



HAL
open science

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

Jeremy Rohmer, Pierre Gehl

► To cite this version:

Jeremy Rohmer, Pierre Gehl. Sensitivity analysis of Bayesian networks to parameters of the conditional probability model using a Beta regression approach. Expert Systems with Applications, In press, <10.1016/j.eswa.2019.113130>. <hal-02408006>

HAL Id: hal-02408006

<https://brgm.hal.science/hal-02408006v1>

Submitted on 12 Dec 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

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

Jeremy Rohmer^{1,*}, Pierre Gehl²

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

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

35

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

38

39 1 Introduction

40 Bayesian Belief Network (BBN) is widely recognized as a valuable modelling tool for expert
41 systems (e.g. Russel et al., 2003). It has been applied in various complex domains, like
42 ecosystems (Milns et al., 2010), genetics and biology (Scutari et al., 2014), industry (Weber et
43 al., 2012), finance forecasting (Malagrino et al., 2018), marine safety (Hänninen et al., 2014),
44 nuclear power plants (Kwag & Gupta 2017), coastal systems (Jäger et al., 2018), multi-hazard
45 risk assessments (Gehl and D’Ayala, 2016), etc. An expert system has the ability to represent
46 and reason on knowledge with the purpose of solving problems and giving advice (as defined
47 by Jackson, 1999), by relying on three components (i.e. knowledge base, observation base,
48 inference engine). Each of these components can benefit from the key features of BBN: (1) its
49 capability to represent expert knowledge and to combine and integrate expert knowledge with
50 information from any kind of sources, including experimental data, historical data, results from
51 numerical simulations, etc.; (2) its high flexibility to model any causal relationships and to
52 explicitly display the relationship among variables using a network-based approach, which can
53 intuitively be understood by experts (Wiegerinck et al., 2010); and (3) its capability to answer
54 probabilistic queries about them and to find out updated knowledge of the state of a subset of
55 variables when other variables (the evidence variables) are observed. For instance, probabilistic
56 queries in reliability assessments may correspond to finding the probability values of some
57 failure cause given the observed damage level of the considered system. See Appendix A for a
58 brief overview. The interested reader can also refer to Jensen (2001) for a complete introduction
59 to BBN.

60 Formally, BBN is based on the graphical representation of the probabilistic relations among
61 random variables by means of a directed acyclic graph composed of nodes (i.e. the states of the
62 random variables) and arcs (i.e. dependency between nodes). See an example in Fig. 2A. The
63 nodes connected by an arc are called the parent and child nodes respectively. One child node
64 may have several parent nodes, meaning that this node is affected by several factors. Similarly,
65 a parent node could have several child nodes, meaning that this factor may have influences on
66 several other factors. Conditional probabilities are the probabilities that reflect the degree of
67 influence of the parent nodes on the child node. The probabilistic dependence (i.e. the relation
68 cause-effect) is represented via a table called a Conditional Probability Table (CPT) when the
69 variables X (nodes) are discrete. In this case, the CPT entries correspond to the probability value
70 $P(X_i = k | Pa(X_i) = j)$ where k denotes the k^{th} possible level (either a discrete value or a
71 category) that node X_i can take given that its parents $Pa(X_i)$ takes the j^{th} possible level. When
72 the variables are continuous, the conditional distribution given its parents can typically be
73 represented by means of a continuous probability distribution. The most popular model is the
74 Gaussian distribution, as follows:

75

$$76 \quad P(X_i = y | Pa(X_i) = \mathbf{x}) = G(y | m_0 + Z\mathbf{x}, S) \quad (\text{Eq. 1})$$

77

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

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

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

91 model varies with variation of the CPM parameters. Subsequently, the results from SA can be
92 used as a basis for parameter tuning, as well as for studying the robustness of the model output
93 to changes in the parameters (Coupé & van der Gaag 2002; Laskey et al. 1995).

94 For discrete BBNs, a widespread SA method relies on the use of sensitivity functions (Coupé
95 & van der Gaag 2002; Castillo et al., 1997), which describe how the considered output
96 probability varies as one CPT entry value is changed. Some recent extensions have been
97 proposed to conduct multi-way SA like the framework by Leonelli et al. (2017), i.e. SA when
98 several CPT entries are allowed to vary. Effects of parameter changes can be described by the
99 Chan–Darwiche (CD) distances (Chan–Darwiche 2002; 2005). The CD distance is used to
100 quantify global changes by measuring how the overall distribution behaves when one (or more)
101 parameter is varied. Likewise, SA approaches have been developed for continuous BBNs
102 (mostly based on linear Gaussian regression) either based on computing partial derivatives
103 (Castillo & Kjærulff 2003) or based on the use of divergence measures (Gómez-Villegas et al.,
104 2007), with generalisation to multi-way SA (Gómez-Villegas et al., 2013).

105 Yet, several limitations of the existing SA methods exist.

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

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

112 (3) The SA is usually performed by focusing on one type of co-variation of the different
113 parameters (like proportional, uniform and order-preserving co-variation, see e.g., Renooij,
114 2014).

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

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

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

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

140

141 **2 Statistical methods**

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

147 **2.1 Overall procedure**

148 The different steps of the procedure hold as follows:

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

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

- 156 - Step 2: estimate the query probability of interest P derived from the inferences using
 157 the BBN. This probability value is the predictand;
- 158 - Step 3: establish the link between C and P using a regression model. We propose to rely
 159 on Beta regression models (see Sect. 2.1) using boosting-based fitting procedure (Sect.
 160 2.2) to deal with the potentially large number of predictors;
- 161 - Step 4: check the adequacy of the Beta model by checking that the residuals are well
 162 approximated by the standard normal distribution (see Sect. 2.4).

163 2.2 Use of Beta regression for BBN sensitivity analysis

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

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

$$175$$

$$176 \quad d(P, a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} P^{a-1}(1 - P)^{b-1} \quad (\text{Eq. 2})$$

177

178 where Γ is the gamma function, (a,b) are the shape parameters. In the following, we preferably
 179 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

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

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

185

$$186 \quad (P(n - 1) + 0.5)/n \quad \text{(Eq. 3)}$$

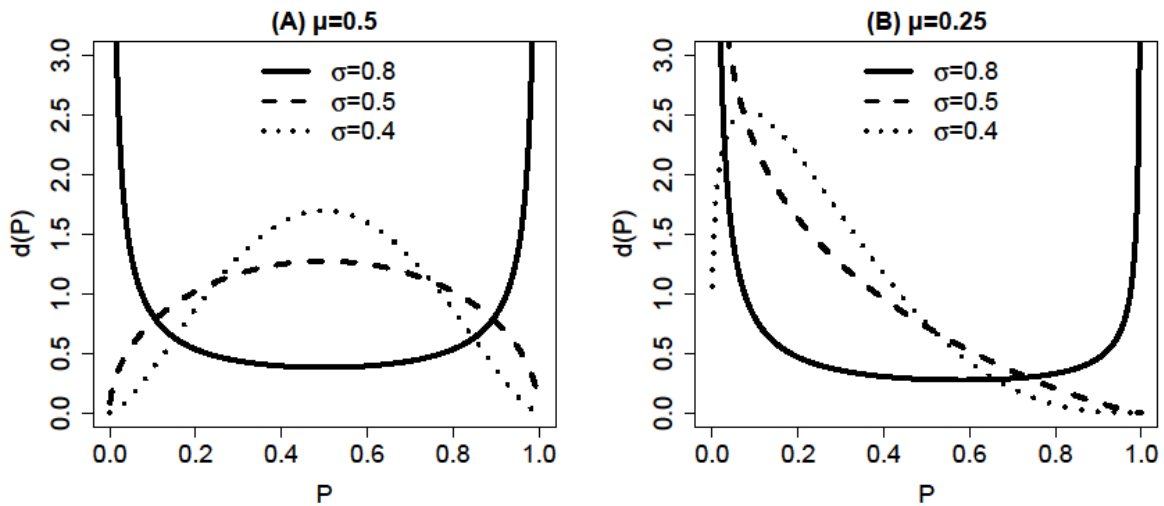
187 where n is the number of samples (i.e. the number perturbations performed to study the
188 sensitivity of P).

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

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

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

209 - studying the evolution of σ enables the BBN practitioners to identify the sources of data
 210 variability (e.g., Smyth and Verbyla, 1999) i.e. to study the influence on the uncertainty
 211 of P . In this manner, the BBN practitioners can identify different probabilistic regimes;
 212 e.g. situations where the probability of interest might switch from low to high values,
 213 i.e. situations of low to high risk. Such situations can be highlighted by low and high
 214 values of μ . Values of σ are then useful to indicate the confidence in the occurrence of
 215 such situations. For instance, high value of μ (high average value of P) together with
 216 low value of σ (low variance) provide strong evidence that changes in the considered
 217 parameter(s) might surely lead to situations of high probability. High values for both
 218 Beta parameters show, however, that the situation of high probability might occur but
 219 only with low confidence (i.e. high uncertainty).
 220



221
 222 *Figure 1. Probability density functions for the beta law. (A) mean $\mu=0.5$; (B) mean $\mu=0.25$.*

223
 224 Given the transformation of P (Eq. 3), BBN sensitivity analysis can be performed using Beta
 225 regression models (step 3 of the procedure) whose parameters are fitted using, for instance,
 226 maximum likelihood techniques (Cribari-Neto and Zeileis 2010). Yet, the functional form of
 227 the predictor-predictand relationship (e.g., quadratic or exponential) can hardly be specified in
 228 advance in our case (except for specific cases, like binary discrete BBN) and the relationships
 229 should preferably be learnt from the data. A possible option can rely on more advanced Beta
 230 distribution regression techniques, for instance using generalized additive models for location,
 231 scale and shape (GAMLSS, Rigby and Stasinopoulos, 2005), which allows deriving smooth

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

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

240 **2.3 Boosted Beta Regression**

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

246

$$247 \quad \theta = \eta_{\theta}(\mathbf{C}) = \beta_0 + \sum_{j=1}^J f_j(c_j|\beta_j) \quad (\text{Eq. 4})$$

248

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

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

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

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

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

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

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

286 **2.4 Model adequacy**

287 Once the BBR model has been fitted, an important aspect is to check the model adequacy (step
288 4 of the procedure), i.e. how well the BBR model is appropriate to describe the randomly
289 generated BBN-derived probabilities. This can be done by analysing the statistics of the
290 residuals and checking whether their distribution is well approximated by the standard normal
291 distribution (see e.g. Rigby and Stasinopoulos, 2005). Yet, contrary to ordinary least square

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

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

301

$$302 \quad R^2 = 1 - \frac{\sum_{i=1}^{i=N} (\hat{q}_i - q_i)^2}{\sum_{i=1}^{i=N} (q_i - \bar{q})^2} \quad (\text{Eq. 5})$$

303

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

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

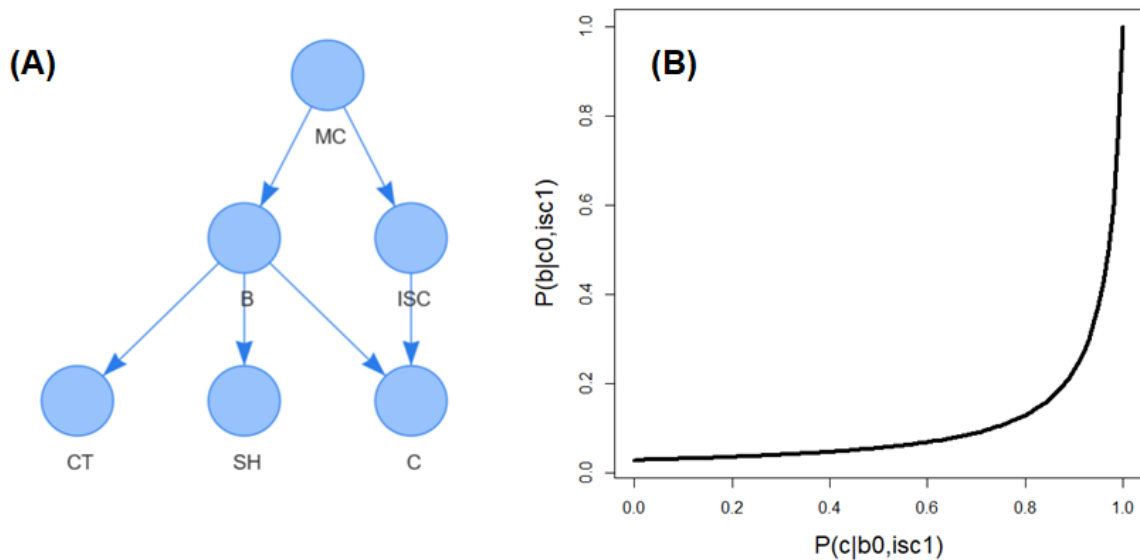
317

318 **3 Synthetic case study**

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

321 **3.1 Description**

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



324
325 *Figure 2. A) Brain tumor structure network; B) Sensitivity function derived by van der Gaag et al. (2013)*
326 *showing the evolution of the probability of interest $P(b|c0,isc1)$ as a function of the CPT entry*
327 *$P(c|b0,isc1)$.*

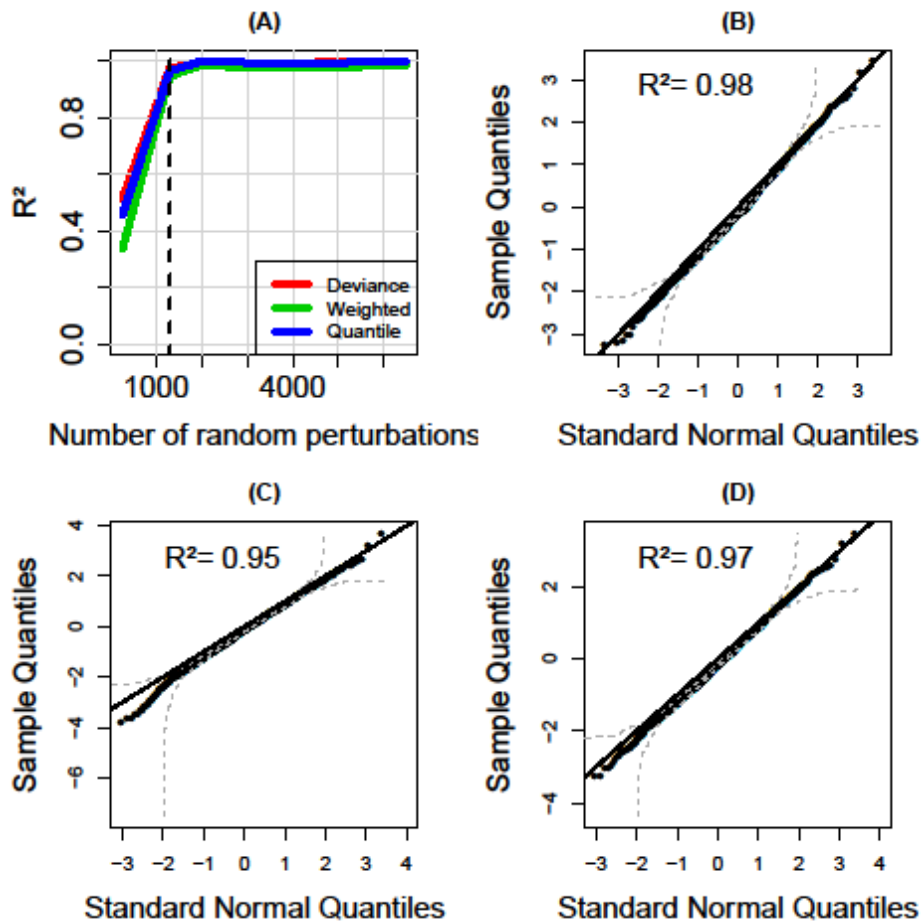
328
329 Node MC refers to metastatic cancer, which may potentially lead to the development of a brain
330 tumour (node B) and may give rise to an increased level of serum calcium (node ISC). The
331 presence of a brain tumour can be established from a CT scan (CT). Another indicator of the
332 presence of a brain tumour can be related to severe headaches (SH). A brain tumour or an
333 increased level of serum calcium are both likely to cause a patient to fall into a coma (C). The
334 conditional probabilistic relationships between the nodes (CPT entries) are provided in Table
335 1. We focus here on the probability $P=P(b|c0,isc1)$, namely the probability to develop brain
336 tumor given the absence of coma and an increased level of serum calcium. The robustness of P
337 is studied with respect to the values of 13 CPT entries (Table 1).

338
339 [Table 1 about here]

340 3.2 Application

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

347 The minimum number of random perturbations of the BBN was chosen by studying the
348 evolution of the R^2 indicator values for the different residuals' formulations (see Sect. 2.3). Fig.
349 3A shows that the convergence can be considered reached for a minimum number of ten times
350 the number of CPT entries, i.e. 1,300 for which the R^2 values all reach very satisfactory values
351 above 95%. The visual inspection of the normal Q-Q plots in Fig. 3B-D confirms the
352 satisfactory adequacy of the BBR model. We can however note some deviations for very low
353 quantile values but this is only indicated by one residuals' formulation (Fig. 3C). Besides,
354 taking into account the 95% confidence band (outlined by dashed lines in Fig. 3C), this
355 discrepancy can be considered low.

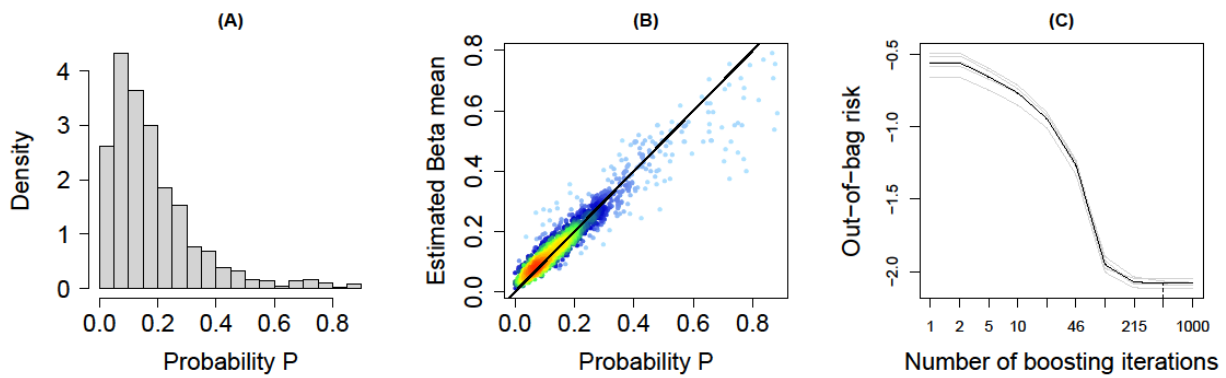


356

357 *Figure 3. (A) Evolution of the R^2 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 R^2 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.



372

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

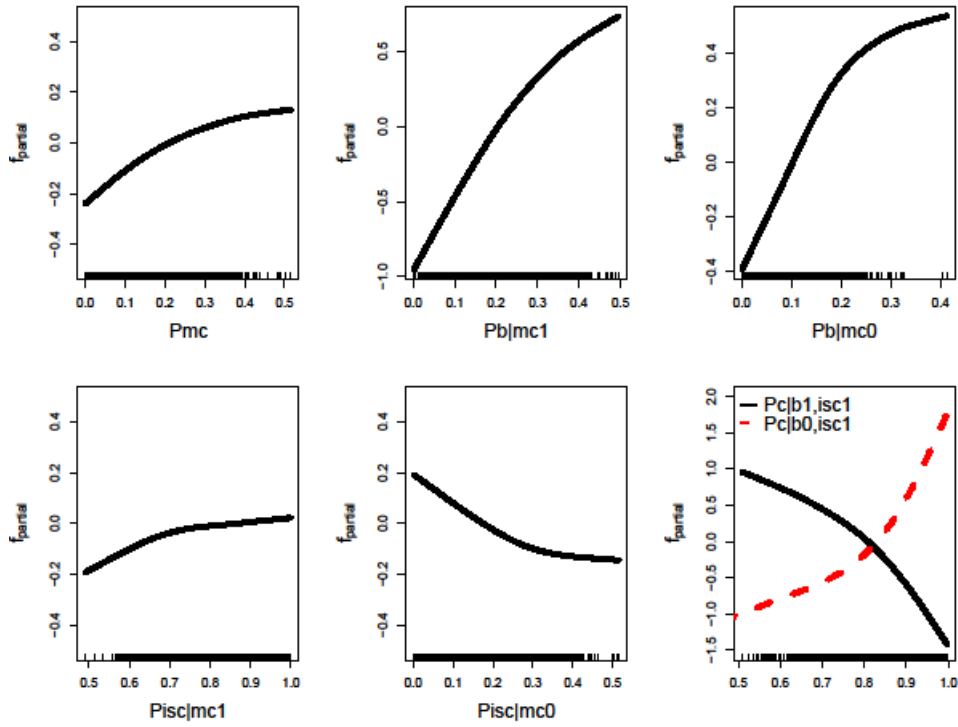
378

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

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

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

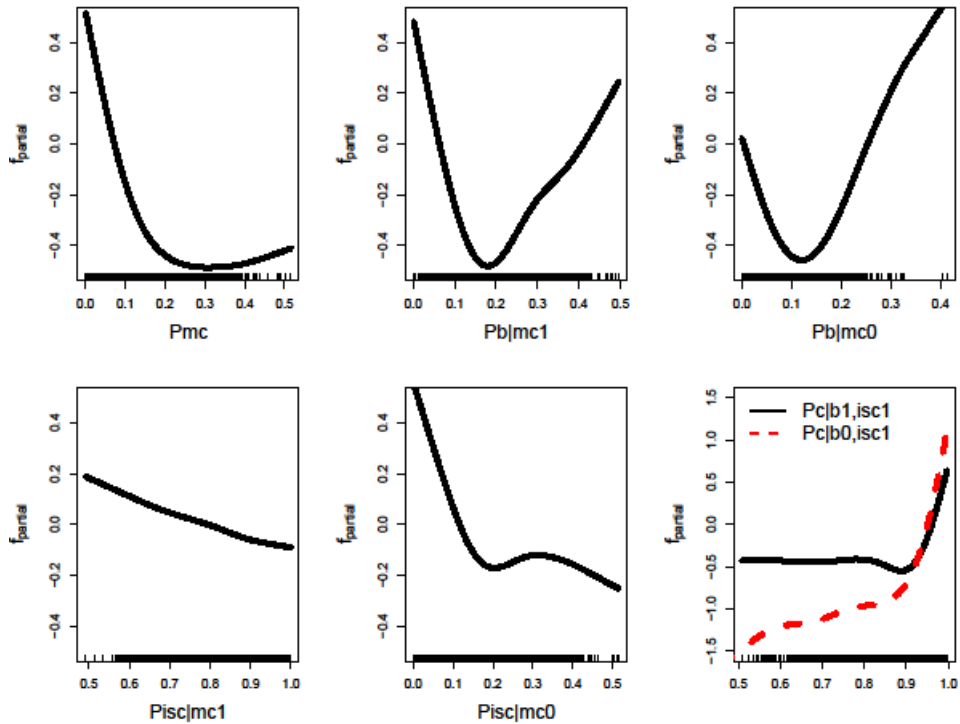
399



400

401 *Figure 5. Partial effect of each CPT entry on μ (logit-transformed) applied to the brain tumor network.*

402 *Note the different scales of the x- and y-axis.*



403

404 *Figure 6. Partial effect of each CPT entry on σ (logit-transformed) applied to the brain tumor network.*

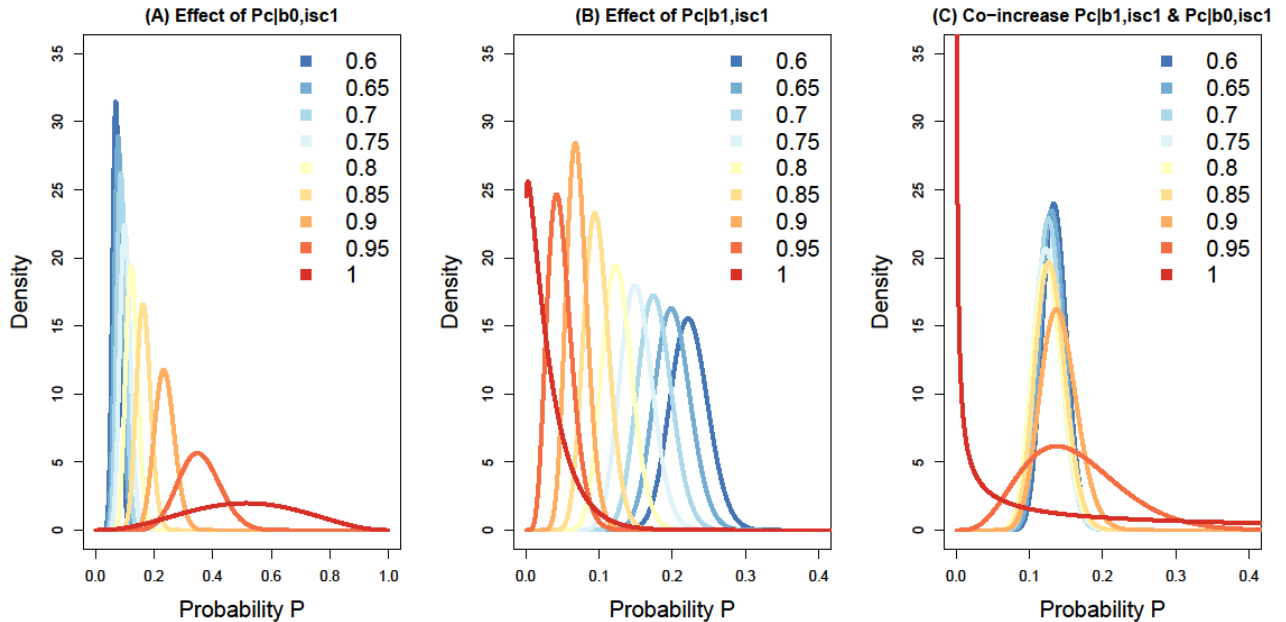
405 *Note the different scales of the x- and y-axis.*

406

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

- 411 - When $P(c|b0,isc1)$ is varied from low to high values (Fig. 7A), the corresponding Beta
 412 density function is translated from low to high values, but with more and more
 413 dispersion around the mode, which is more specifically amplified when $P(c|b0,isc1)$
 414 exceeds value of ~ 0.9 ;
- 415 - When $P(c|b1,isc1)$ is varied from low to high values (Fig. 7B), the effect on the Beta
 416 mode is opposite to the one due to the effect of $P(c|b0,isc1)$. This exemplifies the
 417 decreasing trend outlined in Fig. 5 (bottom right hand corner): the mode is translated
 418 from high to low values with an effect on the dispersion less pronounced than for
 419 $P(c|b0,isc1)$. This is in agreement with the lower partial effect outlined in Fig. 6(bottom
 420 right hand corner) ;

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



425
 426 *Figure 7. (A) Evolution of the Beta law fitted to the P values given increasing values of $P(c|b0,isc1)$*
 427 *from low (0.6) to high values (1.0); (B) Evolution of the Beta law fitted to the P values given increasing*
 428 *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*
 429 *values given increasing values (with same increment) of both $P(c|b1,isc1)$ and $P(c|b0,isc1)$.*

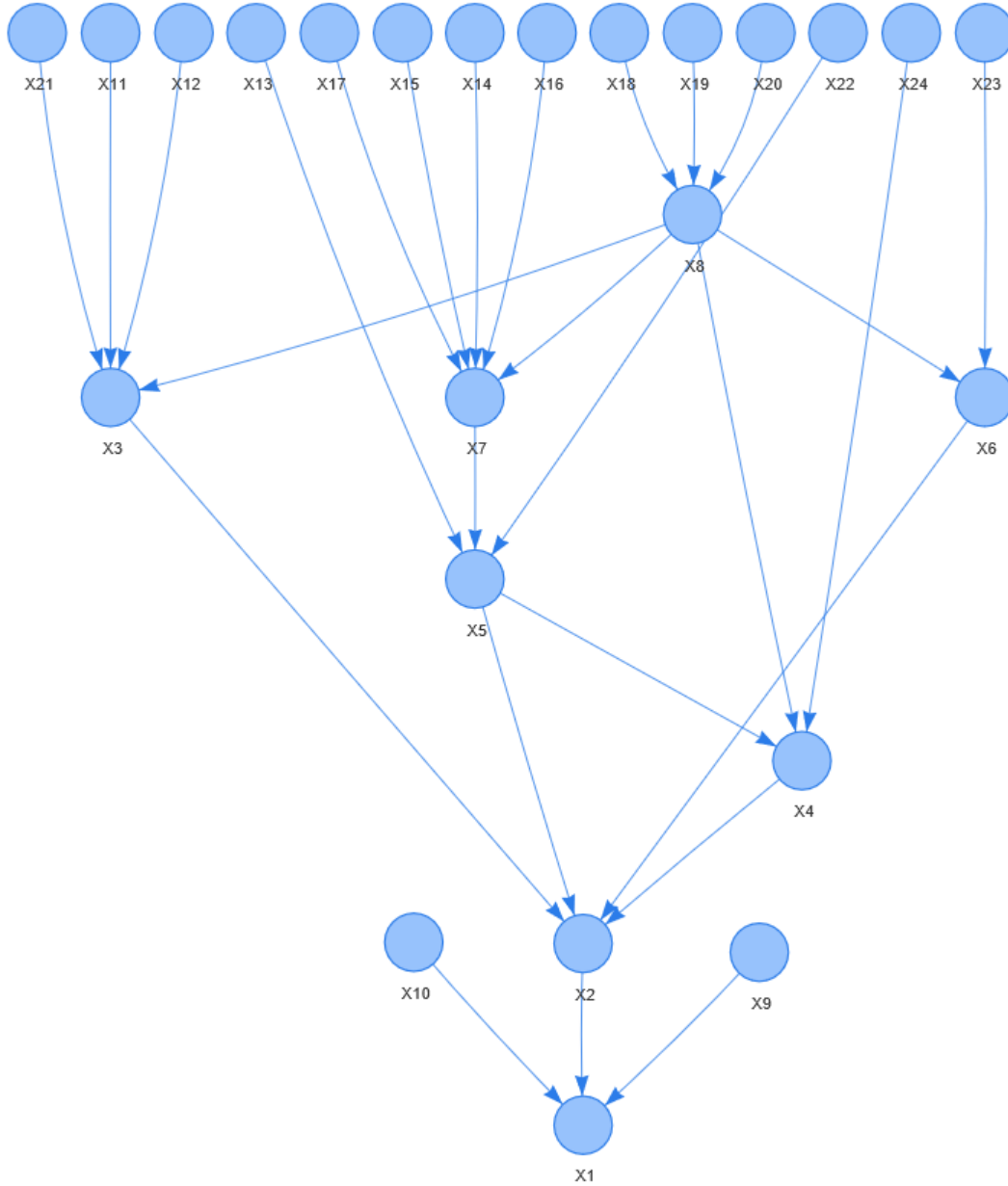
430

431 **4 Real-case applications**

432 In this section, we apply the BBR approach to real cases. First, a linear Gaussian BBN is
 433 considered for assessing the damage of reinforced concrete structures (Castillo and Kjærulff,
 434 2003). This enables us to illustrate a situation where the number of parameters is large enough
 435 to hamper the interpretation, i.e. here more than 40 variables have to be processed (Sect. 4.1).
 436 Second, a discrete BBN is considered for reliability analysis. For this case, analytical sensitivity
 437 functions can hardly be derived since the interest is not the sensitivity to the values of the CPT
 438 entries directly, but the physical parameters, which determine them.

439 **4.1 Reinforced-Concrete BBN**

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



442

443 *Figure 8. Structure of the reinforced concrete network.*

444

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

447

448

[Table 2 about here]

449

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

455 We investigate the robustness of the probability for the damage of the system $X_1 \geq 1$ given the
456 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
457 in Table 3 are randomly perturbed using Gaussian laws truncated at zero (with mean correspond
458 to the values given in Table 3, standard deviation of 0.2 for the regression coefficients and of
459 0.5 for the intercept terms). In total 43 variables are considered.

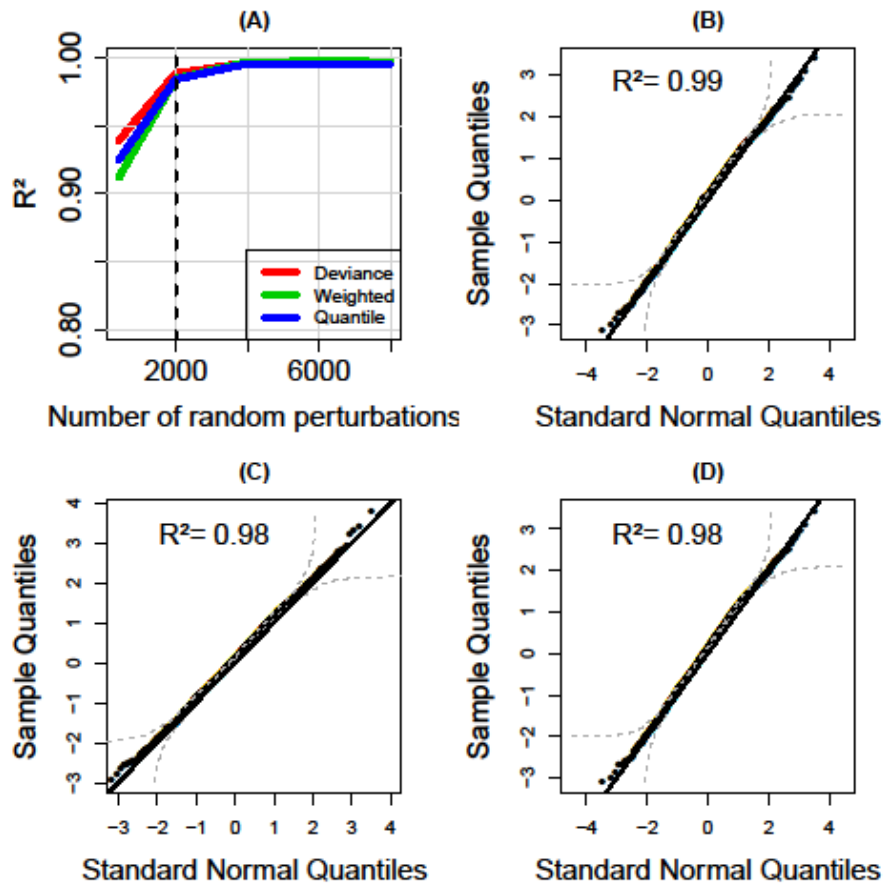
460

461

[Table 3 about here]

462

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

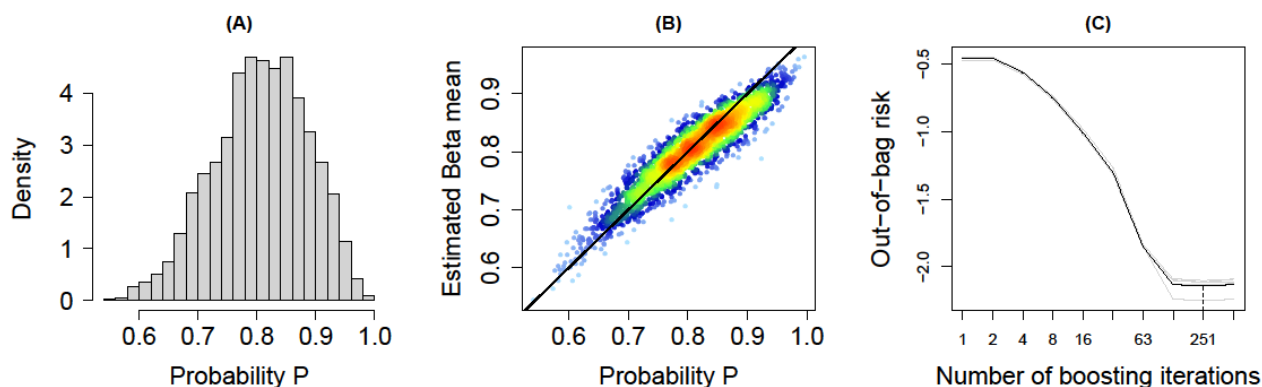


467

468 *Figure 9. (A) Evolution of the R^2 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 R^2 indicator (Eq. 5) is also reported.*

474

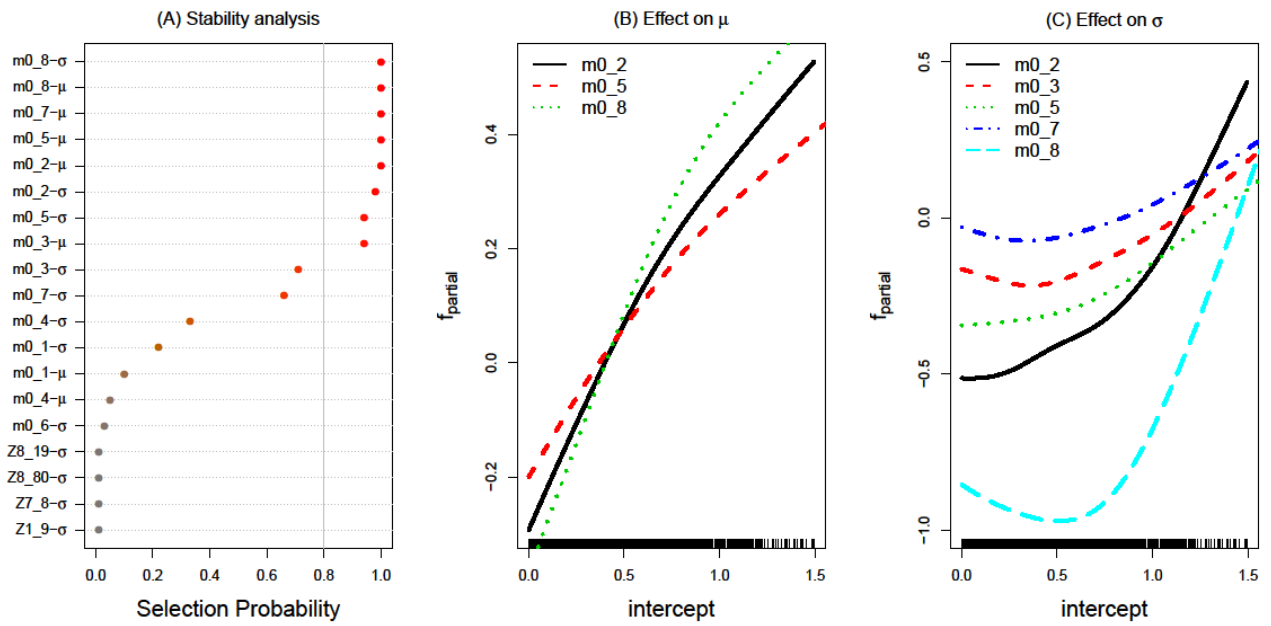
475 Fig. 10A depicts the corresponding histogram of the randomly generated P values (with mean
 476 of 0.8 and standard deviation of 0.08). Fig. 10B shows that the Beta mean is very informative
 477 with a good agreement with P . The optimal stopping iteration of the regression model is selected
 478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



480
 481 *Figure 10. (A) Histogram of randomly generated P values for the concrete network; (B) Comparison*
 482 *between the estimated Beta mean and P (the colours indicate the density of the dots); (C) Evolution of*
 483 *the (out-of-bag) risk estimated using a 5-fold cross-validation procedure as a function of the number of*
 484 *boosting iterations; the optimal stopping iteration is selected as the one minimizing the average risk*
 485 *over all 5 cross-validation iterations (indicated by a vertical dashed line at 251).*

486

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



500

501 *Figure 11. (A) Selection probability for each CPM parameter of the concrete network derived from the*
 502 *stability selection analysis (Appendix C). (B) Partial effect of the intercept m_0 of nodes X_2 , X_5 and X_8 on*
 503 *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*
 504 *parameter (logit-transformed).*

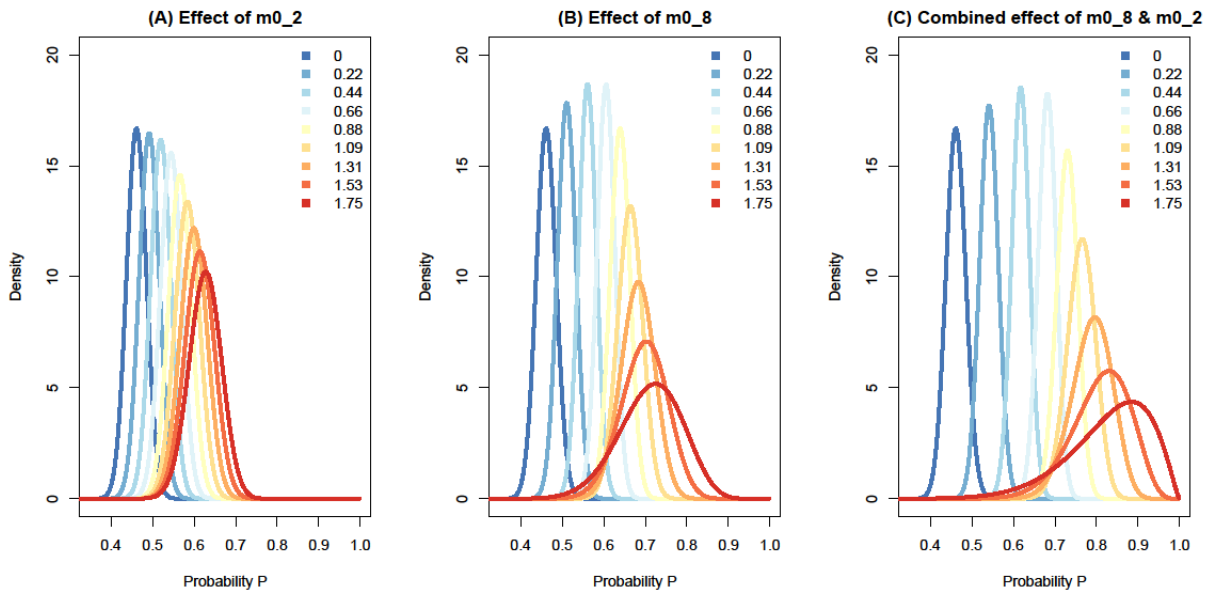
505

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

511 - Fig. 12A shows that the increase of $m_{0,2}$ mainly affects the increase of the best estimate
 512 of P (with only slight effect on the dispersion), whereas Fig. 12B both affects the best
 513 estimate and the dispersion, i.e. the larger $m_{0,8}$, the larger the mean value of P (i.e. the
 514 riskier the situation), but also the larger the uncertainty on P (i.e. the occurrence of the
 515 risky situation remains uncertain);

516 - Fig. 12C shows the evolution of the Beta's distribution when both parameters are
 517 increased similarly: we see that the combined increase amplifies the translation of the
 518 Beta's distribution to very high values, i.e. risky situations (driven by both increasing
 519 effects as shown in Fig. 12B). Despite the increase in σ during this process, the
 520 occurrence of this risky situations remains of moderate-to-high confidence, which

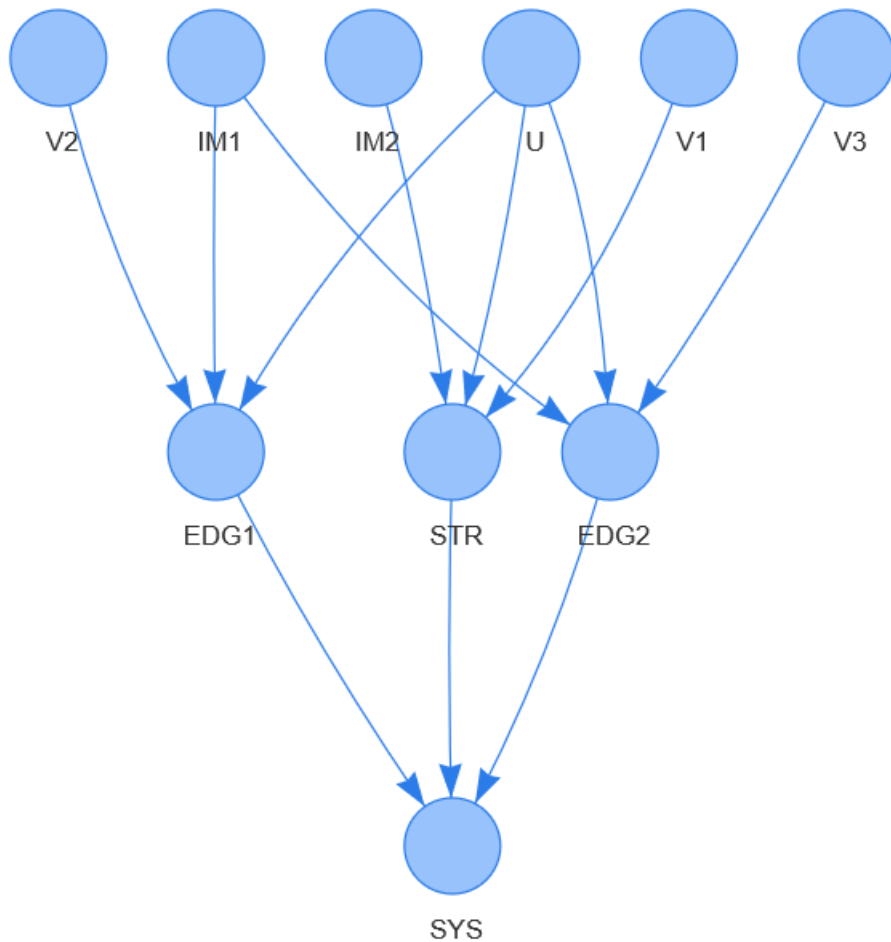
521 suggests that changes in m_{0_8} and m_{0_2} might ultimately lead to the failure of the concrete
522 system.



523
524 *Figure 12. (A) Evolution of the Beta law fitted to the P values given increasing values of the intercept*
525 *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*
526 *given increasing values of the intercept m_0 of node X_8 from low (0.) to high values (1.75); (C) Evolution*
527 *of the Beta law fitted to the P values given increasing values (with same increment) of both intercepts.*
528

529 4.2 BBN-based reliability assessment

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



538

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

543

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

552
$$Z_i = t_i \cdot U + \sqrt{1 - t_i^2} \cdot V_i \quad (\text{Eq. 6})$$

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

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

558

559
$$\beta_c = \frac{\alpha_i - \ln im}{\beta_i} \quad (\text{Eq. 7})$$

560

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

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

568

569
$$\left(z_i = t_i \cdot u + \sqrt{1 - t_i^2} \cdot v_i \right) \leq -\frac{\alpha_i - \ln im}{\beta_i} \quad (\text{Eq. 8})$$

570

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

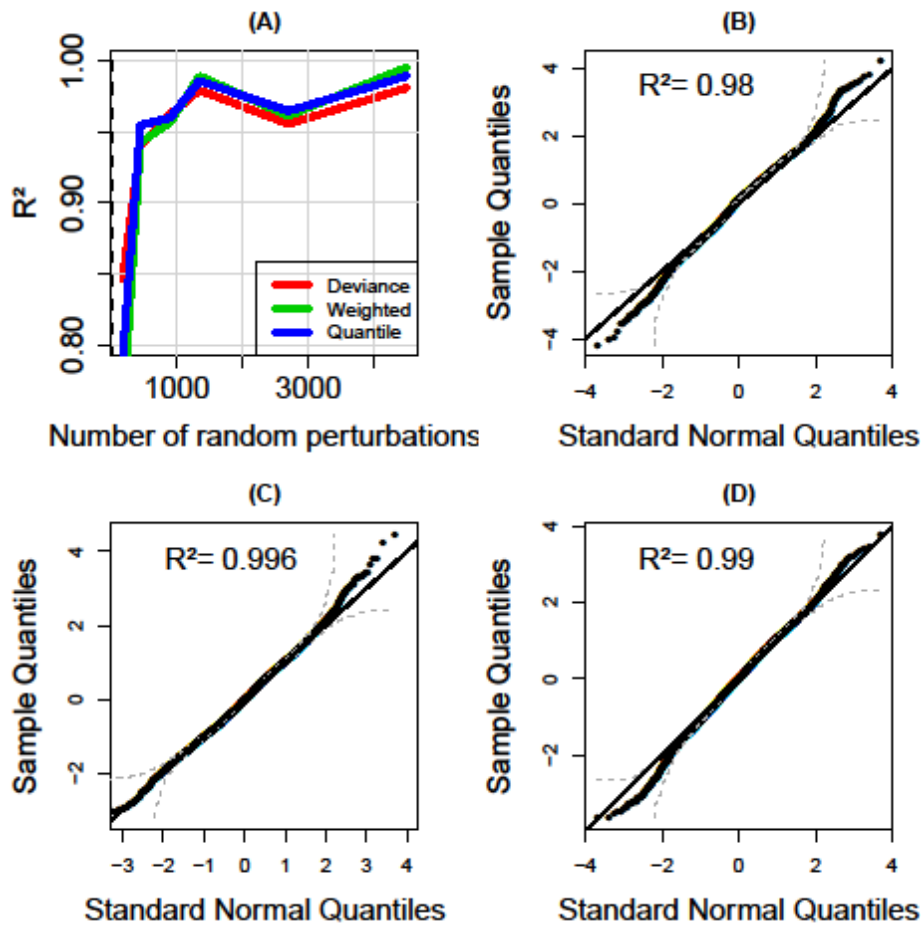
579 by intensity measures IM1 and IM2 of respectively 10 and 12.5 m²/s. In total, 9 parameters are
580 considered.

581

582 *[Table 4 about here]*

583

584 The selection of the minimum number of random perturbations is more difficult than for the
585 brain tumour or for the reinforced concrete case. Fig. 14A shows that at least 1,000 random
586 perturbations are necessary to reach R^2 values above 95%, but due to oscillations in the
587 evolution of the R^2 values, we preferably choose the largest number, i.e. >4,000. The visual
588 inspection of the normal Q-Q plots (Fig. 14C-D) also reveals some deviations for very high and
589 low quantiles (outside the range [-2 - 2]) though it should be noted that they remain within the
590 95% confidence band and with low-to-moderate magnitude. Contrary to the brain tumour case,
591 the three residuals' formulations all agree on the identification of the problem. A deeper
592 analysis show that these deviations correspond here to Beta distributions with very high mean
593 and very low variance values: the sensitivity should be analysed with care for these cases.



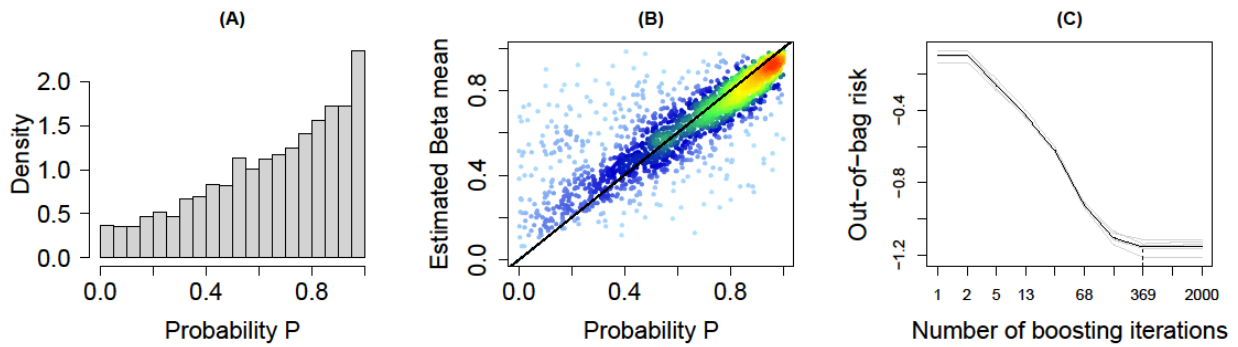
594

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

602

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

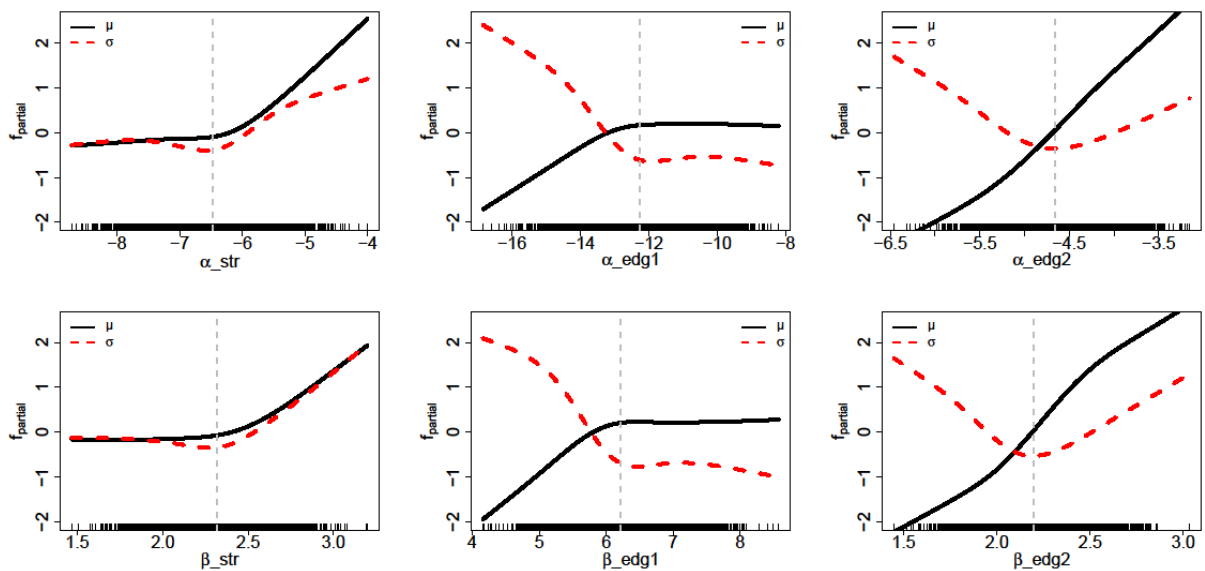
610 on a 5-fold cross validation procedure combined with the noncyclic algorithm as illustrated in
 611 Fig. 15C.



612
 613 *Figure 15. (A) Histogram of randomly generated P values for the NPP reliability network; (B)*
 614 *Comparison between the P values and the estimated mean value provided by the Beta model; (C)*
 615 *Evolution of the (out-of-bag) risk estimated using a 5-fold cross-validation procedure as a function of*
 616 *the number of boosting iterations; the optimal stopping iteration is selected as the one minimizing the*
 617 *average risk over all 5 cross-validation iterations (indicated by a vertical dashed line at 369).*

618

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



623

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

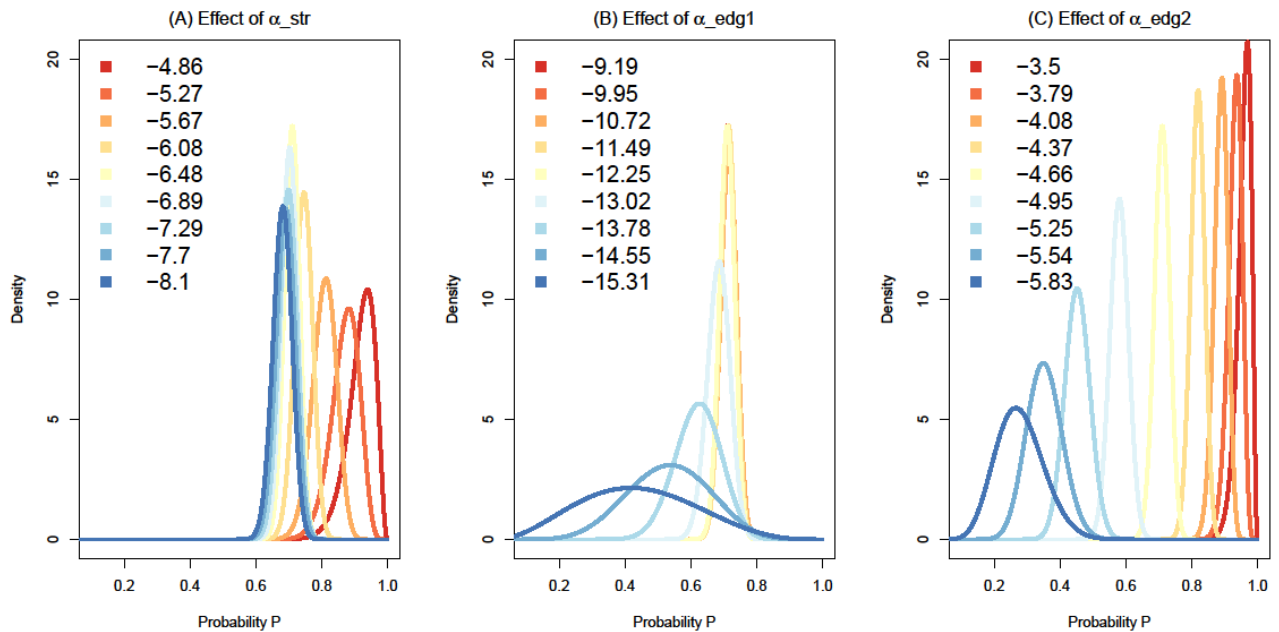
627

628 Several observations can be made:

- 629 - The parameters (α_{str} , β_{str}) of the structural fragility curve (STR, Fig. 16, left) have
630 almost similar effect, whatever the Beta law parameter (logit-transformed); it
631 corresponds to a quasi-bilinear function, which remains quasi-constant below the CPM
632 best estimate then increases above it; from a risk viewpoint, this means that the damage
633 to the structure enhances the occurrence of risky situations, where the system failure's
634 probability P might reach high values. Yet, the effect on σ parameter (logit-transformed)
635 also indicates that the confidence in the occurrence of this probabilistic regime is low;
- 636 - The effects of the parameters (α_{edg2} , β_{edg2}) of the second electrical generator's
637 fragility curve (EDG2, Fig. 16, right) are quasi-linear with respect to the Beta law mean
638 μ , but is non-monotonic with respect to the σ parameter (logit-transformed). This can
639 be schematically understood as a decrease of uncertainty on P (i.e. the Beta law
640 dispersion) when approaching the original value, then an increase above it;
- 641 - The effect of the parameters (α_{edg1} , β_{edg1}) of the first electrical generator's fragility
642 curve (EDG1, Fig. 16, center) is more complex. The effect on the Beta mean μ (logit-
643 transformed) corresponds to a bilinear function, which increases at around the CPM best
644 estimate, then remains quasi-constant above it (i.e. to a lesser extent, it corresponds to
645 the opposite behaviour to the one for STR). The effect on the Beta parameter σ
646 corresponds to an inverted sigmoid, which reaches its lower horizontal plateau above
647 the CPM best estimate. This means that above their original values, parameters (α_{edg1} ,
648 β_{edg1}) little affect the system failure's probability.

649 Fig. 17 illustrates how α_{str} , α_{edg1} and α_{edg2} act differently on the system failure. When
650 increasing α_{str} , the system failure starts increasing only for values well above the CPM best
651 estimates. This is shown by the Beta distributions, which remains "stuck" at moderate values
652 (around 0.6; see for cold colours in Fig. 17A). When increasing α_{edg1} , the resulting Beta
653 distribution remains stuck at high value around 0.70-0.75 (note the evolution of the mode for
654 warm colours in Fig. 17B), hence leading to risky situations. The confidence in the occurrence

655 of such probabilistic regime is high as indicated by the low dispersion of the Beta distributions.
 656 When increasing α_{edg2} , the Beta distribution continues to translate towards very high values
 657 above 0.90. Despite the very low dispersion of the Beta distribution, the confidence remains
 658 here moderate, because this situation corresponds to the one for which the normality of the
 659 residuals is not met (Fig. 14).



660

661 *Figure 17. Evolution of the Beta distributions fitted to the P values given: (A) increasing values of the*
 662 *mean value of the fragility curve α_{str} ; (B) increasing values of α_{edg1} ; (C) increasing values of*
 663 *α_{edg2} .*

664

665 5 Discussion

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

676 (2019) with respect to sensitivity analysis practices). Multi-way SA methods have been
677 proposed (Leonelli et al., 2017; Gómez-Villegas et al., 2013), but they can rapidly become
678 intractable.

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

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

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

696

697 *[Table 5 about here]*

698

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

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

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

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

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

731

732 **6 Summary and future research directions**

733 The rapidly increasing advances of BBN for modelling of expert systems calls for the
734 developments of robust methods for their validation and verification (e.g., Marcot and Penman,
735 2019). One major pillar to fulfil this purpose is sensitivity analysis. The proposed BBR
736 approach broadens the scope of existing BBN sensitivity analysis methods by providing a larger
737 vision (global) on the CPM influence. The approach has the advantage of being generic (it can

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

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

761

762 **Acknowledgements**

763 This study has been carried out within the NARSIS project, which has received funding from
764 the European Union's H2020-Euratom Programme under grant agreement N° 755439. The
765 analysis was performed using R packages *bnlearn* (available at: [https://cran.r-](https://cran.r-project.org/web/packages/bnlearn/index.html)
766 [project.org/web/packages/bnlearn/index.html](https://cran.r-project.org/web/packages/bnlearn/index.html)) for BBN building and queries, *bnviewer*
767 (<https://cran.r-project.org/web/packages/bnviewer/index.html>) for network visualisation, and
768 *gamboostLSS* (available at: <https://cran.r-project.org/web/packages/gamboostLSS/index.html>)
769 for boosting of distributional regression model with BE() function of *gamlss.dist* package
770 (available at: <https://cran.r-project.org/web/packages/gamlss.dist/index.html>).

771

772 **References**

- 773 Bayer, F. M., & Cribari-Neto, F. (2017). Model selection criteria in beta regression
774 with varying dispersion. *Communications in Statistics-Simulation and Computation*, 46(1),
775 729-746.
- 776 Beuzen, T., Marshall, L., & Splinter, K. D. (2018). A comparison of methods for
777 discretizing continuous variables in Bayesian Networks. *Environmental Modelling &*
778 *Software*, 108, 61-66.
- 779 Bühlmann, P., & Hothorn, T. (2007). Boosting algorithms: Regularization, prediction
780 and model fitting (with discussion). *Statistical Science*, 22, 477–522.
- 781 Castillo, E., Gutiérrez, J. M., & Hadi, A. S. (1997). Sensitivity analysis in discrete
782 Bayesian networks. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems*
783 *and Humans*, 27(4), 412-423.
- 784 Castillo, E., & Kjærulff, U. (2003). Sensitivity analysis in Gaussian Bayesian
785 networks using a symbolic-numerical technique *Reliab. Eng. Syst. Saf.*, 79, 139-148.
- 786 Chan, H., & Darwiche, A. (2002). When do numbers really matter? *J. Artif. Intell.*
787 *Res.*, 17, 265-287.
- 788 Chan, H., & Darwiche, A. (2005). A distance measure for bounding probabilistic
789 belief change *Internat. J. Approx. Reason.*, 38, 149-174.
- 790 Chen, S. H., & Pollino, C. A. (2012). Good practice in Bayesian network modelling.
791 *Environmental modelling & software*, 37, 134–45.
- 792 Cooper, G.F., (1984). NESTOR: a Computer-based Medical Diagnostic Aid that
793 Integrates Causal and Probabilistic Knowledge. Report HPP-84-48, Stanford University.
- 794 Coupé, V. M. H., & van der Gaag, L. C. (2002). Properties of sensitivity analysis of
795 Bayesian belief networks *Ann. Math. Artif. Intell.*, 36, 323-356.
- 796 Cribari-Neto, F., & Zeileis, A. (2010). Beta regression in R. *Journal of Statistical*
797 *Software*, 34.
- 798 Doksum, K. A. and G. L. Sievers, (1976). Plotting with confidence: graphical
799 comparisons of two populations. *Biometrika*, 63(3), 421–434.
- 800 Druzdzel, M.J., & van der Gaag, L. (2000). Building Probabilistic Networks: “where
801 do the numbers come from?” *IEEE Trans. Knowl. Data Eng.*, 12(4), 481–486.
- 802 Eilers, P. H. C., & Marx, B. D. (1996). Flexible smoothing with B-splines and
803 penalties. *Statistical Science*, 11, 89–121.

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

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

808 van Der Gaag, L. C., Kuijper, R., Van Geffen, Y. M. & Vermeulen, J. L. (2013).
809 Towards uncertainty analysis of Bayesian Networks. In 25th Benelux Conference on
810 Artificial Intelligence, Delft, The Netherlands.

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

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

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

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

820 Hofner, B., Mayr, A., & Schmid, M. (2016). gamboostLSS: An R package for model
821 building and variable selection in the GAMLSS framework. *Journal of Statistical Software*,
822 74(1), doi:10.18637/jss.v074.i01.

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

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

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

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

832 Kwag, S., & Gupta, A. (2017). Probabilistic risk assessment framework for structural
833 systems under multiple hazards using Bayesian statistics. *Nuclear Engineering and Design*,
834 315, 20-34.

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

837 Malagrino, L. S., Roman, N. T., & Monteiro, A. M. (2018). Forecasting stock market
838 index daily direction: a Bayesian network approach. *Expert Systems with Applications*, 105,
839 11-22.

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

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

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

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

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

851 Gómez-Villegas, M.A., Main, P., & Susi, R. (2007). Sensitivity analysis in Gaussian
852 Bayesian networks using a divergence measure. *Comm. Statist. Theory Methods*, 36(3), 523-
853 539.

854 Gómez-Villegas, M. A., Main, P., & Susi, R. (2013). The effect of block parameter
855 perturbations in Gaussian Bayesian networks: sensitivity and robustness. *Inform. Sci.*, 222,
856 439-458.

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

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

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

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

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

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

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

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

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

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

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

882 Schmid, M., Wickler, F., Maloney, K. O., Mitchell, R., Fenske, N., & Mayr, A.
883 (2013). Boosted beta regression. *PloS one*, 8(4), e61623.

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

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

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

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

892 Renooij, S. (2014). Co-variation for sensitivity analysis in Bayesian networks:
893 Properties, consequences and alternatives. *International Journal of Approximate Reasoning*,
894 55(4), 1022-1042.

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

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

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

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

906

907 **Tables**

908 Table 1. CPT entries for the brain tumor BBN. For instance $P(c|b0,isc1)$ corresponds to the
 909 probability of node C being in state 1 conditioned by the fact that node B is in state 0 and ISC
 910 is in state 1. The meaning of the other probabilities should be understood following this
 911 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	$P(sh b1)=0.80$ $P(sh b0)=0.60$

912

913

914

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	$m_{0_6}=0$	$Z_{6_{23}}=0.3$ $Z_{6_8}=0.7$	$s_{0_6}=1e-4$
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	$\alpha_{STR}, \beta_{STR}$	-6.48, 2.32
Mean and Standard deviation of the log-normal fragility curve for EDG1	$\alpha_{EDG1}, \beta_{EDG1}$	-12.250, 6.22
Mean and Standard deviation of the log-normal fragility curve for EDG2	$\alpha_{EDG2}, \beta_{EDG2}$	-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	<p>It is simple to implement;</p> <p>It requires a low computational effort;</p> <p>The graphical representation is straightforward to interpret.</p>	<p>It is local: it focuses on the influence of one (or multiple) CPT parameters while the other ones are kept constant;</p> <p>It is restricted to the analysis of discrete BBNs;</p> <p>Multi-way SA can rapidly become intractable.</p>
BBR	<p>It provides insight in the global influence;</p> <p>It is simple to implement using random sampling, which is a generic procedure applicable to any types of BBNs;</p> <p>The results are intuitive to interpret based on a graphical presentation;</p> <p>It provides multilevel information on sensitivity;</p> <p>The combination of gradient boosting and stability selection increases the robustness to the number of parameters.</p>	<p>The number of calls to the BBN inference engine can be large (>1,000);</p> <p>The adequacy of the Beta model should be carefully checked;</p> <p>The demonstration of the applicability to very large-scale BBNs with hundreds of nodes and thousands of CPM parameters remains to be done.</p>

927

928

929 **Appendix A: Bayesian Belief Network Analysis**

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

$$934 P(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i | X_1, \dots, X_{i-1}) \quad (\text{Eq. A1})$$

936 This equation simplifies under the conditional independence assumption as:

$$938 P(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i | \text{Pa}(X_i)) \quad (\text{Eq. A2})$$

939 where $\text{Pa}(X_i)$ corresponds to the parent nodes of X_i . For discrete nodes, the value of
940 $P(X_i | \text{Pa}(X_i))$ is the entry of the Conditional Probability Table. For continuous nodes,
941 $P(X_i | \text{Pa}(X_i))$ can be modelled by a continuous probability distribution whose parameters
942 depend on the values of the parent nodes.

943
944 Conditional queries aim at evaluating the conditional probability of some event, e.g. node X_j
945 takes up the value x when new information/observations become available, i.e. given new
946 “evidence” (denoted e), namely $P(X_j=x | e)$. This procedure, termed as query, relies on inference
947 techniques. Exact inference in a BBN is possible, but is generally not possible in large networks.
948 For networks, which contain a large number of nodes, arcs, or have nodes comprised of
949 variables with large numbers of levels, exact inference becomes too computationally intensive.
950 Among the possible approximate inference algorithms, the present study relies on the logic
951 sampling method (Koller and Friedman 2009), which works from the root nodes down to the
952 leaf nodes. First, a random draw from each of the leaf nodes is taken with probabilities equal to
953 the respective levels. The distributions for the next level of nodes can then be marginalized
954 based on the draws obtained from their parents. From these marginal distributions a random
955 draw is again taken respective of the categorical probabilities. This process is repeated until a
956 draw has been taken from every node in the network. This completed case represents a single
957 sample. After many samples have been taken, the samples are subsetted to the cases which
958 match the evidence of interest. Estimated probabilities can then be obtained for any node of
959 interest from this subset of the samples.

960

961

962 **Appendix B: Gradient boosting within GAMLSS framework**

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

965 **B1 Gradient boosting**

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

969 Consider the training dataset $\mathbf{D}=\{\mathbf{C}^{(j)}, P^{(j)}\}_{j=1,\dots,n}$ where \mathbf{C} is the vector of p predictor variables
970 $c_{i=1,\dots,p}$ and P is the variable of interest, whose expected value (possibly its transformed value)
971 is modelled by an additive model as follows:

972

$$973 \quad \eta(\mathbf{C}) = \beta_0 + \sum_{j=1}^J f_j(c_j|\beta_j) \quad (\text{Eq. B1})$$

974

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

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

980

$$981 \quad R = \sum_{j=1}^n \rho(P^{(j)}, \eta(\mathbf{C}^{(j)})) \quad (\text{Eq. B2})$$

982

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

987 Among the different boosting algorithms, we focus on the component-wise gradient boosting
988 approach of Bühlmann and Hothorn (2007). Instead of focusing on the true outcomes $\mathbf{P}=(P^{(1)},$
989 $P^{(2)}, \dots, P^{(n)})$, this procedure aims at fitting simple regression-type base learners $h(\cdot)$ one by one
990 to the negative gradient vector of loss $\mathbf{u}=(u^{(1)}, u^{(2)}, \dots, u^{(n)})$. The objective is to approximate the
991 j^{th} effect $f_j(c_j|\beta_j) = \sum_m h_j(\cdot)$ at iteration m of the algorithm. Formally, \mathbf{u} is evaluated

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

994

$$995 \quad \mathbf{u} = \left(-\frac{\partial}{\partial \eta} \rho(P, \eta) \Big|_{\eta=\hat{\eta}^{[m-1]}(\mathbf{C}^{(j)})} \Big|_{P=P^{(j)}} \right)_{j=1, \dots, n} \quad (\text{Eq. B3})$$

996

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

1000

$$1001 \quad j^* = \underset{j \in \{1, \dots, J\}}{\operatorname{argmin}} \sum_{i=1}^n (u^{(i)} - h_j(\mathbf{C}^{(i)}))^2 \quad (\text{Eq. B4})$$

1002

1003 The selected base-learner is used to update the current predictor model as follows:

1004

$$1005 \quad \hat{\eta}^{[m]} = \hat{\eta}^{[m-1]} + s \cdot h_{j^*}(\mathbf{C}) \quad (\text{Eq. B5})$$

1006 where s is a step length (with typical value of 0.1).

1007

1008 The main tuning parameter is the number of iterations m , which directly determines the
 1009 prediction performance. If m is too large, rich models with large number of predictors and rough
 1010 functional terms will be constructed (hence leading to overfitting), which might hamper the
 1011 interpretability of the resulting model. If m is too low, sparse models with smooth functional
 1012 terms will be constructed but with the danger of missing some important predictor variables. In
 1013 practice, the selection of m can be carried out using cross-validation procedures in order to
 1014 optimize the predictive risk on observations left out from the fitting process i.e. the “out-of-
 1015 bag” risk which corresponds to the negative log-likelihood for the considered probabilistic
 1016 distribution calculated for the “out-of-bag” samples.

1017

1018 **B2 Boosted GAMLSS**

1019 The afore-described algorithm can be applied to the GAMLSS framework by cycling through
 1020 the distribution parameters θ in the fitting process (Thomas et al., 2018: Algorithm 1). In each
 1021 iteration, the best fitting base-learner is evaluated for each distribution parameter, while the

1022 other ones remain fixed. For a probability distribution with two parameters (like the Beta
1023 distribution), the update in the boosting algorithm at iteration m holds as follows:

1024

$$1025 \quad \frac{\partial}{\partial \eta_{\theta_1}} \rho(P, \theta_1^{[m]}, \theta_2^{[m]}) \xrightarrow{\text{update}} \eta_{\theta_1}^{[m+1]} \quad (\text{Eq. B6})$$

$$1026 \quad \frac{\partial}{\partial \eta_{\theta_2}} \rho(P, \theta_1^{[m+1]}, \theta_2^{[m]}) \xrightarrow{\text{update}} \eta_{\theta_2}^{[m+1]} \quad (\text{Eq. B7})$$

1027

1028 In the original cyclic algorithm, (see algorithm 1 described by Thomas et al. (2018)), separate
1029 stopping values have to be specified for each parameter, which motivated the development of
1030 a noncyclic algorithm (algorithm 2 by Thomas et al. (2018)), which avoids optimizing two
1031 different stopping iterations and reducing the optimisation problem from a multi-dimensional
1032 to a one-dimensional problem.

1033

1034

1035 **Appendix C: Stability Selection Analysis**

1036 This appendix describes the procedure for stability selection (Meinhausen and Bühlmann,
1037 2010). Consider p predictor variables $c_{j=1,\dots,p}$ and the predictand P . Based on n observations, the
1038 stability selection with boosting proceeds as follows:

1039 Step 1. Select a random subset of size $\lfloor n/2 \rfloor$ of the data (where $\lfloor n/2 \rfloor$ corresponds to the largest
1040 integer $\leq n/2$);

1041 Step 2. Fit a boosting model and continue to increase the number of boosting iteration until q
1042 base-learners are selected. $\hat{S}_{\lfloor n/2 \rfloor}$ corresponds to the set of selected variables;

1043 Step 3. Repeat steps 1 and 2 for $b=1,\dots,B$;

1044 Step 4. Compute the selection probabilities per base learner as follows:

$$1045 \quad P_j = \frac{1}{B} \sum_{b=1}^B \mathbf{I}_{\{j \in \hat{S}_{\lfloor n/2 \rfloor, b}\}} \quad (\text{Eq. C1})$$

1046

1047 where $\mathbf{I}(A)$ is the indicator function, which reaches 1 if A is true, 0 otherwise.

1048 Step 5. Select all base-learners associated to a selection probability of at least t . The set of stable
1049 selected variables is thus $\hat{S} = \{j: P_j \geq t\}$.

1050 Meinhausen and Bühlmann (2010) studied the error of selecting false positive variables (i.e.
1051 noise variables) and showed that the selection procedure controls the per-family error rate
1052 ($PFER$) and an upper bound is provided as follows:

1053

$$1054 \quad PFER \leq \frac{q^2}{(2t-1)p} \quad (\text{Eq. C2})$$

1055

1056 where q is the number of selected variables per boosting iteration, p is the number of (possible)
1057 predictors and t is the selection threshold. In practice, at least two of these parameters have to
1058 be specified to run the procedure. In the application case of Sect. 4.1, we specified the upper
1059 bound of $PFER$ (set up at 1) and q at 10. An extensive investigation of the applicability of this
1060 procedure for distributional regression within the boosted GAMLSS setting has been carried
1061 out by Thomas et al. (2018).

1062 **Appendix D: Residuals for Beta regression**

1063 The following appendix describes the main formulations of the residuals for Beta regression
1064 based on Pereira (2019) and references therein.

1065 Let us consider the raw response residuals for the i^{th} observation P_i defined as $r_i = P_i - \hat{P}_i$
1066 where \hat{P} is the mean μ predicted by BBR model. For sake of presentation, let us consider the
1067 Beta parameter $\varphi = \frac{1}{\sigma^2} - 1$ where σ is defined in Sect. 2.1.

1068 The deviance residual is defined as:

$$1069 \quad r_i^{\text{deviance}} = r_i \cdot \sqrt{2|L(P, \varphi) - L(\mu, \varphi)|} \quad (\text{Eq. D1})$$

1070 where the bivariate function $L(.,.)$ holds as follows:

$$1071 \quad L(\mu, \varphi) = l\Gamma(\varphi) - l\Gamma(\mu \cdot \varphi) - l\Gamma((1 - \mu) \cdot \varphi) + (\mu \cdot \varphi - 1) \cdot \log(P) + ((1 - \mu) \cdot \varphi - 1) \cdot \log(1 - P) \quad (\text{Eq. D2})$$

1072 where $l\Gamma(.)$ is natural logarithm of the absolute value of the gamma function.

1073

1074 The standardized weighted residual 1 originally introduced by Espinheira et al. (2008) is
1075 defined as follows:

$$1076 \quad r_i^{\text{weighted}} = \frac{P^* - \mu^*}{\sqrt{v}} \quad (\text{Eq. D3})$$

1077 where P^* is the quantile for the logistic distribution with location set up at zero and scale at one,
1078 and

$$1079 \quad \mu^* = \Gamma''(\mu \cdot \varphi) - \Gamma''((1 - \mu) \cdot \varphi) \quad (\text{Eq. D4})$$

$$1080 \quad v = \Gamma'''(\mu \cdot \varphi) - \Gamma'''((1 - \mu) \cdot \varphi) \quad (\text{Eq. D5})$$

1081 where Γ'' and Γ''' are respectively the second and third derivative of the Γ function.

1082

1083 The quantile residual is defined as:

$$1084 \quad r_i^{\text{Quantile}} = \Phi^{-1}(F(P, \mu, \varphi)) \quad (\text{Eq. D6})$$

1085 where Φ^{-1} is the inverse of the cumulative probability function of the standard normal
1086 distribution and F is the cumulative probability function of the Beta law.

1087