Summary

This paper by Delforge et al. titled "Detecting hydrological connectivity using causal inference from time-series: synthetic and real karstic study cases" uses four causal inference methods to obtain insights on hydrologic connectivity in a karstic site. The authors also use a synthetic case study to obtain a preliminary understanding on the performance of the methods prior to applying them to the real-world case study.

Overall, the paper objective and the experimental design are well articulated and set clearly. I also think that causal inference is an important topic that should receive more attention from the hydrologic community, and it is of an interest to the audience of HESS. However, I noted some technical issues on the implementation of causal inference methods and the interpretation of results. Also, the discussion section is lacking as the authors didn't compare their findings to recent papers that utilized causal inference in hydrology. Below I summarize these issues and other major comments followed by minor ones. I urge the authors to pay attention to these comments while revising their manuscript. Generally, I recommend that this paper undergoes major revisions before it can be considered for publication.

Major comments

1- In the application of multivariate causal inference methods (Partial Correlation and Conditional Mutual Information), the authors didn't illustrate whether the history of variables is used in the conditioning set of variables or not. For instance, if one wants to test for the hypothesis that $\boldsymbol{Y}$ causes $\boldsymbol{X}$ using multivariate methods, then ideally the following conditional test should be implemented: $\mathbb{I}(Y_t, X_t \mid Y_{t-1}{:}Y_{t-\tau}, X_{t-1}{:}X_{t-\tau}, Z)$. This conditional test means that one is testing for the statistical relationship between $\boldsymbol{Y}$ and $\boldsymbol{X}$ at time t while conditioning in the history of both variables up to a lag time of τ as well as the set of variables $\boldsymbol{Z}$ which includes all other confounders. This is a crucial point in the implementation of multivariate methods because it removes the effect of autocorrelation. There are many ways of conditioning in the history of variables; for example, the classical Granger causality (Granger, 1969) conditions in the history of X but not Y. Similarly, Transfer Entropy (Schreiber, 2000) does the same, whereas methods such as momentary information transfer (Pompe & Runge, 2011) conditions on both variables. I think that the authors need to mention explicitly what is the conditioning test for both multivariate methods. Please consider adding this information with clear mathematical expressions. Also, the parameter τ is often one of the most important hyperparameters in causal inference methods, so a discussion on the value of this parameter needs to be included. Please note that τ in this context is slightly different than d_{max} which the authors use to set the maximum lag time for testing interactions.

2- Related to the previous point, the reason why conditioning in the history of variables is important is because hydrologic timeseries of variables are often highly autocorrelated which leads to spurious causal links, and this testing removes autocorrelation. This context is important in interpreting the results obtained from causal inference methods either in the synthetic or real-world case studies. The authors used first-order difference of the original time series to remove the effect of seasonality and autocorrelation; however, this is not needed if the implementation of multivariate methods already conditions on the history of variables. Please revise the interpretation of results both in sections 3.2 and 4.2 to highlight this issue of autocorrelation.

3- I found the discussion section to be lacking and it does not report any insightful comparisons to previous work of causal inference in hydrology. For instance, Ombadi et al. 2020 used four causal inference methods on a synthetic and real-world case studies with formal investigation on the impact of sample length, observational and process noise. Some of the methods used in this paper (e.g., CCM and CMI) were also used in that study. It would be important to compare the findings of both studies on the performance of different causal inference methods as this will allow us to build a consensus on the suitability of causal inference methods for hydrologic applications. Although Ombadi et al. 2020 is perhaps the most relevant to this study, there are other studies that used specifically information-theoretic approaches for hydrologic systems characterization such as (Jiang & Kumar, 2019). Please enrich the discussion section by linking the findings of this study to previous work.

4- The record length of the timeseries used in this study is relatively short. This is one of the main challenges that face the application of causal inference methods in hydrology. In my experience, I found that methods based on information theory (e.g., CMI) often needs a long record (~ 2000 – 3000 data points) to provide reasonable performance. The results shown in this study either for the synthetic or real-world case studies are perhaps significantly impacted by the record length yet no discussion was included on the effect of record length. Please enrich the discussion section by highlighting the potential impacts of sample length.

5- Overall, the paper needs to be subjected to proofreading to eliminate grammatical deficiencies and typos as well as to improve the readability. I highlight some of these grammatical errors, unclear sentences and typos in the minor comments below, but these are only few examples and similar corrections should be implemented throughout the paper.

6- Some of the basic information on causal inference methods that was mentioned in section 2 is not accurate and incorrect. For instance, the description of partial correlation in lines 136-138 is not accurate. The authors mention that "partial correlation is like Granger causality". This is quite vague. What the word "like" means specifically here? It is true that partial correlation shares similarities with Granger causality in the sense that both use linear regression to assess interactions while conditioning on potential confounders. However, there are crucial difference too. For instance, Granger causality is technically implemented in a different way than partial correlation with setting both restrictive and unrestrictive regression models, and testing for statistically significant differences using t-test and F-test. These are very crucial differences. Also, on a higher level, the concept of Granger causality is based on what is known as predictive causality and it take into account time precedence. The authors should be careful in introducing the different methods and use precise information. I suggest that the authors revise section 2.1 by writing clear mathematical expressions for each causal inference method, and also be more precise in their description.

Minor comments

1- Lines 315-320 and elsewhere: the instability of CMI here is attributed to missing data. This might be one of the reasons, but I suspect that the main reason is the short record length (see my major comment #4). Also, a possible but unlikely reason is that the instability is the result of changes in the dynamic connectivity. This might be true if the timeseries used in Figure 6 (a, b, c and d) correspond to different hydrologic conditions (wet vs dry). If this latter case is possible, then it is worth of highlight and discussion.

2- Figures 4 & 5: there are several causal links with arrows pointing toward RF (Rainfall)?? Apparently, this is physically incorrect, but I was not able to find any discussion on this in the paper. Are these arrows drawn in the wrong way? Or these are the real results obtained from causal inference methods? If it is the latter case, then this needs to be discussed. In general, this raises a red flag on the accuracy of causal links obtained from the different methods.

3- Lines 228-229: the standard deviation of the noise added to the precipitation signal is unrealistically large!! Even the smallest value used here which is 0.05 of the standard deviation of precipitation is still large. A proportion of 0.001 of the standard deviation of precip is often sufficient to satisfy the condition of causal sufficiency. If process noise is very large, this will impact the results. See Ombadi et al. 2020 for the impact of process noise on the performance of different causal inference methods.

4- Lines 230-231: why only the last year was used? Is it a spin-off period to eliminate the impact of initial conditions or for computational reasons?

5- Lines 252-253: this is perhaps related to the conditioning on the history of variables (see my major comments #1 and #2)

6- Lines 260-262: This is a well-known issue with causal inference methods. You can refer to some studies that pointed to the same issue in evaluating causal inference methods either in hydrology or other fields.

7- Table 4: please replace the abbreviations with the full name (e.g., TP: True positives) or alternatively add this info to the caption of the table so that it can be a standalone component.

8- Figures 4, 5 and 6: I suppose that the numbers in the arrows denote the lag time of interaction in days; however, this was never introduced or mentioned in the captions. Please revise.

9- Lines 27-29: some applications of causal inference in hydrology are missing here. For instance, soil moisture-rainfall feedback (Wang et al., 2018) or differential impact of environmental drivers of evapotranspiration (Ombadi et al., 2020). There are others too if you look in the literature.

10- Lines 35-44: I liked the distinction between structural, functional and effective connectivity. However, from the text, it was not clear what is meant by the effective connectivity and how it differs from the functional one.

11- Line 71: remove “obtained from”. Typo.

12- Line 144: the correct name is transfer entropy not “entropy transfer”

13- Line 150: “computationally expensive and quickly require …”. The sentence is not logically correct. Please revise.

14- Line 152: grammatical error in “at section 3”. It should be “in section 3”

15- Figure 1 caption: when referring to (a) and (b), please remove the parentheses because it is confusing. Only use the parentheses the first time you introduce them.

16- Line 291: replace “As for CCM” with “Similar to CCM”. The sentence does not read well currently.

17- Line 15: this sentence does not read well at all. I understand what you want to convey, but it needs to be rephrased. Something like “…interactions between variables from timeseries only…etc.”

18- Lines 98-99: the description of the parameter alpha_pc is not very clear and intuitive to me. Could you elaborate?

References

1- Pompe, B., & Runge, J. (2011). Momentary information transfer as a coupling measure of time series. Physical Review E, 83(5), 051122.

2- Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, 424-438.

3- Ombadi, M., Nguyen, P., Sorooshian, S., & Hsu, K. L. (2020). Evaluation of methods for causal discovery in hydrometeorological systems. *Water Resources Research*, *56*(7), e2020WR027251.

4- Schreiber, T. (2000). Measuring information transfer. Physical review letters, 85(2), 461. https://doi.org/10.1103/PhysRevLett.85.461

5- Wang, Y., Yang, J., Chen, Y., De Maeyer, P., Li, Z., & Duan, W. (2018). Detecting the causal effect of soil moisture on precipitation using convergent cross mapping. Scientific reports, 8(1), 12171. https://doi.org/10.1038/s41598-018-30669-2

6- Jiang, P., & Kumar, P. (2019). Using Information Flow for Whole System understanding from component dynamics. Water Resources Research, 55(11), 8305-8329.