Nat. Hazards Earth Syst. Sci. Discuss., 3, C2801–C2814, 2015
www.nat-hazards-earth-syst-sci-discuss.net/3/C2801/2015/

Natural Hazards and Earth System Sciences Discussions
Open Access

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.

S. Balbi et al.

stef.balbi@gmail.com

Received and published: 23 December 2015

Please find below the detailed responses (point by point) and attached a .pdf with the initial track changes in the manuscript. Many thanks to the reviewer for the very useful comments.

1. A sketch of the modules, how they interact with each other and which methods are applied in which modules, might help to keep track. 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

Full Screen / Esc
Printer-friendly Version
Interactive Discussion
Discussion Paper

passage I have no idea how this was done. a. What are the deterministic and probabilistic models and how are they coupled? b. How are the BN modules developed (and which modules)? ——————————————————

—————— We acknowledge that sometimes the flow of the article is a bit tortuous and we'll do our best to improve readability by trying to better link the sections. Regarding the request for a sketch, in our view Figure 1 could explain how the modules interact and thus cover this aspect. For not making the graph even more crowded than it is, we decided to explain further characteristics with text. In Section 2.3, Data and model components we explained that: 1. Hazard (H) is a Bayesian model developed out of an equation with 3 inputs (depth, velocity and debris factor). 2. Vulnerability (V) is a Bayesian network developed by elicitation of experts knowledge. (since it's not trained on data we analyze its behavior in Section 3.1). 3. Exposure contains population density including people in the buildings. Point 1 (H) and 2 (V) are combined with a simple V structure Bayesian Network (trained with the data provided by 25 experts' opinions as explained in Section 2.4) that produces 4 output categories representing probabilities of getting affected (these are in fact probability distributions in each pixel of the study area) and Point 3 provides the total number of receptors that could be affected by those probabilities, so the last passage of section 2.3 is just a multiplication that happens in each pixel. We attached the Bayesian Networks in .net format for transparency (in the short comment). We find that an additional figure could be redundant (adding Very little to figure 1 and 3 combined), however if the editor requires it, we are available to provide a figure (or a table) like the attached one.

2. 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 reviewer is right. As mentioned in Section 3.2 Simulated spatial results (e.g. figure 4) are produced both for the mean and for the CV of the probability distribution produced by the BN model, however in the tables (3 and 4) the aggregated results are computed using the mean only. The reason is that while in the first case the emphasis is on communicating expected results and uncertainty per each pixel of the study area, in the tables the emphasis is on the change between scenarios aggregated per district. Additionally there is the issue of aggregating CV per district which is not in the scope of this paper. So the idea is: we present the output and the uncertainty in spatial explicit terms for 1 scenario and 1 output (baseline-injuries), we show that the uncertainty is there but doesn't undermine results (also see the sensitivity analysis) , so we present trade-offs in aggregate terms per district in the tables. We changed the manuscript according to this clarification.

3. How realistic is 100% Improved EWS scenario or what would be realistic improvements? The improved EWS needs a major investment in the computer systems and modeling but more importantly some contractual agreements with information providers such as cell phone companies. ———————————————————————— In short, this might not be very realistic in the first glance as it depends on variety of technical aspects, but we have considered a theoretical situation to prove the benefits of EWS as a very relevant nonstructural risk mitigation measure for the type of receptors under study. So, in our view, what is important is the marginal change from a situation A to a situation B in the spectrum of goodness of the EWS, and not point B itself. For example, we could argue that the same "% of improvement" at the lower end of the spectrum could return even more benefits (e.g. passing from no EWS to a decent one). Finally, even if right now 100% improved EWS might not be very realistic, in coming years it could be with the improvements in technology, better computers, models, data collection etc.

4. Are the special characteristics of Zurich Hotspots in the study area are included in the model? If yes, how? ————————————————

———————————— The model being spatially explicit, hotspots were described to calculate the results so yes they are considered in the model both in terms of vulnerability and exposure. Hotspots' exposure captures the particular people ' s density in the 2 different scenarios (day/night), and vulnerability variables are also spatially explicit as described in section 2.3. If the reviewer is asking if we took into account the feedback of using the storing capacity of the Sihl Lake, triggered by the EWS, on the hazard module, the answer is no, because the vulnerability model is not dynamically linked to the hazard. The hazard module is static and representing a 300 years event.

5. 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? ——————————————————————————- The 1% is forced to avoid having a non observable state (as evidence) with the probability being 0. Basically the questions were consistent with the nodes' state variables. e.g. What do you think about EWS reliability? (with a brief definition of the discrete states (high/moderate/low). A 75% answer means 3 of 4 experts were agreeing on e.g. scope being moderate and one believing in another state.

6. 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 function of risk using Bayesian networks (Bns)." How is this done? How does the resulting BN look like? ——————————————————————————
H and V are the input nodes of a simple BN that produces 3 types of probabilities of being affected by the flood (death, injured or PTSD) in each pixel of the study area. Exposure is just the number of people being present in the study area. (See Fig. 1 in this document)

7. 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 ———————————————————————————- We agree with the reviewer on the terminology here. We deleted that statement that was causing some confusion.

8. 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 structure learning, yet in other passages I get the impression, that only the conditional probabilities are estimated. p. 6621 l. 26-27: ——————————————————————————— We refer to both types of learning structure and/or conditional probabilities. Vogel et al. (2012) is cited to give an example of learning (structure) in the context of flood vulnerability.

9. 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. ———————————————————————————- We fixed the terminology according to the reviewer suggestions

10. 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. ——————————————————————————————— We had our hypothesis about the structure (Hazard > probability of Injury < Vulnerability). We used the PC algorithm as one of the possible methods in GeNIe with the added value of learning the causal structure, so that we could confirm our hypothesis as well as learning the CPTs.

11. 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. ——————————————————————————— Inspired by a function (Eq1) that is used in the literature, we discretized the Bayesian module that could be combined with the vulnerability Bayesian module into a simple V structure trained as described above. We agree that there is a loss of information but it is marginal compared to the gained simplicity when dealing with the surveyed experts. In practice, we wanted to make the model coherent with the questionnaire so that we could also test different levels of hazard's depth, velocity and DF in a second stage. The example the reviewer is mentioning is not calculated correctly. In the questionnaire, moderate hazard was described as "the flood depth is marginal (e.g. < 0.5m), but the water velocity is significant for an average person (e.g.

NHESSD

3, C2801–C2814, 2015

> 2ms−1) and there is some debris factor". In our Hazard BN that would mean Depth is low, Velocity is moderate and DF is present. The leaf node Hazard is a probability distribution: 9% low and 91% moderate as in the picture attached. The equation, assuming for example depth=0.3 and velocity =2.5, would be : HR = 0.3*(2.5 + 0.5) +1 = 1.9; not 3.5 Our approach is consistent with the classes of Table 5 in Ronco et al. (2015) [Hydrol. Earth Syst. Sci., 19, 1561–1576, 2015] except that it is made probabilistic. We can provide the Hazard map if that is necessary or refer to Ronco et al. (2015).

12. Considering the data collected from the expert interviews for the hazard and vulnerability 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. ———————————————————— The first part of the questionnaire explains the modules and the various levels of hazard and vulnerability and the influencing factors then examples are given just as reminder of what e.g. a moderate hazard could be like.

13. 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. ——————————————————————- We don't see any problem in the respondent tuning his own answers, particularly because he/she is asked to provide (subjective) opinions about hypothetical situations. There are in fact many other methods such as Delphi method that

encourage experts to revise their answers for an improved consistency.

14. 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 proceed- ing? How do you avoid to just replicate the provided answers, which will result in an illusive certainty about the derived estimates. ——————————————————————————

——— We agree that this part was confusing. We rephrased in the following way: "The data provided by this panel of experts were used to create a large representative dataset. This dataset was used to train the BNs with bootstrap sampling, so that the contingent probabilities in the learned network approximate the causal structure and probability distribution of the original sample. The dimension of the dataset allowed the use of the PC learning algorithm, a well-established constraint learning algorithm named after its authors, Peter Spirtes and Clark Glymour (Spirtes et al., 2000)." We had to discretize the values into classes before running the PC learning algorithm (bootstrapping is part of the PC). As we explained GeNIe doesn't support continuous variables.

15. 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). ——————————————————

———————————— First paragraph of 3.1: "The Bayesian vulnerability module was developed and tested by the authors and experts participating to the KULTURisk consortium. The foundations of the conceptual model were established during the development of the KULTURisk framework (Mojtahed et al., 2012; Balbi et al., 2012; Giupponi et al., 2013) and are thoroughly documented in Giupponi et al. (2014)." Usually, as we cited other studies above, the vulnerability indicators are aggregated using weights. We intended to point out that in the BN, there is no more need to extract and solicit the weights for aggregation of the indicators. This is because BN implicitly captures these weights by combination of causal structure and the conditional probability distribution. Regarding single most likely outcome and adjustment: this is a common practice when you are building CPT with experts and not from data, as documented in Marcot et al. (2006). The figures in figure 3 are realistic (showing a uniform distribution over the all case study area) but the model uses the evidences coming from the input data in each pixel. The variable states are based on the local data. e.g the percentage of foreigners is given as single value per district and ranges from 21% to 40.6%.

16. 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 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? ————————————————————————— The purpose is to understand the behavior of this module. EWS parameters can produce a sensitivity range of: above 10%, 10% to 5%, and below 5%. See the note to table 2. This means that sensitivity is there, it is significant but not too large to undermine the results. The reference we provided is Frey and Patil (2002). "A conditional confidence analysis (Frey and Patil, 2002) is performed, taking each state of input nodes individually. For every state of the output node (i.e. vulnerability) the range of variation of the marginal probability is computed over all the possible states of the input nodes." It's the classic operation performed to compute the tornado diagrams in sensitivity analysis. We can add a second reference specific to Tornado diagrams in GeNIe. We put the intervals into

a table because it's squeezing the information of 3 (tornado) diagrams and seemed more accessible than the graphs. The sensitivity analysis in SMILE/GeNIe is based on section 4 from the "Making Sensitivity Analysis Computationally Efficient" paper by Kjaerulff and van der Gaag (2000) (available at http://arxiv.org/pdf/1301.3868.pdf). The GeNIe tornado chart shows the top N parameters in the network influencing the specified output (the node/outcome pair) selected from the dropdown list in the upper part of the window. Below the tornado you can change how much the parameters can change (by percent or going to full 0..1 range). Note that each parameter is assumed to be changing separately - we do not perform n-way sensitivity at this point.

17. 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. ———————————————————————————————— This was answered in point 2. Those values were computed using the mean (and not the mode) in each pixel. The problem is that those figures are cumulative per district. While it's difficult to communicate cumulative uncertainty we suggest that the emphasis should be in the variation from one situation to the other (trade-off). As the reviewer noticed we compare only mean values and assume that uncertainty around those means behaves more or less the same (it's not too wrong since EWS parameters can affect only up to 12% the output).

18. 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. ——————————————————— True. We meant that the numbers are comparable.

19. 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? ——————————————————— As mentioned in the note to the Figure 4a (we are sorry because probably the note didn't make it through) High and Low are associated to the mean of probability distribution of the probability to get injured where high is equal to 5.8% per 50m2 and low is 0. It was on purpose in the note because that threshold per se is not very meaningful. We agree that the light blue over the light green background is not very visible. We suggest to change it to the attached.

20. p. 6621, l. 19: I suggest either "to ESTIMATE the actual number" or "to compute the EXPECTED number" 21. p. 6621 l. 27: grammar: "the considered factors ARE expressed" ——————————————————— Corrected as suggested.

22. Formulations in the questionnaire: question 5.4. "H and V being equal": probably it is meant H and V are fixed, not H=V? ——————————————————— Yes, it was meant to say when H and V are fixed or being equal to previous value. Should we change the questionnaire ex-post?

Please also note the supplement to this comment:
http://www.nat-hazards-earth-syst-sci-discuss.net/3/C2801/2015/nhessd-3-C2801-2015-supplement.pdf

---

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

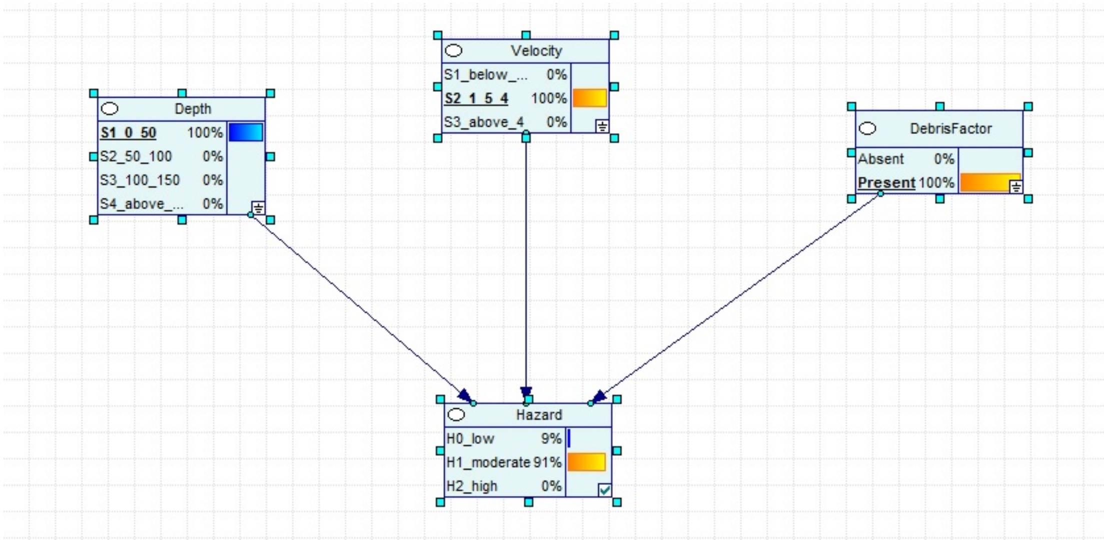

Fig. 1. hazard BN

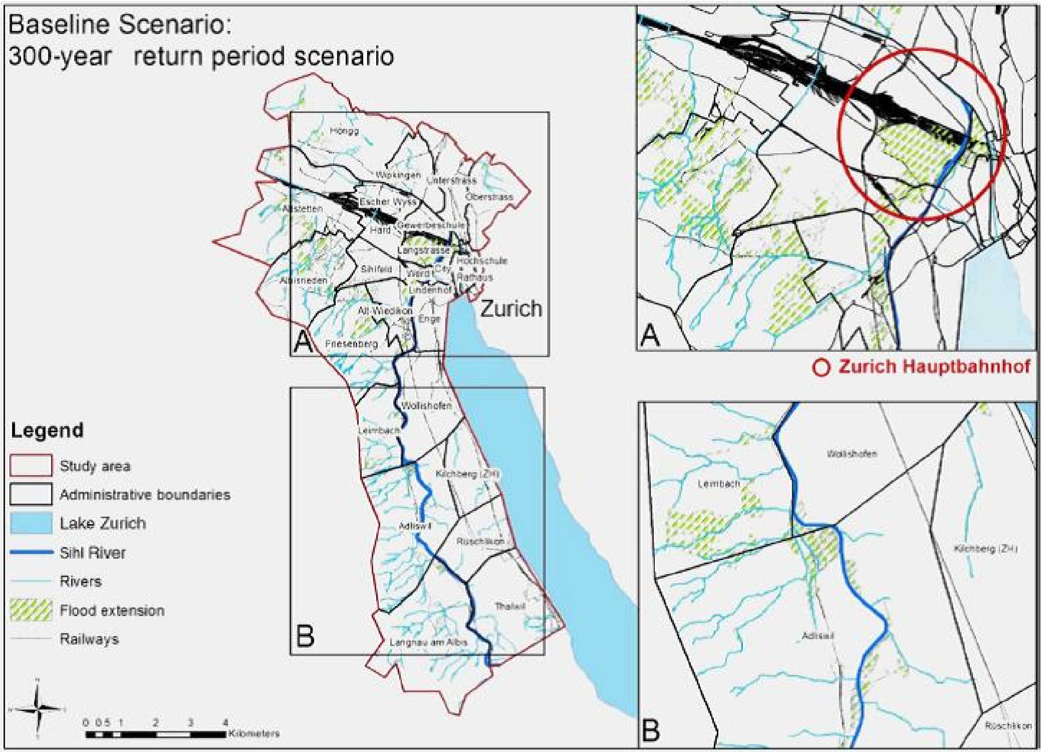

Fig. 2. hazard map from

Interactive Comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

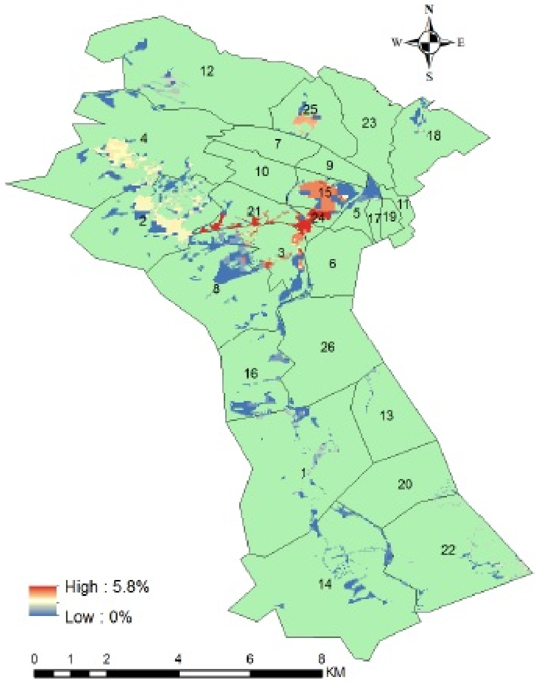

Fig. 3. new figure 4a

NHESSD
3, C2801–C2814, 2015

Interactive Comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper