Sensitivity analysis of Bayesian networks to parameters of the conditional probability model using a Beta regression approach

Jeremy Rohmer, Pierre Gehl

Sensitivity analysis of Bayesian networks to parameters of the conditional probability model using a Beta regression approach

Jeremy Rohmer[1,*], Pierre Gehl[2]

Affiliations :

1: BRGM, 3 av. C. Guillemin - 45060 Orléans Cedex 2 – France, E-mail : j.rohmer@brgm.fr

2: BRGM, 3 av. C. Guillemin - 45060 Orléans Cedex 2 – France, E-mail : p.gehl@brgm.fr

*: Correspondence to: J. Rohmer

Abstract

Ensuring the validity and credibility of Bayesian Belief Network (BBN) as a modelling tool for expert systems requires appropriate methods for sensitivity analysis (SA), in order to test the robustness of the BBN diagnostic and prognostic with respect to the parameterisation of the conditional probability model (CPM). Yet, the most widely used techniques (based on sensitivity functions for discrete BBNs) only provide a local insight on the CPM influence, i.e. by varying only one CPM parameter at a time (or a few of them) while keeping the other ones unchanged. To overcome this limitation, the present study proposes an approach for global SA relying on Beta Regression using gradient boosting (potentially combined with stability selection analysis): it presents the benefit of keeping the presentation intuitive through a graph-based approach, while being applicable to a large number of CPM parameters. The implementation of this approach is investigated for three cases, which cover a large spectrum of situations: (1) a small discrete BBN, used to capture medical knowledge, demonstrates the proposed approach; (2) a linear Gaussian BBN, used to assess the damage of reinforced concrete structures, exemplifies a case where the number of parameters is too large to be easily processed and interpreted (>40 parameters); (3) a discrete BBN, used for reliability analysis of nuclear power plant, exemplifies a case where analytical solutions for sensitivity can hardly be

derived. Finally, provided that the validity of the BBR model is carefully checked, we show that the proposed approach can provide richer information than traditional SA methods at different levels: (i) it selects the most influential parameters; (ii) it provides the functional relation between the CPM parameter and the result of the probabilistic query; and (iii) it identifies how the CPM parameters can lead to situations of high probability, while quantifying the confidence in the occurrence of these situations.

Keywords: Bayesian Network; Sensitivity; Distributional Regression; Beta Distribution; Gradient Boosting; Stability Selection.

1 Introduction

Bayesian Belief Network (BBN) is widely recognized as a valuable modelling tool for expert systems (e.g. Russel et al., 2003). It has been applied in various complex domains, like ecosystems (Milns et al., 2010), genetics and biology (Scutari et al., 2014), industry (Weber et al., 2012), finance forecasting (Malagrino et al., 2018), marine safety (Hänninen et al., 2014), nuclear power plants (Kwag & Gupta 2017), coastal systems (Jäger et al., 2018), multi-hazard risk assessments (Gehl and D'Ayala, 2016), etc. An expert system has the ability to represent and reason on knowledge with the purpose of solving problems and giving advice (as defined by Jackson, 1999), by relying on three components (i.e. knowledge base, observation base, inference engine). Each of these components can benefit from the key features of BBN: (1) its capability to represent expert knowledge and to combine and integrate expert knowledge with information from any kind of sources, including experimental data, historical data, results from numerical simulations, etc.; (2) its high flexibility to model any causal relationships and to explicitly display the relationship among variables using a network-based approach, which can intuitively be understood by experts (Wiegerinck et al., 2010); 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 (the evidence variables) are observed. For instance, probabilistic queries in reliability assessments may correspond to finding the probability values of some failure cause given the observed damage level of the considered system. See Appendix A for a brief overview. The interested reader can also refer to Jensen (2001) for a complete introduction to BBN.

Formally, BBN is based on the graphical representation of the probabilistic relations among random variables by means of a directed acyclic graph composed of nodes (i.e. the states of the random variables) and arcs (i.e. dependency between nodes). See an example in Fig. 2A. The nodes connected by an arc are called the parent 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. The probabilistic dependence (i.e. the relation cause-effect) is represented via a table called a Conditional Probability Table (CPT) when the variables *X* (nodes) are discrete. In this case, the CPT entries correspond to the probability value $\mathrm{P}(X_i = k|\,\mathrm{Pa}(X_i) = j)$ where k denotes the k^{th} possible level (either a discrete value or a category) that node X_i can take given that its parents $\mathrm{Pa}(X_i)$ takes the j^{th} possible level. When the variables are continuous, the conditional distribution given its parents can typically be represented by means of a continuous probability distribution. The most popular model is the Gaussian distribution, as follows:

$$\mathrm{P}(X_i = y|\,\mathrm{Pa}(X_i) = \boldsymbol{x}) = G(y|m_0 + Z\boldsymbol{x}, S) \qquad \text{(Eq. 1)}$$

where *G* is the Gaussian probability distribution whose mean is parameterized by a linear regression model with intercept m_0 and regression coefficients *Z*, and *S* is the conditional variance.

When the nodes are both discrete and continuous, different hybrid techniques exist in the literature (e.g., Murphy 1999; Shenoy, 2006; Beuzen et al., 2018).

Whatever the nature of the nodes, the pillar of any BBN-based results (either evidence propagation or inference) is the specification of the parameters of the conditional probability model (denoted CPM), i.e. the CPT entries or the regression coefficients of the linear Gaussian regression model. This process is recognized in the literature as one of the most delicate part of the BBN development (e.g., Chen & Pollino 2012; Druzdzel & van der Gaag 2000, etc.), which raises the question of confidence in the diagnosis or prognosis derived from the BBN-based expert systems (see discussion by Pitchforth & Mengersen, 2013). In the validation framework of BBN, sensitivity analysis (SA) tools play a major role in order to study how the output of a

model varies with variation of the CPM 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 & van der Gaag 2002; Laskey et al. 1995).

For discrete BBNs, a widespread SA method relies on the use of sensitivity functions (Coupé & van der Gaag 2002; Castillo et al., 1997), which describe how the considered output probability varies as one CPT entry value is changed. Some recent extensions have been proposed to conduct multi-way SA like the framework by Leonelli et al. (2017), i.e. SA when several CPT entries are allowed to vary. Effects of parameter changes can be described by the Chan–Darwiche (CD) distances (Chan–Darwiche 2002; 2005). The CD distance is used to quantify global changes by measuring how the overall distribution behaves when one (or more) parameter is varied. Likewise, SA approaches have been developed for continuous BBNs (mostly based on linear Gaussian regression) either based on computing partial derivatives (Castillo & Kjærulff 2003) or based on the use of divergence measures (Gómez-Villegas et al., 2007), with generalisation to multi-way SA (Gómez-Villegas et al., 2013).

Yet, several limitations of the existing SA methods exist.

(1) The SA methods differ depending on the type of BBN (discrete, continuous, hybrid);

(2) Though simple and efficient to implement, the approach based on sensitivity functions (combined with CD-distance analysis) remains local, because the values of only one parameter of the CPM is varied, while the other ones are kept constant. Multi-way SA methods have been proposed, but can rapidly become intractable. The SA procedure for Gaussian BBN presents the same limitation as being based on partial derivatives;

(3) The SA is usually performed by focusing on one type of co-variation of the different parameters (like proportional, uniform and order-preserving co-variation, see e.g., Renooij, 2014).

In the present study, we propose an alternative approach for SA, whose characteristics should complement the existing ones and overcome the afore-described limitations. The proposed approach should be:

- Global: since the sensitivity of any BBN-based probability of interest is affected by multiple variations in the CPM parameters, the sensitivity is studied in a global manner and the different parameters are allowed to be varied all together;

- Generic: it should apply to any kind of BBN, i.e. discrete, Gaussian or hybrid, with limited restriction on the type of variations of the CPM parameters;
- Robust to the number of parameters: the number of parameters can rapidly increase (in relation with number of nodes), typically reaching several dozens even for a moderate number of BBN nodes, which can hamper the interpretation of any global SA;
- Concise: the presentation of the SA results should be as intuitive as the sensitivity function method by using a graph-based approach.

To do so, we address the problem of SA for BBN with the viewpoint of regression by considering the BBN-based probabilistic queries P as the predictand and the CPM parameters as the predictors (denoted $\boldsymbol{C}$). In contrast to the classical regression model, which restricts the analysis to the expected value of P as a function of $\boldsymbol{C}$, we aim at estimating the full probabilistic variation of P by taking advantages of recent developments for distributional regression (e.g., Koenker et al., 2013). In the subsequent sections, we first describe the principles of the proposed method together with the implementation details (Sect. 2). In Sect. 3, we apply the approach to a small discrete BBN (6 nodes) used to capture medical knowledge (adapted from Cooper 1984) to exemplify the potentialities of the proposed approach. Then, in Sect. 4, two real cases are used to investigate the applicability of the proposed approach. Finally, Sect. 5 discusses the strengths and weaknesses of the proposed approach from the methodological and operational viewpoints.

2 Statistical methods

In this section, we first describe the principles underlying the development of the proposed approach (Sect. 2.1). We then provide the justifications for using Beta regression (Sect. 2.2) for sensitivity analysis of BBNs. Sect. 2.3 gives further technical details on the key ingredient of the procedure, namely the Boosted Beta Regression (BBR) technique. Finally, we describe how to check the adequacy of the BBR model to fit the data (Sect. 2.3).

2.1 Overall procedure

The different steps of the procedure hold as follows:

- Step1: generate perturbations of network's entry values $\boldsymbol{C}$ (i.e. the predictors) using, for instance, some random sampling techniques. For binary nodes, the CPT entries can be

randomly perturbed using truncated Gaussian distributions (van der Gaag et al., 2013) ; for multi-level discrete nodes, it can be based on Dirichlet distributions (e.g., Young et al., 2009) ; for continuous Gaussian nodes, truncated Gaussian distributions can also be used if the purpose is to study the robustness to small-to-moderate variations; see also Gómez-Villegas et al. (2014) for alternative possible probability laws;

- Step 2: estimate the query probability of interest *P* derived from the inferences using the BBN. This probability value is the predictand;
- Step 3: establish the link between ***C*** and *P* using a regression model. We propose to rely on Beta regression models (see Sect. 2.1) using boosting-based fitting procedure (Sect. 2.2) to deal with the potentially large number of predictors;
- Step 4: check the adequacy of the Beta model by checking that the residuals are well approximated by the standard normal distribution (see Sect. 2.4).

2.2 Use of Beta regression for BBN sensitivity analysis

We choose to rely on the Beta regression model (e.g., Ferrari and Cribari-Neto, 2004) due to different difficulties inherent to our case. The first difficulty is related to the nature of the predictand, which lies within the interval [0 ; 1]. This prevents from a direct application of ordinary least squares (linear) regression techniques, because bounded data (such as rates and proportions- or here probability values) are typically heteroskedastic (e.g., Cribari-Neto and Zeileis, 2010), which means that their variance depends on the predictors' values.

Second, the distributions of such data are typically asymmetric, and the normal assumption underlying standard regression models might not be valid in our case. An alternative approach is to use regression models that are based on a probability distribution suitable for handling bounded data. A good candidate is the Beta law, whose density distribution *d* is defined as follows:

$$d(P,a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} P^{a-1}(1-P)^{b-1} \qquad \text{(Eq. 2)}$$

where Γ is the gamma function, (*a*,*b*) are the shape parameters. In the following, we preferably use an alternative parametrisation (μ,σ), where $\mu = \frac{a}{a+b}$ is the mean of *P* and $\sigma = \frac{1}{(a+b+1)^{1/2}}$ is

related to the variance of *P*, i.e. $\mu(1-\mu)\sigma^2$ (e.g., Schmid et al., 2013) The σ-parameter allows covering a large spectrum of density shapes as shown in Fig. 1 given different μ and σ values.

An additional difficulty is the inclusion of the boundary values at 0 and 1, because the density *d* in Eq. 2 is not defined at these values. This can be overcome by means of a simple transformation of *P* (Smithson and Verkuilen, 2006) as follows:

$$(P(n-1)+0.5)/n \qquad \text{(Eq. 3)}$$

where *n* is the number of samples (i.e. the number perturbations performed to study the sensitivity of *P*).

The joint analysis of both Beta parameters (μ,σ) enables the BBN practitioners to investigate the sensitivity of the BBN with respect to two levels. The first level is related to the evolution of μ as a function of the CPM parameters. These evolutions can be interpreted in a similar manner as the sensitivity functions used in the traditional SA for BBN, but it is worth noting that these functions are global; in the sense that they are constructed based on samples where all CPM parameters are allowed to vary all together.

In the present study, we propose to introduce a second level of analysis by allowing the Beta σ parameter to vary as well. This is justified as follows:

- when the Beta σ parameter is incorrectly taken to be constant, some efficiency loss in the fitting process has been reported in the literature (e.g., Bayer and Cribari-Neto, 2017 and references therein);
- a potentially large number of predictors have to be handled in the regression model and we propose to select only one part of the CPM parameters that have influence on the BBN probabilistic query *P* (see Sect. 2.3). This means that for fixed values of these selected parameters, values of *P* can still vary due to perturbations of the CPM parameters that were left out. The σ parameter is here used to describe this type of variability;
- some algorithms for the estimation of BBN probabilistic queries are based on random sampling (like particle filter using logic sampling by Koller and Friedman (2009)), which may introduce some noise in the *P* estimates;

- studying the evolution of σ enables the BBN practitioners to identify the sources of data variability (e.g., Smyth and Verbyla, 1999) i.e. to study the influence on the uncertainty of *P*. In this manner, the BBN practitioners can identify different probabilistic regimes; e.g. situations where the probability of interest might switch from low to high values, i.e. situations of low to high risk. Such situations can be highlighted by low and high values of μ. Values of σ are then useful to indicate the confidence in the occurrence of such situations. For instance, high value of μ (high average value of *P*) together with low value of σ (low variance) provide strong evidence that changes in the considered parameter(s) might surely lead to situations of high probability. High values for both Beta parameters show, however, that the situation of high probability might occur but only with low confidence (i.e. high uncertainty).

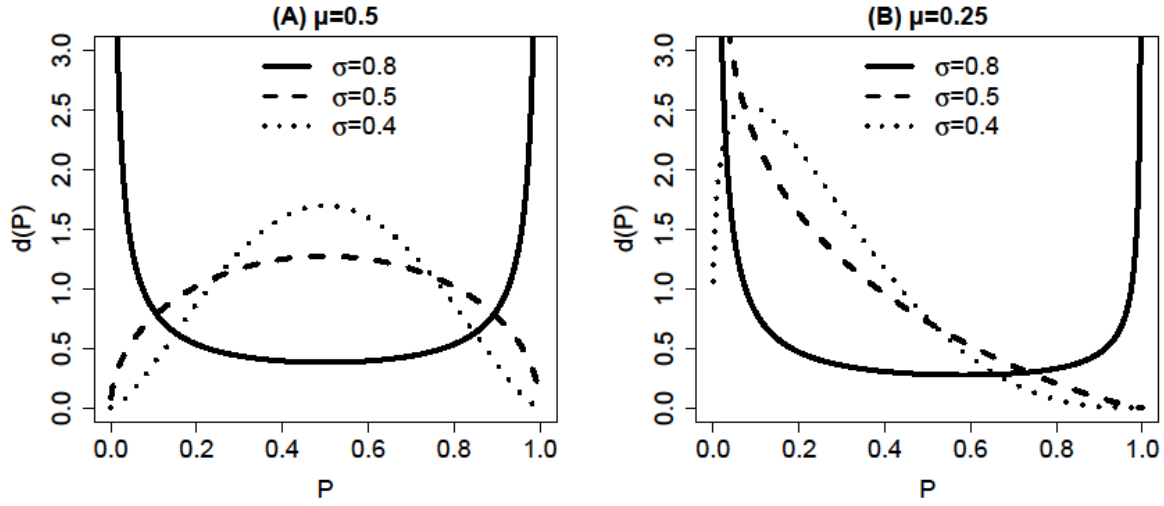


Figure 1. Probability density functions for the beta law. (A) mean μ=0.5; (B) mean μ=0.25.

Given the transformation of *P* (Eq. 3), BBN sensitivity analysis can be performed using Beta regression models (step 3 of the procedure) whose parameters are fitted using, for instance, maximum likelihood techniques (Cribari-Neto and Zeileis 2010). Yet, the functional form of the predictor-predictand relationship (e.g., quadratic or exponential) can hardly be specified in advance in our case (except for specific cases, like binary discrete BBN) and the relationships should preferably be learnt from the data. A possible option can rely on more advanced Beta distribution regression techniques, for instance using generalized additive models for location, scale and shape (GAMLSS, Rigby and Stasinopoulos, 2005), which allows deriving smooth

non-linear functional terms from the data, which correspond to the "partial effects" (more formally introduced in Sect. 2.3). These terms are the key ingredients for sensitivity analysis, because they hold the information of each parameter's individual effect on the considered Beta parameter.

The difficulty in our case is, however, the possible large number of predictors, which can typically exceed several tens in real cases. This situation imposes the use of techniques for variable selection during the fitting process. A possible option is the combination of GAMLSS with boosting-based approach (Mayr et al., 2012), which is detailed in the next section.

2.3 Boosted Beta Regression

In contrast to classical regression model, GAMLSS for Beta probability distribution aims at regressing the Beta parameters $\boldsymbol{\theta}=(\mu,\sigma)$ (or their transformation, e.g. via a log or a logit function) to the p predictor variables $\boldsymbol{C}=(c_1, c_2,\ldots, c_p)$, e.g., the CPT entries for discrete BBN or the regression coefficients of a Gaussian BBN. In the following, we restrict the analysis to a semi-parametric additive formulation as follows:

$$\theta = \eta_\theta(\boldsymbol{C}) = \beta_0 + \sum_{j=1}^{J} f_j(c_j|\beta_j) \qquad \text{(Eq. 4)}$$

where $\eta_\theta(.)$ is the link function that related the considered parameter $\boldsymbol{\theta}=(\mu,\sigma)$ with the predictor variables $\boldsymbol{C}$; $J \leq p$, β_0 is a constant and the functional term $f_j(.)$ corresponds to a univariate smooth non-linear model like regression penalized regression P-spline models (Eilers and Marx 1996) with parameters β_j. These functional terms (termed as partial effect) hold the information of each parameter's individual effect on the considered Beta parameter.

The fitting is performed using the *gamboostLSS* algorithm of Mayr et al. (2012), which uses the Beta log-likelihood function (termed as risk) as an optimization criterion based on the component-wise gradient boosting technique (Bühlmann and Hothorn 2007, see further details in Appendix B). The approach is termed as BBR model (Boosted Beta Regression).

One advantage of using boosting techniques is to perform variable selection during the fitting process, which allows screening the parameters $\boldsymbol{C}$ which hold most information with respect to the conditional distribution of P. This is performed by assessing the individual fits of each

predictor variable, and by updating only the coefficient of the best-fitting predictor variable in each iteration. Variable selection is carried out successively for the mean μ and for the σ parameter.

When the algorithm is stopped, the final model only contains the set of best-fitting predictors. The number of boosting iterations controls the smoothness of the non-linear effects. Low values lead to sparse models with smooth functional terms, whereas large values lead to more complex models with larger number of predictors and rougher functional terms. In practice, the selection of the stopping parameter can be carried out using cross-validation procedures in order to optimize the risk on observations left out (i.e. "out-of-bag") from the fitting process i.e. the out-of-bag risk, which corresponds here to the negative log-likelihood of the Beta distribution calculated for the "out-of-bag" samples. To avoid optimizing two different stopping iterations, i.e. one for each Beta parameter, the procedure can be enhanced using the noncyclic algorithm of Thomas et al. (2018), which allows reducing the optimisation problem from a multi-dimensional to a one-dimensional problem.

In some situations, the resulting BBR model can still remain too rich to be easily interpretable by BBN practitioners. This procedure can be completed by the stability selection analysis (described in Appendix C), which allows screening the most influential variables in the BBR model.

In summary, the boosting-based approach allows to both select a limited number of influential predictors among all network's entry values $\boldsymbol{C}$, and to derive the corresponding partial effects for each Beta parameter. Note that the identified predictors are not necessarily the same for μ and for σ, and can be unique or multiple. The final results of the procedure are the partial effects (Eq. 4), which can directly be used to analyze the sensitivity of the considered Beta parameter to $\boldsymbol{C}$ as illustrated on the application cases (Sect. 3 and 4), by considering two levels of analysis (respectively related to the best estimate of P using μ, and to the uncertainty of P using σ).

2.4 Model adequacy

Once the BBR model has been fitted, an important aspect is to check the model adequacy (step 4 of the procedure), i.e. how well the BBR model is appropriate to describe the randomly generated BBN-derived probabilities. This can be done by analysing the statistics of the residuals and checking whether their distribution is well approximated by the standard normal distribution (see e.g. Rigby and Stasinopoulos, 2005). Yet, contrary to ordinary least square

regression, the raw response residuals $r = P - \mu$ cannot be used, because they do not account for the heteroscedasticity of the model (i.e. the variance of P is a function of the Beta mean μ, see Sect. 2.1).

In the following, we propose to keep the analysis of r in order to give information regarding the capability of the Beta mean to explain P. However, to properly validate the use of the Beta model, we rely on alternative residuals' formulations (see e.g., Pereira, 2019 and references therein). We focus here on three of the most widely-used ones described in Appendix D. The normality of these residuals is investigated by means of the normal Q-Q plot and by computing the coefficient of determination R^2 as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{i=N}(\widehat{q_i} - q_i)^2}{\sum_{i=1}^{i=N}(q_i - \bar{q})^2} \quad \text{(Eq. 5)}$$

where N is the number of quantile levels; q_i is the quantile of standard normal distribution at the i^{th} level; $\bar{q}$ is the mean of the quantile of standard normal distribution over the levels i=1…N; $\widehat{q_i}$ is the quantile of BBR residual at the i^{th} level. The closer R^2 to one, the better the agreement between the BBR residuals' quantiles and the ones of the standard normal distribution, hence the more satisfactory the adequacy of the BBR model. Furthermore, studying the evolution of R^2 as function of the number of random perturbations of $\boldsymbol{C}$ provides an option to estimate the minimum number of required permutations for the BBR model to be valid.

It should however be underlined that the residuals are analysed with the objective of checking that the Beta distribution is an appropriate model to explain the BBN-derived probabilities. Using them for another objective, for instance to compare different probability model families (e.g. Gamma, Gaussian) or to perform predictions, is made difficult by the use of boosting algorithms (see Hofner et al., 2016: Sect. 5.4). For these purposes, the use of the out-of-bag risk is recommended.

3 Synthetic case study

In this section, we consider a BBN of small number of nodes (described in Sect 3.1) to exemplify the functionalities of the proposed approach (Sect. 3.2).

3.1 Description

We focus on the BBN adapted by van der Gaag et al. (2013) from Cooper (1984) in the field of oncology. The network (as depicted in Fig. 2A) is composed of 6 nodes and 6 arcs.

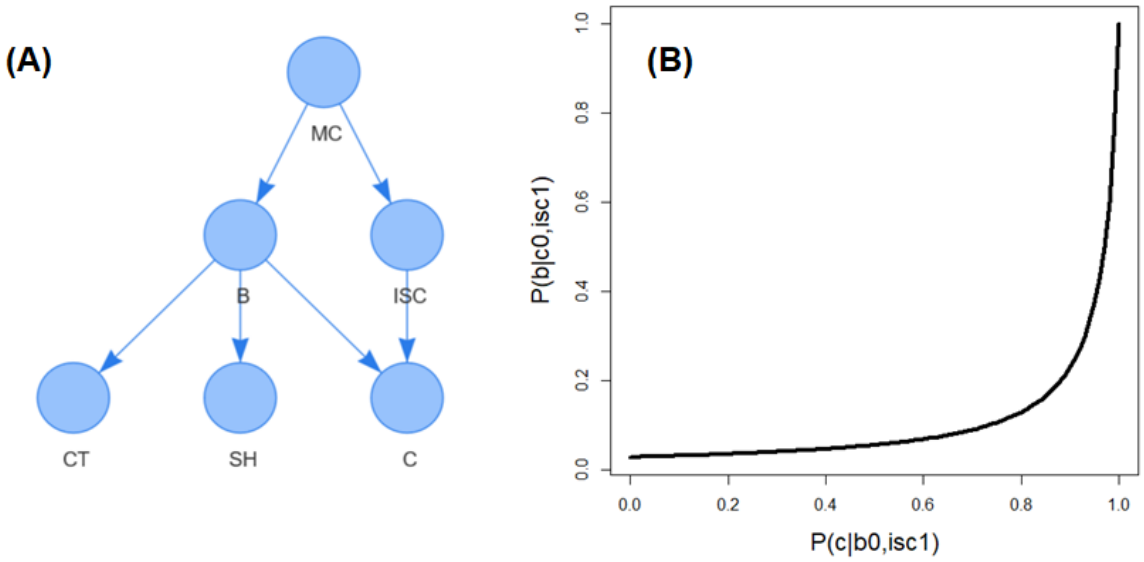


Figure 2. A) Brain tumor structure network; B) Sensitivity function derived by van der Gaag et al. (2013) showing the evolution of the probability of interest P(b|c0,isc1) as a function of the CPT entry P(c|b0,isc1).

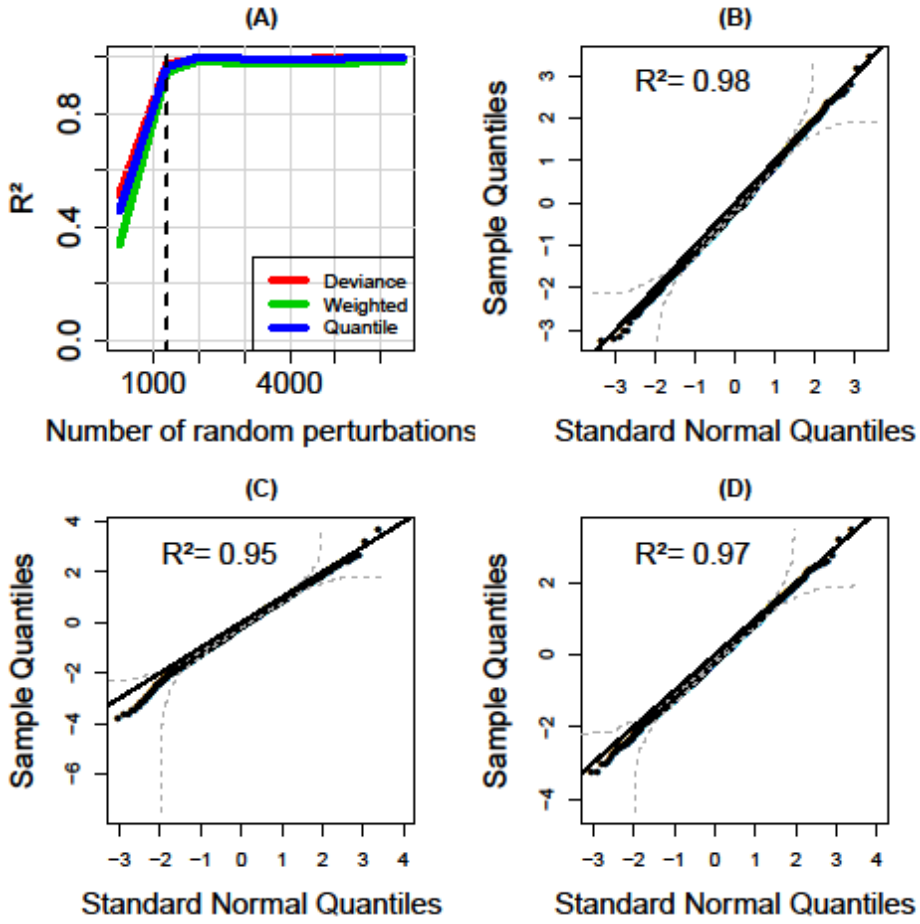
Node MC refers to metastatic cancer, which may potentially lead to the development of a brain tumour (node B) and may give rise to an increased level of serum calcium (node ISC). The presence of a brain tumour can be established from a CT scan (CT). Another indicator of the presence of a brain tumour can be related to severe headaches (SH). A brain tumour or an increased level of serum calcium are both likely to cause a patient to fall into a coma (C). The conditional probabilistic relationships between the nodes (CPT entries) are provided in Table 1. We focus here on the probability P=P(b|c0,isc1), namely the probability to develop brain tumor given the absence of coma and an increased level of serum calcium. The robustness of P is studied with respect to the values of 13 CPT entries (Table 1).

[*Table 1 about here*]

3.2 Application

We now apply the proposed BBR approach. It starts with the random sampling of the 13 CPT entry values using Gaussian distributions (with a standard deviation of 0.1 for all the 13 CPT values) truncated at zero and one as proposed by van der Gaag et al. (2013). For each of the randomly generated values of the CPT entries, the probability of interest *P*, i.e. P(b|c0,isc1), is calculated through approximate Bayesian inference based on particle filter using logic sampling (Koller and Friedman 2009). These values are then transformed using Eq. 3.

The minimum number of random perturbations of the BBN was chosen by studying the evolution of the R^2 indicator values for the different residuals' formulations (see Sect. 2.3). Fig. 3A shows that the convergence can be considered reached for a minimum number of ten times the number of CPT entries, i.e. 1,300 for which the R^2 values all reach very satisfactory values above 95%. The visual inspection of the normal Q-Q plots in Fig. 3B-D confirms the satisfactory adequacy of the BBR model. We can however note some deviations for very low quantile values but this is only indicated by one residuals' formulation (Fig. 3C). Besides, taking into account the 95% confidence band (outlined by dashed lines in Fig. 3C), this discrepancy can be considered low.



356

357 *Figure 3. (A) Evolution of the R² indicator as a function of the number of random perturbations for the*
358 *brain tumor BBN. Three residuals' formulations are considered (see Appendix D). The vertical dashed*
359 *line indicates the selected number of random perturbations. For this number, normal Q-Q plots are*
360 *provided considering different residuals: (B) deviance; (C) standardized weighted; (D) quantile. The*
361 *dashed lines indicate the boundaries of the 95% confidence band based on the Kolmogorov-Smirnov*
362 *statistic (Doksum and Sievers, 1976). The value of the R² indicator (Eq. 5) is also reported.*

363

364 The histogram of the probability of interest *P* values is provided in Fig. 4A with mean and
365 standard deviation at respectively 0.18 and 0.14. Fig. 4B shows that the Beta mean is
366 informative regarding *P* and satisfactorily explains *P*, which is in agreement with the analysis
367 of the Q-Q plots (Fig. 3). Some deviations can however be noticed for very large values, but
368 may be related to the low number of data to perform the fitting (see the light colour in Fig. 4B
369 indicating a low density of dots). The optimal stopping iteration of BBR model is selected by a
370 5-fold cross validation procedure (combined with the noncyclic algorithm of Thomas et al.,
371 2018) as illustrated in Fig. 4C.

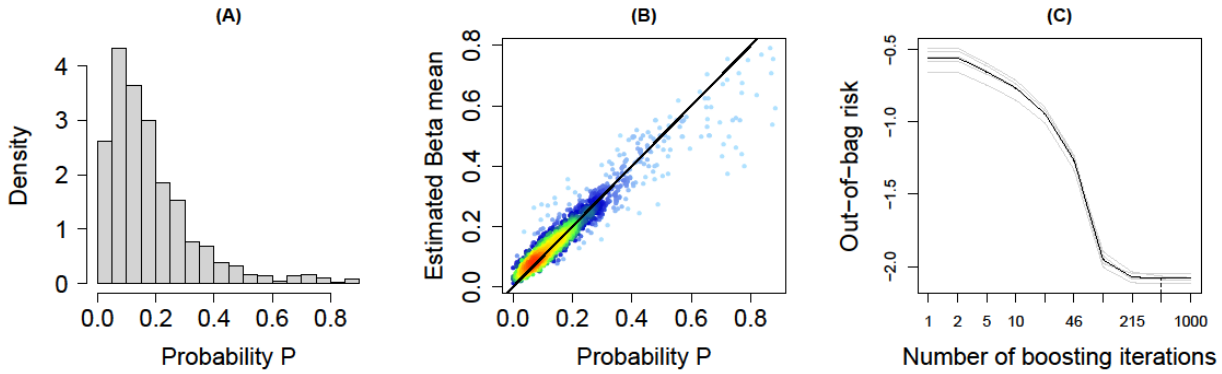


Figure 4. (A) Histogram of randomly generated P values for the brain tumor network; (B) Comparison between the estimated Beta mean and P (the colours indicate the density of the dots); (C) Evolution of the (out-of-bag) risk estimated using a 5-fold cross-validation procedure as a function of the number of boosting iterations; the optimal stopping iteration is selected as the one minimizing the average risk over all 5 cross-validation iterations (indicated by a vertical dashed line at 464).

Figure 5 and 6 respectively depicts the non-linear effects on μ (logit-transformed) and σ (logit-transformed) of the CPT entries selected by the boosting algorithm during the fitting of the BBR model. Different conclusions regarding the sensitivity of *P* to the CPT entries can be drawn:

- Information on the importance ranking of the CPT entries can be derived. We show that only 7 out of the 13 CPT entries have been selected by the boosting algorithm, namely the CPT entries related to the nodes MC, ISC and C in direct relation with node B (see Fig. 2A);
- The individual contributions of the selected CPT entries in Figure 5 and 6 are global in the sense that they are constructed by accounting for the co-variations of all inputs (contrary to the traditional SA using sensitivity functions, which imposes the construction of the function by varying one input one at a time);
- Information on the type of effect the CPT entries can be derived. We show that the CPT entries' effect on the Beta mean μ (logit-transformed) is monotonic but rarely linear. In particular, the effect of P(c|b0,isc1) almost follows an exponential-like trend (outlined by dashed red lines in Fig. 5, bottom right hard corner), which is in agreement with the traditional SA of *P* (Fig. 2B) using a sensitivity function of polynomial form (as derived by van der Gaag et al. 2013);

- The effect on the Beta parameter σ (logit-transformed) is also nonlinear with respect to the selected CPT entries, hence indicating that the CPT entries not only affect the best estimate of *P* but also its precision, i.e. dispersion of the underlying Beta law.

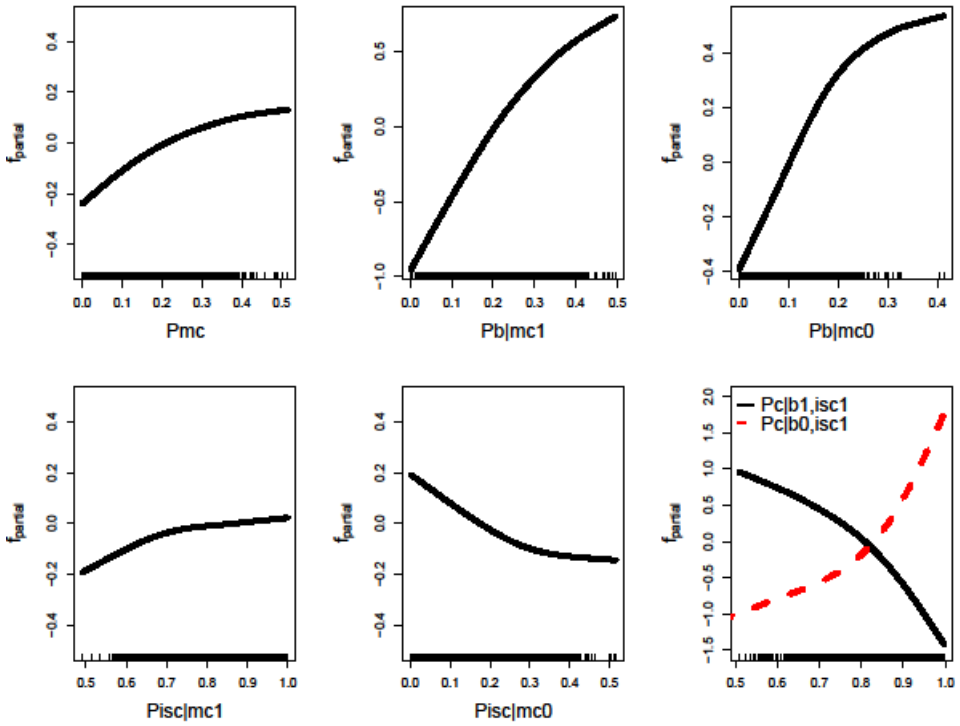


Figure 5. Partial effect of each CPT entry on μ (logit-transformed) applied to the brain tumor network. Note the different scales of the x- and y-axis.

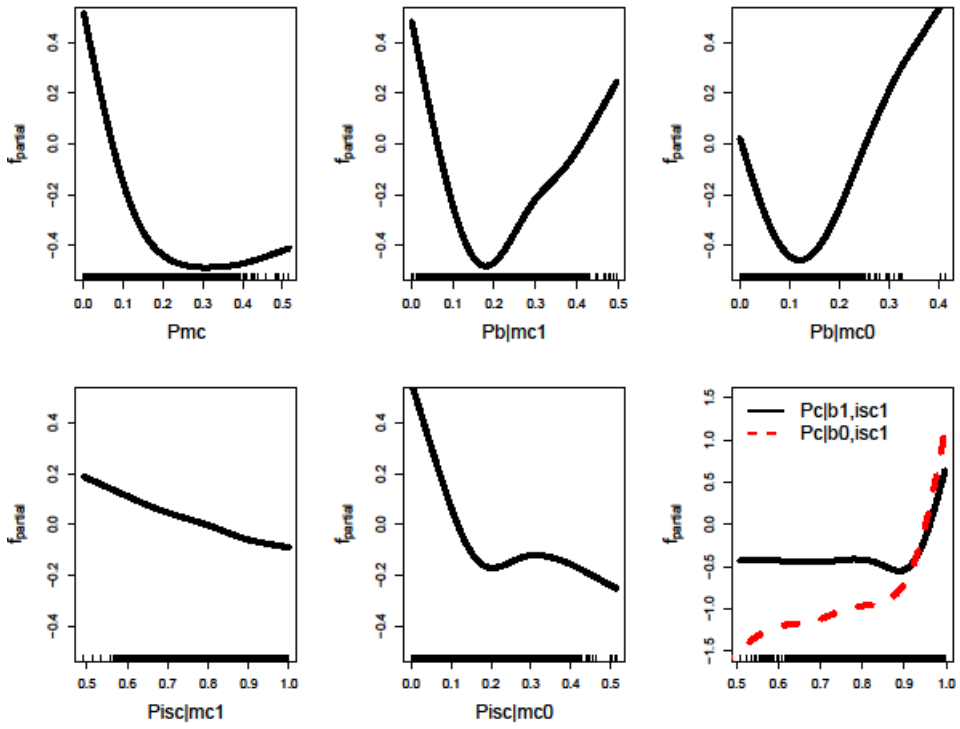


Figure 6. Partial effect of each CPT entry on σ (logit-transformed) applied to the brain tumor network. Note the different scales of the x- and y-axis.

Since the partial effects in Fig. 5 and 6 are global, these can easily be used by varying each CPT entries one-at-a-time or jointly. Figure 7(top) depicts the results of such analysis considering the individual variation of P(c|b1,isc1) and of P(c|b0,isc1) on the Beta density function related to *P* (while keeping the other CPT entry values at their original values as provided in Table 1).

- When P(c|b0,isc1) is varied from low to high values (Fig. 7A), the corresponding Beta density function is translated from low to high values, but with more and more dispersion around the mode, which is more specifically amplified when P(c|b0,isc1) exceeds value of ~0.9;
- When P(c|b1,isc1) is varied from low to high values (Fig. 7B), the effect on the Beta mode is opposite to the one due to the effect of P(c|b0,isc1). This exemplifies the decreasing trend outlined in Fig. 5 (bottom right hand corner): the mode is translated from high to low values with an effect on the dispersion less pronounced than for P(c|b0,isc1). This is in agreement with the lower partial effect outlined in Fig. 6(bottom right hand corner) ;

- Finally, Figure 7C depicts the result of the analysis when both CPT entries are jointly varied. This shows that the impact of both CPT entries is compensated when they are jointly increased similarly: the Beta distributions remains "stuck" around 0.15 unless both CPT entries reach very high values above 0.95.

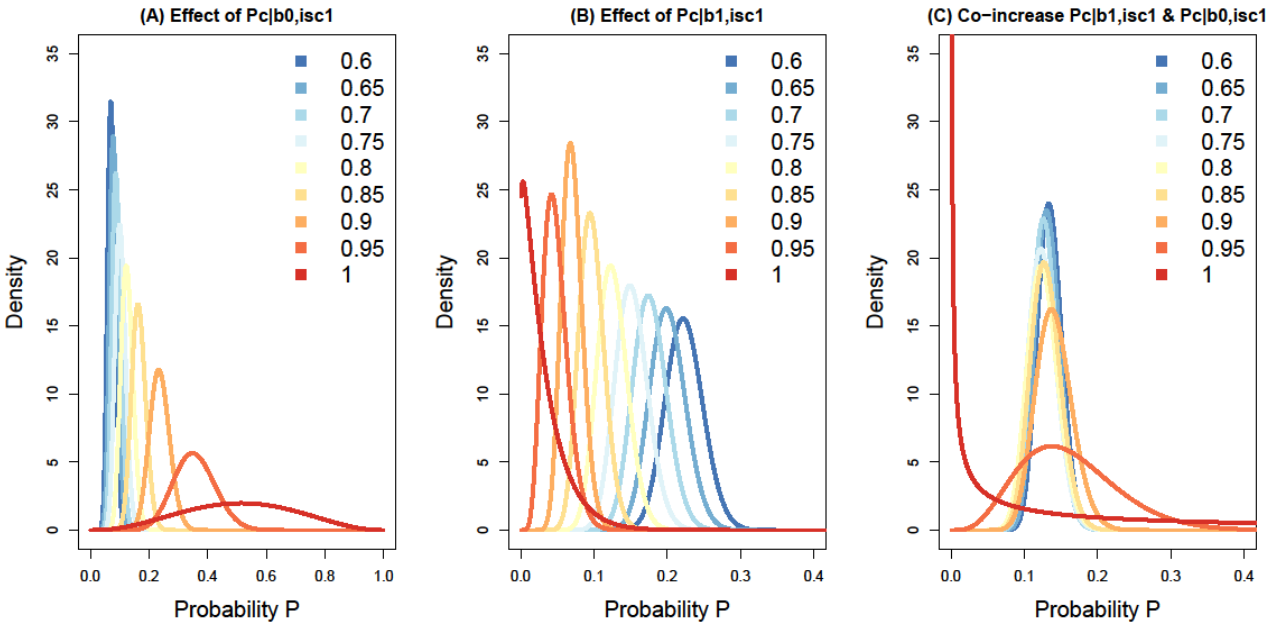


Figure 7. (A) Evolution of the Beta law fitted to the P values given increasing values of P(c|b0,isc1) from low (0.6) to high values (1.0); (B) Evolution of the Beta law fitted to the P values given increasing values of P(c|b1,isc1) from low (0.6) to high values (1.0); (C) Evolution of the Beta law fitted to the P values given increasing values (with same increment) of both P(c|b1,isc1) and P(c|b0,isc1).

4 Real-case applications

In this section, we apply the BBR approach to real cases. First, a linear Gaussian BBN is considered for assessing the damage of reinforced concrete structures (Castillo and Kjærulff, 2003). This enables us to illustrate a situation where the number of parameters is large enough to hamper the interpretation, i.e. here more than 40 variables have to be processed (Sect. 4.1). Second, a discrete BBN is considered for reliability analysis. For this case, analytical sensitivity functions can hardly be derived since the interest is not the sensitivity to the values of the CPT entries directly, but the physical parameters, which determine them.

4.1 Reinforced-Concrete BBN

We investigate the robustness of the BBN presented in Castillo and Kjærulff (2003) for assessing the damage of reinforced concrete structures (Fig. 8).

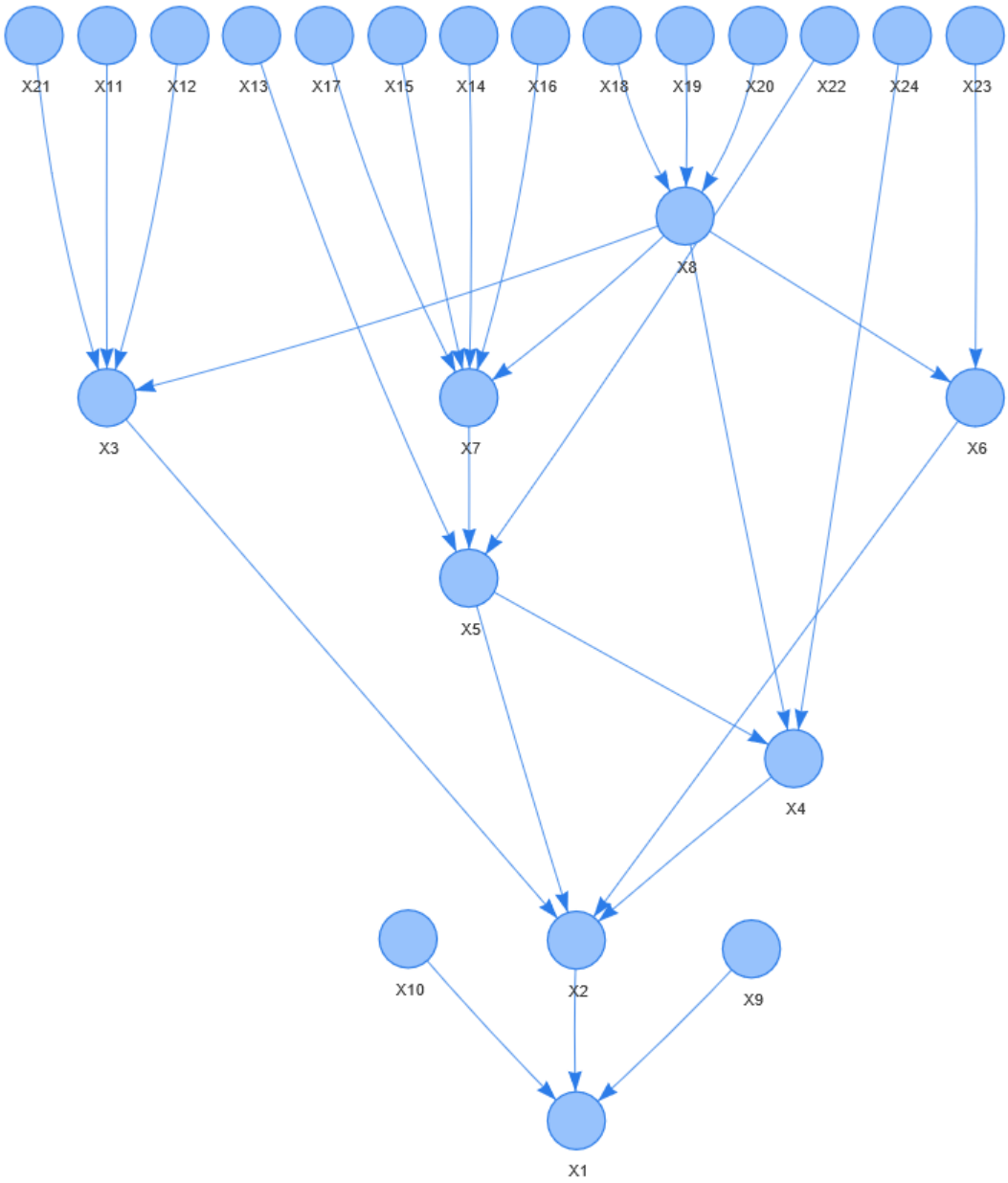


Figure 8. Structure of the reinforced concrete network.

The BBN is composed of 24 continuous nodes, 27 arcs and more than 40 parameters. The random variables (see nodes' meaning in Table 2) are assumed to be normally distributed.

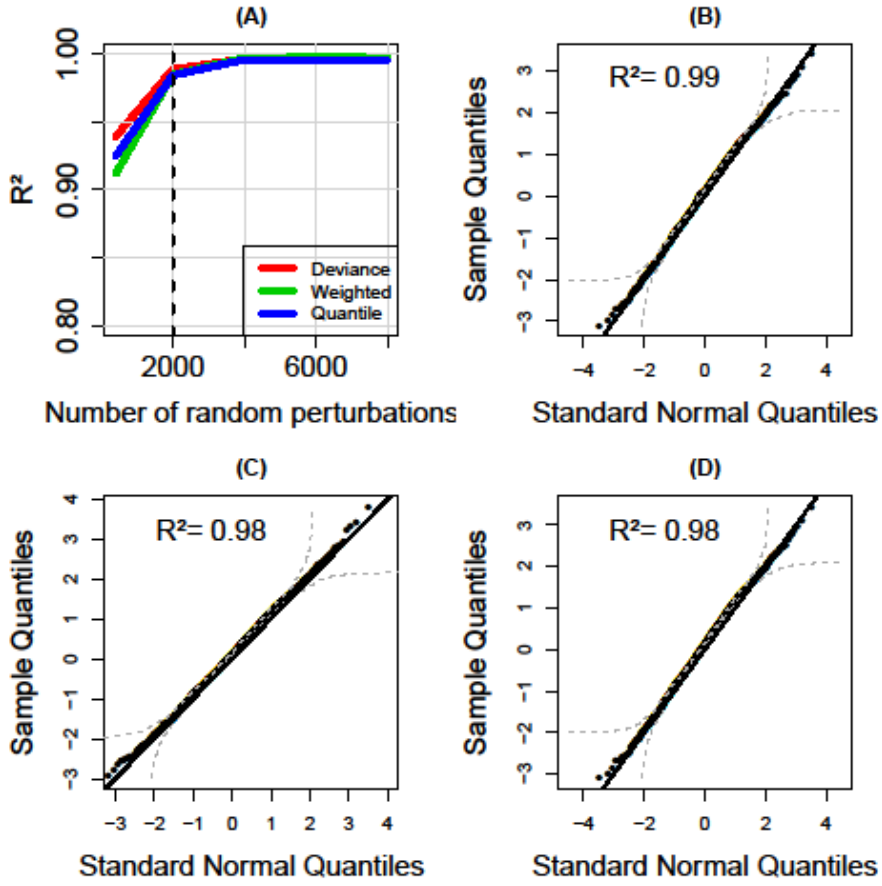
[*Table 2 about here*]

The original parametrization of Castillo and Kjærulff (2003) follows the regression model in Eq. 1 and assumes zero mean for all variables (i.e. nil intercept m_0), and conditional standard of 1.0 for observable nodes (i.e. X_9-X_{24}) and $1e^{-4}$ otherwise (term denoted s_0 in Table 3). Values for the regression coefficients are provided in Table 3. In the following, we refer to these values as the CPM best estimate for the BBN parametrization.

We investigate the robustness of the probability for the damage of the system $X_1 \geq 1$ given the evidence that the beam is weak $X_9 \geq 1$, i.e. $P = P(X_1 \geq 1 | X_9 \geq 1)$. The CPM best estimates provided in Table 3 are randomly perturbed using Gaussian laws truncated at zero (with mean correspond to the values given in Table 3, standard deviation of 0.2 for the regression coefficients and of 0.5 for the intercept terms). In total 43 variables are considered.

[*Table 3 about here*]

Fig. 9A depicts the evolution of the R^2 indicators considering the different residuals' formulations. This shows that a minimum of 2,000 random perturbations is required to consider the BBR model adequate with R^2 values above 98%. This is also confirmed by the visual inspection of the normal Q-Q plot in Fig. 9B-D.



467

468 *Figure 9. (A) Evolution of the R² indicator as a function of the number of random perturbations for the*
469 *reinforced-concrete BBN. Three residuals' formulations are considered (see Appendix D). The vertical*
470 *dashed line indicates the selected number of random perturbations. For this number, normal Q-Q plots*
471 *are provided considering different residuals: (B) deviance; (C) standardized weighted; (D) quantile.*
472 *The dashed lines indicate the boundaries of the 95% confidence band based on the Kolmogorov-Smirnov*
473 *statistic (Doksum and Sievers, 1976). The value of the R² indicator (Eq. 5) is also reported.*

474

475 Fig. 10A depicts the corresponding histogram of the randomly generated *P* values (with mean
476 of 0.8 and standard deviation of 0.08). Fig. 10B shows that the Beta mean is very informative
477 with a good agreement with *P*. The optimal stopping iteration of the regression model is selected
478 based on a 5-fold cross validation procedure combined with the noncyclic algorithm as
479 illustrated in Fig. 10C.

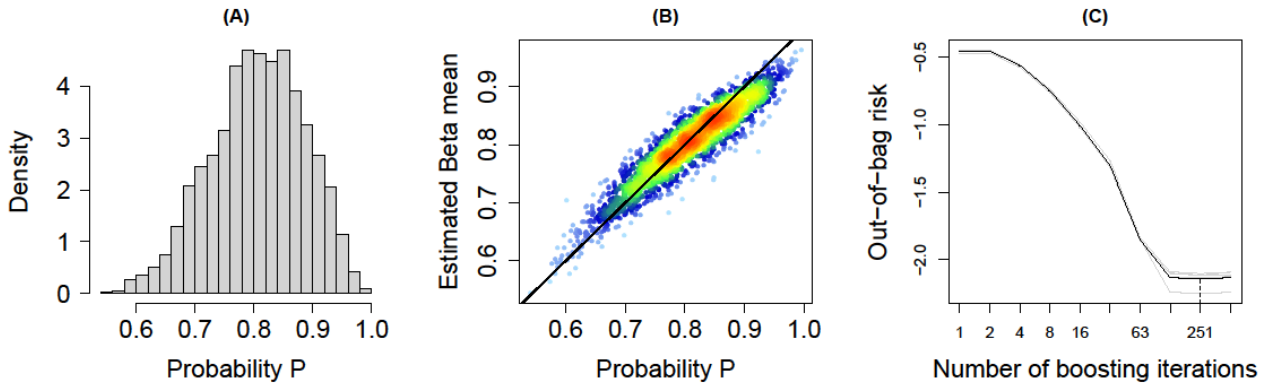


Figure 10. (A) Histogram of randomly generated P values for the concrete network; (B) Comparison between the estimated Beta mean and P (the colours indicate the density of the dots); (C) Evolution of the (out-of-bag) risk estimated using a 5-fold cross-validation procedure as a function of the number of boosting iterations; the optimal stopping iteration is selected as the one minimizing the average risk over all 5 cross-validation iterations (indicated by a vertical dashed line at 251).

The number of parameters of the conditional probability model is larger (almost four times) than the case described in Sect. 3. In this case, despite the regularisation associated to the boosting algorithm, the resulting BBR model can still remain too rich to be easily interpretable by BBN practitioners. The direct application of the boosting algorithm ends up here with more than 30 variables, i.e. analysing each of the partial effect may here not be practical. Therefore, the analysis is completed by the stability selection analysis (described in Appendix C) to further screen the most influential variables in the BBR model. Fig. 11A depicts the selection probabilities for each CPM parameter with a threshold at 0.8 (see parametrisation in Appendix C). This shows that only the intercept terms for node X_2, X_5 and X_8 appear to be influential with respect to the Beta mean μ (logit-transformed), whereas only the intercept terms for node X_2, X_3, X_5, X_7 and X_8 appear to be influential with respect to the Beta parameter σ (logit-transformed). Interestingly, the nodes which affect the probability of interest are not necessarily the ones in direct connections with X_1 (see Fig. 8).

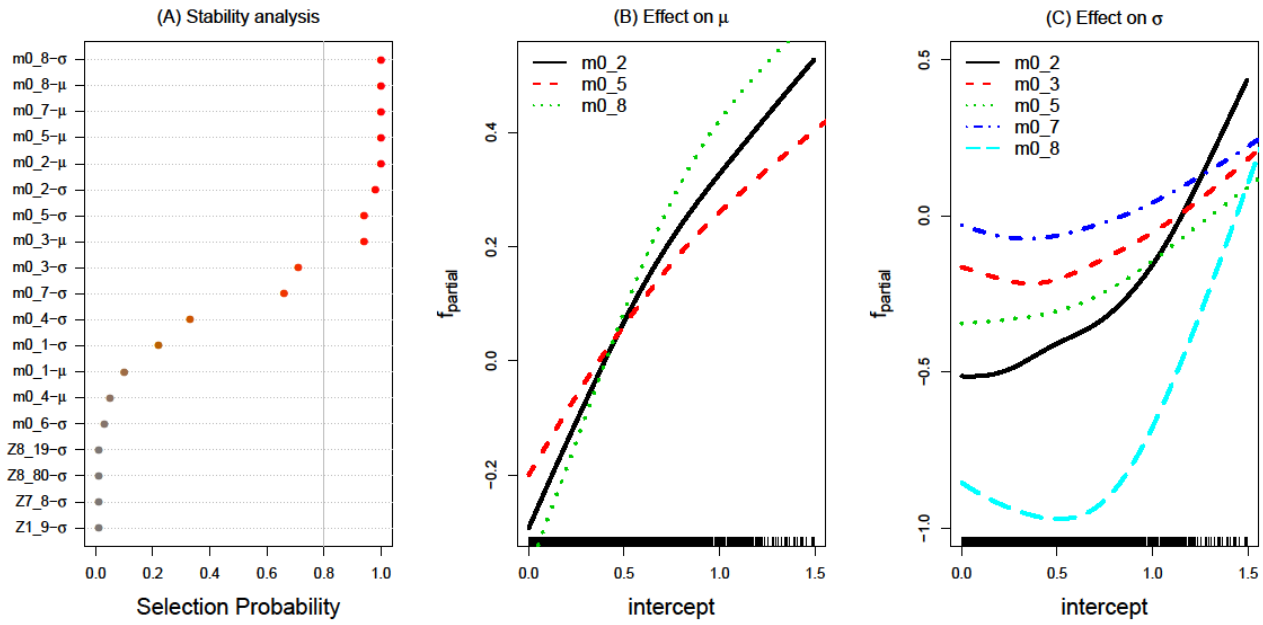


Figure 11. (A) Selection probability for each CPM parameter of the concrete network derived from the stability selection analysis (Appendix C). (B) Partial effect of the intercept m_0 of nodes X_2, X_5 and X_8 on μ (logit-transformed). (C) Partial effect of the intercept m_0 of nodes X_2, X_3, X_5, X_7 and X_8 on the σ parameter (logit-transformed).

The non-linear effects for each selected CPM parameters are depicted in Fig. 11B,C. This shows that the individual effect mainly corresponds to an increasing monotonic functions considering both parameters of the Beta distribution; with a higher nonlinear effect of m_0 of node X_8. This is confirmed by the Beta distribution's evolution with respect to the intercept parameter of node X_2 and X_8 (Fig. 12A, B).

- Fig. 12A shows that the increase of m_{0_2} mainly affects the increase of the best estimate of P (with only slight effect on the dispersion), whereas Fig. 12B both affects the best estimate and the dispersion, i.e. the larger m_{0_8}, the larger the mean value of P (i.e. the riskier the situation), but also the larger the uncertainty on P (i.e. the occurrence of the risky situation remains uncertain);
- Fig. 12C shows the evolution of the Beta's distribution when both parameters are increased similarly: we see that the combined increase amplifies the translation of the Beta's distribution to very high values, i.e. risky situations (driven by both increasing effects as shown in Fig. 12B). Despite the increase in σ during this process, the occurrence of this risky situations remains of moderate-to-high confidence, which

suggests that changes in m_{0_8} and m_{0_2} might ultimately lead to the failure of the concrete system.

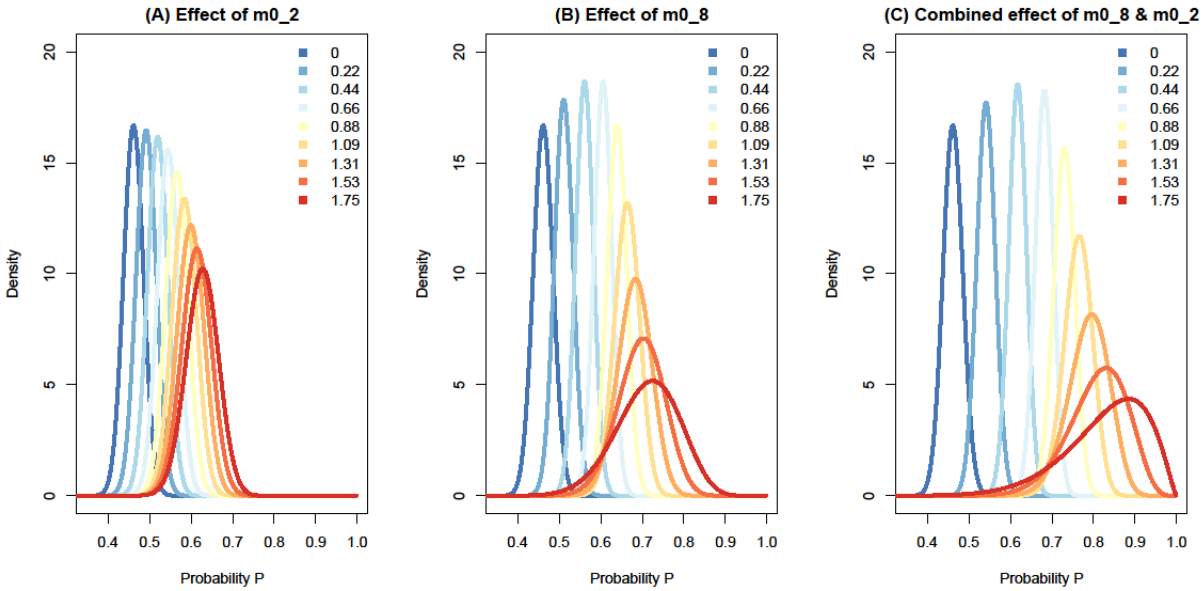


Figure 12. (A) Evolution of the Beta law fitted to the P values given increasing values of the intercept m_0 of node X_2 from low (0.) to high values (1.75); (B) Evolution of the Beta law fitted to the P values given increasing values of the intercept m_0 of node X_8 from low (0.) to high values (1.75); (C) Evolution of the Beta law fitted to the P values given increasing values (with same increment) of both intercepts.

4.2 BBN-based reliability assessment

Figure 13 depicts the BBN constructed by Gehl and Rohmer (2018) for studying the problem of station blackout (node SYS) following an earthquake at a given nuclear power plant (NPP) sub-system. The earthquake event is characterised by two intensity measures (nodes IM1 and IM2 respectively corresponding to the peak ground acceleration and the spectral acceleration at the first vibration period of the structure). The NPP sub-system is composed of a 5-story reinforced-concrete structure hosting two emergency diesel generators (EDGs). Three damage events (STR for structural damage and EDG for failure of the anchorage of the generators) are considered.

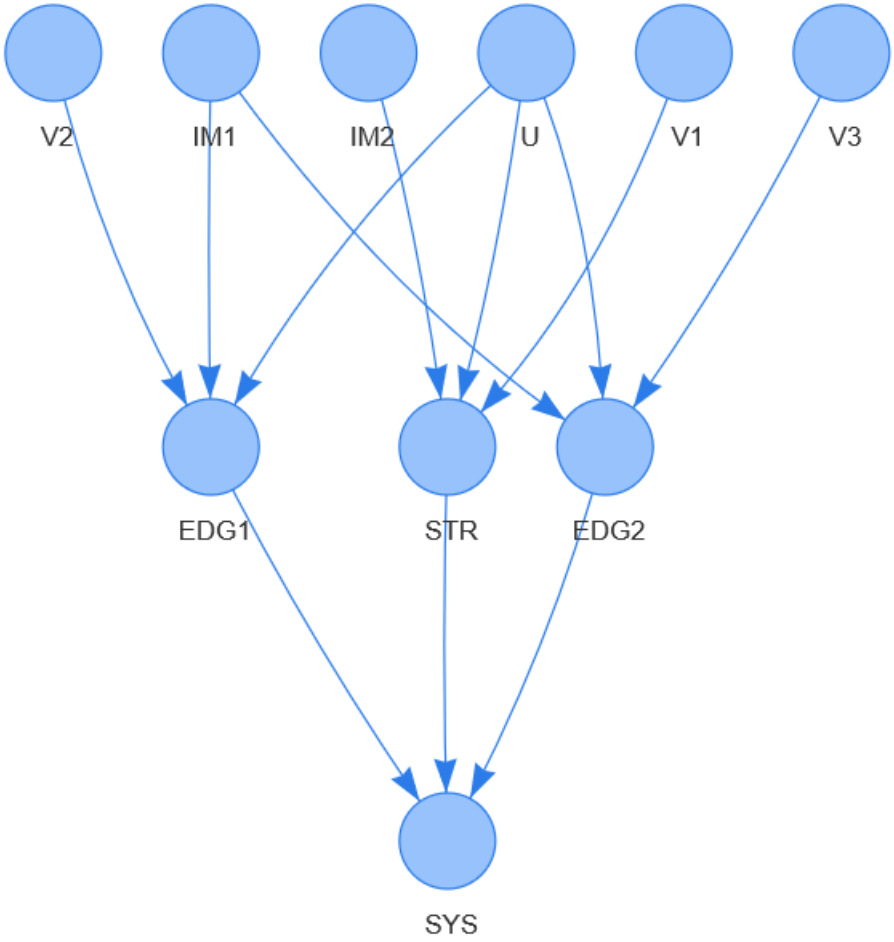


Figure 13. Structure of the network for studying the NPP subsystem reliability; IM1 and IM2 are deterministic nodes representing intensity measure characterizing the earthquake event; Nodes U, V1, V2 and V3 are zero-centered Gaussian random variables used to incorporate correlation. Only the sensitivity to the parameters for the blue nodes is investigated.

The conditional probabilities of the failure states of STR, EDG1 and EDG2 are estimated through fragility functions (i.e. probabilistic model that relates the failure probability to the intensity measure IM), which are derived from non-linear seismic time histories applied to the structure. Due to the observed statistical dependence between the failure events (i.e., due to the common seismic loading applied to the three components), auxiliary variables U, V1, V2 and V3 (described by a standard normal distribution) are added to the BBN following the approach by Gehl and D'Ayala (2016). These variables result from the Dunnett-Sobel decomposition of the correlation coefficients between the safety factors of the components as follows:

$$Z_i = t_i \cdot U + \sqrt{1 - t_i^2} \cdot V_i \quad \text{(Eq. 6)}$$

where Z_i is the standard score of the safety factor of component i (Z_1 corresponds to STR, Z_2 to EDG1, and Z_3 to EDG2) and t_i is the Dunnett-Sobel coefficient that approximates the correlation between the failure events.

By definition, the failure of component i occurs if $Z_i \leq -\beta_c$, where β_c is the reliability index expressed as follows:

$$\beta_c = \frac{\alpha_i - \ln im}{\beta_i} \quad \text{(Eq. 7)}$$

where α_i and β_i are the fragility parameters (median and standard dispersion) of component i, under the common assumption of a cumulative lognormal distribution for the fragility function; im is the value of the intensity measure of interest (IM1 for EDG1 and EDG2, IM2 for STR, see Fig. 13).

Based on this framework, the CPT of the failure event of a component i is built by considering all combinations of discretized values $\{im; u; v_i\}$, and by checking the following condition (Gehl and D'Ayala, 2016):

$$\left(z_i = t_i \cdot u + \sqrt{1 - t_i^2} \cdot v_i\right) \leq -\frac{\alpha_i - \ln im}{\beta_i} \quad \text{(Eq. 8)}$$

For the sensitivity analysis, we do not consider the value of the CPT entries directly but the physical parameters, which determine them, namely the parameters of the fragility curves, (mean α and standard deviation β). The numerical values of the physical parameters considered in this application are detailed in Table 4. The fragility curves' parameters were randomly perturbed via a zero-centered Gaussian noise with standard deviation that is 10% of the original values (termed as CPM best estimates in the following). For the parameters of Dunnett-Sobel decomposition, the Gaussian noise is truncated at one. The probability of interest is here the probability of the sub-system failure, i.e. P=P(SYS=1) given an earthquake event characterised

by intensity measures IM1 and IM2 of respectively 10 and 12.5 m²/s. In total, 9 parameters are considered.

[*Table 4 about here*]

The selection of the minimum number of random perturbations is more difficult than for the brain tumour or for the reinforced concrete case. Fig. 14A shows that at least 1,000 random perturbations are necessary to reach R^2 values above 95%, but due to oscillations in the evolution of the R^2 values, we preferably choose the largest number, i.e. >4,000. The visual inspection of the normal Q-Q plots (Fig. 14C-D) also reveals some deviations for very high and low quantiles (outside the range [-2 - 2]) though it should be noted that they remain within the 95% confidence band and with low-to-moderate magnitude. Contrary to the brain tumour case, the three residuals' formulations all agree on the identification of the problem. A deeper analysis show that these deviations correspond here to Beta distributions with very high mean and very low variance values: the sensitivity should be analysed with care for these cases.

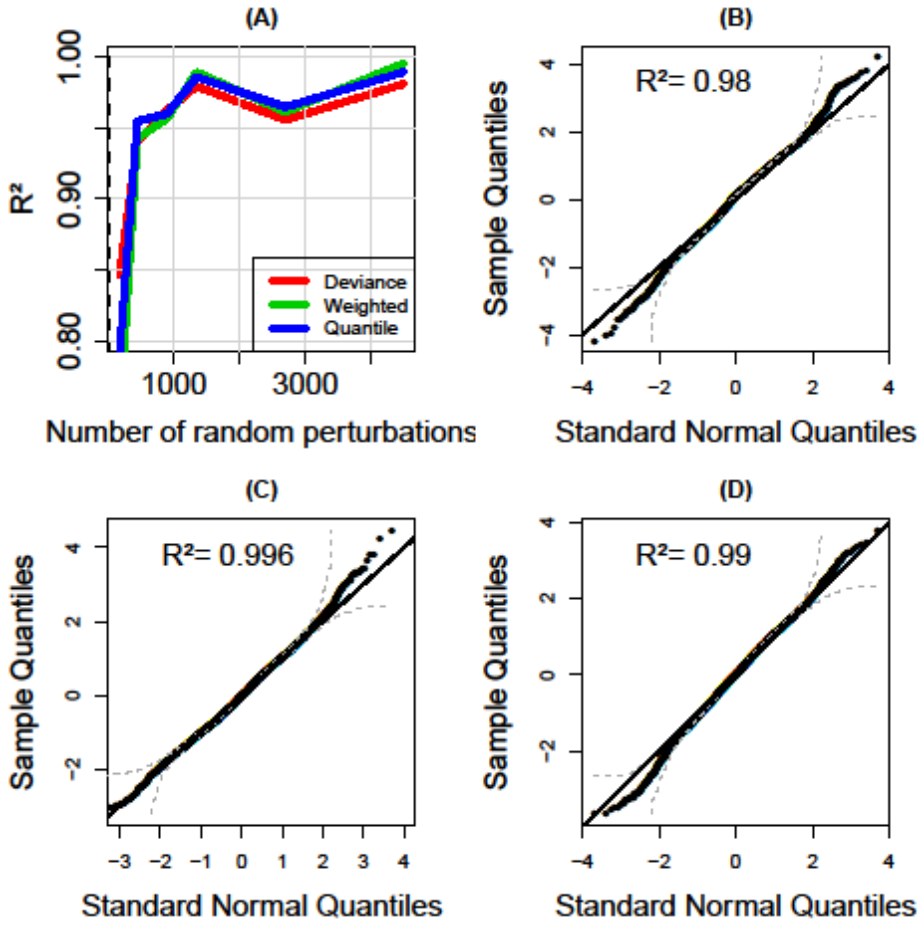


Figure 14. (A) Evolution of the R^2 indicator as a function of the number of random perturbations for the BBN-based reliability assessment. Three residuals' formulations are considered (see Appendix D). The vertical dashed line indicates the selected number of random perturbations. For this number, normal Q-Q plots are provided considering different residuals: (B) deviance; (C) standardized weighted; (D) quantile. The dashed lines indicate the boundaries of the 95% confidence band based on the Kolmogorov-Smirnov statistic (Doksum and Sievers, 1976). The value of the R^2 indicator (Eq. 5) is also reported.

The histogram of *P* values is provided in Fig. 15A with mean and standard deviation at respectively 0.65 and 0.26. Contrary to the reinforced-concrete BBN, the agreement with *P* is less satisfactory (Fig. 15B). A major part of the *P* values (see the warm colour indicating the density of dots in fig. 15B) appear to be reproduced by the Beta mean, but the scatter plot remains disperse. This type of analysis can be considered a complement to the analysis of the Q-Q plots (Fig. 14) and supports a cautious attitude with respect to the conclusions drawn from the sensitivity analysis. The optimal stopping iteration of the regression model is selected based

on a 5-fold cross validation procedure combined with the noncyclic algorithm as illustrated in Fig. 15C.

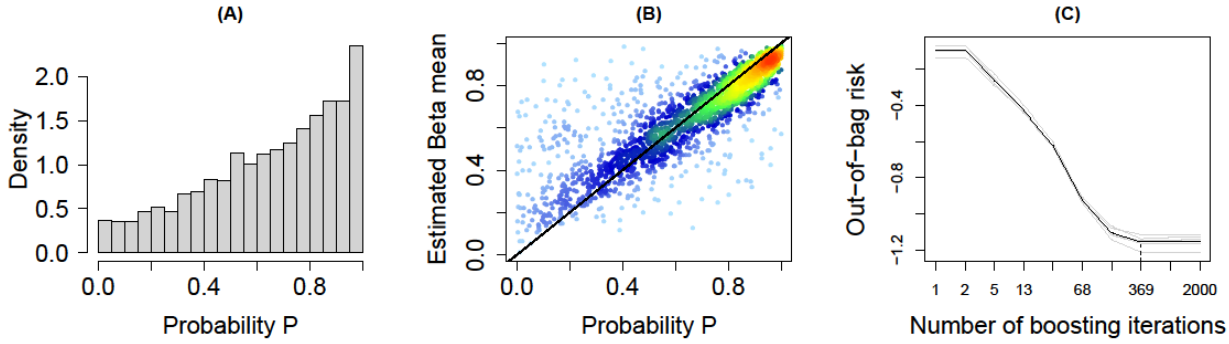


Figure 15. (A) Histogram of randomly generated P values for the NPP reliability network; (B) Comparison between the P values and the estimated mean value provided by the Beta model; (C) Evolution of the (out-of-bag) risk estimated using a 5-fold cross-validation procedure as a function of the number of boosting iterations; the optimal stopping iteration is selected as the one minimizing the average risk over all 5 cross-validation iterations (indicated by a vertical dashed line at 369).

The application of the boosting algorithm reveals that only the parameters of the fragility curves are selected. These are the only parameters identified as significant by the BBR approach, i.e. the correlation parameters $t_{1\text{-}3}$ are discarded by the algorithm. Figure 16 provides the corresponding partial effects for both Beta law parameters.

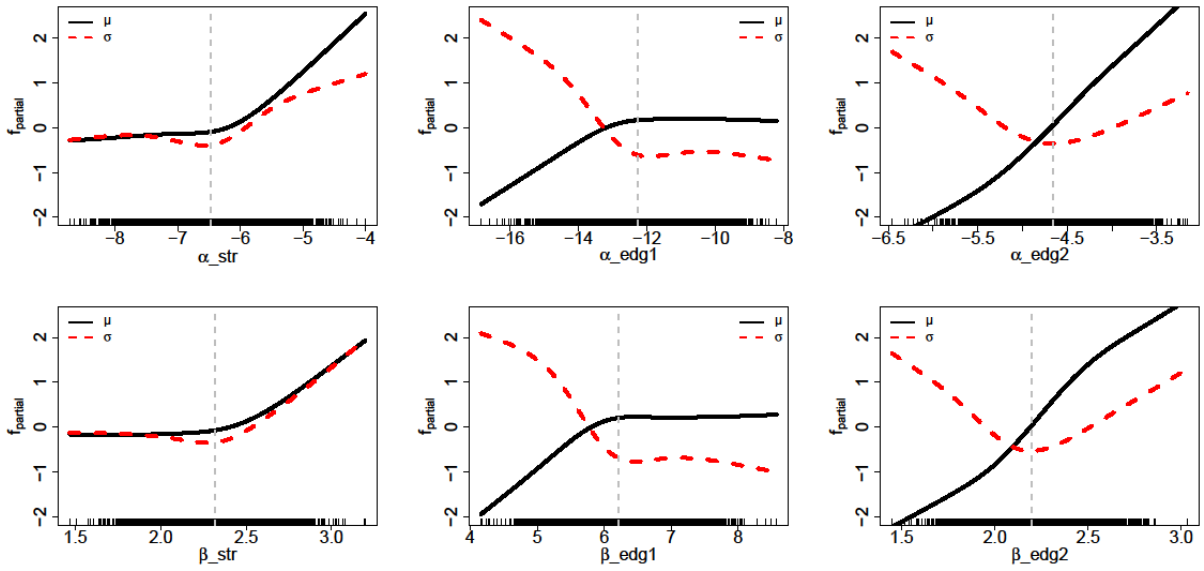


Figure 16. Partial effect of the fragility curve's parameters on μ (logit-transformed) (black straight line) and on σ (logit-transformed) (red dashed line). The vertical grey-coloured dashed line indicates the CPM best estimates of Table 4.

Several observations can be made:

- The parameters (α_str, β_str) of the structural fragility curve (STR, Fig. 16, left) have almost similar effect, whatever the Beta law parameter (logit-transformed); it corresponds to a quasi- bilinear function, which remains quasi-constant below the CPM best estimate then increases above it; from a risk viewpoint, this means that the damage to the structure enhances the occurrence of risky situations, where the system failure's probability *P* might reach high values. Yet, the effect on σ parameter (logit-transformed) also indicates that the confidence in the occurrence of this probabilistic regime is low;
- The effects of the parameters (α_edg2, β_edg2) of the second electrical generator's fragility curve (EDG2, Fig. 16, right) are quasi-linear with respect to the Beta law mean μ, but is non-monotonic with respect to the σ parameter (logit-transformed). This can be schematically understood as a decrease of uncertainty on *P* (i.e. the Beta law dispersion) when approaching the original value, then an increase above it;
- The effect of the parameters (α_edg1, β_edg1) of the first electrical generator's fragility curve (EDG1, Fig. 16, center) is more complex. The effect on the Beta mean μ (logit-transformed) corresponds to a bilinear function, which increases at around the CPM best estimate, then remains quasi-constant above it (i.e. to a lesser extent, it corresponds to the opposite behaviour to the one for STR). The effect on the Beta parameter σ corresponds to an inverted sigmoid, which reaches its lower horizontal plateau above the CPM best estimate. This means that above their original values, parameters (α_edg1, β_edg1) little affect the system failure's probability.

Fig. 17 illustrates how α_str, α_edg1 and α_edg2 act differently on the system failure. When increasing α_str, the system failure starts increasing only for values well above the CPM best estimates. This is shown by the Beta distributions, which remains "stuck" at moderate values (around 0.6; see for cold colours in Fig. 17A). When increasing α_edg1, the resulting Beta distribution remains stuck at high value around 0.70-0.75 (note the evolution of the mode for warm colours in Fig. 17B), hence leading to risky situations. The confidence in the occurrence

of such probabilistic regime is high as indicated by the low dispersion of the Beta distributions. When increasing α_edg2, the Beta distribution continues to translate towards very high values above 0.90. Despite the very low dispersion of the Beta distribution, the confidence remains here moderate, because this situation corresponds to the one for which the normality of the residuals is not met (Fig. 14).

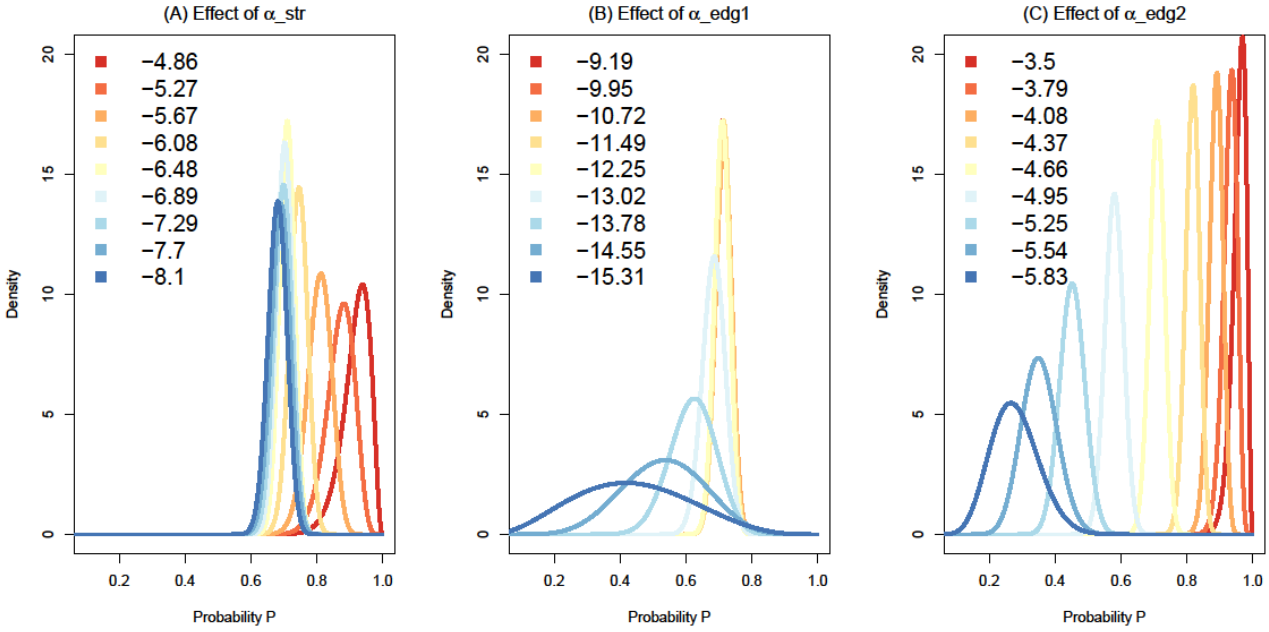


Figure 17. Evolution of the Beta distributions fitted to the P values given: (A) increasing values of the mean value of the fragility curve α_str; (B) increasing values of α_edg1; (C) increasing values of α_edg2.

5 Discussion

Ensuring the validity and credibility of increasingly complex BBN-based expert systems (Pitchforth & Mengersen, 2013; Kleemann et al., 2017; Marcot and Penman, 2019) requires a broad vision on the sensitivity to the CPM parameters. As outlined in the introduction, the most widely used approach is based on sensitivity functions for discrete BBNs, and on partial derivatives for continuous BBNs. Reasons are the intuitive interpretation of the results (thanks to the graphical presentation) and the simplicity of the implementation. Yet, these approaches only provide information on the local influence, in the sense that the parameters are varied "one-at-a-time". This means that it can only provide a restricted vision on sensitivity, because the exploration of the sensitivity remains limited to a few CPM parameters, while the domain of the other CPM parameters is left mostly unexplored (as thoroughly discussed by Saltelli et al.

(2019) with respect to sensitivity analysis practices). Multi-way SA methods have been proposed (Leonelli et al., 2017; Gómez-Villegas et al., 2013), but they can rapidly become intractable.

In this context, the proposed BBR approach can be considered more complete by providing global insight on the sensitivity by means of the partial effects (Fig. 4,10,13,14), which directly show the individual, possibly non-linear, effects of all CPM parameters. In Sect. 4, we have shown how the BBR approach can provide richer information than existing methods at multiple levels:

- Level (1): it selects the most influential parameters and it screens those that are of negligible influence (information that is generally unknown a priori);
- Level (2): it provides the functional relation between the CPM parameters and the result of the probabilistic query. The proposed approach builds on a graphical presentation of the results, which eases the interpretation;
- Level (3): the practitioner can study how the combined effect of the CPM parameters can lead to different probabilistic regimes (e.g., situations of high probability values) by studying the evolution of the Beta mean. Furthermore, the practitioner can measure the confidence in the occurrence of the identified probabilistic regimes, by studying the effect on the Beta σ parameter.

Table 5 provides a summary of the strengths and weaknesses of the BBR approach with comparison to the approach based on sensitivity functions.

[*Table 5 about here*]

From an operational viewpoint, we tested the applicability of the BBR approach using different application cases covering a large spectrum of situations, namely (1) a small discrete BBN, used to capture medical knowledge, to exemplify the functionalities; (2) a linear Gaussian BBN, used to assess the damage of reinforced concrete structures, exemplifies a case where the number of parameters is too large to be easily processed and interpreted (>40 parameters); (3) a discrete BBN, used for reliability analysis of nuclear power plant, exemplifies a case where analytical solutions for sensitivity can hardly be derived. To ease the implementation, we proposed a procedure based on random experiments, i.e. the BBN-based results are derived

from randomly generated values of the CPM parameters. This procedure is generic, because it can be applied to any CPM parameters whatever the type of the considered BBN: discrete (with CPT entries), continuous (with regression coefficients), or hybrid.

However, it should be underlined that providing information on the global effect of the CPM parameters comes at the expense of a larger computational effort and complexity. The number of calls to the BBN inference engine is larger (typically >1,000) compared to the sensitivity functions, which may become problematic when the inference is difficult (hence time consuming) to perform.

The pillar of the BBR approach is the adequacy of the BBR model to explain the BBN-derived probabilities. This adequacy should be carefully checked, and we propose here to investigate the statistics of the residuals. In some situations, this might reveal situations where the interpretation should be conducted with care as exemplified by the third application case. The difficulties encountered for this case is also related to the characteristics of the sensitivity analysis. Contrary to the brain tumour and to the reinforced concrete BBN, this case does not consider the value of the CPM parameters directly, but the physical parameters that influence them. This adds a level of complexity to the problem and make the BBR model more difficult to fit.

Finally, it should be noted that the second application case shows that the BBR approach is robust even in the presence a large number of parameters (which can rapidly grow as a function of the number of BBN nodes). Yet, the applicability to complex cases with several hundreds of nodes (and thousands of parameters), like the pathfinder network (Heckerman et al., 1992), remains an open question especially regarding two aspects: (1) the capability of the combined "gradient boosting- stability selection analysis" to handle so many terms; (2) the complexity of the interpretation of the partial effects.

6 Summary and future research directions

The rapidly increasing advances of BBN for modelling of expert systems calls for the developments of robust methods for their validation and verification (e.g., Marcot and Penman, 2019). One major pillar to fulfil this purpose is sensitivity analysis. The proposed BBR approach broadens the scope of existing BBN sensitivity analysis methods by providing a larger vision (global) on the CPM influence. The approach has the advantage of being generic (it can

be applied to any kind of BBN, i.e. discrete, Gaussian or hybrid), and robust to the number of parameters (that can rapidly increase, typically reaching several dozens, even for moderate number of BBN nodes).

Bringing the BBR approach to a fully operational state, raises different questions that should be addressed in future work. The first line of future research should concentrate on the intensification of the applicability tests using very large-scale BBNs. An important aspect to be tested is the applicability to highly constraining situations where the number of variables largely exceeds the size of the training database. In such cases, the performance of the stability selection analysis should be more extensively investigated (see Meinshausen & Bühlmann, 2010; Thomas et al., 2018). The second research direction should concentrate on improving the interpretability of the results. The graphical representation of the partial effects is a strength of the BBR approach but might lose its conciseness as the number of functional terms selected as important largely increases (>100). We should take advantage of advances in distributional regression methods that rely on trees like the one proposed by Grün et al. (2012) for Beta regression. The presentation of the results using a network is expected to provide a more concise and understandable presentation of the results. Finally, it should be underlined that the current work has focused on only one part of the problem of uncertainties in BBNs. The CPM parameters constitute only one ingredient for BBN development; the second one being the DAG specification, which 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). To address the whole spectrum of uncertainties in BBN building, sensitivity methods both covering DAG and CPM learning would be beneficial. Again, a solution relying on trees is worth investigating, as recently proposed to deal with psychometric networks (Jones et al., 2019).

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. The analysis was performed using R packages *bnlearn* (available at: https://cran.r-project.org/web/packages/bnlearn/index.html) for BBN building and queries, *bnviewer* (https://cran.r-project.org/web/packages/bnviewer/index.html) for network visualisation, and *gamboostLSS* (available at: https://cran.r-project.org/web/packages/gamboostLSS/index.html) for boosting of distributional regression model with BE() function of *gamlss.dist* package (available at: https://cran.r-project.org/web/packages/gamlss.dist/index.html).

References

Bayer, F. M., & Cribari-Neto, F. (2017). Model selection criteria in beta regression with varying dispersion. Communications in Statistics-Simulation and Computation, 46(1), 729-746.

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.

Bühlmann, P., & Hothorn, T. (2007). Boosting algorithms: Regularization, prediction and model fitting (with discussion). Statistical Science, 22, 477–522.

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.

Castillo, E., & Kjærulff, U. (2003). Sensitivity analysis in Gaussian Bayesian networks using a symbolic-numerical technique Reliab. Eng. Syst. Saf., 79, 139-148.

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 Internat. J. Approx. Reason., 38, 149-174.

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

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.

Cribari-Neto, F., & Zeileis, A. (2010). Beta regression in R. Journal of Statistical Software, 34.

Doksum, K. A. and G. L. Sievers, (1976). Plotting with confidence: graphical comparisons of two populations. Biometrika, 63(3), 421–434.

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.

Eilers, P. H. C., & Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. Statistical Science, 11, 89–121.

Espinheira, P. L., Ferrari, S. L. P., Cribari-Neto, F., (2008). On beta regression residuals. Journal of Applied Statistics, 35 (4), 407-419.

Ferrari, S., & Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. Journal of Applied Statistics, 31(7), 799-815.

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.

Gehl, P., & Rohmer, J. (2018). Vector intensity measures for a more accurate reliability assessment of NPP sub-systems. In International Conference on Technological Innovations in Nuclear Civil Engineering, Saclay, France.

Grün, B., Kosmidis, I., & Zeileis, A. (2012). Extended Beta Regression in R: Shaken, Stirred, Mixed, and Partitioned. Journal of Statistical Software, 48(11), 1–25.

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

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

Hofner, B., Mayr, A., & Schmid, M. (2016). gamboostLSS: An R package for model building and variable selection in the GAMLSS framework. Journal of Statistical Software, 74(1), doi:10.18637/jss.v074.i01.

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.

Jones, P.J., Mair, P., Simon, T., & Zeileis, A. (2019). Network Model Trees, OSF ha4cw, OSF Preprints. doi:10.31219/osf.io/ha4cw

Kleemann, J., Celio, E., & Fürst, C., (2017). Validation approaches of an expert-based Bayesian Belief Network in Northern Ghana, West Africa. Ecol. Model., 365, 10–29.

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

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.

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

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

Mayr, A., Fenske, N., Hofner, B., Kneib, T., & Schmid, M. (2012). Generalized additive models for location, scale and shape for high dimensional data - a flexible approach based on boosting. Journal of the Royal Statistical Society, Series C, 61, 403–427.

Meinshausen, N., & Bühlmann, P. (2010). Stability selection. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 72(4), 417-473.

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

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

Gómez-Villegas, M.A., Main, P., & Susi, R. (2007). Sensitivity analysis in Gaussian Bayesian networks using a divergence measure. Comm. Statist. Theory Methods, 36(3), 523-539.

Gómez-Villegas, M. A., Main, P., & Susi, R. (2013). The effect of block parameter perturbations in Gaussian Bayesian networks: sensitivity and robustness. Inform. Sci., 222, 439-458.

Heckerman, D., Horwitz, E. & Nathwani, B. (1992). Towards Normative Expert Systems: Part I. The Pathfinder Project. Methods of Information in Medicine, 31, 90-105.

Jackson, P. (1999). Introduction to expert systems, third edition, Addison-Wesley.

Koenker, R., Leorato, S., Peracchi, F. (2013). Distributional vs. Quantile Regression. Technical Report 300, Centre for Economic and International Studies, University of Rome Tor Vergata, Rome, Italy.

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

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

Murphy, K. P. (1999). A variational approximation for Bayesian networks with discrete and continuous latent variables. In 15th conference on Uncertainty in artificial intelligence. Morgan Kaufmann Publishers Inc.

Pereira, G. H. (2019). On quantile residuals in beta regression. Communications in Statistics-Simulation and Computation, 48(1), 302-316.

Pitchforth, J., & Mengersen, K. (2013). A proposed validation framework for expert elicited Bayesian Networks. Expert Systems with Applications, 40(1), 162-167.

Rigby, R. A., & Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape. Journal of the Royal Statistical Society: Series C (Applied Statistics), 54(3), 507-554.

Russell, S., Norvig, P., Canny, J., Malik, J., & Edwards, D. (2003). Artificial intelligence: a modern approach. Prentice Hall, Englewood Cliffs.

Saltelli, A., Aleksankina, K., Becker, W., Fennell, P., Ferretti, F., Holst, N., Li, S., & Wu, Q. (2019). Why so many published sensitivity analyses are false: A systematic review of sensitivity analysis practices. Environmental modelling & software, 114, 29-39.

Schmid, M., Wickler, F., Maloney, K. O., Mitchell, R., Fenske, N., & Mayr, A. (2013). Boosted beta regression. PloS one, 8(4), e61623.

Smithson, M., & Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. Psychological Methods, 11, 54–71.

Smyth, G., & Verbyla, A. (1999). Adjusted likelihood methods for modelling dispersion in generalized linear models. Environmetrics, 10(6), 695–709.

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

Shenoy, P.P. (2006). Inference in hybrid Bayesian networks using mixtures of Gaussians Uncertainty in artificial intelligence, AUAI Press, Corvallis, 428-436.

Renooij, S. (2014). Co-variation for sensitivity analysis in Bayesian networks: Properties, consequences and alternatives. International Journal of Approximate Reasoning, 55(4), 1022-1042.

Thomas, J., Mayr, A., Bischl, B., Schmid, M., Smith, A., and Hofner, B. (2018), Gradient boosting for distributional regression - faster tuning and improved variable selection via noncyclical updates. Statistics and Computing, 28, 673-687.

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.

Wiegerinck, W., Kappen, B., & Burgers, W. (2010). Bayesian networks for expert systems: Theory and practical applications. In Interactive collaborative information systems. Springer, Berlin, Heidelberg, pp. 547-578.

Young, J., Graham, P., & Penny, R. (2009). Using Bayesian networks to create synthetic data. Journal of Official Statistics, 25(4), 549-567.

Tables

Table 1. CPT entries for the brain tumor BBN. For instance P(c|b0,isc1) corresponds to the probability of node C being in state 1 conditioned by the fact that node B is in state 0 and ISC is in state 1. The meaning of the other probabilities should be understood following this example.

Node	Conditional probability
MC	P(mc)=0.20
B	P(b|mc1)=0.20 P(b|mc0)=0.05
ISC	P(isc|mc1)=0.80 P(isc|mc0)=0.20
C	P(c|b1)=0.95 ; P(c|b1,isc1)=0.80 P(c|b1,isc0)=0.80 P(c| b0,isc0)=0.05
CT	P(ct|b1)=0.95 P(ct|b0)=0.10
SH	P(sh|b1)=0.80 P(sh|b0)=0.60

915 Table 2. Definition of the CPT entry in the concrete BBN.

Node	Description
X_1	Damage assessment
X_2	Cracking state
X_3	Cracking state in shear domain
X_4	Steel corrosion
X_5	Cracking state in flexure domain
X_6	Shrinkage cracking
X_7	Worst cracking in flexure domain
X_8	Corrosion state
X_9	Weakness of the beam
X_{10}	Deflection of the beam
X_{11}	Position of the worst shear crack
X_{12}	Breadth of the worst shear crack
X_{13}	Position of the worst flexure crack
X_{14}	Breadth of the worst flexure crack
X_{15}	Length of the worst flexure cracks
X_{16}	Cover
X_{17}	Structure age
X_{18}	Humidity
X_{19}	PH value in the air
X_{20}	Content of chlorine in the air
X_{21}	Number of shear cracks
X_{22}	Number of flexure cracks
X_{23}	Shrinkage
X_{24}	Corrosion

916 Table 3. Definition of the CPM parameters of the concrete BBN; for instance m_{0_1}, is the
917 intercept of the regression model (Eq. 1) linking node X_1 to its parents; Z_{1_9} is the regression
918 coefficient that models the relation between node X_1 and X_9; s_{0_1} is the corresponding standard
919 deviation.

Node	Intercept	Regression coefficient	Standard Deviation
X_1	m_{0_1}=0	Z_{1_9}=0.3 Z_{1_2}=2 Z_{1_10}=0.	s_{0_1}=1e-4
X_2	m_{0_2}=0	Z_{2_5}=0.7 Z_{2_4}=0.5 Z_{2_6}=0.3 Z_{2_3}=0.7	s_{0_2}=1e-4
X_3	m_{0_3}=0	Z_{3_8}=0.3 Z_{3_21}=0.5 Z_{3_12}=0.9 Z_{3_11}=0.7	s_{0_2}=1e-4
X_4	m_{0_4}=0	Z_{4_5}=0.3 Z_{4_24}=0.7 Z_{4_8}=0.7	s_{0_4}=1e-4
X_5	m_{0_5}=0	Z_{5_22}=0.5 Z_{5_7}=0.9 Z_{5_13}=0.7	s_{0_5}=1e-4
X_6	m_{0_6}=0	Z6_23=0.3 Z6_8=0.7	s_{0_6}=1e-4
X_7	m_{0_7}=0	Z_{7_17}=0.4 Z_{7_16}=0.4 Z_{7_15}=0.6 Z_{7_14}=0.6 Z_{7_8}=0.6	s_{0_7}=1e-4

X_8	m_{0_8}=0	Z_{8_20}=0.8 Z_{8_19}=0.9 Z_{8_18}=0.5	s_{0_8}=1e-4

920

921

922 Table 4. Parameters for constructing the CPT of the BBN-based reliability assessment

Parameter	Symbol	Original value
Mean and Standard deviation value of the log-normal fragility curve for STR	α_{STR}, β_{STR}	-6.48, 2.32
Mean and Standard deviation of the log-normal fragility curve for EDG1	α_{EDG1}, β_{EDG1}	-12.250, 6.22
Mean and Standard deviation of the log-normal fragility curve for EDG2	α_{EDG2}, β_{EDG2}	-4.66, 2.20
Dunnett-Sobel decomposition's parameters	$t_{1,2,3}$	0.914, 0.942, 0.999

923

924

925 Table 5. Summary of the strengths and weaknesses of the BBR method and of the approach
926 based on sensitivity functions.

Approach	Strengths	Weaknesses
Sensitivity function	It is simple to implement; It requires a low computational effort; The graphical representation is straightforward to interpret.	It is local: it focuses on the influence of one (or multiple) CPT parameters while the other ones are kept constant; It is restricted to the analysis of discrete BBNs; Multi-way SA can rapidly become intractable.
BBR	It provides insight in the global influence; It is simple to implement using random sampling, which is a generic procedure applicable to any types of BBNs; The results are intuitive to interpret based on a graphical presentation; It provides multilevel information on sensitivity; The combination of gradient boosting and stability selection increases the robustness to the number of parameters.	The number of calls to the BBN inference engine can be large (>1,000); The adequacy of the Beta model should be carefully checked; The demonstration of the applicability to very large-scale BBNs with hundreds of nodes and thousands of CPM parameters remains to be done.

927
928

Appendix A: Bayesian Belief Network Analysis

The analysis of Bayesian Belief Network relies on conditional probabilities. Consider $X_{i=1,\ldots,n}$ the *n* nodes of the BBN. The joint probability distribution can be expressed using conditional probability as:

$$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | X_1, \ldots, X_{i-1}) \quad \text{(Eq. A1)}$$

This equation simplifies under the conditional independence assumption as:

$$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | \mathrm{Pa}(X_i)) \quad \text{(Eq. A2)}$$

where $\mathrm{Pa}(X_i)$ corresponds to the parent nodes of X_i. For discrete nodes, the value of $P(X_i|\mathrm{Pa}(X_i))$ is the entry of the Conditional Probability Table. For continuous nodes, $P(X_i|\mathrm{Pa}(X_i))$ can be modelled by a continuous probability distribution whose parameters depend on the values of the parent nodes.

Conditional queries aim at evaluating the conditional probability of some event, e.g. node X_j takes up the value *x* when new information/observations become available, i.e. given new "evidence" (denoted *e*), namely $P(X_j=x \mid e)$. This procedure, termed as query, relies on inference techniques. Exact inference in a BBN is possible, but is generally not possible in large networks. For networks, which contain a large number of nodes, arcs, or have nodes comprised of variables with large numbers of levels, exact inference becomes too computationally intensive. Among the possible approximate inference algorithms, the present study relies on the logic sampling method (Koller and Friedman 2009), which works from the root nodes down to the leaf nodes. First, a random draw from each of the leaf nodes is taken with probabilities equal to the respective levels. The distributions for the next level of nodes can then be marginalized based on the draws obtained from their parents. From these marginal distributions a random draw is again taken respective of the categorical probabilities. This process is repeated until a draw has been taken from every node in the network. This completed case represents a single sample. After many samples have been taken, the samples are subsetted to the cases which match the evidence of interest. Estimated probabilities can then be obtained for any node of interest from this subset of the samples.

Appendix B: Gradient boosting within GAMLSS framework

We first describe the principles of model-based gradient boosting (Sect. B1), and describe how to apply it within the GAMLSS framework (Sect. B2).

B1 Gradient boosting

This supervised learning technique (e.g., Bühlmann and Hothorn, 2007) combines an ensemble of simple regression models (termed as base-learners), such as linear regression models or regression splines of low degree of freedom, to estimate complex functional relationships.

Consider the training dataset $\boldsymbol{D}=\{\boldsymbol{C}^{(j)}, P^{(j)}\}_{j=1,\ldots,n}$ where $\boldsymbol{C}$ is the vector of p predictor variables $c_{i=1,\ldots,p}$ and P is the variable of interest, whose expected value (possibly its transformed value) is modelled by an additive model as follows:

$$\eta(\boldsymbol{C}) = \beta_0 + \sum_{j=1}^{J} f_j(c_j|\beta_j) \qquad \text{(Eq. B1)}$$

where $J \leq p$, β_0 is a constant intercept and the additive effects $f_j(c_j|\beta_j)$ are pre-defined univariate base-learners, which typically correspond to (semi-)parametric effects with parameter vector β_j.

To estimate the parameters β_j, the boosting algorithm minimizes the empirical risk R, which corresponds to the sum of loss function ρ over all training data:

$$R = \sum_{j=1}^{n} \rho(P^{(j)}, \eta(\boldsymbol{C}^{(j)})) \qquad \text{(Eq. B2)}$$

where the loss function ρ can take different forms, such as the quadratic loss $(P^{(j)} - \eta(\boldsymbol{C}^{(j)}))^2$, which corresponds to the ordinary least square regression or more generally, it can correspond to the negative log-likelihood of the distribution of the variable of interest P (in the case considered here, this corresponds to the Beta distribution).

Among the different boosting algorithms, we focus on the component-wise gradient boosting approach of Bühlmann and Hothorn (2007). Instead of focusing on the true outcomes $\boldsymbol{P}=(P^{(1)}, P^{(2)},\ldots, P^{(n)})$, this procedure aims at fitting simple regression-type base learners $h(.)$ one by one to the negative gradient vector of loss $\boldsymbol{u}=(u^{(1)}, u^{(2)},\ldots, u^{(n)})$. The objective is to approximate the j^{th} effect $f_j(c_j|\beta_j) = \sum_m h_j(.)$ at iteration m of the algorithm. Formally, $\boldsymbol{u}$ is evaluated

considering the current estimated (i.e. iteration *m*-1) additive predictor model $\hat{\eta}^{[m-1]}(\boldsymbol{C}^{(j)})$ as follows:

$$\boldsymbol{u} = \left(-\frac{\partial}{\partial \eta} \rho(P, \eta)_{|_{\eta=\hat{\eta}^{[m-1]}(\boldsymbol{C}^{(j)}), P=P^{(j)}}} \right)_{j=1,\ldots,n} \quad \text{(Eq. B3)}$$

In every boosting iteration, each base-learner is fitted separately to the negative gradient vector. The best-fitting base-leaner is then selected based on the residual sum of squares with respect to $\boldsymbol{u}$ as follows:

$$j^* = \underset{j \in 1,\ldots,J}{\operatorname{argmin}} \sum_{i=1}^{n} (u^{(i)} - h_j(\boldsymbol{C}^{(i)}))^2 \quad \text{(Eq. B4)}$$

The selected base-leaner is used to update the current predictor model as follows:

$$\hat{\eta}^{[m]} = \hat{\eta}^{[m-1]} + s.h_{j*}(\boldsymbol{C}) \quad \text{(Eq. B5)}$$

where s is a step length (with typical value of 0.1).

The main tuning parameter is the number of iterations *m*, which directly determines the prediction performance. If *m* is too large, rich models with large number of predictors and rough functional terms will be constructed (hence leading to overfitting), which might hamper the interpretability of the resulting model. If *m* is too low, sparse models with smooth functional terms will be constructed but with the danger of missing some important predictor variables. In practice, the selection of *m* can be carried out using cross-validation procedures in order to optimize the predictive risk on observations left out from the fitting process i.e. the "out-of-bag" risk which corresponds to the negative log-likelihood for the considered probabilistic distribution calculated for the "out-of-bag" samples.

B2 Boosted GAMLSS

The afore-described algorithm can be applied to the GAMLSS framework by cycling through the distribution parameters θ in the fitting process (Thomas et al., 2018: Algorithm 1). In each iteration, the best fitting base-learner is evaluated for each distribution parameter, while the

other ones remain fixed. For a probability distribution with two parameters (like the Beta distribution), the update in the boosting algorithm at iteration *m* holds as follows:

$$\frac{\partial}{\partial \eta_{\theta_1}} \rho\left(P, \theta_1^{[m]}, \theta_2^{[m]}\right) \xrightarrow{\text{update}} \eta_{\theta_1}^{[m+1]} \quad \text{(Eq. B6)}$$

$$\frac{\partial}{\partial \eta_{\theta_2}} \rho(P, \theta_1^{[m+1]}, \theta_2^{[m]}) \xrightarrow{\text{update}} \eta_{\theta_2}^{[m+1]} \quad \text{(Eq. B7)}$$

In the original cyclic algorithm, (see algorithm 1 described by Thomas et al. (2018)), separate stopping values have to be specified for each parameter, which motivated the development of a noncyclic algorithm (algorithm 2 by Thomas et al. (2018)), which avoids optimizing two different stopping iterations and reducing the optimisation problem from a multi-dimensional to a one-dimensional problem.

Appendix C: Stability Selection Analysis

This appendix describes the procedure for stability selection (Meinhausen and Bühlmann, 2010). Consider p predictor variables $c_{j=1,\ldots,p}$ and the predictand *P*. Based on n observations, the stability selection with boosting proceeds as follows:

Step 1. Select a random subset of size [n/2] of the data (where [n/2] corresponds to the largest integer ≤ n/2);

Step 2. Fit a boosting model and continue to increase the number of boosting iteration until q base-learners are selected. $\hat{S}_{\left[\frac{n}{2}\right]}$ corresponds to the set of selected variables;

Step 3. Repeat steps 1 and 2 for b=1,…,B;

Step 4. Compute the selection probabilities per base learner as follows:

$$P_j = \frac{1}{B}\sum_{b=1}^{B} \mathbf{I}_{\{j \in \hat{S}_{\left[\frac{n}{2}\right],b}\}} \quad \text{(Eq. C1)}$$

where **I**(A) is the indicator function, which reaches 1 if A is true, 0 otherwise.

Step 5. Select all base-learners associated to a selection probability of at least t. The set of stable selected variables is thus $\hat{S} = \{j: P_j \geq t\}$.

Meinhausen and Bühlmann (2010) studied the error of selecting false positive variables (i.e. noise variables) and showed that the selection procedure controls the per-family error rate (*PFER*) and an upper bound is provided as follows:

$$PFER \leq \frac{q^2}{(2t-1)p} \quad \text{(Eq. C2)}$$

where q is the number of selected variables per boosting iteration, p is the number of (possible) predictors and t is the selection threshold. In practice, at least two of these parameters have to be specified to run the procedure. In the application case of Sect. 4.1, we specified the upper bound of *PFER* (set up at 1) and q at 10. An extensive investigation of the applicability of this procedure for distributional regression within the boosted GAMLSS setting has been carried out by Thomas et al. (2018).

Appendix D: Residuals for Beta regression

The following appendix describes the main formulations of the residuals for Beta regression based on Pereira (2019) and references therein.

Let us consider the raw response residuals for the i[th] observation P_i defined as $r_i = P_i - \widehat{P_i}$ where $\widehat{P}$ is the mean μ predicted by BBR model. For sake of presentation, let us consider the Beta parameter $\varphi = \frac{1}{\sigma^2} - 1$ where σ is defined in Sect. 2.1.

The deviance residual is defined as:

$$r_i^{deviance} = r_i . \sqrt{2|L(P, \varphi) - L(\mu, \varphi)|} \quad \text{(Eq. D1)}$$

where the bivariate function $L(.,.)$ holds as follows:

$$L(\mu, \varphi) = l\Gamma(\varphi) - l\Gamma(\varphi . \mu) - l\Gamma\big((1-\mu).\varphi\big) + (\mu.\varphi - 1).\log(P) + \big((1-\mu).\varphi - 1\big).\log(1-P) \quad \text{(Eq. D2)}$$

where $l\Gamma(.)$ is natural logarithm of the absolute value of the gamma function.

The standardized weighted residual 1 originally introduced by Espinheira et al. (2008) is defined as follows:

$$r_i^{weighted} = \frac{P^* - \mu^*}{\sqrt{v}} \quad \text{(Eq. D3)}$$

where P^* is the quantile for the logistic distribution with location set up at zero and scale at one, and

$$\mu^* = \Gamma''(\mu.\varphi) - \Gamma''((1-\mu).\varphi) \quad \text{(Eq. D4)}$$

$$v = \Gamma'''(\mu.\varphi) - \Gamma'''((1-\mu).\varphi) \quad \text{(Eq. D5)}$$

where Γ'' and Γ''' are respectively the second and third derivative of the Γ function.

The quantile residual is defined as:

$$r_i^{Quantile} = \Phi^{-1}(F(P, \mu, \varphi)) \quad \text{(Eq. D6)}$$

where Φ^{-1} is the inverse of the cumulative probability function of the standard normal distribution and F is the cumulative probability function of the Beta law.