Hydrology and Earth System Sciences Discussions

Hydrol. Earth Syst. Sci. Discuss., referee comment RC3
https://doi.org/10.5194/hess-2021-445-RC3, 2021

Comment on hess-2021-445

Anonymous Referee #2

Referee comment on "Detecting hydrological connectivity using causal inference from time series: synthetic and real karstic case studies" by Damien Delforge et al., Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2021-445-RC3, 2021

This paper focuses on causal inference in a hydrologic system, and compares four different causal inference frameworks that range between bivariate versus multivariate and linear versus nonlinear. The authors introduce several aspects and assumptions related to causal inference, such as confounding factors, nonlinear and threshold interactions, effective versus functional connectivity, and the potential for missing drivers. They then apply the four methods to a synthetic hydrology problem in which two reservoirs are connected by meteorologic forcing, such that they appear causally connected if the forcing is not accounted for. Next, they apply the methods to an actual dataset in a karst system. They conclude that while the nonlinear methods detect nonlinear interactions, they also miss some interactions, and should be used carefully or in combination with multiple methods to more robustly detect causal interactions.

This was an interesting paper that should be of interest to the hydrologic community and anyone interested in applying a causal inference method to their data. However I have several comments on the methods, interpretation of the results, and the writing in general and would suggest “moderate” revisions prior to publication.

Major comments:

The lack of statistically significant links in the CMI method made more sense to me when I noticed the sparse dataset (e.g. fewer than 100 data points) that is available for the karst system analysis. In general, I think this method would not be expected to produce robust results given this amount of data, due to the high dimensional pdfs involved for information theory measures. With this, I think the interpretation should not be that CMI performed "worse" in some way, but that it has higher need for data length as a much higher dimensional approach, especially relative to the bivariate methods. Additionally, it seems like the bivariate methods actually do utilize the full amount of data available for any two variables, which means that the comparison is even less "fair" – it might be better to only consider the time window in which there is data available for all (or some, like the P1, P2, and P3 cases in Figures 5 and 6) variables. I think this aspect should be made more clear at the least, and possibly deserves a change in the time windows used for the comparison.

I also have a question about the statistical significance. For example, in some information theory based studies, we use a shuffled surrogates method for statistical significance of a given link, which would differ from a p-value in a correlation analysis. If the method for identifying a statistically significant link varies at all between methods, this also needs to be apparent.

For the synthetic case, I see you used a lot longer dataset than you have available for the real karst case study. This could lead to the better performance for the higher dimensional or multivariate methods – it might be useful to test the synthetic case for a much smaller dataset and observe or confirm whether these methods start to lose their detection of links. This could better show that the CMI/ParCorr types of methods do have better performance, but only when given a lot of data. I think the differences between the methods might make them inherently difficult to compare, but these are some things that could improve the attempt.

Finally, there are several places with strange phrasing, or where a term is introduced before it is defined, so there is momentary confusion on whether a reference is missing or the sentence is relevant. I am highlighting some of these that I noticed in the minor line-by-line comments below.

Minor comments:

Line 7: "appears unstable" relates to my comment on data length…I think the instability is at least partially due to a very small dataset. Either way, it is not very clear what this term means within the abstract.

Line 15: "between variables from variables" did not make sense to me

Line 28: cross-scale

Line 35-45: This paragraph seemed scattered, and I did not come out of it with a clear understanding of "effective connectivity" in particular. Suggest to revise

Line 41: "process-based water flows" – I'm not sure if there are non-process-based flows of water?

Line 45: "progressive constraint" was not clear to me

Line 50: "heterogeneity" instead of "hiddenness"?

Line 60: I'm not sure about the sentence "nonlinearity is imputed to nonlinear hydrological processes", seems redundant

Line 65: would be good to re-define CCF here

Line 74: What do you mean by "to appreciate the results" – to compare with the results, or validate them?

Line 85: "Being multivariate" – I'm not sure that a multivariate approach inherently deals with confounding effects. For example, a multiple linear regression is multivariate, but does not do any type of conditioning on confounding variables…

Line 93: PC and MCI are brought up, and then defined later – would be better to re-arrange such that we are not wondering what they are.

Line 99: "not preselection" to "no preselection"?

Line 103: reference for causal sufficiency? In general, this is a good point for any analysis, you particularly reference it for CMI, could state that this hypothesis really underlies all your methods…

Section 2.2.1: I felt like you did some discussion previously that was particular to each method, but then you have these sections for each method separately. I would move some of the above material at the beginning of 2.1 into these sections directly, and save the "causal sufficiency" aspect at the beginning as it applies to any method.

Line 121: "overall good performance of this value" is vague, do you mean for the synthetic study, or the real study, or in general?

Line 136: I don't think you have introduced Granger Causality

Line 153: Is two weeks of computation for a single processor? I figure it would take different amounts of time depending on whether you used a laptop or a server, etc, so could make this more clear.

Line 176: It seems like for 2014-2017 time-series, there would be more than 465 time steps, implying the presence of gaps. This also comes into play in terms of your total data length for the multivariate methods. Basically, the counts in Table 2 make it seem like there is more data available than there actually is, when you start comparing multiple datasets (with 48 data points being the total overlap).

Line 204: “problematic case of the common cause” – after this, you define what this means, but as it is, the phrase seems a little mysterious, like a Sherlock Holmes story.

Line 210: haven’t defined Qb’ yet

Line 230: This is a “synthetic study” but the years make it seem like you are using actual data from your real study?

Line 234: What is a differenced dataset? This comes up a few times and I’m not completely sure what it is…whether it is the increment or something done in the modeling process.

Line 270: “If causality is hard to infer…” is not a great sentence, excusing a complicated figure and telling us it actually makes sense. You could just remove this.

Line 297: “is be removed”

Line 338: “evanescent singularity” is unclear