. The maximization of wave amplification poses a potential risk of structure overtopping that can damage the rearward face of the experiment equipment or to bring unexpected consequences. The probability of having this event is quantified through the network presented in Figure 6 that takes into account the epistemic uncertainty characterizing the measurements of the amplified waves running up the structure. In addition to that, the wave properties, as well as the OWC configurations studied in the experiment, were considered as influencing factors for the occurrence of structure overtopping.

Since imprecise variables are employed in this simple study, the network for this problem is defined in the toolbox for Credal networks node by node, in the form of computational objects. Once all the variables are defined, an object of the class *CredalNetwork* is created to store the information regarding the nodes on the network, so further calculations can be performed. Then, the method *makeGraph* is invoked to display the network shown in Figure 6. The nodes are defined as follows.

Different harbour wall inclinations and lengths were tested to study their effects on wave amplification. In order to select the case to be studied with the BN, the *Experiment_case* node was defined as an interval node containing the maximum and minimum length ratio, for the variation of harbour wall lengths experiment, or the maximum and minimum angle for the remaining case. The values in the intervals are defined in such way that they can cover all the possible values used in the experiment.

Wave properties were used in the Bayesian network as follows, the *Crest_period* interval node contains the information of the mean level-up-to-down-crossing time of the incident waves. The crest periods were changed from 1 to 2.5 s intervals of 0.25 s. Meanwhile, the *Wave_height* node has the information of the wave height used in the experiments. The waves studied were 0.03, 0.06 and 0.09 m high for each of the harbour wall configurations. It has to be noted that the *Wave_height* will influence directly how high the amplified wave is.

According to some authors, as shown by Allsop review (Allsop et al. 1985), wave amplification phenomenon (or wave run-up) in steep structures slopes follows approximately Rayleigh distribution. For this reason, the *Wave_amplification* is assumed to follow this probability distribution with a scale parameter based on the experimental results obtained from the wave amplification measurements. However, the parameters to define this variable in the experiment are uncertain due to the lack of probabilistic data (i.e. only one measurement was taken for a given value of wave crest period, wall inclination or length). So, the use of p-boxes becomes handy. The *Wave_amplification* p-boxes were defined in such manner that all the possible values tested experimentally were enclosed in the p-box of each case. These values are presented in the Table 1, for the given harbour wall length and inclination configurations as well as each of the different wave heights, respectively.

In this work, the Owen equation proposed, by Mase (Mase et al. 2013), for overtopping discharge ratio Q is used to describe the *Overtopping* node and the probability of exceedance, $P(Z)$, of the maximum admissible wave overtopping is given as:

$$Z = Q_{max} - \left(A T_p g H_s \right) e^{-B \frac{R_c}{T_p \sqrt{g H_s}}} \tag{8}$$

where A and B are dimensionless empirical coefficients depending on the slope ratio of the structure. In this case, A and B are given as 0.0079 and 20.12 (Owen 1980), respectively. H_s is considered as the significant wave height of the amplified waves in meters, g the gravitational acceleration and, R_c the structure freeboard. If the condition

Table 1. Wave amplification p-boxes defined with the Rayleigh-distribution scale parameter for harbour-wall length (a_{length}) and inclination ($a_{inclination}$) experiments.

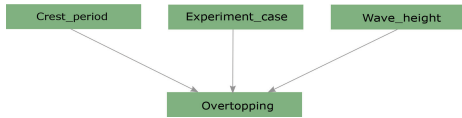| Wave height (m)	a_{length}	$a_{inclination}$	Distribution
0.03	[0.038, 0.077]	[0.045, 0.08]	Rayleigh
0.06	[0.082, 0.142]	[0.11, 0.157]	Rayleigh
0.09	[0.121, 0.213]	[0.183, 0.237]	Rayleigh

Figure 7. Credal network after reduction process for OWC experiment.

Table 2. Occurrence of structure overtopping for wall length case. The bounded values in *Overtopping*$_{a,b,c}$, correspond to the *Wave_amplification* p-boxes in each of the *Wave_height* cases, i.e. 0.03, 0.06 and 0.09 *m*, respectively.

Period (*s*)	*Overtopping*$_a$	*Overtopping*$_b$	*Overtopping*$_c$
1	[0, 0.013]	[0.016, 0.273]	[0.165, 0.555]
1.25	[0, 0.011]	[0.19, 0.279]	[0.165, 0.563]
1.5	[0, 0.01]	[0.021, 0.272]	[0.164, 0.551]
1.75	[0, 0.012]	[0.2, 0.269]	[0.159, 0.557]
2	[0, 0.01]	[0.02, 0.268]	[0.156, 0.562]
2.25	[0, 0.013]	[0.2, 0.275]	[0.161, 0.555]
2.5	[0, 0.009]	[0.2, 0.263]	[0.164, 0.557]

$P(Z \leq 0)$ is exceeded the event is considered a failure of the system meaning a wave overtopping of the OWC for a given maximum overtopping rate Q_{max}. Once all the nodes are defined in the toolbox, the network is reduced using the Adaptive Line Sampling method. In the Figure 7 can be observed the reduced network used for this study. The overtopping events are given in the continuous space of the *Wave_amplification* node for each given state of *Experiment_case*.

In this simple network there is only one p-box employed per simulation (regarding the *Wave_amplification* node), so the lower bound would correspond to the minimum amplification that a wave under those conditions can have. Thus, overtopping probability resulting from this calculation will be the minimum probability that this variable can have, assuming monotonicity in the system. In order words, given the lower bound of *Wave_amplification* the lower bound of *Overtopping* will be found. The same reasoning is used for the case of the upper bounds.

The lower and upper bounds of the overtopping probability are displayed in the Table 2, for the case of harbour wall length experiment, and in Table 3, for the case of harbour wall inclination experiment. In each *Wave Height* column in the tables are stored the overtopping probability bounds computed for each height, i.e., 0.03, 0.06 and 0.09 m for each of the different crest periods (T_p). This was done in order to provide a combinatorial study of the overtopping probability change for all the experimental set ups.

It can be observed that T_p does not influence significantly the structure overtopping occurrence. In fact, changes in overtopping results obtained from

Table 3. Occurrence of structure overtopping for wall inclination case. The bounded values in *Overtopping*$_{a,b,c}$, correspond to the *Wave_amplification* p-boxes in each of the *Wave_height* cases, i.e. 0.03, 0.06 and 0.09 *m*, respectively.

Period (*s*)	*Overtopping*$_a$	*Overtopping*$_b$	*Overtopping*$_c$
1	[0, 0.017]	[0.115, 0.34]	[0.455, 0.623]
1.25	[0, 0.015]	[0.109, 0.342]	[0.451, 0.623]
1.5	[0, 0.015]	[0.111, 0.337]	[0.456, 0.631]
1.75	[0, 0.017]	[0.114, 0.343]	[0.453, 0.629]
2	[0, 0.015]	[0.11, 0.341]	[0.449, 0.621]
2.25	[0, 0.017]	[0.111, 0.34]	[0.455, 0.623]
2.5	[0, 0.017]	[0.113, 0.341]	[0.455, 0.621]

the different crest periods may be due to the randomness of the simulation functions used. A major influence comes mainly from the *Wave_height* and *Experiment_case*. It is logical that higher waves will increase the height of the amplified wave, in consequence, the probability of an overtopping event will increase as well. Comparing the results from both experiment arrangements, it can be resolved that the wall inclination factor increases the most the probability of overtopping occurrence. The wave amplification factor of 0.237 corresponding to a wall inclination of $3\pi/4$ with 0.09 *m* wave height can be referred as the worse case scenario. This can be useful when the values are only scaled up to prototype dimensions. For instance, a structure overtopping assessment can be provided without performing any experimental work, as long as the physical behavior of the variables preserves the same probabilistic distribution.

4 CONCLUSIONS

From the case study presented here, an approximation to the worst case scenario was found with the use of a very simple Bayesian network with the implementation of p-boxes and interval variables (credal sets). The original network containing continuous, interval and discrete variables was reduced using structural reliability methods (adaptive line sampling) to a simpler network containing only crisp probabilities. Since the values used as input in the p-boxes and in the interval variables contain all the possible cases in the experiment, none of the those should overcome the maximum probability given. So, epistemic uncertainty affecting this experiment can be quantified with this method. If specific data (different from that considered here) regarding any of the variables in the experiment, inside the bounds, are given, the overtopping probability results can be provided for that specific case and they will remain below or be equivalent than the maximum value achieved.

The implementation of imprecise data in Bayesian networks is a very necessary tool in engineer-

ing. However, the methods currently available are computationally expensive. Most importantly, when systems with a large number of variables are studied, the algorithms may suffer combinatorial explosion. Adaptive line sampling will be further studied to deal with bounded failure probabilities. In addition to that, random set theory (combined with Subset Simulation to improve calculation speed) appears as a good option to the limitations of the current tool, since the latter method is specially useful for small probability cases and good performance in both low- and high-dimensional spaces as well as in nonlinear limit functions.

REFERENCES

Allsop, N.W.H., L. Franco, & P.J. Hawkes (1985). Wave run-uP on steep slopes-A literature review. *Hydraulics Research 1*, 1–28.

Castillo, E., C. Solares, & P. Gomez (1997). Tail uncertainty analysis in complex systems. *Artificial Intelligence 96*(2), 395–419.

Cozman, F.G. (2000). Credal networks. *Artificial Intelligence 120*(2), 199–233.

Cruz, J. (2008). *Ocean Wave Energy Current Status and Future Perspectives*. Bristol, U.K.: Springer Berlin Heidelberg.

Daniel Raj, D., V. Sundar, & S. Sannasiraj (2016a). Experimental Investigation on Optimizing the Projecting Sidewalls of an Oscillating Water Column. *Proc., 9th Int. Conf. on Coastal Port Engineering in Developing Countries (Pianc-Copedec IX) IX*(c), 1–11.

Daniel Raj, D., V. Sundar, & S. Sannasiraj (2016b). Optimizing the harbor wall inclination in an oscillating water column. In *3rd International Conference on Coastal Zone Engineering And Management In The Middle East*, Dubai,UAE.

de Angelis, M., E. Patelli, & M. Beer (2015). Advanced Line Sampling for efficient robust reliability analysis. *Structural Safety 52*(PB), 170–182.

Falcão, A.F.d.O. (2010). Wave energy utilization: A review of the technologies. *Renewable and Sustainable Energy Reviews 14*(3), 899–918.

Ferson, S., V. Kreinovich, L. Ginzburg, D.S. Myers, & K. Sentz (2003). Constructing Probability Boxes and DempsterShafer Structures. *Sandia National Laboratories* (January), 143.

Hamada, M.S., A.G. Wilson, C.S. Reese, & H.F. Martz (2008). *Bayesian Reliability*. Springer Series in Statistics. New York, NY: Springer New York.

Hasofer, a. M. & N.C. Lind (1974). An exact and invariant first order reliability format. *Journal of the Engineering Mechanics Division ASCE 100*(August), 111–121.

Hosseini, S. & K. Barker (2016). A Bayesian network model for resilience-based supplier selection. *International Journal of Production Economics 180*, 68–87.

Hudson, L.D., B.S. Ware, K. Blackmond, & S.M. Mahoney (2002). An Application of Bayesian Networks to Antiterrorism Risk Management for Military Planners. *Digital Sand-box, Inc.*, 8.

Jensen, F.V. & T.D. Nielsen (2007). *Bayesian Networks and Decision Graphs*. Information Science and Statistics. New York, NY: Springer New York.

Kang, C. & M. Golay (1999, jul). A Bayesian belief network-based advisory system for operational availability focused diagnosis of complex nuclear power systems. *Expert Systems with Applications 17*(1), 21–32.

Korb, K.B. & A.E. Nicholson (2004). *Bayesian Artificial Intelligence*. Boca Raton, U.S.A.: Chapman & Hall/CRC.

Mase, H., T. Tamada, T. Yasuda, T.S. Hedges, & M.T. Reis (2013). Wave Runup and Overtopping at Seawalls Built on Land and in Very Shallow Water. *Journal of Waterway, Port, Coastal, and Ocean Engineering 139*(October), 346–357.

Owen, M.W. (1980). Design of seawalls allowing for wave over-topping. *Report Ex 924*(June), 39.

Patelli, E. (2015). COSSAN: A Multidisciplinary Software Suite for Uncertainty Quantification and Risk Management. In R. Ghanem (Ed.), *Handbook of Uncertainty Quantificatio*, pp. 1–69. Switzerland: Springer International Publishing.

Patelli, E., S. Tolo, H. George-Williams, J. Sadeghi, R. Rocchetta, M. de Angelis, & M. Broggi (2018). OpenCossan 2.0: an efficient computational toolbox for risk, reliability and resilience analysis. In *Proceedings of the joint ICVRAM ISUMA UNCERTAINTIES conference*.

Pearl, J. (1991). Probabilistic Reasoning in Intelligent Systems.

Spiegelhalter, D.J. (1987). A unified approach to imprecision and sensitivity of beliefs in expert systems. In *Uncertainty in Artificial Intelligence*, Seattle, WA, USA, pp. 199–208.

Straub, D. & A. Der Kiureghian (2010). Bayesian Network Enhanced with Structural Reliability Methods: Methodology. *Journal of Engineering Mechanics*.

Tolo, S., E. Patelli, & M. Beer (2016a, jun). Risk Assessment of Spent Nuclear Fuel Facilities Considering Climate Change. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering 3*(2), G4016003.

Tolo, S., E. Patelli, & M. Beer (2016b). Robust vulnerability analysis of nuclear facilities subject to external hazards. *Stochastic Environmental Research and Risk Assessment*, 1–24.

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

Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities* (1st ed.). London, U.K.: Champan and Hall.

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