**Natural Hazards
and Earth System
Sciences**

Open Access

Discussions

The application of Bayesian networks in natural hazard analyses

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

Institute of Earth and Environmental Sciences, University of Potsdam, Germany

[*]*Invited contribution by K. Vogel, recipient of the Outstanding Student Poster (OSP)
Award 2012.*

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

NHESSD

1, 5805–5854, 2013

**The application of
Bayesian networks in
natural hazard
analyses**

K. Vogel et al.

Title Page

Abstract	Introduction
Conclusions	References
Tables	Figures

|◄ | ►|
◄ | ►
Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Abstract

In natural hazards we face several uncertainties due to our lack of knowledge and/or the intrinsic randomness of the underlying natural processes. Nevertheless, deterministic analysis approaches are still widely used in natural hazard assessments, with the pitfall of underestimating the hazard with potentially disastrous consequences.

In this paper we show that the Bayesian network approach offers a flexible framework for capturing and expressing a broad range of different uncertainties as those encountered in natural hazard assessments. Although well studied in theory, the application of Bayesian networks on real-world data is often not straightforward and requires specific tailoring and adaption of existing algorithms. We demonstrate by way of three case studies (a ground motion model for a seismic hazard analysis, a flood damage assessment, and a landslide susceptibility study) the applicability of Bayesian networks across different domains showcasing various properties and benefits of the Bayesian network framework.

We offer suggestions as how to tackle practical problems arising along the way, mainly concentrating on the handling of continuous variables, missing observations, and the interaction of both. We stress that our networks are completely data-driven, although prior domain knowledge can be included if desired.

1 Introduction

Natural hazards such as earthquakes, tsunamis, floods, landslides or volcanic eruptions all share – despite of differing causes, triggers, and effects – many of the same model- and decision theoretic questions. The underlying natural processes are often complex, while the number of potential influencing factors is large. The single and joint effects of the driving forces are not necessarily completely understood, potentially introducing a large degree of uncertainty, which impacts any quantitative analysis. Additionally, the observation on the basis of which inference is made is often sparse,

inaccurate and incomplete, adding yet another layer of uncertainty on top. Thus various sources of uncertainty accumulate and interact in a non-trivial fashion. Indeed, this is one of the reasons why probabilistic approaches are often avoided. Such probabilistic approaches often seem intimidating given the large number of parameters required for such a proceeding, where the analytical tools needed for a rigorous handling of uncertainties are unknown or at least not readily available.

Deterministic 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 would be wrong. 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 are still 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).

In this paper we consider Bayesian networks (BNs) that combine probability theory with graph theory, thus allowing for an intuitive, consistent, and rigorous way of quantifying uncertainties. The (in-)dependencies between the involved variables relevant to a particular hazard domain are translated into a graph structure that enables improved understandings and direct insights into the relationships and workings of a natural hazard system. Moreover, as BNs capture the joint distribution of all variables of a given domain, they may be used for expressing any conditional probability distribution of interest, thereby helping to answer quantitative questions on specific scenarios or process response chains.

In recent years BNs have already successfully been employed in a wide range of earth sciences, including automatic classifications of seismic signals (e.g. Riggelsen et al., 2007), tsunami early warning (e.g. Blaser et al., 2009, 2011), Probabilistic Seismic Hazard Analyses (e.g. Kuehn et al., 2011), and earthquake induced landslide

susceptibility (e.g. Song et al., 2012). Here we additionally highlight three applications of BNs in natural hazard assessments that illustrate the flexibility of BNs. Alongside we discuss problems that arise when BNs are learned from real world data. The handling of continuous variables and incomplete observations is the chief problem here. Most approaches we use are derived from existing ideas and standard approaches, but their combination is challenging, since the individual methods are highly interacting.

In the next section we give an introduction into BNs, which is followed by several applications in natural hazards research. In Sect. 3 we develop a seismic ground motion model based on a synthetic data set, which serves to showcase some typical BN properties. We demonstrate an option to deal with continuous variables, making no prior assumptions about their distributional family. In Sect. 4 data that were collected after the 2002 and 2005/06 flood events in the German Elbe and Danube catchments are used to learn a BN for flood damage assessments. This incomplete data set requires a treatment of missing observations, which is especially challenging in combination with continuous variables. A last application 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, thus revealing important variable interactions that are often overlooked in landslide studies. This application further illustrates the model uncertainty related to BN learning. Conclusions and scope for future work perspectives are presented in Sect. 6.

2 Bayesian networks

We operate in the world of probability theory and adhere to the corresponding probability axioms. We treat all random quantities pertaining to a particular hazard domain, including the covariates, as true random variables, each associated with a (conditional distribution), $p(\cdot|\cdot)$. The interactions between those variables, be it measurables, observations or otherwise, therefore really boils down to the question of

The application of Bayesian networks in natural hazard analyses

K. Vogel et al.

⏮ · ⏭

◀ · ▶

Back · Close

Full Screen / Esc

how the distributions of those random variables interact (at all times in accordance with the axioms of probability theory). A BN is a convenient framework for capturing such interactions, both from an intuitive, but also from a computational point of view. We refer to Koller and Friedman (2009) for a more detailed description of BNs in general and will restrict ourselves here to a few key aspects of the formalism.

2.1 Properties and benefits

As previously mentioned, BNs combine probability theory with graph theory, representing all involved variables, $\boldsymbol{X} = \{X_1, \ldots, X_k\}$, as nodes in a directed acyclic graph, DAG; these variables will coincide with the specific variables of a hazard domain of interest. The arcs of the DAG point from the variables in the parent set, $\boldsymbol{X}_{\mathrm{Pa}(i)}$, to X_i (see Fig. 1), stating that X_i directly depends on (is influenced by, is affected by, ...) $\boldsymbol{X}_{\mathrm{Pa}(i)}$. The probability distribution of X_i is defined conditional on its parents set, $p\left(X_i | \boldsymbol{X}_{\mathrm{Pa}(i)}\right)$. Formally, a BN for $\boldsymbol{X}$ is a pair (DAG, $\boldsymbol{\theta}$), the structure, DAG and the parameters, $\boldsymbol{\theta} = \bigcup\{\theta_{x_i | x_{\mathrm{Pa}(i)}} = p\left(x_i | x_{\mathrm{Pa}(i)}\right)\}$,[1] which describes a joint distribution factorizing as,

$$P\left(\boldsymbol{X} | \mathrm{DAG}, \boldsymbol{\theta}\right) = \prod_{i=1}^{k} p\left(X_i | \boldsymbol{X}_{\mathrm{Pa}(i)}\right). \tag{1}$$

Figure 2 is an example of a simple BN. Although not immediately clear from Eq. (1), a BN is characterized by many attractive properties which we may profit from in a natural hazard setting; here we mention a few:

Prop. 1 – Graphical representation: The interactions of the variables of the entire "system" are encoded through the DAG. The BN structure thus provides information

---

[1]Here we assume discrete variables for which the set of parameters corresponds to the conditional (point) probabilities for each combination of states. For continuous variables the design of the parameters depends on the particular densities $p(\cdot|\cdot)$.

about the underlying processes and the way various variables communicate and share "information" as it spreads around the network. Formally speaking the DAG encodes independence assumptions between variables.

Prop. 2 – Use prior knowledge: The intuitive interpretation of a BN makes it possible to define the BN beforehand based on prior knowledge; alternatively it may be learned from data, or even a combination of the two (cast as Bayesian statistical problem; see later for details): pose a prior BN and update it based on observations.

Prop. 3 – Identify relevant variables: By learning the BN from data we may identify the variables that are (according to the data) relevant. If several "islands" or isolated single nodes (not connected) appear, then it indicates that these variables potentially may be irrelevant.

Prop. 4 – Capture uncertainty: Uncertainty can easily be propagated throughout the network, between any nodes/variables; we effectively compute/estimate probability distributions rather than single point estimates.

Prop. 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/all conditional distribution(s) of interest, in any directions (including forensic and inverse reasoning), e.g. we may ask for an observed damage, what is a likely intensity of the damage causing event. A detailed example for reasoning is given in Sect. 4.3.

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

Prop. 6 – Use incomplete observations: During predictive inference (i.e., computing a conditional distribution) incomplete observations of data do not pose a problem to BNs. By the virtue of the probability axioms, it merely impacts the overall uncertainty involved.

In the course of the paper we will refer to these 1–6 properties to make clear 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, and consequently the properties listed above at times seem rather theoretical/academic. In the sections to come we will by way of examples shed some light on how typical natural hazard problems can be formulated around BNs. We take a data-driven stance, and the aim is thus to learn BNs from observations collected; the next section will introduce some key aspects related to inducing BNs.

2.2 Learning Bayesian networks

Data based BN learning can be seen as an exercise in finding a pair (DAG, $\boldsymbol{\theta}$) which according to the BN decomposition in Eq. (1) could have been "responsible for generating the data". For this we traverse the space of pairs looking for the BN which maximizes a given fitness score; 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 appear in literature derived from different theoretical principles and ideas. We opt for a Bayesian approach to learn BNs[2]. This has several advantages and allows us to elegantly combine prior knowledge and data, i.e. we have encoded our prior belief on the BN space (DAG, $\boldsymbol{\Theta}$) as $P(\text{DAG}, \boldsymbol{\Theta})$, we observe data $\mathbf{d}$ and obtain the revised distribution $P(\text{DAG}, \boldsymbol{\Theta}|\mathbf{d})$ based upon which we select the most probable BN. More specifically, we may rewrite the joint posterior as,

$$\underbrace{P(\text{DAG}, \boldsymbol{\Theta}|\mathbf{d})}_{\text{posterior}} \propto \underbrace{P(\mathbf{d}|\text{DAG}, \boldsymbol{\Theta})}_{\text{likelihood}} \underbrace{P(\boldsymbol{\Theta}, \text{DAG})}_{\text{prior}}. \tag{2}$$

The prior distribution allows us to compensate sparse data, artifacts, bias, etc. but also allows us to assign domain specific prior preferences to certain BNs before seeing data (Prop. 2). In the following applications we work with a non-informative prior, which

---

[2]Note that BNs are not necessarily to be interpreted from a Bayesian statistical perspective.

at the same time acts as a penalty term that regularizes the DAG complexity and avoids over-fitting. The likelihood term decomposes according to Eq. (1). Detailed descriptions for prior and likelihood term are given in Appendix A1.

Using the BN MAP (Bayesian Network Maximum A Posteriori) score we can maximize the right side of Eq. (2) (the product of the prior and the likelihood as given above) as to obtain the "best" BN given data; see Riggelsen (2008) for details with regard to search strategy, hyper-parameter selection, etc.

In the following section we learn a ground motion model, which is used in probabilistic seismic hazard analysis, as a BN; the used data set is synthetically generated. This section serves as an illustration of the BN formalism "in action" and will also present some theoretical and practical problems along with potential solutions in the context of BN learning. In the sections thereafter we will use real data.

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 certain ground motion at a certain site within a certain 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 log-normally 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. 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). Is no reliable

prior knowledge available, we work with a non-informative prior and the learned graph structure provides insight into the dependence structure of the variables and helps for a better understanding of the underlying mechanism. 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.

3.1 The data

The event situation is described by the predictor variables $X_{-Y} = \{M, R, \text{SD}, Q_0, \kappa_0,$ $V_\text{S}30\}$, which are explained in Table 1. We generate a synthetic data set consisting of 10 000 records. The ground motion parameter, Y, is the horizontal peak ground acceleration (PGA). It is generated by a so called *stochastic model* (Boore, 2003), which distorts the shape of a random time series according to physical principles and obtains 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 1).

Since the stochastic model does not have nice analytical properties and its usage is non-trival and time consuming, surrogate models, which describe the stochastic model in a more abstract sense (e.g. regressions), are usually used in PSHA. 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 a completely data-driven and distribution-free learning. In the following subsection we describe an automatic discretization, which is part of the BN learning procedure and takes the dependencies 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

To keep the information loss during the discretization small the number of the intervals and their boundaries have to be chosen carefully. A fine-grained discretization will result in a sparsely connected BN due to regularization constrains, while a rough discretization may not provide the required degree of resolution.

Instead of discretizing as pre-processing step to the Bayesian network learning, we use a multivariate discretization approach, that accounts for the BN structure. The "optimal" discretization, Λ, thus depends on the DAG of the BN and the observed (continuous) data, $\mathbf{d}^c$. Analogue to Sect. 2.2, we again cast the problem in a Bayesian framework searching for the combination of (DAG, $\boldsymbol{\theta}$, Λ) that has the highest posterior probability given continuous data,

$$P\left(\text{DAG},\Theta,\Lambda|\mathbf{d}^c\right) \propto P\left(\mathbf{d}^c|\text{DAG},\Theta,\Lambda\right) P\left(\text{DAG},\Theta,\Lambda\right). \tag{3}$$

Λ is defined by a set of interval boundary points for all variables and bins the original data $\mathbf{d}^c$ yielding $\mathbf{d}$. Expanding on an idea by Monti and Cooper (1998) we assume that all communication/information floss between the variables can be captured by their discrete representations. For given $\mathbf{d}$ and Λ is $\mathbf{d}^c$ consequently independent of DAG and Θ and the likelihood for observing $\mathbf{d}^c$ (for a given discretization, network structure and parameters) can be written as,

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

and Eq. (3) decomposes into

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

The likelihood term is 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, by Λ defined interval have the same probability. 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.

3.2.1 Learned ground motion model

Since we generated the data ourselves, we know which (in-)dependencies the involved variables should adhere to; this is expected to be reflected in the BN DAG we learn from the synthetic data. Due to data construction are the predictors variables $M, R, \text{SD}, Q_0, \kappa_0, V_S 30$ independent from each other and PGA depends on the predictors. Figure 3 shows the dependence structure of the variables. The converging edges at PGA indicate that the predictors become conditional dependent for given PGA. This means for given PGA they carry information about each other, e.g. 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 to the reconstruction of the dependencies from the data, which is done in the following.

The network we found to maximize $P(\text{DAG}, \boldsymbol{\Theta}, \Lambda | \mathbf{d}^c)$ for the 10 000 synthetic, seismic data records is shown in Fig. 4. The corresponding found discretization is plotted in Fig. 5, 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. For discrete/discretized variables the number of parameters needed for the definition of $p(X_i | \boldsymbol{x}_{\mathrm{Pa}(i)})$ in Eq. (1) corresponds to the number of possible state combinations for $(X_i, \boldsymbol{x}_{\mathrm{Pa}(i)})$. Taking the learned discretization shown in Fig. 5, the BN of the data generating process (Fig. 3) is defined by 3858 parameters, 3840 needed alone for the description of $p(\mathrm{PGA} | M, R, \mathrm{SD}, Q_0, \kappa_0, V_{\mathrm{S}}30)$. 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. 4) consists of only 387 parameters and still captures the most relevant dependencies.

Figure 6 shows the lnPGA 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 given PGA SD contains information about R and vice versa. The dependencies between PGA and the remaining predictors are much less distinctive, such that the conditional dependencies between the predictors are negligible and the edges can be reversed for the benefit of parameter reduction. The connection to $V_{\mathrm{S}}30$ is even neglected at all, 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 expresses the independencies (not the dependencies) between the variables (Prop. 2). This means each (conditional) independence statement encoded in the DAG must be true, while encoded dependence relations must not hold per se. In turn this implies that each dependence holding for the data should be encoded in the DAG. The learning approach applied here fulfils the task quite well, detecting the relevant dependencies, 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. 5). PGA is split into 8 intervals, distance and stress drop into 4 and 5, while the other variables consist of only 2 to 3 intervals.

3.3 Approximation of continuous distributions with mixtures of exponentials

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 (Prop. 5) of all involved variables, primary focus is in this context on a single variable, for which the accuracy of the prediction made with a BN is limited by the resolution of the discretization learned. For the BN shown above, the discretization of the target variable into 8 intervals enables a quite precise approximation of the continuous distribution, but this is not necessarily always the case. Complex network structures and smaller data sets used for the BN learning lead to a coarser discretization of the variables. To still enable precise estimations 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 to use *mixtures of truncated exponentials* (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, marginalization (Langseth et al., 2009). To approximate a conditional distribution $p(X_i|\boldsymbol{x}_{\mathrm{Pa}(i)})$ with MTEs, we partition the domain $\Omega_{(X_i,\boldsymbol{x}_{\mathrm{Pa}(i)})}$ into hypercubes $D_1,\ldots,D_L$ and define the density within each hypercube, D_l, such that it follows the form

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

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.

3.3.1 Prediction performance

We conduct a 10-fold cross validation to evaluate the prediction performance of the BN compared to the regression approach: the complete dataset is divided into 10 disjoint subsamples of which in each trial one is defined as test set, 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

$$f\left(\boldsymbol{X}_{-Y}\right) = a_0 + a_1 M + a_2 M \cdot \ln \mathrm{SD} + (a_3 + a_4 M)\ln\sqrt{a_5^2 + R^2} + a_6 \kappa R + a_7 V_{\mathrm{S}}30 + a_8 \ln \mathrm{SD}$$

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

We compare the regression approach in terms of prediction performance to the BN with discretized variables and with MTE approximations. For this purpose we consider for each model the conditional density distributions of $\ln \mathrm{PGA}$ given the other variables and inspect how much probability each density assigns to the real $\ln \mathrm{PGA}$ value. For the regression approach the conditional density follows a normal distribution, $\mathcal{N}\left(f\left(\boldsymbol{X}_{-Y}\right), \sigma^2\right)$, while it is defined via the DAG and the parameters $\boldsymbol{\theta}$ using the BN-models. Table 2 shows for each test set the conditional density of the observed $\ln \mathrm{PGA}$ averaged over the individual records. An other measure for the prediction performance is the mean squared error of the estimates for $\ln \mathrm{PGA}$ (Table 3). We define the mean values of the conditional densities as point estimates for $\ln \mathrm{PGA}$. For the regression model the estimate i.e. corresponds to $f(\boldsymbol{x}_{-Y})$.

Even though the discretization of $\ln \mathrm{PGA}$ is relative precise in the discrete BNs (8 intervals in each trial, except for the first trial, where $\ln \mathrm{PGA}$ is split into 7 intervals),

the MTE-approximation of the continuous conditional distributions improves the prediction performance of the BN. Still it does not entirely match the precision of the regression function. Anyhow, the prediction performances are in 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 totally data driven. Further profits the regression approach in this example from the fact that the target variable ($\ln \mathrm{PGA}$) 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 into all directions (Prop. 5), 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 (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 variety of factors (Thieken et al., 2005), stage-damage functions are still widely used. This is due to the fact that 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 2005/06 flood events in the Elbe and Danube catchments in Germany (Thieken et al., 2005; Elmer et al., 2010) offer a unique opportunity to learn about the driving forces of flood damage from a BN perspective. Figure 7 shows the investigated catchments. The data set consists of 1135 records of discrete and continuous variables describing the flooding situation, building and household characteristics, precaution, warning and emergency measures and building damage. Table 4 lists the 29 considered variables allocated to their domains. In Sect. 3.2 we already 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, and 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.

4.2 Handling of incomplete records

To learn the BN we again maximize the joint posterior for the given data, Eq. (3). The calculation of the joint posterior requires for each variable X_i the number of counts for each combination of states for $(X_i, x_{\mathrm{Pa}(i)})$. 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 principled 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. Friedman, 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 for predicting a missing variable X_i based on the observations of its Markov Blanket (MB), $\boldsymbol{X}_{\mathrm{MB}(i)}$. The Markov Blanket identifies the set of variables directly influencing X_i, i.e. parents, children and parents of children. For illustration see Fig. 8a. 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; although maybe not immediately clear, this follows directly from the BN factorization of the joint distribution. However, for predictive approximation purposes we choose to always ignore the impact from outside the MB. Hence, the predictions for X_i based on observed data factorizes according to the DAG in Fig. 8a, as,

$$P(X_i|\boldsymbol{X}_{\mathrm{MB}(i)}, \boldsymbol{\theta}, \mathrm{DAG}) \propto \theta_{X_i|\boldsymbol{x}_{\mathrm{Pa}(i)}} \prod_{j \in \mathrm{Ch}(i)} \theta_{X_j|\boldsymbol{x}_{\mathrm{Pa}(j)}}, \tag{6}$$

where Ch(i) are the variable indices of the children of X_i. The prediction of X_i requires inference in the BN (albeit very simple) where correct estimates of all $\boldsymbol{\theta}$ are assumed. However, these in general can not be given without resorting to iterative procedures. To avoid this we first define a slightly modified version of the predictive function based on DAG′ which is derived from DAG by allocating edges, pointing to X_i, to each variable in its MB; see Fig. 8. The resulting DAG′ preserves all dependencies given in DAG and

can alternatively be used for the prediction of X_i,

$$P(X_i | \boldsymbol{x}_{\mathrm{Pa}(i)}^{\mathrm{DAG}'}, \boldsymbol{\theta}^{\mathrm{DAG}'}, \mathrm{DAG}') \stackrel{\mathrm{def}}{=} \theta_{X_i | \boldsymbol{x}_{\mathrm{Pa}(i)}}^{\mathrm{DAG}'}. \tag{7}$$

For this predictive distribution we need to estimate the parameters $\theta_{X_i | \boldsymbol{x}_{\mathrm{Pa}(i)}}^{\mathrm{DAG}'}$. However, this we can approximately estimate without using iterative methods by a *similar cases* approach[3], described in Appendix A2. A detailed description for the generation of the predictive distribution is given in Riggelsen (2006); Vogel et al. (2013).

To avoid confusions, we want to stress that the MBs of the individual variables change during the BN-learning procedure. The prediction of the missing values has to be updated for each change of the DAG.

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 on 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. 9a). Applying 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. 9b). These relations, especially the first one, match with experts expectations and speak for an improvement of the learned BN structure.

---

[3]Note that more parameters are required for the derived predictive distribution Eq. (7), but now at least *all* influencing variables are considered jointly, whereas in the former case Eq. (6) we would have estimated the parameters based on the parent sets only if a non-iteraive similar cases approach would have been used.

Using the graphical representation (Prop. 1), as mentioned in Sect. 2.1 the learned DAG (Fig. 9b) 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. Especially the domains "precaution" and "flood parameters" are densely connected with the building damage and should be included in any damage assessment (Prop. 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 requirement for complete observations does not hold for BNs (Prop. 6). The prediction of the building damage, e.g., depends only on the variables of its Markov Blanket (marked with a bold frame in Fig. 9). Is the observation of the Markov Blanket variables 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 use of many variables thus now provides additional knowledge and proves to be an advantage.

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

The prediction of the relative building loss is of primary interest in flood damage assessments. Similar to our proceeding for the ground motion modeling, 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 type, building quality, contamination and private precaution. While sdf and FLEMOps+r give point estimates, the BN gives us a distribution for rloss and thus reveals the uncertainty of the prediction (Prop. 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 5 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 to model the whole distribution of the target variable and to conduct the prediction even though not all variables are observed.

4.3.2 Example for inference: impact of precaution

As an example of reasoning (Prop. 5), we consider the effect of precaution on the building loss. Figure 10 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. 10a: all other variables are unknown and summed out) and for a specific flood event (Fig. 10b: $7.5\,\mathrm{m} \leq$ water depth $< 96.5\,\mathrm{m}$; $82\,\mathrm{h} \leq$ duration $< 228\,\mathrm{h}$; $1 \leq$ 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 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 just considered example and ask for

a given observed damage in a specific or general event situation, what is a likely state of precaution. Forensic reasoning might be of interest e.g. for insurance companies.

5 Landslides

So far we assumed the existence of a unique model that explains the data best, but in practical problems there may be many models almost as good as the best, i.e. explaining 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 work on landslides, which are an other 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 are suitable for sufficiently predicting 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 %, which raises the question why landslides still continue to cause massive losses despite this seemingly high predictive accuracy (Korup and Stolle, 2013). Moreover, success rates do not show any significant increase over the last ten years regardless of the number of landslide data or predictors used. 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

We use a database of nearly 300 000 landslide deposit areas that derive mostly from deep-seated slow-moving failures and earthflows throughout Japan. This inventory was compiled by the National Research Institute for Earth Science and Disaster Prevention NIED (http://lsweb1.ess.bosai.go.jp/gis-data/index.html), and is one of the largest of its kind available. We have compiled a number of geological, climatic, and topographic metrics 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 have been used in modified form in other studies (Table 6). While all of these candidate predictors may be physically related to slope instability, part of this choice of predictor composition is intentionally arbitrary in order to learn more about its effects on BN learning and structure. The final data set used for the BN learning consists of 553 records, where $\sim 0.4\%$ of the data are missing.

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 focussing 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. receive a similar score according to Eq. (2).

An additional source of uncertainty can be attributed to 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 (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 ten 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 11 gives a summarized representation of the BN DAG structures. The frequency with which an edge between two variables is learned is encoded by the widths (by scaling it accordingly). Despite of the difference 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 get a score 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 the 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 of (or rather thanks to) the DAG structural differences we can glean some instructive insights from the learned BNs. 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 by other interactions. To get an understanding which variables are most relevant for predicting landslide-affected terrain, we coded the variables in Fig. 11 by the frequency with that 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 been invoked traditionally 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 regionalised bedrock river sinuosity or short-term (10 yr 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 link 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 learned BNs part of the MB. Also, the role of lithology, expressed by the fractions of different rock types outcropping in a given area form a highly interacting cluster, where the information about accretionary complexes, i.e.

heavily tectonized and welded remnants of former island arcs, seems to be of major importance for the landslide prediction. Alternatively it is either the fraction of plutonic, sedimentary or volcanic rocks that is part of the MB.

The learned BN structures are counter-intuitive compared to many other susceptibility models that traditionally emphasize hillslope inclination and topographic relief. Further studies must show, if the found dependencies are regional artifacts or valid on a larger scale. Nevertheless, these findings 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. Equally important, 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. Alongside 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 transfered 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

⏮ | ⏭

◀ | ▶

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

(tree augmented) Naive Bayes networks or an adapted score for the network learning that would weight more 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. Learning the DAG structure and parameters from data, BNs require no prior domain knowledge. Yet if available, it can be exploited via the prior term, which is part of the scoring function. The 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 based on observations from related variables. The prediction can be updated as soon as new information about so far missing variables becomes available.

Capturing the uncertainty and providing a probability distribution instead of a point estimate, BNs provide a valuable contribution on the basis of which decision making should be made. Allowing for inference they also enable detailed examinations of specific scenarios. Thus Bayesian Networks may thus be used for an improved communication between scientist and public authorities and help for a better evaluation of hazards in general.

Appendix A

A1 BN learning: MAP score

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

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

which corresponds to the MAP score of the BN.

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

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

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

The joint prior distribution decomposes as $P(\boldsymbol{\Theta}, \mathrm{DAG}) = P(\boldsymbol{\Theta}|\mathrm{DAG})P(\mathrm{DAG})$. For the applications in this paper we assume that all DAGs are equally likely and consequently define $P(\mathrm{DAG})$ to be uniform over the space of BN structures. $P(\boldsymbol{\Theta}|\mathrm{DAG})$ 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

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

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

MAP score extension for continuous variables

Are continuous variables 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:

$$P(\mathrm{DAG}, \mathbf{\Theta}, \Lambda | \mathbf{d}^c) \propto P(\mathbf{d}^c | \mathrm{DAG}, \mathbf{\Theta}, \Lambda) P(\mathrm{DAG}, \mathbf{\Theta}, \Lambda).$$

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

$$P(\mathrm{DAG}, \mathbf{\Theta}, \Lambda | \mathbf{d}^c)$$

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

The likelihood term is defined as in Eq. (A1) and the prior decomposes into $P(\mathbf{\Theta} | \mathrm{DAG}, \Lambda) P(\Lambda | \mathrm{DAG}) P(\mathrm{DAG})$. For $P(\Lambda | \mathrm{DAG})$ and $P(\mathrm{DAG})$ we assume uniform distributions analogue as for the original MAP score and we define $P(\mathbf{\Theta} | \mathrm{DAG}, \Lambda)$ as product Dirichlet again Eq. (A2). For the continuous data term we define

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

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 $\theta^{\mathrm{DAG}'}_{X_i | \mathbf{x}_{\mathrm{Pa}(i)}}$ from incomplete data. Instead of using

the unobserved statistics $n(\cdot)$, we rely on counts of similar cases here. The statistics, $s(x_i, \boldsymbol{x}_{\text{Pa}(i)})$, is 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 $\boldsymbol{x}_{\text{Pa}(i)}$; for each record where $(X_i, \boldsymbol{x}_{\text{Pa}(i)})$ is fully observed we add 1 to $s(x_i, \boldsymbol{x}_{\text{Pa}(i)})$. For an incomplete observed parents set we count the possible completions of the record and add 1/number-of-possible-completions to $s(x_i, \boldsymbol{x}_{\text{Pa}(i)})$. Refer to Fig. 12 for an example. We now estimate the parameters

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

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

References

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. 5807

Blaser, L., Ohrnberger, M., Riggelsen, C., Babeyko, A., and Scherbaum, F.: Bayesian networks for tsunami early warning, Geophys. J. Int. 185, 1431–1443, 2011. 5807

Boore, D.: Simulation of ground motion using the stochastic method, Pure Appl. Geophys., 160, 635–676, 2003. 5813

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. 5827

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. 5820, 5824

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. 5821

Friedman, N.: The Bayesian structural EM algorithm, Fourteenth conference on Uncertainty in artificial intelligence, 24–26 July 1998, Madison, WI, 129–138, 1998. 5821

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. 5827

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. 5827

Koller, D. and Friedman, N.: Probabilistic Graphical Models: Principles and Techniques, The MIT Press, 2009. 5809

Korup, O. and Stolle, A.: Landslide Prediction from Machine Learning, Geology Today, submitted, 2013. 5825

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. 5807

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. 5817

Langseth, H., Nielsen, T. D., Rumí, R., and Salmerón, A.: Inference in hybrid Bayesian networks, Reliab. Eng. Syst. Safe., 94, 1499–1509, 2009. 5817

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. 5818

Little, R. and Rubin, D.: Statistical Analysis with Missing Data, vol. 4, Wiley, New York, 1987. 5820

The application of Bayesian networks in natural hazard analyses

K. Vogel et al.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◄ | ►

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

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. 5819, 5823

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. 5814

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. 5814, 5817

Riggelsen, C.: MCMC learning of Bayesian network models by Markov blanket decomposition, in: Machine Learning: ECML 2005, Springer, Berlin, Heidelberg, 329–340, 2005. 5827

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. 5821, 5822

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. 5812

Riggelsen, C., Ohrnberger, M., and Scherbaum, F.: Dynamic Bayesian networks for real-time classification of seismic signals, 11th European conference on Principles and Practice of Knowledge Discovery in Databases, 17–21 September 2007, Warsaw, Poland, 565–572, 2007. 5807

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. 5808

Tanner, M. and Wong, W.: The calculation of posterior distributions by data augmentation, J. Am. Stat. Assoc., 82, 528–540, 1987. 5821

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. 5819, 5820

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, Granada, Spain, 19–21 September 2012, 347–354, 2012. 5815, 5822

The application of Bayesian networks in natural hazard analyses

K. Vogel et al.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

|◄ | ►|

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

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 Reliability, 16–20 June 2013, New York, NY, 2013. 5815, 5822

The application of Bayesian networks in natural hazard analyses

K. Vogel et al.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◄ | ►

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Table 1. Variables used for the ground motion model.

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

Table 2. Results of a 10-fold cross validation to test the prediction performance of the BN (discrete and MTE approximations of the conditional distributions) and the regression approach. It shows for each trial the average of the conditional density for the observed $\ln$ PGA value.

Averaged conditional density		
$BN_{discrete}$	BN_{MTE}	Regression
0.237	0.320	0.331
0.240	0.297	0.329
0.239	0.298	0.331
0.218	0.255	0.323
0.216	0.260	0.339
0.222	0.257	0.339
0.215	0.252	0.332
0.243	0.317	0.330
0.212	0.249	0.328
0.243	0.315	0.331

Table 3. Results of a 10-fold cross validation to test the prediction performance of the BN (discrete and MTE approximations of the conditional distributions) and the regression approach. It shows for each trial the mean squared error of the predicted $\ln \text{PGA}$.

Mean squared error		
BN_{discrete}	BN_{MTE}	Regression
1.021	0.749	0.663
1.197	0.963	0.680
1.082	0.821	0.673
1.262	0.951	0.723
1.201	0.851	0.629
1.298	1.059	0.625
1.297	1.077	0.672
1.149	0.713	0.701
1.343	1.161	0.692
1.169	0.841	0.666

Table 4. Variables used for the flood damage assessment. C: continuous, O: ordinal, N: nominal.

Variable	Scale and range
Flood parameters	
Water depth	C: 248 cm below ground to 670 cm above ground
Inundation duration	C: 1 to 1440 h
Flow velocity indicator	O: 0 = still to 3 = high velocity
Contamination indicator	O: 0 = no contamination to 6 = heavy contamination
Return period	C: 1 to 848 yr
Warning and emergency measures	
Early warning lead time	C: 0 to 336 h
Quality of warning	O: 1 = receiver of warning knew exactly what to do to 6 = receiver of warning had no idea what to do
Indicator of flood warning source	N: 0 = no warning to 4 = official warning through authorities
Indicator of flood warning information	O: 0 = no helpful information to 11 = many helpful information
Lead time period elapsed without using it for emergency measures	C: 0 to 335 h
Emergency measures indicator	O: 1 = no measures undertaken to 17 = many measures undertaken
Precaution	
Precautionary measures indicator	O: 0 = no measures undertaken to 38 = many, efficient measures undertaken
Perception of efficiency of private precaution	O: 1 = very efficient to 6 = not efficient at all
Flood experience indicator	O: 0 = no experience to 9 = recent flood experience
Knowledge of flood hazard	N (yes/no)
Building characteristics	
Building type	N (1 = multifamily house, 2 = semi-detached house, 3 = one-family house)
Number of flats in building	C: 1 to 45 flats
Floor space of building	C: 45 to 18000 m^2
Building quality	O: 1 = very good to 6 = very bad
Building value	C: 92 244 to 3 718 677 €
Socio-economic factors	
Age of the interviewed person	C: 16 to 95 yr
Household size, i.e. number of persons	C: 1 to 20 people
Number of children (< 14 yr) in household	C: 0 to 6
Number of elderly persons (> 65 yr) in household	C: 0 to 4
Ownership structure	N (1 = tenant; 2 = owner of flat; 3 = owner of building)
Monthly net income in classes	O: 11 = below 500€ to 16 = 3000€ and more
Socioeconomic status according to Plapp (2003)	O: 3 = very low socioeconomic status to 13 = very high socioeconomic status
Socioeconomic status according to Schnell et al. (1999)	O: 9 = very low socioeconomic status to 60 = very high socioeconomic status
Flood loss	
rloss – loss ratio of residential building	C: 0 = no damage to 1 = total damage

Table 5. Mean squared errors of a 5-fold cross-validation, where the relative building loss is predicted with the three models stage-damage function, FLEMOps + r and Bayesian Networks.

	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

Table 6. Variables used for 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]
Regionalised 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 levelling 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).
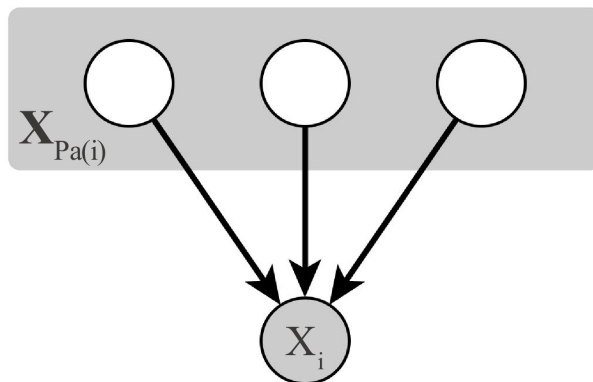
Fig. 1. Illustration of a parent set $x_{\mathrm{Pa}(i)}$ of X_i in a BN.
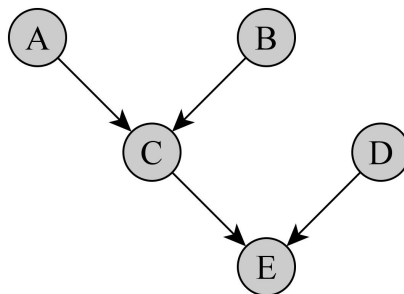
$$P(A,B,C,D,E) = \mathsf{p}(E|C,D)\,\mathsf{p}(D)\,\mathsf{p}(C|A,B)\,\mathsf{p}(B)\,\mathsf{p}(A)$$

Fig. 2. Example for the decomposition of the joint distribution according to a DAG.
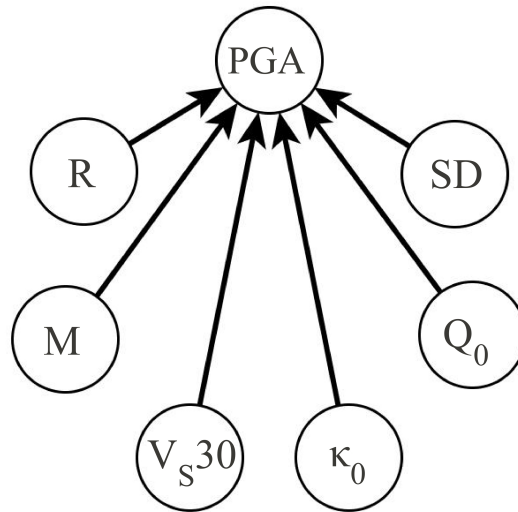
Fig. 3. Theoretic BN, capturing dependencies of the data generating model.
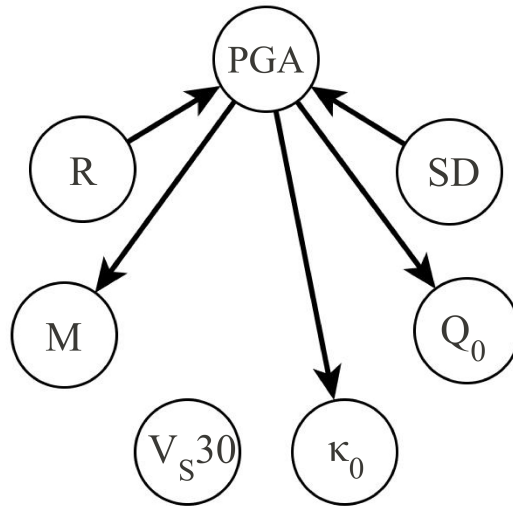
Fig. 4. BN learned from the generated synthetic data.

The application of Bayesian networks in natural hazard analyses

K. Vogel et al.

Full Screen / Esc
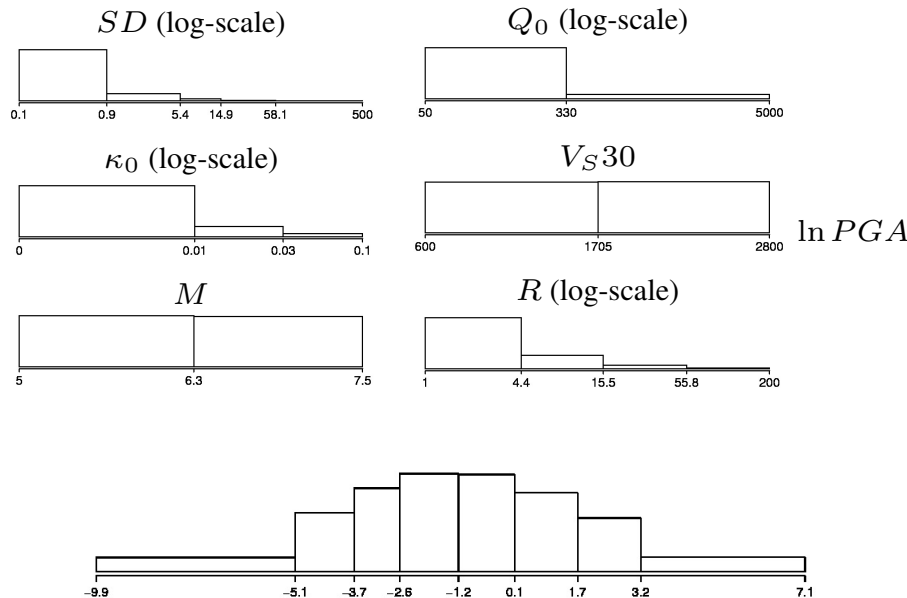
Fig. 5. Marginal distribution of the variables, discretized according to the discretization found for the BN in Fig. 4. The number of intervals per variable ranges from 2 to 8.
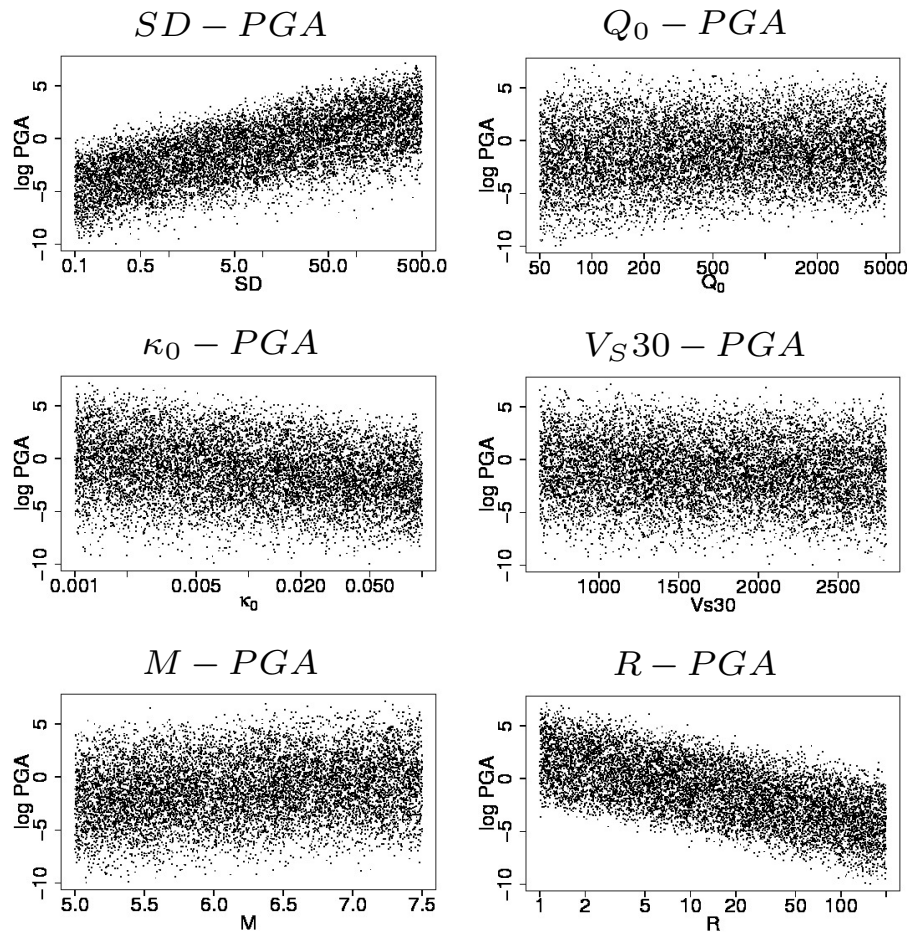
Fig. 6. Dependencies between the predictor variables and ln PGA.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◄ | ►

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion
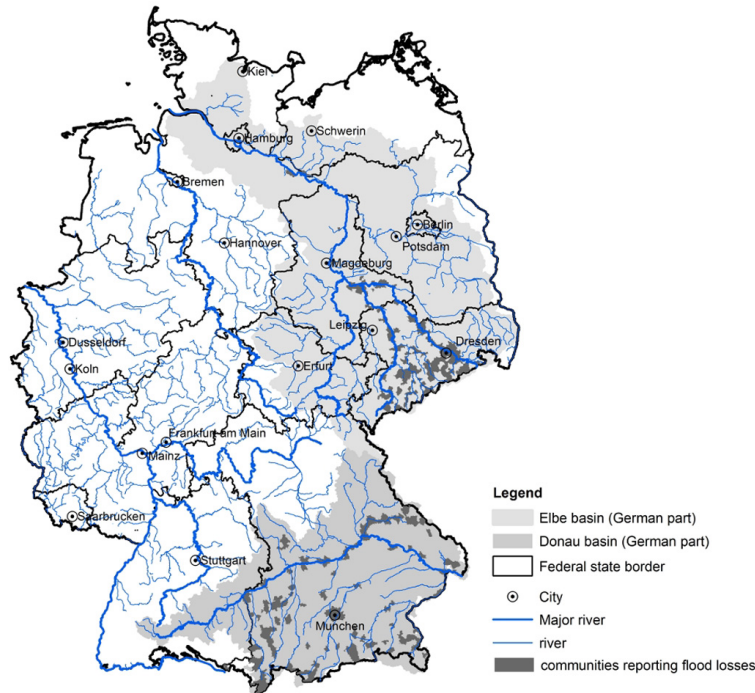
Fig. 7. Investigated catchments and location of communities reporting losses from the 2002, 2005 and 2006 flood events in the Elbe and Danube catchments.

Full Screen / Esc

Printer-friendly Version

Interactive Discussion
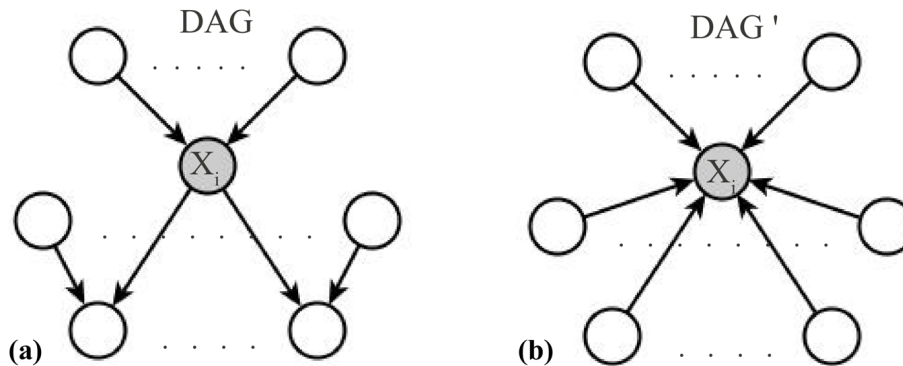
Fig. 8. **(a)** Markov Blanket of X_i; **(b)** new model DAG$'$ derived from DAG (all edges present in the Markov Blanket of X_i in DAG are directed towards X_i in DAG$'$).
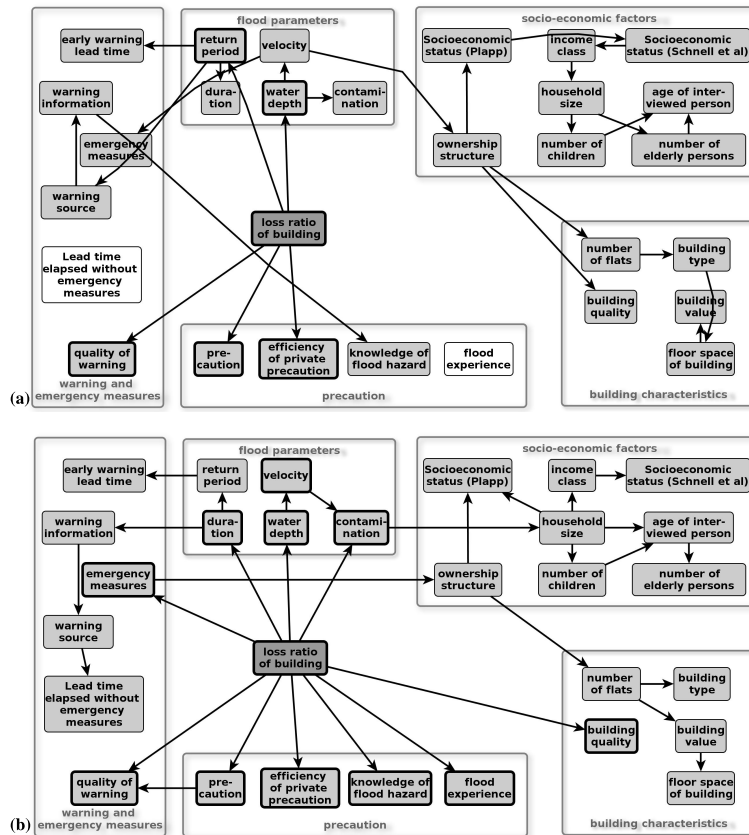
Fig. 9. BNs learned for flood damage assessment. **(a)** Missing values are randomly replaced. **(b)** Missing values are estimated using the Markov Blanket predictor. Nodes with a bold frame belong to the Markov Blanket of relative building loss and are thus assumed to have direct impact on the flood loss.
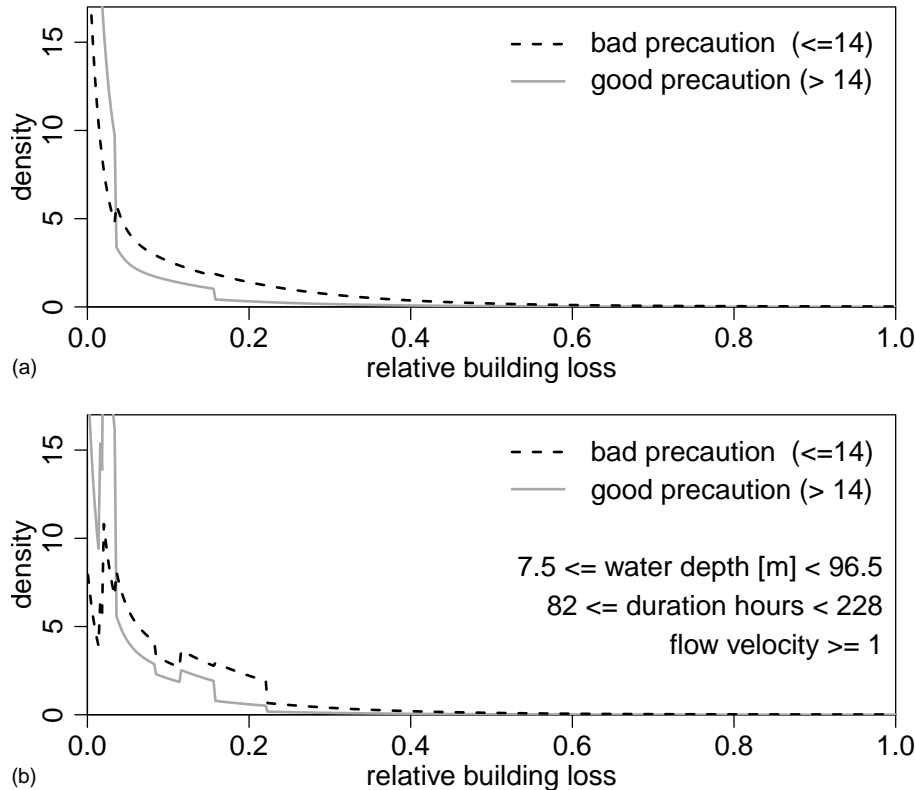
Fig. 10. (a) Distribution of the building loss depending on the precaution. Flood specific parameters as well as other parameters are unknown and summed out. **(b)** Distribution of the building loss depending on the precaution for a specific flood situation, where water depth, duration and flow velocity are known. Other parameters are unknown and summed out.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

|◄ | ►|

◄ | ►

Back | Close

Full Screen / Esc

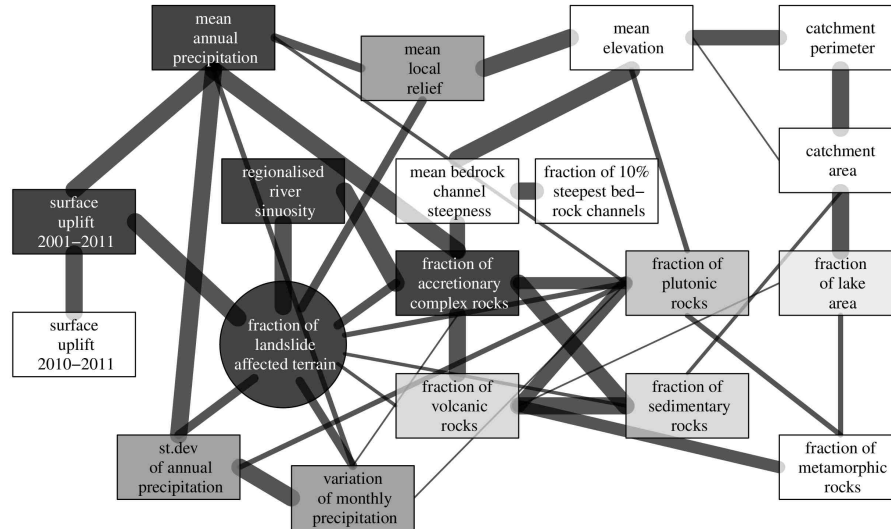Printer-friendly Version

Interactive Discussion

Fig. 11. Summary of ten learned network structures, 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 by the frequency with that they occur as part of the Markov Blanket of fraction of landslide affected terrain (circular node shape), where darker hues indicate more frequent occurrence.

DAG' for C

Estimate

$\hat{\theta}_{C=t|A=t,B=f}$:

$s(t,(t,f)) = \frac{7}{4}$

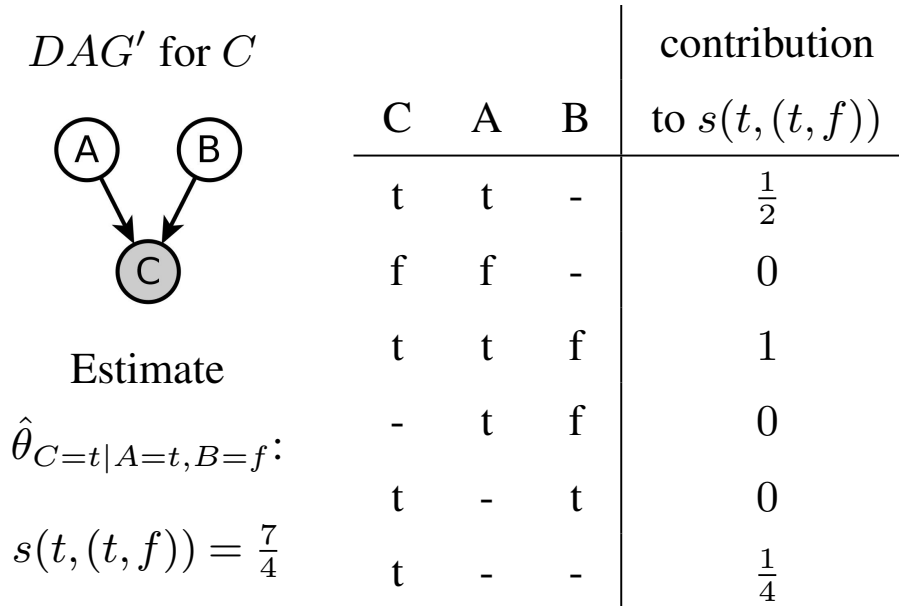| C	A	B	contribution to $s(t,(t,f))$
t	t	-	$\frac{1}{2}$
f	f	-	0
t	t	f	1
-	t	f	0
t	-	t	0
t	-	-	$\frac{1}{4}$

Fig. 12. Illustration for 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 depending A and B. The three variables take the values t and f. An exemplary data set is given in the table on the right together with the contribution for each record to $s(C = t,(A = t,B = f))$.