

Interactive comment on “Estimating extreme river discharges in Europe through a Bayesian Network” by Dominik Paprotny and Oswaldo Morales Nápoles

Dominik Paprotny and Oswaldo Morales Nápoles

d.paprotny@tudelft.nl

Received and published: 25 August 2016

We would like thank the referee for the time spent in reviewing our article and the valuable comments. Below, we list all the comments and our response.

Page 5, line 3. (VII). From the Bayesian network, a conditional/marginal distribution of the max discharge data can be obtained. But how a particular value of the annual maxima is obtained? Is it the median?

-> As we mention in section 2.5, page 11, we calculate the mean of the conditional probability distribution of annual maxima of discharge.

Page 9, line 5. It is not clear to me which empirical marginal distribution is used for the

C1

variables considered in Fig3. In particular, I am curious on which marginal distribution is used for Max discharge data. Whether it is an extreme value distribution (or skewed distribution) or some other distribution? Please explain the reason if just a Normal distribution is used, because the extreme data may not fit well.

-> The distributions are all empirical, i.e. non-parametric. We use the usual estimator of the cumulative probability distribution; please see the supplement to this reply for the equation. The conditional probability distribution of annual maxima of discharge is empirical as well. We will clarify this in the text. No marginal Normal distribution is used; it is only the joint distribution that is modelled through a Normal, or Gaussian, copula.

Section 2.5 and Fig 3. Seven variables are considered as parents of max discharge in the manuscript, which means that the max discharge is “predicted” using 7 variables. Since the network is not created automatically, manually developing a model with seven predictors could cause the problem of overfitting (a good fit but with poor predictability). The model validation (section 3.1) showed the fit is good, but the “risk” of overfitting is not discussed. Usually a cross-validation need to be considered, but I understand that it is hard to conduct such analysis for the max discharge with limited data. I think it may worth to put some discussion of overfitting in the manuscript.

-> By standards of Bayesian Networks, a model with 8 nodes is relatively small. One way to verify overfitting is by testing the model in other geographical zones. It should be noted that in the manuscript we perform a split-sample validation and achieved the same results. We will amend the last paragraph of the discussion to mention overfitting.

Some references are missing, such as Hanea et al. 2006 (page 9, line 28) and Mutua 1994 (page 11 line 26). Please check carefully for the rests.

-> We thank the reviewer for pointing us to this error. List of references will be revised and corrected; we identified two more errors. The missing citations are:

C2

- Hanea, A. M., Kurowicka, D., Cooke, R. M.: Hybrid Method for Quantifying and Analyzing Bayesian Belief Nets, *Qual. Reliab. Engng. Int.*, 22, 709–729, doi:10.1002/qre.808, 2006.
- Moriasi, D., Arnold, J., Van Liew, M., Binger, R., Harmel, R., and Veith T.. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *T. ASABE*, 50, 885–900, 2007.
- Mutua, F. M.: The use of the Akaike Information Criterion in the identification of an optimum flood frequency model, *Hydrolog. Sci. J.*, 39, 235–244, doi:10.1080/02626669409492740, 1994.

Please also note the supplement to this comment:

<http://www.hydrol-earth-syst-sci-discuss.net/hess-2016-250/hess-2016-250-AC2-supplement.pdf>

Interactive comment on *Hydrol. Earth Syst. Sci. Discuss.*, doi:10.5194/hess-2016-250, 2016.