Nat. Hazards Earth Syst. Sci. Discuss., 1, C2360–C2363, 2014
www.nat-hazards-earth-syst-sci-discuss.net/1/C2360/2014/

Natural Hazards and Earth System Sciences
Discussions

Interactive comment on "The application of Bayesian networks in natural hazard analyses" *by* K. Vogel et al.

K. Vogel et al.

kvog@geo.uni-potsdam.de

Received and published: 22 January 2014

Referee: I find the title misleading, as this paper is focusing entirely on learning Bayesian networks (BNs). Based on the title, I expected a broader overview on BNs, in particular addressing also the substantial body of work that has been carried out on modeling natural hazards with BNs based on an a-priori understanding of the (in-)dependencies among the involved variables. Also, there is little discussion of how the BN is ultimately used in the process (with the exception of the landslide example). For this reason, the title of the paper should be slightly modified to narrow down the scope, e.g. to "Bayesian Network Learning for Natural Hazard Analyses" or "Applications of Bayesian Network Learning for Natural Hazard Analyses".

Authors: Good point. We refrain here from providing an exhaustive review of the literature in this field, and limit ourselves to the process of learning the BNs given a number of scenarios that often arise when confronted with natural hazards data. "Bayesian network learning for natural hazard analyses " would be a better choice indeed. We believe that this process of learning is far from non-trivial and often incompletely explained in the existing literature, and take this as ample motivation for our contribution.

Referee: Overall, the authors use references rather sparingly. Since the paper, due to it covering a wide range of applications, can only briefly touch upon certain aspects, providing additional references would be helpful. This applies in particular to the theory, where additional references should be given whenever the details of the implementation are not provided in the paper (eg. Markov blanket, non-informative prior used for P(theta|DAG), etc). This would strongly help the reader who is interested in implementing the procedures.

Authors: OK. This can be improved

Referee: In line 28, please clarify why you think that such tools are not known or available. After all, probabilistic methods are used by many.

Authors: With "analytical tools needed for a rigorous handling of uncertainties" we mean tools, that take all different sources of uncertainties (intrinsic randomness, model uncertainty, uncertainty in observations, ...) into account. Probabilistic methods that are mentioned in the literature usually focus on one (or at least limited) kind(s) of uncertainty. Also BNs do not solve this issue entirely, as we have mentioned, for example, in section 5.2.

Referee: In line 74 it is stated that "The interactions between those variables, be it

measurables, observations or otherwise, therefore really boils down to the question of how the distributions of those random variables interact". I think see what the authors want to say (that all information is contained in the joint probability distribution), but the sentence can easily be misunderstood by people that are less fond of probability. Think of reformulating this sentence.

Authors: OK, we have rephrased this statement to: "Therefore, all pertinent information on the interactions between those variables, be it measurables, observations or otherwise, is contained in their joint probability distribution."

Referee: In line 83/84 you write that the meaning of the arcs is "depends on (is influenced by, is affected by)". All these statements imply causality. However, the learned BN structures are not necessarily causal; they just encode a dependence structure that – hopefully – is consistent with the true but (following your philosophy) unknown causal structure of the problem.

Authors: We agree about the misleading formulation. The sentence could be changed to: "The arcs of the DAG point from the variables in the parent set, X_Pa(i) , to X_i (see Fig. 1), indicating a direct dependency between X_i and X_Pa(i)." We would not go into more detail about causality at this point, since it is picked up later on in context with the ground motion example.

Referee: Line 173: the "a so called stochastic model" sounds a bit weird to me. A large number b of different stochastic models are used throughout PSHA, so referring to a specific model by this name does not make sense. Please state more clearly with which model you create the data.

Authors: We have rewritten this to: "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. "

Referee: Paragraph starting with line 250: This is an important point. I suggest making an example to illustrate this to the reader.

Authors: OK.

Referee: Second paragraph of section 5: What is the definition of success rate in landslide prediction? Are these statistical predictions, and of what kind? And what scale? (BTW, I strongly agree with the critical view that the authors have on claimed successes.)

Authors: The success rate in the majority of studies is simply defined as the percentage of correctly identified (or "predicted") locations that were subject to slope instability in the past. In other words, the success rates cited refer to the "true positives" rate obtained from various machine learning schemes applied to studying landslide susceptibility. Obviously, this metric is not the most optimal to characterise a model's predictive skill, precision, and sharpness. The types of statistical methods are manifold, and so is the scale of the underlying, although most studies have focused on areas encompassing several tens to thousands of square kilometres.

Thanks for the language corrections. We have changed these accordingly.

Interactive comment on Nat. Hazards Earth Syst. Sci. Discuss., 1, 5805, 2013.