Natural Hazards
and Earth System
Sciences

Open Access

Bayesian network learning for natural hazard analyses

K. Vogel[1,*], **C. Riggelsen**[2], **O. Korup**[1], **and F. Scherbaum**[1]

[1]Institute of Earth and Environmental Sciences, University of Potsdam, Germany
[2]Pivotal Software Inc., Palo Alto, USA
Invited contribution by K. Vogel, recipient of the Outstanding Student Poster (OSP) Award 2012.

Correspondence to: K. Vogel (kvog@geo.uni-potsdam.de)

Abstract. Modern natural hazards research requires dealing with several uncertainties that arise from limited process knowledge, measurement errors, censored and incomplete observations, and the intrinsic randomness of the governing processes. Nevertheless, deterministic analyses are still widely used in quantitative hazard assessments despite the pitfall of misestimating the hazard and any ensuing risks.

In this paper we show that Bayesian networks offer a flexible framework for capturing and expressing a broad range of uncertainties encountered in natural hazard assessments. Although Bayesian networks are well studied in theory, their application to real-world data is far from straightforward, and requires specific tailoring and adaptation of existing algorithms. We offer suggestions as how to tackle frequently arising problems in this context and mainly concentrate on the handling of continuous variables, incomplete data sets, and the interaction of both. By way of three case studies from earthquake, flood, and landslide research, we demonstrate the method of data-driven Bayesian network learning, and showcase the flexibility, applicability, and benefits of this approach.

Our results offer fresh and partly counterintuitive insights into well-studied multivariate problems of earthquake-induced ground motion prediction, accurate flood damage quantification, and spatially explicit landslide prediction at the regional scale. In particular, we highlight how Bayesian networks help to express information flow and independence assumptions between candidate predictors. Such knowledge is pivotal in providing scientists and decision makers with well-informed strategies for selecting adequate predictor variables for quantitative natural hazard assessments.

1 Introduction

Natural hazards such as earthquakes, tsunamis, floods, landslides, or volcanic eruptions have a wide range of differing causes, triggers, and consequences. Yet the art of predicting such hazards essentially addresses very similar issues in terms of model design: the underlying physical processes are often complex, while the number of influencing factors is large. The single and joint effects of the driving forces are not always fully understood, which introduces a potentially large degree of uncertainty into any quantitative analysis. Additionally, observations that form the basis for any inference are often sparse, inaccurate and incomplete, adding yet another layer of uncertainty. For example, Merz et al. (2013) point out the various sources of uncertainty (scarce data, poor understanding of the damaging process, etc.) in the context of flood damage assessments, while Berkes (2007) calls attention to the overall complexity of human–environment systems, as well as the importance of understanding underlying uncertainties to improve resilience. Similarly, Bommer and Scherbaum (2005) discuss the importance of capturing uncertainties in seismic hazard analyses to balance between investments in provisions of seismic resistance and possible consequences in the case of insufficient resistance.

Nevertheless, deterministic approaches are still widely used in natural hazards assessments. Such approaches rarely provide information on the uncertainty related to parameter estimates beyond the use of statistical measures of dispersion such as standard deviations or standard errors about empirical means. However, uncertainty is a carrier of information to the same extent as a point estimate, and ignoring it or dismissing it as simply an error may entail grave consequences. Ignoring uncertainties in quantitative hazard

appraisals may have disastrous effects, since it often leads to over- or underestimates of certain event magnitudes. Yet deterministic approaches persist as the state of the art in many applications. For example, tsunami early warning systems evaluate pre-calculated synthetic databases and pick out the scenario that appears closest to a given situation in order to estimate its hazard (Blaser et al., 2011). Recently developed models for flood damage assessments use classification approaches, where the event under consideration is assigned to its corresponding class, and the caused damage is estimated by taking the mean damage of all observed events belonging to the same class (Elmer et al., 2010). In seismic hazard analysis the usage of regression-based ground motion models is common practice, restricting the model to the chosen functional form, which is defined based on physical constrains (Kuehn et al., 2009).

In this paper we consider Bayesian networks (BNs), which we argue are an intuitive, consistent, and rigorous way of quantifying uncertainties. Straub (2005) underlines the large potential of BNs for natural hazard assessments, heralding not only the ability of BNs to model various interdependences but also their intuitive format: the representation of (in)dependences between the involved variables in a graphical network enables improved understandings and direct insights into the relationships and workings of a natural hazard system. The conditional relationships between dependent variables are described by probabilities, from which not only the joint distribution of all variables but any conditional probability distribution of interest can be derived. BNs thus endorse quantitative analyses of specific hazard scenarios or process-response chains.

In recent years, BNs have been used in avalanche risk assessment (e.g., Grêt-Regamey and Straub, 2006), tsunami early warning (e.g., Blaser et al., 2009, 2011), earthquake risk management (e.g., Bayraktarli and Faber, 2011), probabilistic seismic hazard analysis (e.g., Kuehn et al., 2011), and earthquake-induced landslide susceptibility (e.g., Song et al., 2012). Aguilera et al. (2011) give an overview of applications of BNs in the environmental sciences between 1990 and 2010, and conclude that the potential of BNs remains underexploited in this field. This is partly because, even though BNs are well studied in theory, their application to real-world data is not straightforward. Handling of continuous variables and incomplete observations remains the key problem. This paper aims to overcome these challenges. Our objective is to briefly review the technique of learning BNs from data, and to suggest possible solutions to implementation problems that derive from the uncertainties mentioned above. We use three examples of natural hazard assessments to discuss the demands of analyzing real-world data, and highlight the benefits of applying BNs in this regard.

In our first example (Sect. 3), we develop a seismic ground motion model based on a synthetic data set, which serves to showcase some typical BN properties. In this context we demonstrate a method to deal with continuous variables with-
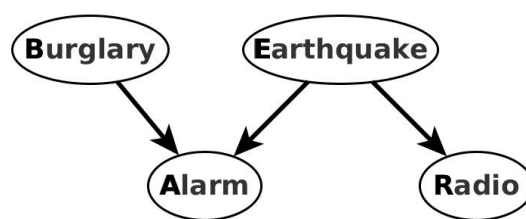
Figure 1. The figure shows the BN for the burglary example. The graph structure illustrates the dependence relations of the involved variables: the alarm can be triggered by a burglary or earthquake. An earthquake might be reported in the radio newscast. The joint distribution of all variables can be decomposed into the product of its conditionals accordingly: $P(B, E, A, R) = P(B) P(E) P(A|B, E) P(R|E)$.

out any prior assumptions on their distributional family. In Sect. 4 we use data that were collected after the 2002 and 2005/2006 floods in the Elbe and Danube catchments, Germany, to learn a BN for flood damage assessments. This example is emblematic of situations where data are incomplete, and requires a treatment of missing observations, which can be challenging in combination with continuous variables. Our final example in Sect. 5 deals with a regional landslide susceptibility model for Japan, where we investigate how the same set of potential predictors of slope stability may produce nearly equally well performing, though structurally different, BNs that reveal important and often overlooked variable interactions in landslide studies. This application further illustrates the model uncertainty related to BN learning.

2 Bayesian networks (BNs)

The probabilistic framework of BNs relies on the theorem formulated by Reverend Thomas Bayes (1702–1761), and expresses how to update probabilities in light of new evidence (McGrayne, 2011). By combining probability theory with graph theory, BNs depict probabilistic dependence relations in a graph: the nodes of the graph represent the considered random variables, while (missing) edges between the nodes illustrate the conditional (in)dependences between the variables. Textbooks often refer to the burglary alarm scenario for a simple illustration of BNs (Pearl, 1998). In this example, the alarm of your home may not only be triggered by burglary but also by earthquakes. Moreover, earthquakes have a chance to be reported in the news. Figure 1 shows the dependence relations of these variables as captured by a BN. Now, imagine you get a call from your neighbor notifying you that the alarm went off. Supposing the alarm was triggered by burglary, you drive home. On your way home you hear the radio reporting a nearby earthquake. Even though burglaries and earthquakes may be assumed to occur independently, the radio announcement changes your belief in the burglary, as the earthquake "explains away" the alarm. BNs

Table 1. Conditional probabilities in the burglary example, giving the conditional probabilities for *earthquake* (*e*), *burglary* (*b*), *alarm* (*a*), and *earthquake reported* (*r*). The parameters that define the conditional distributions correspond for discrete variables to the conditional (point) probabilities. Note that the conditional probability values for *no earthquake* ($\overline{e}$), *no burglary* ($\overline{b}$), etc. can be derived from the fact that the conditionals sum up to 1.

$\theta_e = p(e)$	$= 0.001$	$\theta_{a|e,b} = p(a|e,b)$	$= 0.98$
$\theta_b = p(b)$	$= 0.01$	$\theta_{a|e,\overline{b}} = p(a|e,\overline{b})$	$= 0.95$
$\theta_{r|e} = p(r|e)$	$= 0.95$	$\theta_{a|\overline{e},b} = p(a|\overline{e},b)$	$= 0.95$
$\theta_{r|\overline{e}} = p(r|\overline{e})$	$= 0.001$	$\theta_{a|\overline{e},\overline{b}} = p(a|\overline{e},\overline{b})$	$= 0.03$

offer a mathematically consistent framework to conduct and specify reasonings of such kind. A detailed introduction to BNs is provided in Koller and Friedman (2009) and Jensen and Nielsen (2001), while Fenton and Neil (2012) offers easy and intuitive access. In this paper we restrict ourselves to several key aspects of the BN formalism.

2.1 Properties and benefits

Applying BNs to natural hazard assessments, we define the specific variables of the hazard domain to be the nodes in a BN. In the following we denote this set of random variables as $X = \{X_1, \ldots, X_k\}$. The dependence relations between the variables are encoded in the graph structure, generating a directed acyclic graph (DAG). The directions of the edges define the flow of information, but do not necessarily indicate causality. As we shall see in subsection "Learned ground motion model" of Sect. 3.2, it may prove beneficial to direct edges counterintuitively in order to fulfill regularization constraints. The set of nodes from which edges are directed to a specific node, X_i, is called the parent set, $X_{\mathrm{Pa}(i)}$, of X_i (see Fig. 2). Table 2 summarizes the notations used in this paper.

Apart from the graph structure, a BN is defined by conditional probabilities that specify the dependence relations encoded in the graph structure. The conditional probability distribution for each variable, X_i, is given conditioned on its parent set: $p\left(X_i|X_{\mathrm{Pa}(i)}\right)$. For simplification we restrict ourselves here to discrete variables for which θ is the set of conditional (point) probabilities for each combination of states for X_i and $X_{\mathrm{Pa}(i)}$: $\theta = \{\theta_{x_i|x_{\mathrm{Pa}(i)}} = p(x_i|x_{\mathrm{Pa}(i)})\}$. The conditional probabilities for the burglary BN example are given in Table 1. For continuous variables, the design of the parameters depends on the family of distributions of the particular densities $p(\cdot|\cdot)$.

Given the BN structure (DAG) and parameters (θ), it follows from the axioms of probability theory that the joint distribution of all variables can be factorized into a product of conditional distributions:

$$P(X|\mathrm{DAG}, \theta) = \prod_{i=1}^{k} p\left(X_i|X_{\mathrm{Pa}(i)}\right). \tag{1}$$
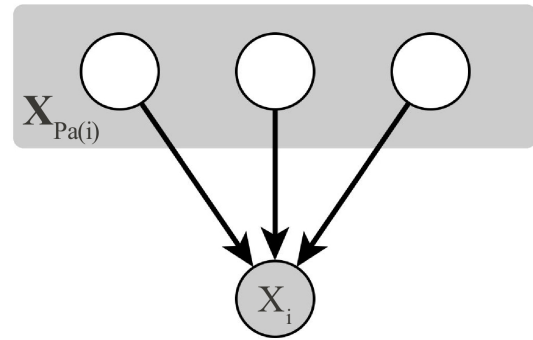
Figure 2. Illustration of a parent set in a BN. $X_{\mathrm{Pa}(i)}$ is the parent set of X_i.

Further, applying Bayes theorem, $P(A|B) = \frac{P(A,B)}{P(B)} = \frac{P(B|A)\,P(A)}{P(B)}$, each conditional probability of interest can be derived. In this way a BN is characterized by many attractive properties that we may profit from in a natural hazard setting, including the following properties:

- *Property 1 – graphical representation*: the interactions of the variables of the entire "system" are encoded in the DAG. The BN structure thus provides information about the underlying processes and the way various variables communicate and share "information" as it is propagated through the network.

- *Property 2 – use prior knowledge*: the intuitive interpretation of a BN makes it possible to define the BN based on prior knowledge; alternatively it may be learned from data, or even a combination of the two (cast as Bayesian statistical problem) by posing a prior BN and updating it based on observations (see below for details).

- *Property 3 – identify relevant variables*: by learning the BN from data we may identify the variables that are (according to the data) relevant; "islands" or isolated single unconnected nodes indicate potentially irrelevant variables.

- *Property 4 – capture uncertainty*: uncertainty can easily be propagated between any nodes in the BN; we effectively compute or estimate probability distributions rather than single-point estimates.

- *Property 5 – allow for inference*: instead of explicitly modeling the conditional distribution of a predefined target variable, the BN captures the joint distribution of all variables. Via inference, we can express any given or all conditional distribution(s) of interest, and reason in any direction (including forensic and inverse reasoning): for example, for a given observed damage we may infer the likely intensity of the causing event. A detailed example for reasoning is given in Sect. 4.3.

Table 2. Summary of notations used in this paper.

Notation	Meaning
X_i	a specific variable
x_i	a realization of X_i
$\boldsymbol{X} = \{X_1, \ldots, X_k\}$	set of the considered variables
$\boldsymbol{X}_{\mathrm{Pa}(i)}$	parent set of X_i
$\boldsymbol{x}_{\mathrm{Pa}(i)}$	a realization of the parent set
$\boldsymbol{X}_{-Y}$	all variables but Y
DAG	directed acyclic graph (graph structure)
$p(X_i \mid \boldsymbol{X}_{\mathrm{Pa}(i)})$	conditional probability of a variable conditioned on its parent set
$\theta_{x_i \mid \boldsymbol{x}_{\mathrm{Pa}(i)}}$	parameter that defines the probability for x_i given $\boldsymbol{x}_{\mathrm{Pa}(i)}$
$\boldsymbol{\theta} = \left\{ \theta_{x_i \mid \boldsymbol{x}_{\mathrm{Pa}(i)}} \right\}$	set of model parameters that defines the conditional distributions
Θ	random variable for the set of model parameters
BN: (DAG, $\boldsymbol{\theta}$)	Bayesian network, defined by the pair of structure and parameters
$\boldsymbol{d}$	discrete/discretized data set that is used for BN learning
$\boldsymbol{d}^{\mathrm{c}}$	(partly) continuous data set that is used for BN learning
Λ	discretization that bins the original data $\boldsymbol{d}^{\mathrm{c}}$ into $\boldsymbol{d}$
$\boldsymbol{X}_{\mathrm{MB}(i)}$	set of variables that form the Markov blanket of X_i (Sect. 4.2)
Ch(i)	variable indices of the children of X_i (Sect. 4.2)

Note that inference in BNs is closed under restriction, marginalization, and combination, allowing for fast (close to immediate) and exact inference.

- *Property 6 – use incomplete observations*: during predictive inference (i.e., computing a conditional distribution), incomplete observations of data are not a problem for BNs. By virtue of the probability axioms, it merely impacts the overall uncertainty involved.

In the following we will refer to these properties 1–6 in order to clarify what is meant. For "real-life" modeling problems, including those encountered in natural hazard analysis, adhering strictly to the BN formalism is often a challenging task. Hence, the properties listed above may seem unduly theoretical. Yet many typical natural hazard problems can be formulated around BNs by taking advantage of these properties. We take a data-driven stance and thus aim to learn BNs from collected observations.

2.2 Learning Bayesian networks

Data-based BN learning can be seen as an exercise in finding a BN which, according to the decomposition in Eq. (1), could have been "responsible for generating the data". For this we traverse the space of BNs (Castelo and Kocka, 2003) looking for a candidate maximizing a fitness score that reflects the "usefulness" of the BN. This should however be done with careful consideration to the issues always arising in the context of model selection, i.e., over-fitting, generalization, etc. Several suggestions for BN fitness scoring are derived from different theoretical principles and ideas (Bouckaert, 1995). Most of them are based on the maximum likelihood estimation for different DAG structures according to Eq. (1).

In this paper we opt for a Bayesian approach to learn BNs (note that BNs are not necessarily to be interpreted from a Bayesian statistical perspective). Searching for the most probable BN, (DAG, $\boldsymbol{\theta}$), given the observed data, $\boldsymbol{d}$, we aim to maximize the BN MAP (Bayesian network maximum a posteriori) score suggested by Riggelsen (2008):

$$\underbrace{P(\mathrm{DAG}, \Theta \mid \boldsymbol{d})}_{\text{posterior}} \propto \underbrace{P(\boldsymbol{d} \mid \mathrm{DAG}, \Theta)}_{\text{likelihood}} \; \underbrace{P(\Theta, \mathrm{DAG})}_{\text{prior}}. \qquad (2)$$

The likelihood term decomposes according to Eq. (1). The prior encodes our prior belief in certain BN structures and parameters. This allows us to assign domain specific prior preferences to specific BNs before seeing the data (Property 2) and thus to compensate for sparse data, artifacts, bias, etc. In the following applications we use a non-informative prior, which nevertheless fulfills a significant function. Acting as a penalty term, the prior regularizes the DAG complexity and thus avoids over-fitting. Detailed descriptions for prior and likelihood term are given in Appendix A1 and Riggelsen (2008).

The following section illustrates the BN formalism "in action" and will also underscore some theoretical and practical problems along with potential solutions in the context of BN learning. We will learn a ground motion model, which is used in probabilistic seismic hazard analysis, as a BN; the data are synthetically generated. Subsequently, we consider two other natural hazard assessments where we learn BNs from real-world data.

Table 3. Variables used in the ground motion model and the corresponding distributions used for the generation of the synthetic data set which is used for BN learning.

X_i	Description	Distribution$_{[range]}$
	Predictors	
M	Moment magnitude of the earthquake	$\mathcal{U}_{[5,7.5]}$
R	Source-to-site distance	$\text{Exp}_{[1\,km,\,200\,km]}$
SD	Stress released during the earthquake	$\text{Exp}_{[0\,bar,\,500\,bar]}$
Q_0	Attenuation of seismic wave amplitudes in deep layers	$\text{Exp}_{[0\,s^{-1},\,5000\,s^{-1}]}$
κ_0	Attenuation of seismic wave amplitudes near the surface	$\text{Exp}_{[0\,s,\,0.1\,s]}$
V_S30	Average shear-wave velocity in the upper 30 m	$\mathcal{U}_{[600\,m\,s^{-1},\,2800\,m\,s^{-1}]}$
	Ground motion parameter	
PGA	Horizontal peak ground acceleration	According to the stochastic model (Boore, 2003)

3 Seismic hazard analysis: ground motion models

When it comes to decision making on the design of high-risk facilities, the hazard arising from earthquakes is an important aspect. In probabilistic seismic hazard analysis (PSHA) we calculate the probability of exceeding a specified ground motion for a given site and time interval. One of the most critical elements in PSHA, often carrying the largest amount of uncertainty, is the ground motion model. It describes the conditional probability of a ground motion parameter, Y, such as *(horizontal) peak ground acceleration*, given earthquake- and site-related predictor variables, X_{-Y}. Ground motion models are usually regression functions, where the functional form is derived from expert knowledge and the ground motion parameter is assumed to be lognormally distributed: $\ln Y = f(X_{-Y}) + \epsilon$, with $\epsilon \sim \mathcal{N}(0, \sigma^2)$. The definition of the functional form of $f(\cdot)$ is guided by physical model assumptions about the single and joint effects of the different parameters, but also contains some ad hoc elements (Kuehn et al., 2011). Using the Bayesian network approach there is no prior knowledge required per se, but if present it can be accounted for by encoding it in the prior term of Eq. (2). If no reliable prior knowledge is available, we work with a non-informative prior, and the learned graph structure provides insight into the dependence structure of the variables and helps in gaining a better understanding of the underlying mechanism (Property 1). Modeling the joint distribution of all variables, $X = \{X_{-Y}, Y\}$, the BN implicitly provides the conditional distribution $P(Y|X_{-Y}, \text{DAG}, \Theta)$, which gives the probability of the ground motion parameter for specific event situations needed for the PSHA (Property 5).

3.1 The data

The event situation is described by the predictor variables $X_{-Y} = \{M, R, \text{SD}, Q_0, \kappa_0, V_S30\}$, which are explained in Table 3. We generate a synthetic data set consisting of
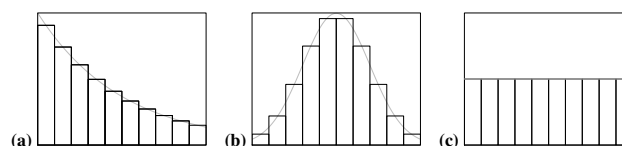

Figure 3. When working with continuous variables, we have to make assumptions about the functional form of the probability distributions (gray), e.g., **(a)** exponential, **(b)** normal, and **(c)** uniform. Thus we restrict the distributions to certain shapes that may not match reality. In contrast, using a discrete multinomial distribution (black), each continuous distribution can be approximated and we avoid prior restrictions on the shape. Rather the shape is learned from the data by estimating the probability for each interval.

10 000 records. The ground motion parameter, Y, is the horizontal peak ground acceleration (PGA). It is generated by a so-called *stochastic model* which is described in detail by Boore (2003). The basic idea is to distort the shape of a random time series according to physical principles and thus to obtain a time series with properties that match the ground-motion characteristics. The predictor variables are either uniform ($\mathcal{U}$) or exponentially (Exp) distributed within a particular interval (see Table 3).

The stochastic model does not have good analytical properties, and its usage is non-trivial and time consuming. Hence, surrogate models, which describe the stochastic model in a more abstract sense (e.g., regressions), are used in PSHA instead. We show that BNs may be seen as a viable alternative to the classical regression approach. However, before doing so, we need to touch upon some practical issues arising when learning BNs from continuous data.

For continuous variables we need to define the distributional family for the conditionals $p(\cdot|\cdot)$ and thus make assumptions about the functional form of the distribution. To avoid such assumptions and "let the data speak", we discretize the continuous variables, thus allowing for
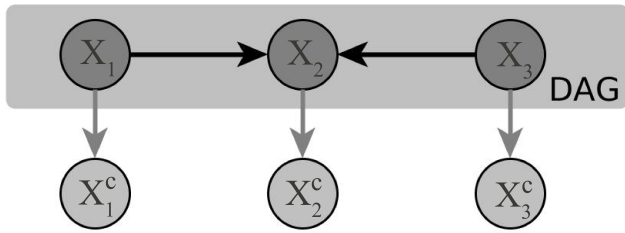
Figure 4. Representation of the dependency assumptions in the discretization approach: the dependency relations of the variables are captured by their discrete representations (gray-shaded area). A continuous variable, X_i^c, depends only on its discrete counterpart, X_i.

completely data-driven and distribution-free learning (see Fig. 3). In the following subsection we describe an automatic discretization, which is part of the BN learning procedure and takes the dependences between the single variables into account. However, the automatic discretization does not necessarily result in a resolution that matches the requirements for prediction purposes or decision support. To increase the potential accuracy of predictions, we approximate, once the network structure is learned, the continuous conditionals with *mixtures of truncated exponentials* (MTE), as suggested by Moral et al. (2001). More on this follows in Sect. 3.3.

3.2 Automatic discretization for structure learning

The range of existing discretization procedures differs in their course of action (supervised vs. unsupervised, global vs. local, top-down vs. bottom-up, direct vs. incremental, etc.), their speed and their accuracy. Liu et al. (2002) provide a systematic study of different discretization techniques, while Hoyt (2008) concentrates on their usage in the context with BN learning. The choice of a proper discretization technique is anything but trivial as the different approaches result in different levels of information loss. For example, a discretization conducted as a pre-processing step to BN learning does not account for the interplay of the variables and often misses information hidden in the data. To keep the information loss small, we use a multivariate discretization approach that takes the BN structure into account. The discretization is defined by a set of interval boundary points for all variables, forming a grid. All data points of the original continuous (or partly continuous) data set, d^c, that lie in the same grid cell, correspond to the same value in the discretized data set, d. In a multivariate approach, the "optimal" discretization, denoted by Λ, depends on the structure of the BN and the observed data, d^c. Similar to Sect. 2.2, we again cast the problem in a Bayesian framework searching for the combination of (DAG, θ, Λ) that has the highest posterior probability given the data,

corresponds according to Monti and Cooper (1998) to
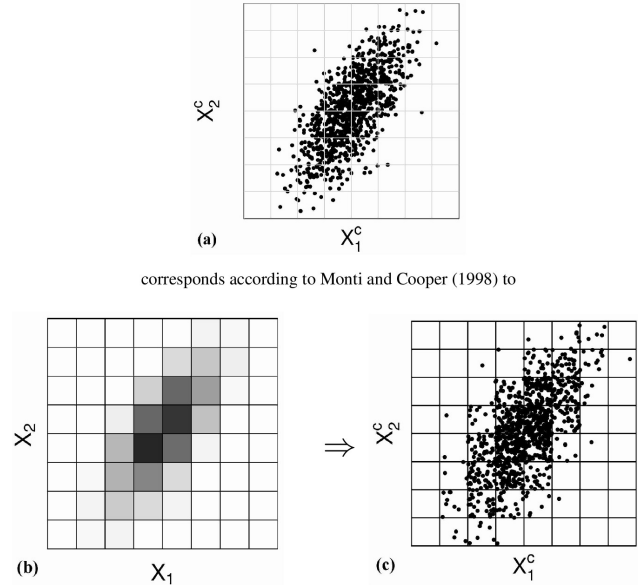
Figure 5. For the discretization approach each multivariate continuous distribution (**a**) is characterized by a discrete distribution that captures the dependence relations (**b**) and a continuous uniform distribution over each grid cell (**c**). For exemplification assume we consider two dependent, continuous variables: X_1^c and X_2^c. (**a**) shows a possible realization of a corresponding sample. According to Monti and Cooper (1998) we now assume that we can find a discretization, such that the resulting discretized variables X_1 and X_2 capture the dependence relation between X_1^c and X_2^c. This is illustrated by (**b**), where the shading of the grid cells corresponds to their probabilities (which are defined by θ). A darker color means that we expect more realizations in this grid cell. Further, we say that, within each grid cell, the realizations are uniformly distributed, as illustrated in (**c**).

$$\underbrace{P\left(\text{DAG}, \Theta, \Lambda | d^c\right)}_{\text{posterior}} \propto \underbrace{P\left(d^c | \text{DAG}, \Theta, \Lambda\right)}_{\text{likelihood}} \underbrace{P\left(\text{DAG}, \Theta, \Lambda\right)}_{\text{prior}}. \quad (3)$$

Let us consider the likelihood term: expanding on an idea by Monti and Cooper (1998), we assume that all communication/flow of information between the variables can be captured by their discrete representations (see Fig. 4) and is defined by the parameters θ. Thus only the distribution of the discrete data d depends on the network structure, while the distribution of the continuous data d^c is, for given d, independent of the DAG (see Figs. 4 and 5). Consequently the likelihood for observing d^c (for a given discretization, network structure and parameters) can be written as

$$P\left(d^c | \text{DAG}, \Theta, \Lambda\right) = P\left(d^c | d, \Lambda\right) P\left(d | \text{DAG}, \Theta, \Lambda\right) \quad (4)$$

and Eq. (3) decomposes into

$$P\left(\text{DAG}, \Theta, \Lambda | d^c\right) \propto \underbrace{P\left(d^c | d, \Lambda\right)}_{\text{continuous data}} \underbrace{P(d | \text{DAG}, \Theta, \Lambda)}_{\text{likelihood (discrete)}}$$
$$\underbrace{P(\text{DAG}, \Theta, \Lambda)}_{\text{prior}}.$$
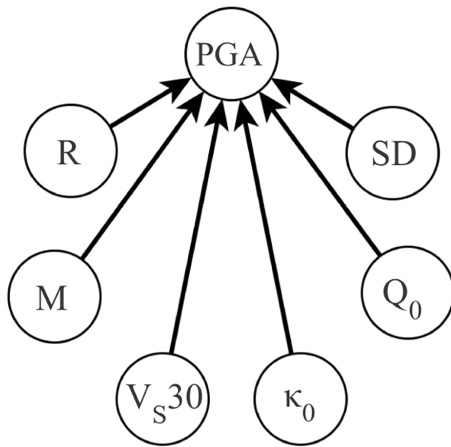
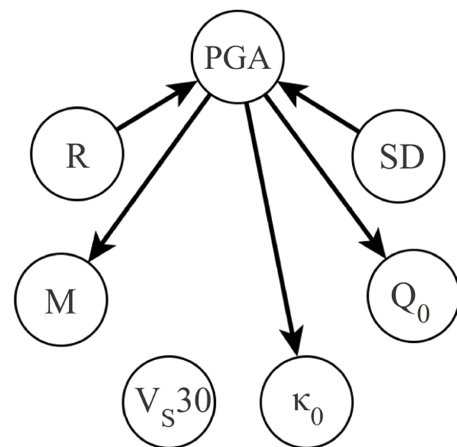**Figure 6.** Theoretic BN for the ground motion model. It captures the known dependences of the data-generating model.
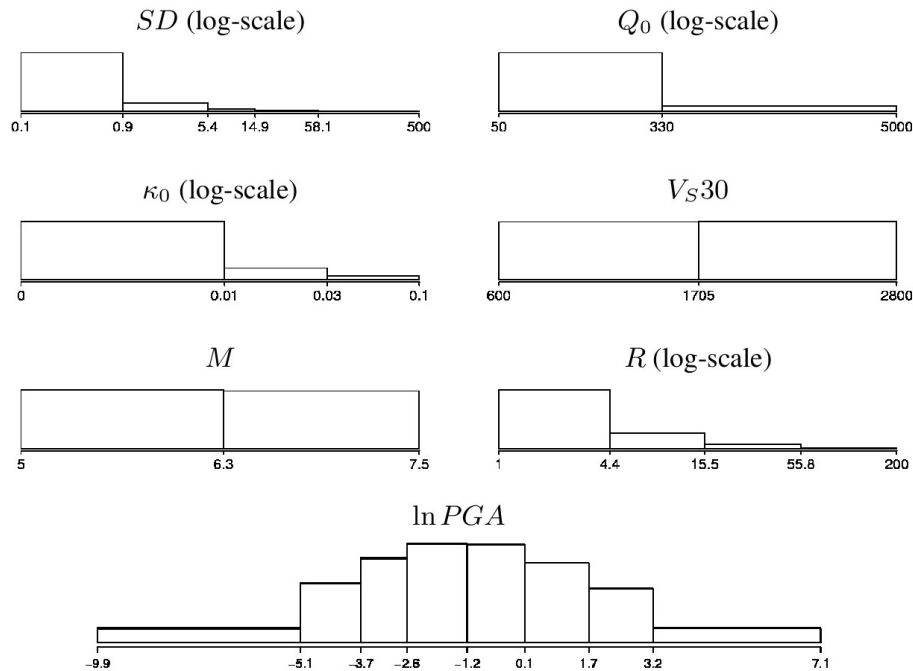

Figure 7. BN for the ground motion model learned from the generated synthetic data. It captures the most dominant dependences. Less distinctive dependences are neglected for the sake of parameter reduction.

The likelihood (discrete) term is now defined as for the separate BN learning for discrete data (Sect. 2.2), and we use a non-informative prior again. For the continuous data, we assume that all continuous observations within the same interval defined by Λ have the same probability (Fig. 5). More information about the score definition can be found in the Appendix A1, and technical details are given in Vogel et al. (2012, 2013). In the following we discuss the BN and discretization learned from the synthetic seismic data set.

Learned ground motion model

Since we generated the data ourselves, we know which (in)dependences the involved variables should adhere to; this is expected to be reflected in the BN DAG we learn from the synthetic data (Property 1, 3). Due to data construction, the predictor variables M, R, SD, Q_0, κ_0, and V_S30 are independent of each other and PGA depends on the predictors. Figure 6 shows the dependence structure of the variables. The converging edges at PGA indicate that the predictors become conditionally dependent for a given PGA. This means that, for a given PGA, they carry information about each other; for example, for an observed large PGA value, a small stress drop indicates a close distance to the earthquake. The knowledge about the dependence relations gives the opportunity to use the seismic hazard application for an inspection of the BN learning algorithm regarding the reconstruction of the dependences from the data, which is done in the following.

The network that we found to maximize $P(\text{DAG}, \Theta, \Lambda | d^c)$ for the 10 000 synthetic seismic data records is shown in Fig. 7. The corresponding discretization that was found is plotted in Fig. 8, which shows the marginal distributions of the discretized variables. The learned BN differs from the original one, mainly due to regularization constraints as we will explain in the following: as mentioned in Sect. 2, the joint distribution

of all variables can be decomposed into the product of the conditionals according to the network structure (see Eq. 1). For discrete/discretized variables, the number of parameters needed for the definition of $p(X_i | X_{\text{Pa}(i)})$ in Eq. (1) corresponds to the number of possible state combinations for (X_i, $X_{\text{Pa}(i)}$). Taking the learned discretization shown in Fig. 8, the BN of the data-generating process (Fig. 6) is defined by 3858 parameters, 3840 needed alone for the description of $p(\text{PGA} | M, R, \text{SD}, Q_0, \kappa_0, V_S30)$. A determination of that many parameters from 10 000 records would lead to a strongly over-fitted model. Instead we learn a BN that compromises between model complexity and its ability to generate the original data. The BN learned under these requirements (Fig. 7) consists of only 387 parameters and still captures the most relevant dependences.

Figure 9 shows the $\ln$ PGA values of the data set plotted against the single predictors. A dependence on stress drop (SD) and distance (R) is clearly visible. These are also the two variables with remaining converging edges on PGA, revealing that, for a given PGA, SD contains information about R and vice versa. The dependences between PGA and the remaining predictors are much less distinctive, such that the conditional dependences between the predictors are negligible and the edges can be reversed for the benefit of parameter reduction. The connection to V_S30 is neglected completely, since its impact on PGA is of minor interest compared to the variation caused by the other predictors.

Note that the DAG of a BN actually maps the independences (not the dependences) between the variables. This means that each (conditional) independence statement encoded in the DAG must be true, while encoded dependence relations must not hold per se (see Fig. 10 for explanation). In turn this implies that each dependence holding for the data should be encoded in the DAG. The learning approach

Figure 8. Marginal distribution of the variables included in the ground motion model, discretized according to the discretization learned for the BN in Fig. 7. The number of intervals per variable ranges from 2 to 8.

applied here fulfills the task quite well, detecting the relevant dependences, while keeping the model complexity at a moderate level.

The model complexity depends not only on the DAG but also on the discretization. A complex DAG will enforce a small number of intervals, and a large number of intervals will only be chosen for variables with a strong influence on other variables. This effect is also visible for the learned discretization (Fig. 8). PGA is split into eight intervals, distance and stress drop into four and five, respectively, and the other variables consist of only two to three intervals.

3.3 Approximation of continuous distributions with mixtures of exponentials (MTEs)

A major purpose of the ground motion model is the prediction of the ground motion ($\ln$ PGA) based on observations of the predictors; hence, although the BN captures the joint distribution (Property 5) of all involved variables, the focus in this context is on a single variable. The accuracy of the prediction is limited by the resolution of the discretization learned for the variable. For the BN shown above, the discretization of the target variable into eight intervals enables a quite precise approximation of the continuous distribution, but this is not the case per se. Complex network structures and smaller data sets used for BN learning lead to a coarser discretization of the variables. To enable precise estimates, we may search for alternative approximations of the (or at least some, in particular the primary variable(s) of interest)

continuous conditional distributions once the BN has been learned.

Moral et al. (2001) suggest using MTEs for this purpose, since they allow for the approximation of a variety of functional shapes with a limited number of parameters (Langseth and Nielsen, 2008) and they are closed under the operations used for BN inference: restriction, combination, and marginalization (Langseth et al., 2009). The basic idea is to approximate conditional distributions $p(X_i|X_{\mathrm{Pa}(i)})$ with a combination/mixture of truncated exponential distributions. For this purpose the domain $\Omega_{(X_i, X_{\mathrm{Pa}(i)})}$ is partitioned into hypercubes $D_1, \ldots, D_L$, and the density within each hypercube, D_l, is defined such that it follows the form

$$p_{\downarrow D_l}\left(X_i|X_{\mathrm{Pa}(i)}\right) = a_0 + \sum_{j=1}^{J} a_j\, e^{b_j X_i + c_j^T X_{\mathrm{Pa}(i)}}. \tag{5}$$

The determination of the hypercubes and the number of exponential terms in each hypercube as well as the estimation of the single parameters is done according to the maximum likelihood approach described in Langseth et al. (2010). In the following we show how the MTE approximation improves the BN prediction performance compared to the usage of the discretized variables, and we compare the results to those from a regression approach.

Prediction performance

We conduct a 10-fold cross validation to evaluate the prediction performance of the BN compared to the regression
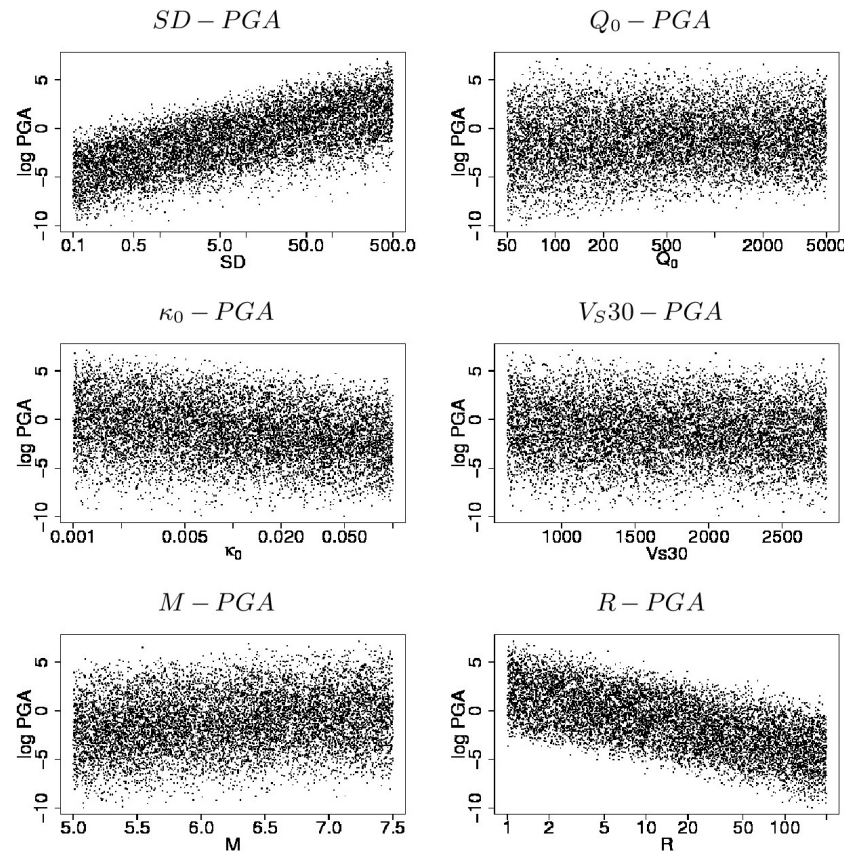
Figure 9. The individual panels show the dependences between the predictor variables M, R, SD, Q_0, κ_0, and $V_S 30$ and the target variable $\ln \text{PGA}$ by plotting the data used to learn the BN for ground motion modeling.

approach: the complete data set is divided into 10 disjoint subsamples, of which one is defined as a test set in each trial while the others are used to learn the model (regression function or BN). The functional form of the regression function is determined by expert knowledge based on the description of the Fourier spectrum of seismic ground motion and follows the form

$$f(X_{-Y}) = a_0 + a_1 M + a_2 M \cdot \ln \text{SD} + (a_3 + a_4 M)$$
$$\ln \sqrt{a_5^2 + R^2} + a_6 \kappa R + a_7 V_S 30 + a_8 \ln \text{SD},$$

with $\kappa = \kappa_0 + t^*$, $t^* = \frac{R}{Q_0 V_{sq}}$ and $V_{sq} = 3.5 \text{ km s}^{-1}$.

We compare the regression approach in terms of prediction performance to the BN with discretized variables and with MTE approximations. For this purpose we determine the conditional density distributions of $\ln \text{PGA}$ given the predictor variables for each approach and consider how much probability it assigns to the real $\ln \text{PGA}$ value in each observation. For the regression approach the conditional density follows a normal distribution, $\mathcal{N}(f(X_{-Y}), \sigma^2)$, while it is defined via the DAG and the parameters θ using the BN models. Table 4a shows for each test set the conditional density value of the observed $\ln \text{PGA}$ averaged over the individ-

ual records. Another measure for the prediction performance is the mean squared error of the estimates for $\ln \text{PGA}$ (Table 4b). Here the point estimate for $\ln \text{PGA}$ is defined as the mean value of the conditional density. For example, in the regression model the estimate corresponds to $f(x_{-Y})$.

Even though the discretization of $\ln \text{PGA}$ is relative precise using the discrete BNs (eight intervals in each trial, except for the first trial, where $\ln \text{PGA}$ is split into seven intervals), the MTE approximation of the conditional distributions improves the prediction performance of the BN. Still, it does not entirely match the precision of the regression function. However, the prediction performances are on the same order of magnitude, and we must not forget that the success of the regression approach relies on the expert knowledge used to define its functional form, while the structure of the BN is learned in a completely data-driven manner. Further the regression approach profits in this example from the fact that the target variable ($\ln \text{PGA}$) is normally distributed, which is not necessarily the case for other applications. Focusing on the prediction of the target variable the regression approach also does not have the flexibility of the BN, which is designed to capture the joint distribution of all variables and thus allows for inference in all directions (Property 5),
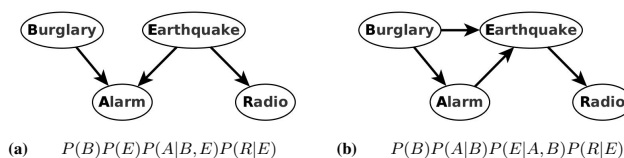
Table 4. Results of a 10-fold cross validation to test the prediction performance of the BN (with discrete and MTE approximations of the conditional distributions) and the regression approach. **(a)** contains the calculated conditional densities for the observed ln PGA values averaged over each trial. **(b)** contains the mean squared error of the predicted ln PGA for each trial.

(a) $P(B)P(E)P(A|B,E)P(R|E)$ (b) $P(B)P(A|B)P(E|A,B)P(R|E)$

Figure 10. The graph structure of a BN dictates how the joint distribution of all variables decomposes into a product of conditionals. Thus for a valid decomposition each independence assumption mapped into the BN must hold. Usually this applies to a variety of graphs, i.e., the complete graph is always a valid independence map as it does not make any independence assumption. **(a)** and **(b)** show two valid BN structures and the corresponding decompositions for the burglary example. The independence assumptions made in both BNs hold; however **(b)** does not capture the independence between earthquakes and burglaries. An independence map that maps all independences **(a)** is called a perfect map, yet perfect maps do not exist for all applications. Furthermore, for parameter reduction it might be beneficial to work with an independence map that differs from the perfect map.

(a) Averaged conditional density		
	$BN_{discrete}$	BN_{MTE}	Regression
1	0.237	0.320	0.331
2	0.240	0.297	0.329
3	0.239	0.298	0.331
4	0.218	0.255	0.323
5	0.216	0.260	0.339
6	0.222	0.257	0.339
7	0.215	0.252	0.332
8	0.243	0.317	0.330
9	0.212	0.249	0.328
10	0.243	0.315	0.331
Avg.	0.229	0.282	0.331

| (b) Mean squared error		
	$BN_{discrete}$	BN_{MTE}	Regression
1	1.021	0.749	0.663
2	1.197	0.963	0.680
3	1.082	0.821	0.673
4	1.262	0.951	0.723
5	1.201	0.851	0.629
6	1.298	1.059	0.625
7	1.297	1.077	0.672
8	1.149	0.713	0.701
9	1.343	1.161	0.692
10	1.169	0.841	0.666
Avg.	1.202	0.919	0.672

variety of factors (Thieken et al., 2005), stage–damage functions are still widely used. This is because the number of potential influencing factors is large and the single and joint effects of these parameters on the degree of damage are largely unknown.

4.1 Real-life observations

The data collected after the 2002 and 2005/2006 flood events in the Elbe and Danube catchments in Germany (see Fig. 11) offer a unique opportunity to learn about the driving forces of flood damage from a BN perspective. The data result from computer-aided telephone interviews with flood-affected households, and contain 1135 records for which the degree of damage could be reported. The data describe the flooding and warning situation, building and household characteristics, and precautionary measures. The raw data were supplemented by estimates of return periods, building values, and loss ratios, as well as indicators for flow velocity, contamination, flood warning, emergency measures, precautionary measures, flood experience, and socioeconomic factors. Table 5 lists the 29 variables allocated to their domains. A detailed description of the derived indicators and the survey is given by Thieken et al. (2005) and Elmer et al. (2010). In Sect. 3.2 we dealt with the issue of continuous data when learning BNs; here we will apply the methodology presented there. However, in contrast to the synthetic data from the previous section, many real-world data sets are, for different reasons, lacking some observations for various variables. For the data set at hand, the percentage of missing values is below 20 % for most variables, yet for others it reaches almost 70 %. In the next subsection we show how we deal with the missing values in the setting of the automatic discretization described in Sect. 3.2 when learning BNs.

as exemplified in Sect. 4.3. Additional benefits of BNs, like their ability to make use of incomplete observations, will be revealed in the following sections, where we investigate real-world data.

4 Flood damage assessment

In the previous section we dealt with a fairly small BN (a few variables/nodes) and a synthetic data set. In this section we go one step further and focus on learning a larger BN from real-life observations on damage caused to residential buildings by flood events. Classical approaches, so-called stage–damage functions, relate the damage for a certain class of objects to the water stage or inundation depth, while other characteristics of the flooding situation and the flooded object are rarely taken into account (Merz et al., 2010). Even though it is known that the flood damage is influenced by a

Table 5. Variables used in the flood damage assessment and their corresponding ranges. C: continuous; O: ordinal; N: nominal.

Variable	Scale and range	Percentage of missing data
	Flood parameters	
Water depth	C: 248 cm below ground to 670 cm above ground	1.1
Inundation duration	C: 1 to 1440 h	1.6
Flow velocity indicator	O: 0 = still to 3 = high velocity	1.1
Contamination indicator	O: 0 = no contamination to 6 = heavy contamination	0.9
Return period	C: 1 to 848 years	0
	Warning and emergency measures	
Early warning lead time	C: 0 to 336 h	32.3
Quality of warning	O: 1 = receiver of warning knew exactly what to do to 6 = receiver of warning had no idea what to do	55.8
Indicator of flood warning source	N: 0 = no warning to 4 = official warning through authorities	17.4
Indicator of flood warning information	O: 0 = no helpful information to 11 = many helpful information	19.1
Lead time period elapsed without using it for emergency measures	C: 0 to 335 h	53.6
Emergency measures indicator	O: 1 = no measures undertaken to 17 = many measures undertaken	0
	Precaution	
Precautionary measures indicator	O: 0 = no measures undertaken to 38 = many efficient measures undertaken	0
Perception of efficiency of private precaution	O: 1 = very efficient to 6 = not efficient at all	2.9
Flood experience indicator	O: 0 = no experience to 9 = recent flood experience	68.6
Knowledge of flood hazard	N (yes/no)	32.7
	Building characteristics	
Building type	N: (1 = multifamily house, 2 = semi-detached house, 3 = one-family house)	0.1
Number of flats in building	C: 1 to 45 flats	1.2
Floor space of building	C: 45 to 18 000 m^2	1.9
Building quality	O: 1 = very good to 6 = very bad	0.6
Building value	C: €92 244 to 3 718 677	0.2
	Socioeconomic factors	
Age of the interviewed person	C: 16 to 95 years	1.6
Household size, i.e., number of persons	C: 1 to 20 people	1.1
Number of children (< 14 years) in household	C: 0 to 6	10.1
Number of elderly persons (> 65 years) in household	C: 0 to 4	7.6
Ownership structure	N: (1 = tenant; 2 = owner of flat; 3 = owner of building)	0
Monthly net income in classes	O: 11 = below EUR 500 to 16=EUR 3000 and more	17.6
Socioeconomic status according to Plapp (2003)	O: 3 = very low socioeconomic status to 13 = very high socioeconomic status	25.5
Socioeconomic status according to Schnell et al. (1999)	O: 9 = very low socioeconomic status to 60 = very high socioeconomic status	31.7
	Flood loss	
rloss – loss ratio of residential building	C: 0 = no damage to 1 = total damage	0
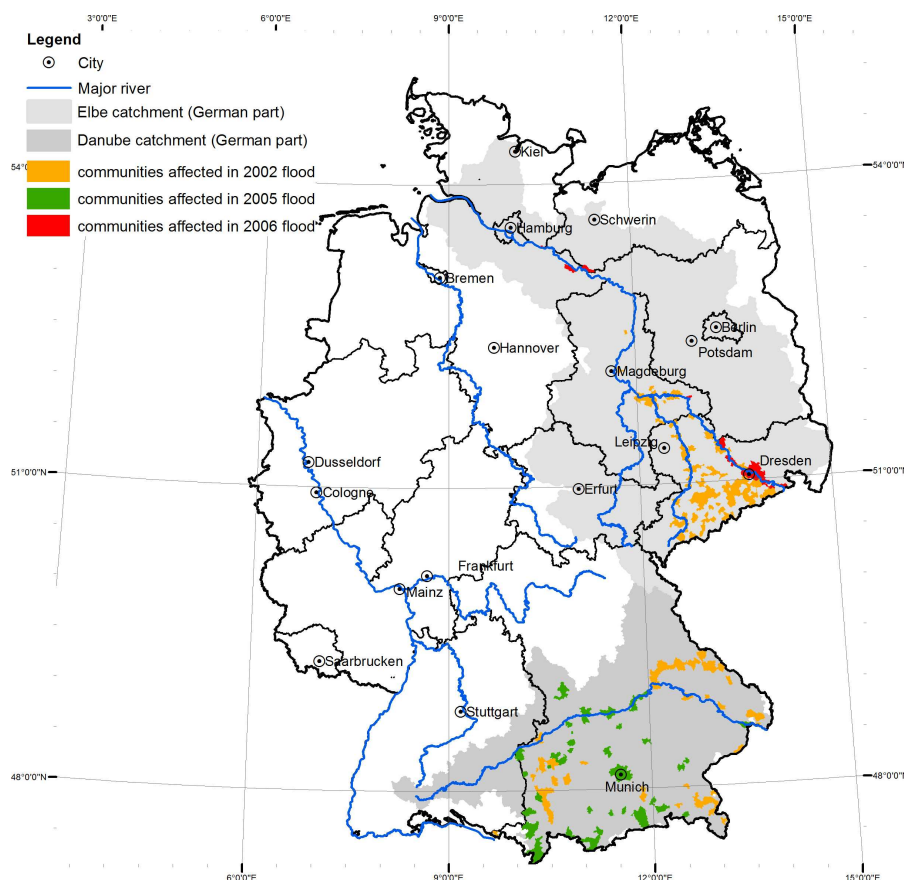
Figure 11. Catchments investigated for the flood damage assessment and location of communities reporting losses from the 2002, 2005, and 2006 floods in the Elbe and Danube catchments (Schroeter et al., 2014).

4.2 Handling of incomplete records

To learn the BN, we again maximize the joint posterior for the given data (Eq. 3). This requires the number of counts for each combination of states for $(X_i, X_{\text{Pa}(i)})$, considering all variables, $i = 1, \ldots, k$ (see Appendix A1). However this is only given for complete data, and for missing values it can only be estimated by using expected completions of the data. We note that a reliable and unbiased treatment of incomplete data sets (no matter which method is applied) is only possible for missing data mechanisms that are *ignorable* according to the *missing (completely) at random* (M(C)AR) criteria as defined in Little and Rubin (1987), i.e., the absence/presence of a data value is independent of the unobserved data. For the data sets considered in this paper, we assume the MAR criterion to hold and derive the predictive function/distribution based on the observed part of the data in order to estimate the part which is missing.

In the context of BNs a variety of approaches has been developed to estimate the missing values (so-called "imputation"). Most of these principled approaches are iterative algorithms based on expectation maximization (e.g., Fried-

man, 1997, 1998) or stochastic simulations (e.g., Tanner and Wong, 1987). In our case we already have to run several iterations of BN learning and discretization, each iteration requiring the estimation of the missing values. Using an iterative approach for the missing value prediction will thus easily become infeasible. Instead we use a more efficient albeit approximate method, using the *Markov blanket predictor* developed by Riggelsen (2006).

The idea is to generate a predictive function which enables the prediction of a missing variable X_i based on the observations of its Markov blanket (MB), $X_{\text{MB}(i)}$. The Markov blanket identifies the variables that directly influence X_i, i.e., the parents, and children of X_i, as well as the parents of X_i's children. An example is given in Fig. 12. Assuming the MB is fully observed, it effectively blocks influence from all other variables, i.e., the missing value depends only on its MB. When some of the variables in the MB are missing, it does not shield off X_i. However, for predictive approximation purposes, we choose to always ignore the impact from outside the MB. Hence, the prediction of X_i based on the observed data reduces to a prediction based on the observations of the
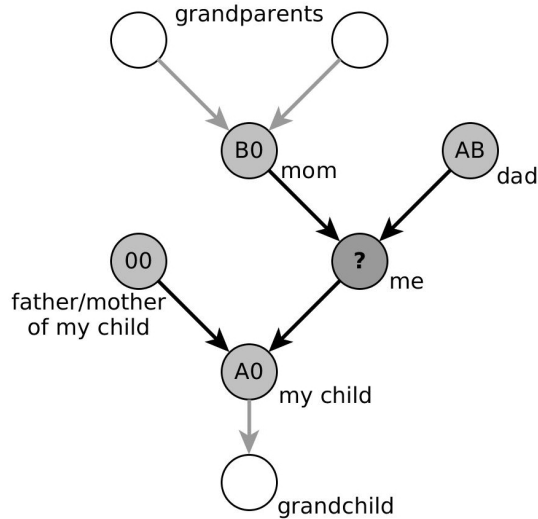
Figure 12. Illustration of a Markov blanket (gray-shaded nodes) on a blood group example: let us assume that I do not know my blood group for some reason, but I know the genotypes of my relatives. The genotypes of my parents provide information about my own blood group specification – in the pictured example they restrict the list of opportunities to the four options: AB, A0, B0 and BB – as well as the genotype of my child reveals information, excluding BB from the list of possible options. Considering the genotype of the father/mother of my child alone does not provide any information about my blood type (our blood groups are independent from each other), but together with the information about our child it again restricts the list of opportunities, leaving only AB and A0 as possible options (conditioned on our child our blood groups become dependent). All these variables (blood type of my parents, my children, and the parents of my children) provide direct information about the considered variable (my blood type) and form its Markov blanket. If I know the values of the Markov blanket, further variables do not provide any additional information. For example, knowing the genotypes of my parents, the knowledge about my grandparents does not deliver any further information about myself (the information is "blocked" by my parents). Yet, if the blood type of my parents is unknown, the information about my grandparents can "flow" and provides new insights.

MB and factorizes according to the DAG in Fig. 13a:

$$P\left(X_i | X_{\mathrm{MB}(i)}, \boldsymbol{\theta}, \mathrm{DAG}\right) \propto \theta_{X_i | X_{\mathrm{Pa}(i)}} \prod_{j \in \mathrm{Ch}(i)} \theta_{X_j | X_{\mathrm{Pa}(j)}}, \qquad (6)$$

where $\mathrm{Ch}(i)$ are the variable indices for the children of X_i. Thus the prediction of X_i requires, according to Eq. (6), inference in the BN (albeit very simple) where correct estimates of $\boldsymbol{\theta}$ are assumed. These in general can not be given without resorting to iterative procedures. To avoid this we define a slightly modified version of the predictive function, for which we define all variables that belong to the MB of X_i to be the parents of X_i in a modified DAG′ (see Fig. 13 for illustration). Thus $X_{\mathrm{Pa}(i)}^{\mathrm{DAG}'}$ corresponds to $X_{\mathrm{MB}(i)}^{\mathrm{DAG}}$. The resulting DAG′ preserves all dependences given in DAG and can
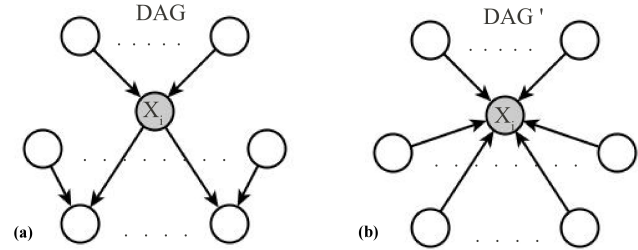
Figure 13. (a) The Markov blanket of X_i comprises its parents and children, as well as the parents of its children. The prediction of missing values is based on the observations of the variables in the Markov blanket. To avoid inference that requires unknown parameters, the subgraph of DAG that spans the Markov blanket (a) is modified by directing all edges towards X_i, receiving the DAG′ pictured in (b).

alternatively be used for the prediction of X_i,

$$P\left(X_i | X_{\mathrm{Pa}(i)}^{\mathrm{DAG}'}, \boldsymbol{\theta}^{\mathrm{DAG}'}, \mathrm{DAG}'\right) \stackrel{\mathrm{def}}{=} \theta_{X_i | X_{\mathrm{Pa}(i)}}^{\mathrm{DAG}'}. \qquad (7)$$

For this predictive distribution we need to estimate the parameters $\theta_{X_i | X_{\mathrm{Pa}(i)}}^{\mathrm{DAG}'}$. Note that more parameters are required for the newly derived predictive distribution, but now at least *all* influencing variables are considered jointly and an iterative proceeding can be avoided. The parameters are estimated with a *similar-cases* approach, which is described in Appendix A2. A detailed description for the generation of the predictive distribution is given in Riggelsen (2006) and Vogel et al. (2013).

It is worth noting that, as the MBs of variables change during the BN learning procedure, the prediction of missing values (depending on the MB) needs to be updated as well.

4.3 Results

Coming back to the flood damage data, we have three variables with more than one-third of the observations missing: flood experience (69 % missing), warning quality (56 % missing) and lead time elapsed without emergency measures (54 % missing). In a first "naive" application (Vogel et al., 2012), no special attention was paid to a proper treatment of missing values; the missing values were simply randomly imputed, resulting in the isolation of two variables (flood experience and lead time elapsed) in the network; no connection to any other variable was learned (Fig. 14a). With application of the Markov blanket predictor, the situation changes and a direct connection from the relative building damage, rloss, to flood experience is found, as well as a connection between warning source and elapsed lead time (Fig. 14b). These relations, especially the first one, match with experts' expectations and speak for an improvement in the learned BN structure.
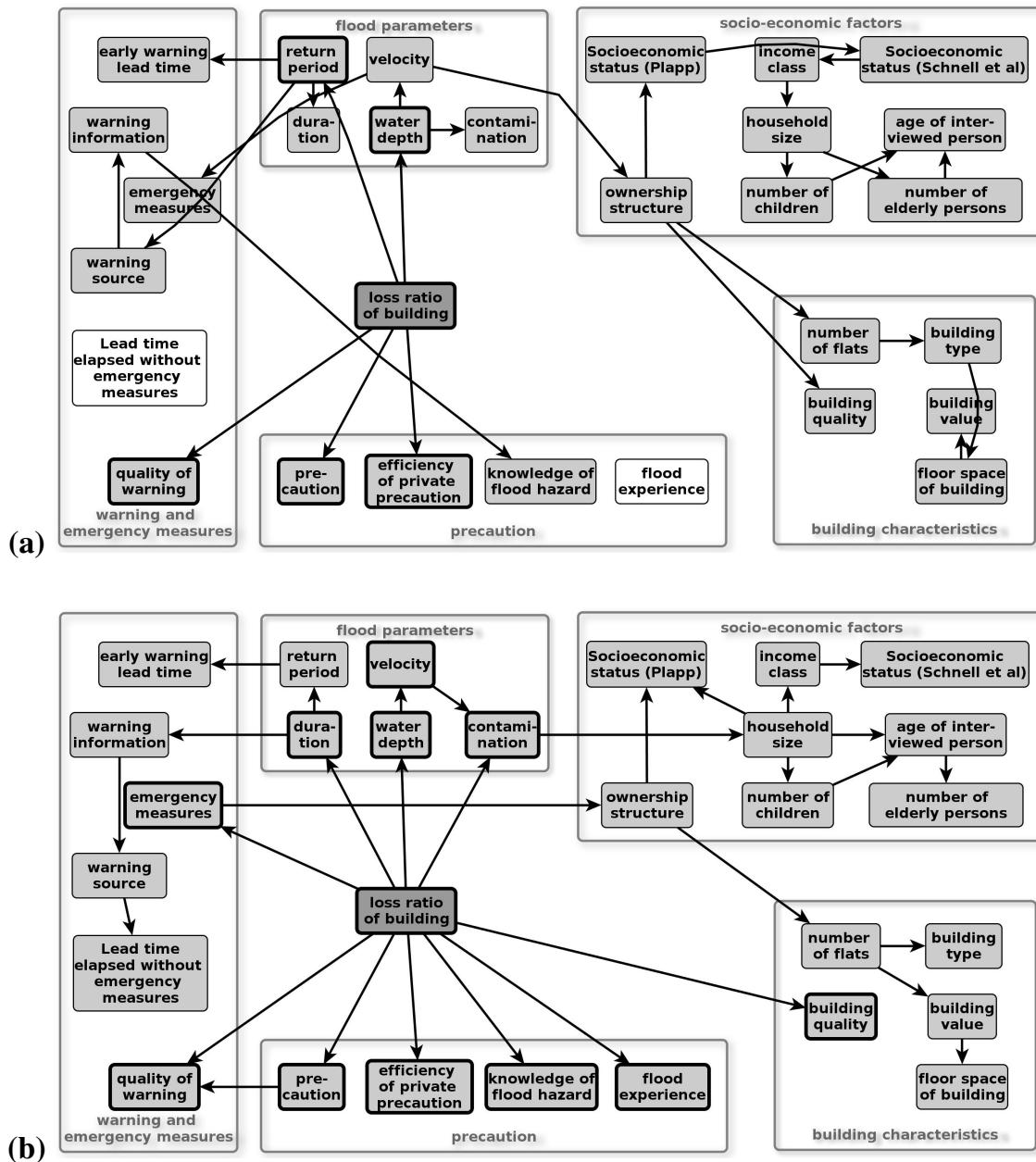
Figure 14. BNs learned for flood damage assessments, showing the effect of the applied missing value estimator. The algorithm used to learn **(a)** replaces missing values randomly, while the one used to learn **(b)** applies the Markov blanket predictor for the estimation of missing values. Nodes with a bold frame belong to the Markov blanket of *relative building loss* and are thus assumed to have a direct impact on the caused flood damage.

Using the graphical representation (Property 1), as mentioned in Sect. 2.1, the learned DAG (Fig. 14b) gives insight into the dependence relations of the variables. It reveals a number of direct links connecting the damage-describing variable with almost all subdomains. This supports the demand for improved flood damage assessments that take several variables into account (Merz et al., 2010). Moreover, the DAG shows which variables are the most relevant for

the prediction of rloss. The domains "precaution" and "flood parameters" in particular are densely connected to building damage and should be included in any damage assessment (Property 3).

Existing approaches for flood damage assessments usually consider fewer variables and an employment of a large number of variables is often considered as disadvantageous, since complete observations for all involved variables are rare. The
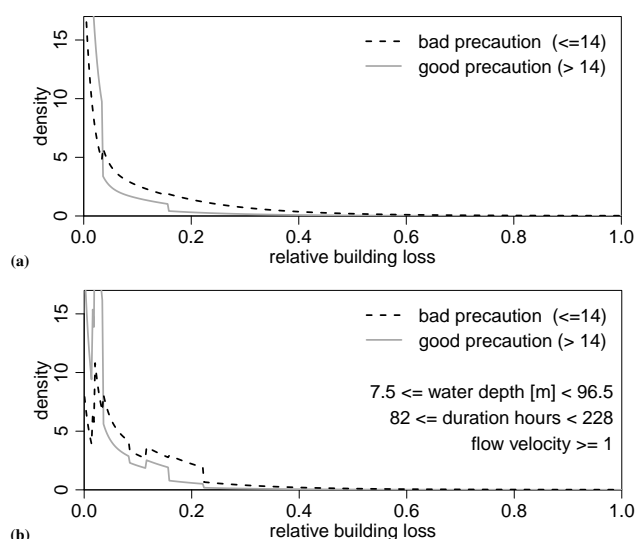
Figure 15. (a) Conditional distribution of the building loss conditioned on the precaution. Flood-specific parameters as well as other parameters are unknown and summed out. **(b)** Conditional distribution of the building loss depending on precaution for a specific flood situation: water depth, duration, and flow velocity are known. Other parameters are unknown and summed out.

requirement for complete observations does not hold for BNs (Property 6). The prediction of the building damage, for example, depends only on the variables of its Markov blanket (marked with a bold frame in Fig. 14). If the observation of the Markov blanket variables is incomplete (not all variables are observed at inference time), information from outside the Markov blanket "flows" into the prediction by indirectly marginalizing (summing) missing variables out. The inclusion of many variables thus provides additional knowledge and proves to be an advantage of BNs.

Moreover, the capability of BNs to predict from incomplete observations enables us to make predictions at an early stage of an event, employing only the information that is present at any given time. The prediction can subsequently be updated as new information becomes available.

4.3.1 Prediction performance

As for the seismic hazard example in Sect. 3, the prediction of a certain target variable is likewise of particular interest in flood damage assessments. Similar to our previous proceeding (Sect. 3.3) we approximate the distribution of the target variable with mixtures of truncated exponentials, thus achieving a better resolution for the distribution of interest. The resulting prediction performance of the BN is compared to currently used flood damage assessment approaches, namely the stage–damage function (sdf) and the FLEMOps + r model (Elmer et al., 2010), which was developed from the same data set, estimating the building damage based on water depth, flood frequency, building

Table 6. Results of a 5-fold cross validation to compare the prediction performance of the three models used in the flood damage assessment: stage–damage function, FLEMOps + r, and Bayesian networks. For each trial, the table contains the mean squared error of the estimated fraction of building damage (in log scale).

	sdf	FLEMOps + r	BN_{MTE}
1	0.0111	0.0108	0.0104
2	0.0133	0.0114	0.0116
3	0.0161	0.0145	0.015
4	0.0200	0.0194	0.0169
5	0.0166	0.0150	0.0163
Avg.	0.0154	0.0142	0.014

type, building quality, contamination, and private precaution. While sdf and FLEMOps + r give point estimates, the BN delivers a distribution for rloss and thus reveals the uncertainty of the prediction (Property 4). Especially when it comes to decision making, the identification of uncertainty is a major advantage of the BN. However, to allow for model comparison, we reduce the distribution provided by the BN to its mean value, which we define to be the estimate of rloss. Table 6 shows the mean squared error of a 5-fold cross validation for the three model approaches. The prediction performance of the BN is comparable to the one of the FLEMOps + r, while the BN has the additional advantage of modeling the whole distribution of the target variable and conducting the prediction even though not all variables are observed.

4.3.2 Example for inference: impact of precaution

As an example of reasoning (Property 5), we consider the effect of precaution on the building loss. Figure 15 shows the distribution of the building loss for a good precaution (precautionary measures indicator > 14) and a bad precaution (precautionary measures indicator ≤ 14) in a general case (Fig. 15a: all other variables are unknown and summed out) and for a specific flood event (Fig. 15b: 7.5 m ≤ water depth < 96.5 m; 82 h ≤ duration < 228 h; 1 ≤ velocity). We may appreciate how a good precaution increases the chance for no or only small building losses.

Similar investigations may support the identification of efficient precautionary measures, not only in the context of flood events but also for natural hazards in general. They may also help to convince authorities or private persons to undertake the suggested precautions. Using the flexibility of BNs and their ability to model specific situations, BNs may thus contribute to a better communication between scientists and non-scientific stakeholders. BNs can also be used for forensic reasoning, i.e., we can turn around the direction of reasoning in the example just considered and ask what a likely state of precaution is for a given observed damage in a specific or general event situation. Forensic reasoning might be of

interest, for instance, for insurance companies. Forensic reasoning might be of interest, for instance, for insurance companies.

5 Landslides

So far we assumed the existence of a unique model that explains the data best. In practical problems, however, there may be many models almost as good as the best, i.e., ones that explain the data similarly well. This results in an uncertainty about which BN structure to use. We consider this problem in our last application, where we apply BN learning to landslides, which are another ubiquitous natural hazard in many parts of the world.

A key theme in many landslide studies is the search for those geological, hydroclimatological, topographic, and environmental parameters that sufficiently predict the susceptibility to slope failure in a given region. A wide range of multivariate data analysis techniques has been proposed to meet this challenge. Amongst the more prominent methods are logistic regression, artificial neural networks, and Bayesian weights of evidence. The popularity of such methods is only matched by their seeming success: a recent review of 674 scientific papers on the topic indicates that most reported success rates are between 75 and 95 % (Korup and Stolle, 2014), where in the majority of studies the success rate is defined as the percentage of correctly (true positives and true negatives) identified locations that were subject to slope instability in the past. This raises the question as to why landslides still continue to cause massive losses despite this seemingly high predictive accuracy. Moreover, success rates do not show any significant increase over the last 10 years regardless of the number of landslide data or predictors used (Korup and Stolle, 2014). An often overlooked key aspect in these analyses is the potential for correlated or interacting predictor candidates. Few studies have stringently explored whether this likely limitation is due to physical or statistical (sampling) reasons.

5.1 Data

The landslide data are taken from an inventory of $\sim 300\,000$ digitally mapped landslide deposit areas across the Japanese islands (Korup et al., 2014). These landslides were mapped systematically mostly from stereographic image interpretation of air photos, and compiled by the National Research Institute for Earth Science and Disaster Prevention NIED (http://lsweb1.ess.bosai.go.jp/gis-data/index.html). The dominant types of failure in this database are deep-seated slow-moving earthflows and more rapid rockslides. The mapped size range of the deposits from these landslides spans from 10^2 to $10^7\,\mathrm{m}^2$ footprint area and is distinctly heavy tailed (Korup et al., 2012). Many of the landslide deposits are covered by vegetation. Individual deposits do not carry any time-stamp information, and so the inventory contains both historic and prehistoric slope failures, likely containing landslides up to several thousands of years old. Smaller rockfalls or soil slips are not included. Similarly, the inventory contains no data on specific trigger mechanisms (such as earthquakes, rainfall, or snowmelt), the dominant type of materials mobilized, or absolute age information for the bulk of individual landslides. In this context, the data nicely reflect common constraints that scientists encounter when compiling large landslide databases from remote sensing data covering different time slices. Yet this type of inventory is frequently used as a key input for assessing and mapping regional landslide susceptibility from a number of statistical techniques, including BNs. However, data-driven learning of BNs containing landslide information has, to the best of our knowledge, not been attempted before. We have compiled a number of geological, climatic, and topographic metrics for individual catchments throughout the Japanese islands to test their influence on the average fraction of landslide-affected terrain that we computed within a 10 km radius. Most of our candidate predictors (Table 7) have been used in modified form in other studies (Korup et al., 2014). While all of these candidate predictors may be physically related to slope instability, our choice of predictors is intentionally arbitrary in order to learn more about their effects on BN learning and structure. The final data set used for the BN learning consists of landslide and predictor data that we averaged at the scale of 553 catchments that are up to $10^3\,\mathrm{km}^2$ large, and that we sampled randomly from the drainage network across Japan. This averaging approach produced $\sim 0.4\,\%$ missing data in the subset, and aptly simulates further commonly encountered constraints on the quality of large landslide inventories.

5.2 Uncertainty in BN structure

Ideally, a given model should adequately encapsulate natural phenomena such as the causes and triggers of slope instability. However, there may be several equally well poised, but competing, models because of the intrinsic uncertainty tied to the governing processes. In practice we also face other limitations that prevent us from focusing on one single best model. The finite number of observations we have at our disposal for learning, and the fact that it is unclear which relevant predictor variables to consider for landslide prediction, implies that several models may be justifiable. This is a general problem when attempting to formally model natural systems. In our case this means that several BNs might explain the data (almost) equally well, i.e., they receive a similar score according to Eq. (2).

An additional source of uncertainty stems from the structure learning algorithm used to maximize the score defined in Eq. (2) or – for continuous variables – in Eq. (3). For infinite data sets the algorithm terminates according to *Meek's conjecture* in the (unique) optimal equivalence class of DAGs

Table 7. Variables used in the landslide model.

Name	Definition	Unit
Mean elevation	Average of elevation values within catchment boundaries	[m]
Catchment area	Log-transformed catchment area	[a.u.]
Catchment perimeter	Total length of catchment divides	[m]
Mean local topographic relief	Maximum elevation difference in a 10 km radius	[m]
Mean annual precipitation[a]	Based on interpolated rainfall station data (reference period 1980–2010)	[mm]
Mean coefficient of variation of annual precipitation[a]	Based on interpolated rainfall station data, with standard deviation divided by mean (reference period 1980–2010)	[1]
Mean coefficient of variation of monthly precipitation[a]	Based on interpolated rainfall station data, with standard deviation divided by mean (reference period 1980–2010)	[1]
Mean surface uplift 2001–2011[b]	GPS-derived accumulated surface uplift 2001–2011	[m]
Mean surface uplift 2010–2011[b]	GPS-derived accumulated surface uplift 2010–2011	[m]
Mean fraction of 10 % steepest bedrock channels	Average fraction of 10 % steepest channels per unit length of bedrock-river drainage network in a 10 km radius, based on an arbitrarily set reference concavity $\theta = 0.45$	[1]
Mean bedrock channel steepness	Average of channel steepness index per reach length, based on an arbitrarily set reference concavity $\theta = 0.45$	[1]
Regionalized river sinuosity	Average bedrock-channel sinuosity weighted by drainage network length in a 10 km radius calculated as the flow length of a given channel segment divided by its shortest vertex distance	[1]
Fraction of volcanic rocks[c]	Fraction of catchment area underlain by volcanic rocks	[1]
Fraction of lakes	Fraction of catchment area covered by lakes	[1]
Fraction of plutonic rocks[c]	Fraction of catchment area underlain by plutonic rocks	[1]
Fraction of sedimentary rocks[c]	Fraction of catchment area underlain by sedimentary rocks	[1]
Fraction of accretionary complex rocks[c]	Fraction of catchment area underlain by accretionary complex rocks	[1]
Fraction of metamorphic rocks[c]	Fraction of catchment area underlain by metamorphic rocks	[1]
Median area of landslide-affected terrain	Fraction of landslide terrain per unit catchment area within a 10 km radius calculated using an inventory of mostly prehistoric landslide-deposit areas	[1]

[a] Calculated using data provided by the Japan Meteorological Agency (JMA, http://www.jma.go.jp/jma/indexe.html).
[b] Calculated from secular high-precision leveling data (Kimura et al., 2008).
[c] Calculated using the seamless digital geological map of Japan (1 : 200 000) available from the Geological Survey of Japan (https://gbank.gsj.jp/seamless).

(Chickering, 2002), but this does not necessarily hold for finite data sets, incomplete observations and a search space extended by the discretization. The algorithm for the traversal of the BN hypothesis space contains stochastic elements and may get stuck in local optima, providing slightly different results for different runs.

To analyze this random behavior, we run the BN learning and discretization algorithm 10 times on the same data set of landslide data. We do not expect to end up with the same BN in each trial, as the constraints to meet Meek's conjecture are not fulfilled. Instead, we are more interested in documenting how strongly the results differ from each other.

Figure 16 gives a summarized representation of the BN DAG structures. The frequency with which an edge between two variables is learned is encoded according to its widths (by scaling it accordingly). Despite the differences in DAG structures, all learned BNs seem to model the data-generating process almost equally well, which can be gathered from the score obtained by Eq. (3): for the BNs learned, we observed scores between $-64\,364.42$ and $-64\,253.98$. This is a promising result, since it indicates that, even though the algorithm gets stuck in local maxima, the quality of the results

does not differ significantly. This supports the assumption that the quality of the learned BN is not seriously affected by random effects of the learning algorithm. Multiple runs of the algorithm on other data sets confirm this assumption.

In literature on BN learning (and on model learning based on data in general), ideas of how to handle several competing, but all justifiable, BNs have been investigated. Friedman et al. (1999) use bootstrap sampling to learn BNs from different variations of the data set. Based on those they develop a confidence measure on features of a network (e.g., the presence of an edge or membership of a node to a certain Markov blanket). A Bayesian approach is presented by Friedman and Koller (2000) and Riggelsen (2005), who approximate the Bayesian posterior on the DAG space using a Markov chain Monte Carlo approach. An adaptation of these methods for the extended MAP score introduced in this paper is left for future work.

5.3 Results

Despite (or rather thanks to) the DAG structural differences, we can glean some instructive insights from the learned BNs.
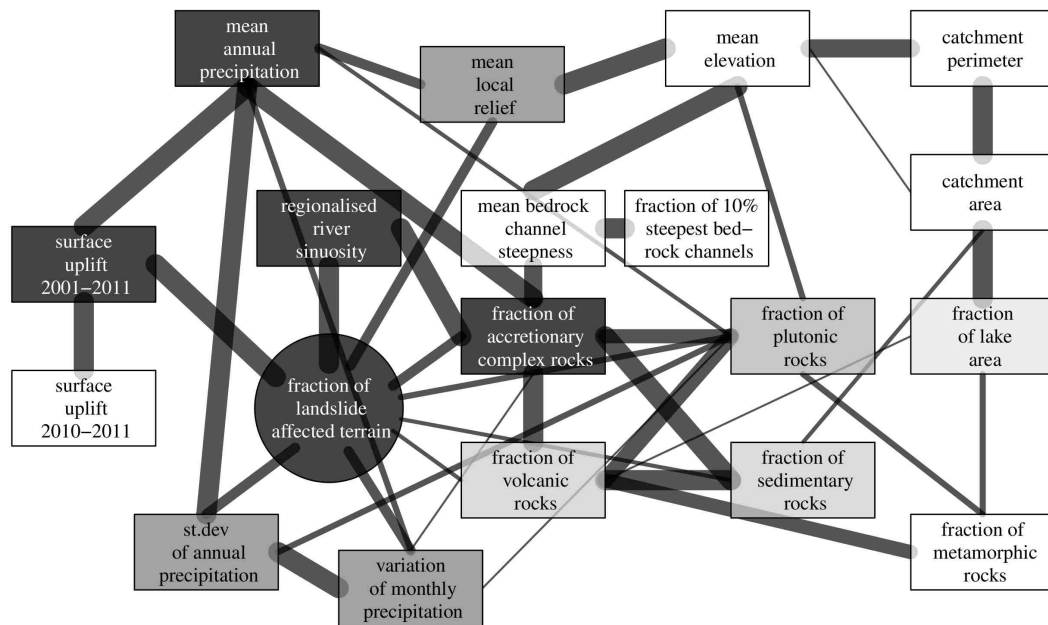
Figure 16. Summary of 10 learned network structures modeling landslides susceptibility, all based on the same data set. Arrow widths between the variables are scaled to the number of times they occur in the learned BNs. Likewise, we color-coded the variables according to the frequency with which they occur as part of the Markov blanket of *fraction of landslide-affected terrain* (circular node shape), where darker hues indicate more frequent occurrences.

The fact that we can learn something about the landslide-affected terrain from several BN structures indicates that the different predictors are highly interacting, and that a missed link between two variables can often be compensated for by other interactions. To understand which variables are most relevant for the prediction of landslide-affected terrain, we coded the variables in Fig. 16 according to the frequency at which they occur as part of the target variable's Markov blanket, where darker hues indicate more frequent occurrences.

Perhaps the most surprising aspect of the learned BNs is that only few of the predictors that have traditionally been invoked to explain landslide susceptibility are duly represented in the Markov blanket. These include mean annual precipitation (part of the MB in each run) – including some derivatives such as precipitation variability (either annual or monthly variation is part of the MB) – and mean local topographic relief (part of the MB in half of the runs).

Instead, predictors such as regionalized bedrock river sinuosity or short-term (10-year cumulative) surface uplift derived from a dense network of GPS stations seem to provide relevant information about landslide-affected terrain in Japan. Bedrock river sinuosity may reflect the ability of rivers to carve more pronounced meanders in rocks with closely spaced defects. Therefore, sinuosity could be linked to first order to important rock-mass properties that govern the abundance of landslides. However, the link to contemporary surface uplift is less clear. Many of Japan's currently subsiding areas are limited to low-relief forearc areas, which feature

fewer landslides – hence an indirect connection via topography seems plausible. Yet predictors such as mean elevation or bedrock channel steepness (as a proxy of fluvial erosion and undercutting of hillslopes) play largely subdued roles in the MB of the learned BNs. Also, the role of lithology seems to be of major importance for the landslide prediction. In our data, lithology is expressed by the fractions of different rock types outcropping in a given area, which form a highly interacting cluster. Here the information about accretionary complexes, i.e., heavily tectonized and welded remnants of former island arcs, is always part of the MB. Furthermore, it is either the fraction of plutonic, sedimentary, or volcanic rocks that is part of the MB.

The learned BN structures are counterintuitive compared to many other susceptibility models that traditionally emphasize hillslope inclination and topographic relief (Korup and Stolle, 2014). Further studies may wish to elucidate whether the dependences contained in the BNs are regional artifacts or valid on a larger scale. Nevertheless, our results illustrate that the BN approach may reveal novel and unexpected insights into regional landslide prediction by highlighting unusual links between predictor variables that other multivariate models may not show as clearly. What is equally important is that BNs underscore which predictors may yield sufficient predictive potential should others not be available.

6 Conclusions

The Bayesian network approach is a powerful framework to capture uncertainties and probabilistic elements in natural hazard assessments. We demonstrated its flexible applicability in seismic hazard, flood damage, and landslide susceptibility analyses. In addition, we discussed the handling of continuous data and incomplete observations, as well as the uncertainty about the model structure, i.e., challenges that may arise when BNs are learned from real-world data. Our suggested way of dealing with these problems is fully data driven and can thus easily be transferred to other domains.

Since the interest of most natural hazard assessment is in the prediction of a certain target variable, we compared the prediction performance of the BNs learned for the seismic hazard and flood damage application to currently used models. In both cases the BNs perform reasonable well. This is especially promising, since the BNs are designed to capture the joint distribution of all variables and thus put similar effort into the prediction of each variable, whereas alternative models focus on predicting the target variable solely. For a better prediction performance, we might think of different graphical models that share the focus on the target variable. These could include (tree-augmented) naive Bayes networks or an adapted score for the network learning that puts more weight on the target variable. Thus learned networks may also be more reliable in the identification of the variables relevant for the prediction, but will fail to capture the overall picture of dependence relations.

Working with BNs, we profit from several attractive properties inherent to the BN framework. No prior domain knowledge is required, as the DAG structure and parameters can be learned from data. Yet, if available, expert knowledge can be exploited via the prior term, which is part of the scoring function. Discovered (in)dependence relations help us to understand the underlying process and to identify (ir)relevant variables. An intuitive understanding is supported by the graphical representation of BNs, although the same data may produce different graphs with comparable performance. This highlights the potential for new insights into interactions between large sets of candidate predictors. The ability of BNs to predict from incomplete observations allows for hazard estimations at an early stage of an event. Using inference we can estimate missing values of a variable based on the observations of neighboring variables. The prediction can be updated as soon as new information becomes available about variables that are missing so far.

BNs capture the uncertainty and provide a probability distribution instead of a point estimate. Consequently they provide a valuable contribution on the basis of which decision making should be made. Moreover, BNs allow for inference, and thus they enable detailed examinations of specific scenarios. Bayesian networks may thus be used for improved communication between scientists and public authorities and may help in creating a better assessment of natural hazards that does not shy away from any uncertainties involved.

Appendix A:

A1   BN learning: MAP score

For a data-based BN learning we search for the pair (DAG, $\boldsymbol{\theta}$) that has the highest probability for the observed data and thus maximizes the joint posterior

$$\underbrace{P(\text{DAG}, \boldsymbol{\Theta}|\boldsymbol{d})}_{\text{posterior}} \propto \underbrace{P(\boldsymbol{d}|\text{DAG}, \boldsymbol{\Theta})}_{\text{likelihood}} \underbrace{P(\boldsymbol{\Theta}, \text{DAG})}_{\text{prior}},$$

which corresponds to the MAP score of the BN (Riggelsen, 2008).

Investigating the single components of the score, we find that the likelihood term $P(\boldsymbol{d}|\text{DAG}, \boldsymbol{\Theta})$ is the product of Eq. (1) for every independent sample; hence, for complete data, this is

$$P(\boldsymbol{d}|\text{DAG}, \boldsymbol{\theta}) = \prod_{l=1}^{|\boldsymbol{d}|}\prod_{i=1}^{k} p\left(x_i^{(l)}|\boldsymbol{x}_{\text{Pa}(i)}^{(l)}\right) = \prod_{i,x_i,\boldsymbol{x}_{\text{Pa}(i)}} \theta_{x_i|\boldsymbol{x}_{\text{Pa}(i)}}^{n(x_i,\boldsymbol{x}_{\text{Pa}(i)})}, \quad (A1)$$

where $x_i^{(l)}$ is the observed value of X_i in the lth record and $\boldsymbol{x}_{\text{Pa}(i)}^{(l)}$ is the corresponding observation of the parent set. This reduces the score to a function of $n(x_i, \boldsymbol{x}_{\text{Pa}(i)})$, the number of counts (occurrences) of the values $(x_i, \boldsymbol{x}_{\text{Pa}(i)})$ in the data.

The joint prior distribution decomposes as $P(\boldsymbol{\Theta}, \text{DAG}) = P(\boldsymbol{\Theta}|\text{DAG})P(\text{DAG})$. For the applications in this paper we assume that all DAGs are equally likely and consequently define $P(\text{DAG})$ to be uniform over the space of BN structures. $P(\boldsymbol{\Theta}|\text{DAG})$ we define as a non-informative prior, acting at the same time as a penalty term. For discrete data, this is a (product) Dirichlet distribution given by

$$P(\boldsymbol{\theta}|\text{DAG}) = \prod_{i=1}^{k}\prod_{\boldsymbol{x}_{\text{Pa}(i)}} \frac{\Gamma\left(\sum_{x_i} \alpha\left(x_i, \boldsymbol{x}_{\text{Pa}(i)}\right)\right)}{\prod_{x_i}\Gamma\left(\alpha\left(x_i, \boldsymbol{x}_{\text{Pa}(i)}\right)\right)} \prod_{x_i} \theta_{x_i|\boldsymbol{x}_{\text{Pa}(i)}}^{\alpha\left(x_i, \boldsymbol{x}_{\text{Pa}(i)}\right)-1}, \quad (A2)$$

where $\alpha(\cdot)$ are so-called hyperparameters, primarily governing the regularization (to avoid over-fitting).

MAP score extension for continuous variables
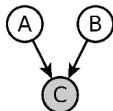
If continuous variables are contained in the data set, we learn a discretization, Λ, parallel to the BN. The here used extended MAP score corresponds to the joint posterior of BN and discretization:

$$P\left(\text{DAG}, \boldsymbol{\Theta}, \Lambda|\boldsymbol{d}^{\text{c}}\right) \propto P(\boldsymbol{d}^{\text{c}}|\text{DAG}, \boldsymbol{\Theta}, \Lambda)P(\text{DAG}, \boldsymbol{\Theta}, \Lambda).$$

Considering Eq. (4), we rewrite the equation above as

$$P(\text{DAG}, \boldsymbol{\Theta}, \Lambda|\boldsymbol{d}^{\text{c}}) \propto \underbrace{P\left(\boldsymbol{d}^{\text{c}}|\boldsymbol{d}, \Lambda\right)}_{\text{continuous data}} \underbrace{P(\boldsymbol{d}|\text{DAG}, \boldsymbol{\Theta}, \Lambda)}_{\text{likelihood (discrete)}} \underbrace{P(\text{DAG}, \boldsymbol{\theta}, \Lambda)}_{\text{prior}}.$$

The likelihood term is defined as in Eq. (A1), and the prior decomposes into $P(\boldsymbol{\Theta}|\text{DAG}, \Lambda)\,P(\Lambda|\text{DAG})\,P(\text{DAG})$.

DAG' for C					contribution to $s(t, (t, f))$
		C	A	B	
		t	t	-	$\frac{1}{2}$
		f	f	-	0
Estimate		t	t	f	1
$\hat{\theta}_{C=t|A=t, B=f}$:		-	t	f	0
$s(t, (t, f)) = \frac{7}{4}$		t	-	t	0
		t	-	-	$\frac{1}{4}$

Figure A1. Illustration of the calculation of $s(\cdot)$ used for the parameter estimation in DAG$'$. The graph on the left shows a DAG$'$ for the estimation of C conditioned on A and B. The three variables take the values t and f. An example data set is given in the table on the right together with the contribution for each record to s ($C=t$, ($A=t$, $B=f$)).

For $P(\Lambda|\text{DAG})$ and $P(\text{DAG})$, we assume uniform distributions analogue as for the original MAP score and we define $P(\boldsymbol{\Theta}|\text{DAG}, \Lambda)$ as product Dirichlet again (Eq. A2). For the continuous data term we define

$$P(\boldsymbol{d}^{\text{c}}|\boldsymbol{d}, \Lambda) = \prod_{i}\prod_{x_i}\left(\frac{1}{n(x_i)}\right)^{n(x_i)},$$

which corresponds to the assumption that all continuous observations are equally likely within the same interval. The joint posterior has a closed form as a function of $n(\cdot)$.

A2   Similar-cases approach

The estimation of a missing value as described in Sect. 4.2 requires, according to Eq. (7), the prediction of the parameter $\theta_{x_i|\boldsymbol{x}_{\text{Pa}(i)}}^{\text{DAG}'}$ from incomplete data. Instead of using the unobserved statistics $n(\cdot)$, we rely on counts of similar cases here. The statistics, $s(x_i, \boldsymbol{x}_{\text{Pa}(i)})$, are a weighted count of all records where $X_i = x_i$ and the observed part of the parents set $\boldsymbol{X}_{\text{Pa}(i)}$ matches with $\boldsymbol{x}_{\text{Pa}(i)}$. This means that we add 1 to $s(x_i, \boldsymbol{x}_{\text{Pa}(i)})$ for each record where $(x_i, \boldsymbol{x}_{\text{Pa}(i)})$ is fully observed. For each matching record with an incompletely observed parents set, we count the possible completions for the missing parents and add 1 divided by the number of possible completions to $s(x_i, \boldsymbol{x}_{\text{Pa}(i)})$; refer to Fig. A1 for an example. Finally we estimate the parameters

$$\hat{\theta}_{x_i|\boldsymbol{x}_{\text{Pa}(i)}}^{\text{DAG}'} = \frac{s\left(x_i, \boldsymbol{x}_{\text{Pa}(i)}\right)}{\sum_{x_i} s\left(x_i, \boldsymbol{x}_{\text{Pa}(i)}\right)},$$

fully defining the predictive distribution for X_i (Eq. 7).

References

Aguilera, P. A., Fernández, A., Fernández, R., Rumí, R., and Salmerón, A.: Bayesian networks in environmental modelling, Environ. Modell. Softw., 26, 1376–1388, doi:10.1016/j.envsoft.2011.06.004, 2011.

Bayraktarli, Y. Y. and Faber, M. H.: Bayesian probabilistic network approach for managing earthquake risks of cities, Georisk, 5, 2–24, doi:10.1080/17499511003679907, 2011.

Berkes, F.: Understanding uncertainty and reducing vulnerability: lessons from resilience thinking, Nat. Hazards, 41, 283–295, doi:10.1007/s11069-006-9036-7, 2007.

Blaser, L., Ohrnberger, M., Riggelsen, C., and Scherbaum, F.: Bayesian Belief Network for Tsunami Warning Decision Support, Lect. Notes. Artif. Int., 5590, 757–768, doi:10.1007/978-3-642-02906-6_65, 2009.

Blaser, L., Ohrnberger, M., Riggelsen, C., Babeyko, A., and Scherbaum, F.: Bayesian networks for tsunami early warning, Geophys. J. Int., 185, 1431–1443, 2011.

Bommer, J. and Scherbaum, F.: Capturing and Limiting Groundmotion Uncertainty in Seismic Hazard Assessment, Directions in Strong Motion Instrumentation, Nato Science Series: IV: Earth and Environmental Sciences, 58, 25–40, doi:10.1007/1-4020-3812-7_2, 2005.

Boore, D.: Simulation of ground motion using the stochastic method, Pure Appl. Geophys., 160, 635–676, 2003.003.

Bouckaert, R. R.: Bayesian belief networks: from construction to inference, Ph.D. thesis, University Utrecht, Netherlands, 1995.

Castelo, R. and Kocka, T.: On inclusion-driven learning of Bayesian networks, J. Mach. Learn. Res., 4, 527–574, 2003.

Chickering, D. M.: Optimal structure identification with greedy search, J. Mach. Learn. Res., 3, 507–554, 2002.

Elmer, F., Thieken, A. H., Pech, I., and Kreibich, H.: Influence of flood frequency on residential building losses, Nat. Hazards Earth Syst. Sci., 10, 2145–2159, doi:10.5194/nhess-10-2145-2010, 2010.

Fenton, N. and Neil, M.: Risk assessment and decision analysis with bayesian networks, CRC Press, Inc. Boca Raton, FL, USA, 2012.

Friedman, N.: Learning belief networks in the presence of missing values and hidden variables, Fourteenth International Conference on Machine Learning, July 1997, Nashville, TN, 125–133, 1997.

Friedman, N.: The Bayesian structural EM algorithm, Fourteenth conference on Uncertainty in artificial intelligence, 24–26 July 1998, Madison, WI, 129–138, 1998.

Friedman, N. and Koller, D.: Being Bayesian about network structure, Sixteenth conference on Uncertainty in artificial intelligence, 30 June–3 July 2000, Stanford, CA, 201–210, 2000.

Friedman, N., Goldszmidt, M., and Wyner, A.: Data analysis with Bayesian networks: a bootstrap approach, Fifteenth conference on Uncertainty in artificial intelligence, 30 July–1 August 1999, Stockholm, Sweden, 196–205, 1999.

Grêt-Regamey, A. and Straub, D.: Spatially explicit avalanche risk assessment linking Bayesian networks to a GIS, Nat. Hazards Earth Syst. Sci., 6, 911–926, doi:10.5194/nhess-6-911-2006, 2006.

Hoyt, P. J.: Discretization and Learning of Bayesian Networks using Stochastic Search, with Application to Base Realignment and Closure (BRAC), Ph.D. thesis, George Mason University, Fairfax, VA, 2008

Jensen, F. and Nielsen, T.: Bayesian Networks and Decision Graphs, Springer, New York, USA, 2001.

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

Korup, O. and Stolle, A.: Landslide Prediction from Machine Learning, Geol. Today, 30, 26–33, 2014.

Korup, O., Gorum, T., Hayakawa, Y.: Without power? Landslide inventories in the face of climate change, Earth Surf. Proc. Land., 37, 92–99, 2012.

Korup, O., Hayakawa, Y., Codilean, A. T., Matsushi, Y., Saito, H., Oguchi, T., Matsuzaki, H.: Japan's Sediment Flux to the Pacific Ocean Revisited, Earth-Sci. Rev., 135, 1–16, 2014.

Kuehn, N., Scherbaum, F., and Riggelsen, C.: Deriving empirical ground-motion models: Balancing data constraints and physical assumptions to optimize prediction capability, B. Seismol. Soc. Am., 99, 2335–2347, 2009.

Kuehn, N. M., Riggelsen, C., and Scherbaum, F.: Modeling the joint probability of earthquake, site, and ground-motion parameters using Bayesian networks, B. Seismol. Soc. Am., 101, 235–249, 2011.

Langseth, H. and Nielsen, T. D.: Parameter estimation in mixtures of truncated exponentials, 4th European Workshop on Probabilistic Graphical Models, 17–19 September 2008, Hirtshals, Denmark, 169–176, 2008.

Langseth, H., Nielsen, T. D., Rumí, R., and Salmerón, A.: Inference in hybrid Bayesian networks, Reliab. Eng. Syst. Safe., 94, 1499–1509, 2009.

Langseth, H., Nielsen, T. D., Rumí, R., and Salmerón, A.: Parameter estimation and model selection for mixtures of truncated exponentials, Int. J. Approx. Reason., 51, 485–498, 2010.

Little, R. and Rubin, D.: Statistical Analysis with Missing Data, vol. 4, Wiley, New York, 1987.

Liu, H., Hussain, F., Tan, C. L., and Dash, M.: Discretization: An Enabling Technique, Data Mining Knowl. Discov., 6, 393–423, 2002.

McGrayne, S. B.: The Theory that Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, & Emerged Triumphant from Two Centuries of Controversy, Yale University Press, Yale, 2011.

Merz, B., Kreibich, H., Schwarze, R., and Thieken, A.: Review article "Assessment of economic flood damage", Nat. Hazards Earth Syst. Sci., 10, 1697–1724, doi:10.5194/nhess-10-1697-2010, 2010.

Merz, B., Kreibich, H., and Lall, U.: Multi-variate flood damage assessment: a tree-based data-mining approach, Nat. Hazards Earth Syst. Sci., 13, 53–64, doi:10.5194/nhess-13-53-2013, 2013.

Monti, S. and Cooper, G. F.: A multivariate discretization method for learning Bayesian networks from mixed data, Fourteenth conference on Uncertainty in artificial intelligence, 24–26 July 1998, Madison, WI, 404–413, 1998.

Moral, S., Rumí, R., and Salmerón, A.: Mixtures of truncated exponentials in hybrid Bayesian networks, in: Symbolic and Quantitative Approaches to Reasoning with Uncertainty, edited by: Benferhat, S. and Besnard, P., Springer, Berlin, Heidelberg, 156–167, 2001.

Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann, San Francisco, CA, USA, 1998.

Plapp, S. T.: Wahrnehmung von Risiken aus Naturkatastrophen: Eine empirische Untersuchung in sechs gefährdeten Gebieten Süd- und Westdeutschlands, edited by: Werner, U., Verlag Versicherungswirtschaft, Karlsruhe, Germany, 2003 (in German).

Riggelsen, C.: MCMC learning of Bayesian network models by Markov blanket decomposition, in: Machine Learning: ECML 2005, Springer, Berlin, Heidelberg, 329–340, 2005.

Riggelsen, C.: Learning Bayesian networks from incomplete data: an efficient method for generating approximate predictive distributions, SIAM International conf. on data mining, 20–22 April 2006, Bethesda, Maryland, 130–140, 2006.

Riggelsen, C.: Learning Bayesian networks: a MAP criterion for joint selection of model structure and parameter, Eighth IEEE International Conference on Data Mining, 15–19 December 2008, Pisa, Italy, 522–529, 2008.

Schnell, R., Hill, P. B., and Esser, E.: Methoden der empirischen Sozialforschung, 6th Edn., Walter De Gruyter, New York, 535 pp., 1999.

Schroeter, K., Kreibich, H., Vogel, K., Riggelsen, C., Scherbaum, F., and Merz, B.: How useful are complex flood damage models?, Water Resour. Res., 50, 3378–3395, doi:10.1002/2013WR014396, 2014.

Song, Y., Gong, J., Gao, S., Wang, D., Cui, T., Li, Y., and Wei, B.: Susceptibility assessment of earthquake-induced landslides using Bayesian network: a case study in Beichuan, China, Comput. Geosci., 42, 189–199, 2012.

Straub, D.: Natural hazards risk assessment using Bayesian networks, 9th International Conference on Structural Safety and Reliability, 20–23 June 2005, Rome, Italy, 2005.

Tanner, M. and Wong, W.: The calculation of posterior distributions by data augmentation, J. Am. Stat. Assoc., 82, 528–540, 1987.

Thieken, A. H., Müller, M., Kreibich, H., and Merz, B.: Flood damage and influencing factors: new insights from the August 2002 flood in Germany, Water Resour. Res., 41, W12430, doi:10.1029/2005WR004177, 2005.

Vogel, K., Riggelsen, C., Merz, B., Kreibich, H., and Scherbaum, F.: Flood damage and influencing factors: a Bayesian network perspective, 6th European Workshop on Probabilistic Graphical Models, 19–21 September 2012, Granada, Spain, 347–354, 2012.

Vogel, K., Riggelsen, C., Scherbaum, F., Schröter, K., Kreibich, H., and Merz, B.: Challenges for Bayesian network learning in a flood damage assessment application, 11th International Conference on Structural Safety and Reliability, 16–20 June 2013, New York, NY, 2013.