Hydrol. Earth Syst. Sci. Discuss.,
doi:10.5194/hess-2016-250-RC1, 2016

Hydrology and Earth System Sciences
Discussions

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

Anonymous Referee #1

Received and published: 26 July 2016

The paper combines Copulas with the Bayesian network framework aiming to construct a precise multivariate model for extreme river discharges. The approach is very challenging and requires a good understanding of both methodological concepts, Bayesian networks and Copulas. The paper is well structured and organized, especially considering the bunch of information and background knowledge that needs to be explained/referred to. Yet, some proceedings/issues are not clear to me.

section 2.4:

The choice or the Gaussian copula could be better justified, e.g. are there any physical explanations?

- Mention in the main text the two types of copulas you are testing as well.

- Briefly explain in the supplement, why you choose the selected types for comparison and not other/further representatives of no/lower/upper tail dependence. Are there any physical explanations? The test only shows, that Gaussian performs better than the gumbel and clayton, but it might still be a bad choice. Supplement 3 even indicates that the Gaussian copula is not a very good choice.

I do not understand how you get and use the conditional rank correlations for continuous distributions. The references you provide only give a short explanation and refer to further literature again. Is there some standard literature on the definition of conditional rank correlations for continuous distributions? To my understanding the conditional rank correlation depends on the state of additional parents, yet in fig. 3 you only give single numbers (no conditioning is visible). Yet, if the rank correlation is independent of the states of the other parents, you miss to model the joint effects and could use a naive Bayes instead (which would be far more simple than going all the way over BNs and copulas). On the other hand, if the conditional rank correlation depends on other parents states, there must be a way to calculate it for each possible conditioning state (to be able to perform inference) or it must be determined for a discrete approximation of the conditioning variable (which increases the number of required parameters significantly; similar as for using discrete BNs from the beginning). Maybe you could comment on this in the discussion forum.

section 2.6:

I find this subsection difficult to read and suggest to revise the sentence structure of this section.

Why do you use 30-year time periods? What would be the effect of using shorter/longer time periods?

Please mention the distributions you are testing in addition to GEV. GEV performs best

compared to what?

section 3.1:

p.13, l. 5: "2-year discharge has the same performance as Q_MAMX" <- where does this information come from?

Why do you suddenly have 4 different time periods? Section 2.6 mentions only 3.

The regional performance seems to depend strongly on the number of stations used per region

You mention several times 'the model performance remains acceptable'. What is your understanding of acceptable?

How do you explain the better performance of the BN quantified from the smaller dataset of 917 records?

p. 18:

You might extend your comment on using different types of copulas: How suitable is the Gaussian to model all interactions? How well does it fit the data (are there objective measures)? Would it be possible to use different types of copulas in the same BN and thus find a better description of each interaction? Which other types of copulas could be useful to check? What do you expect, to which extend could the model be improved, by using different types of copulas?

figure 3 and supplement 4:

Why do you use a discrete BN in the shown examples for inference? This does not correspond with your objective to find/use a continous BN with a low number of parameters. The discrete conditional distributions you show in the supplement, are not smooth. I guess, this should not happen, if you stick to continuous representations.

Typos and minor issues:

p. 6, l.30: the influence of DIFFERENCES models

p. 8, l.14: need to explained

p. 8, l.22: a set of noTes and arcs

p. 9, l. 1: which is the actually case

p. 9, l. 3: to be precise, the actual number of required conditional probabilities/parameters is a bit smaller, since some parameters result from probability theory (the parameters that describe a distribution for a specific condition have to sum up to one)

p. 9, l.28: Hanea 2006 is missing in the list of references

p.10, l.27: values these climate variables

p.11, l. 9: This variable is influence

p.11, l.11-12: check complete sentence

p.11, l.21: allowed to performed

p.15, l. 8: median return periods are show

p.17, l.12: I would not consider this fact as "evidenced", but rather as indicated

p.18, l. 3: Potential incorporation different time spans

p.18, l.21: non-Gaussian copula would a better model

p.20, l. 6

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