Nat. Hazards Earth Syst. Sci. Discuss., 3, C2400–C2406, 2015
www.nat-hazards-earth-syst-sci-discuss.net/3/C2400/2015/

Natural Hazards and Earth System Sciences
Discussions

Interactive comment on "A spatial Bayesian network model to assess the benefits of early warning for urban flood risk to people" *by* S. Balbi et al.

Anonymous Referee #1

Received and published: 30 November 2015

General comment:

The paper applies a Bayesian network approach to study the potential benefits of an improved Early Warning System (EWS), i.e. in term of damages to people, in the study area of the lower Sihl valley (Switzerland). I judge the suggested consideration of human vulnerability as important and little studied aspect of flood damages. Since values as the ability to cope, the capacity to adapt, etc are difficult to measure, the study has to deal with many uncertainties. Further uncertainties result from model uncertainty. Bayesian networks are an adequate approach for the study, since they allow to model the related uncertainties and to provide information about the uncertainties in the re-

sults. Yet, the paper reveals some inconsistencies in the application of the approach and has some weaknesses in the documentation. At times the structure of the paper is difficult to follow, due to the consideration of the different modules of the model and jumping between them. A sketch of the modules, how they interact with each other and which methods are applied in which modules, might help to keep track. Eventhough there could be more references to previous/following section (we will discuss this in more detail in section . . .). The same applies to the described methods. The concept of BNs is described in "2.2. Methods", but the section misses the information about the algorithm/methods applied in this study. This information in turn is provided in "2.3 Data and model components". Some steps remain to be just briefly addressed. E.g. in section 2.3. last passage I have no idea how this was done. What are the deterministic and probabilistic models and how are they coupled? How are the BN modules developed (and which modules)? On the other hand the description of study area and its specific characteristics is quite precise (even though I do not see, where these characteristics are captured in the model). In general graphics of the overall BN and of the subnetworks would improve the understanding.

Specific Comments:

The abstract claims that the approach provides estimates of model uncertainty and probability distributions of all outputs. Even though the authors emphasize the importance of providing uncertainties, the uncertainties are mostly ignored in the discussion of the results. Instead the reduction of fatalities and injuries for the improved scenario are provided in single numbers. Those numbers are not very reliable, considering the different sources of uncertainty integrated in the model. I.e. the reported reduction of fatalities by 75% is based on 4 estimated fatalities in the baseline scenario compared to 1 estimated fatality in the improved scenario. How large are the uncertainties related to these numbers? Even small variations lead to large effects in the fraction of the two compared scenarios.

The studied improved EWS assumes "maximum theoretical effectiveness ": to 100%

NHESSD

3, C2400–C2406, 2015

completely sufficient in reliability, lead time and scope. How realistic is such a scenario or what would be realistic improvements?

Section 2.1. and 3.2 gives a detailed description of Zurich's hotspots in case of a major flood event (e.g. the Sihl River flowing directly under the heavily frequented main station) and describes the opportunity to increase the buffering capacity of the Sihl Lake by releasing water (without passing through the power plant) at least one day ahead of the serious event. Are these special characteristics of the study area included in the model? If yes, how?

The evaluation of the existing (baseline) EWS is based on 4 expert interviews. Some more information (e.g. in the appendix) about the provided information/questionnaire would be nice. I do not understand how the provided percentages are extracted from the answers. How is a percentage of 1% possible if 4 experts are questioned. Considering the aspect of uncertainty, I am especially interested in the consistence of the expert opinions?

Could you be more precise about the single steps of the method? E.g. p. 6621, l. 16-17: "Hazard, vulnerability, and exposure are integrated into a single func- tion of risk using Bayesian networks (Bns)." How is this done? How does the resulting BN look like?

The description of the BN method (section 2.2) has some weaknesses, leaving the impression that the authors are not very familiar with the BN concept. p. 6622, l. 1-2: Input nodes are not necessarily nodes without parents (and vice versa). The term "prior probability" is used in a wrong context. Prior probability expresses knowledge prior to the observation of data. It is not the probability distribution of the input nodes. p. 6622 l. 5 - . . . : "BNs can be constructed through expert opinion or by learning the conditional probability distributions from the data." To me it is not clear, if the authors refer to the construction of the BN (including learning the graph structure) or only to learning the conditional probabilities. The provided reference (Vogel et al. 2012) deals with

NHESSD

3, C2400–C2406, 2015

structure learning, yet in other passages I get the impression, that only the conditional probabilities are estimated. p. 6621 l. 26-27: BNs do no necessarily reflect causalities, that is only the case for causal networks. p. 6622 l. 1: The probability distributions in a BN are not necessarily defined over a finite number of states, e.g. if continuous distributions are considered. Yet, in most applications discrete variables with a finite number of states are considered. p. 6621, l. 22: More a formal issue: Actually in a BN the INdependencies are represented in the graph.

Further in section 2.3 again it is not clear if yon learn the structure of the BN or only the conditional probabilities. On page 6625 l. 28 you write “causal structure”, which indicates you learn only the probabilities. Yet, in the next sentence you write, you use the PC learning algorithm, which is an algorithm to learn the structure. Besides, Sprites et al (2000), which you cite in that context, writes about the PC algorithm “on sample data the procedure takes unnecessary risks”. Why do you use it anyhow? In general, please be more specific about your proceeding. How do you determine the graph structure (expert knowledge?)? Which method do you use to estimate the conditional probabilities? How do you include expert knowledge? Show some results.

Concerning the hazard Bayesian module (section 2.3), I do not see, in which sense the module is Bayesian. The determination of the hazard rate looks quite deterministic to me. Can you comment on that? Further I do not understand to which purpose it is necessary to discretize (and thus loose information) at that point. An easy example could help to understand the calculation of the hazard rate. Which values are inserted into the equation (e.g. for a depth of 80cm is the value 80 inserted or a value, that corresponds to the class; which values correspond to which class? Considering the example on page 6625, l. 15 (depth: 1st class, velocity 2nd class, debris factor), I would calculate a hazard factor of 1*(2+0.5)+1=3.5, which corresponds to a major and not a moderate hazard. For illustration it would be nice, if a hazard map for the study area could be provided.

Considering the data collected from the expert interviews for the hazard and vulnera-

NHESSD
3, C2400–C2406, 2015

bility modules, I have some doubts concerning their informative value. In the questionnaire the different levels of hazard and vulnerability are explained by a specific example. Yet, the described example is only one representative of the considered class. For a different specification of the same hazard or vulnerability class (e.g. a moderate hazard represented by a higher water depth, but lower velocity and no debris factor) the expert's evaluation of the situation might change. Further, the expert is asked to check his answers for consistency. In that way the expert is motivated to tune his answers accordingly. Instead the consistence of the expert's answers should be checked by an independent person. Even in the 2nd case chances are high, that experts knowingly or unknowingly, manipulate their answers in order to be consistent. That does not mean, that the answers are reliable as well. The questionnaires are used "to create a larger representative data set through bootstrapping" (page 6625, l. 26) Could you be more precise here? How is the bootstrapping conducted/ how did you sample? Are the answers of experts discretized? If yes, how and when (before or after bootstrapping)? What do you accomplish by bootstrapping? You do not receive new information by sampling from given data. How do you hope to improve your results by that proceeding? How do you avoid to just replicate the provided answers, which will result in an illusive certainty about the derived estimates.

I do not really understand, what is done in section 3.1. Is the vulnerability completely module taken from the KULTURisk consortium or are there modifications in the current study? What are preference weights (p. 6627, l.16)? How are they used? Why is a single most likely outcome (and not a distribution) considered (p. 6627, l. 21)? How are the probabilities adjusted to represent reasonable probability distribution (p. 6627, l. 25)? Are the distributions provided in figure 3 realistic (e.g. probabilities for fraction of disabled people is 1/3 for below 5%, 5-15% and above 15% each; similar for old people and foreigners).

The proceeding in the sensitivity study is unknown to me. Maybe you could provide a reference. What is an acceptable sensitivity (p.6629, l.1)? To my understanding

NHESSD

3, C2400–C2406, 2015

the study provides information about the effects of the studied input variables on the vulnerability. The early warning parameters do NOT? appear to be very sensitive (p. 6629, l. 2). What is than the purpose of that paper?

Results and conclusions: Despite the mentioned importance of capturing uncertainties, uncertainties are hardly considered in the in the last sections. E.g. table 3 provides numbers of injuries and fatalities for the studied districts. Those numbers are provided in integers, thus I assume the correspond to the mode of the distribution (not the mean value = expected value). I.e. in the consideration of fatalities the conclusion referred from these numbers can be very misleading and unreliable. There is no information about the uncertainty provided. Instead the reduction of fatalities by 75% is mentioned at least 3 times in the paper, without any mentioning of uncertainties. In my opinion, statements like "with a probability of x% we have to expect more than y fatalities in the baseline scenario/improved scenario" would be more justified.

The probabilistic approach exceeds the estimated number of injured people by 30% in comparison to the deterministic approach and the number of fatalities by over 50% (p. 6633, l. 22-23). I would not call that a match.

Minor issues:

The figures are hard to read. Quality should be improved. Figure 4. a) the color scale is not well chosen. It is difficult to distinguish the different levels of blue. What does high and low mean in numbers?

p. 6621, l. 19: I suggest either "to ESTIMATE the actual number" or "to compute the EXPECTED number"

p. 6621 l. 27: grammer: "the considered factors ARE expressed"

Formulations in the questionnaire: question 5.4. "H and V being equal": probably it is meant H and V are fixed, not H=V?

NHESSD

3, C2400–C2406, 2015

Interactive comment on Nat. Hazards Earth Syst. Sci. Discuss., 3, 6615, 2015.

NHESSD

3, C2400–C2406, 2015