

Hydrol. Earth Syst. Sci. Discuss., author comment AC3  
<https://doi.org/10.5194/hess-2021-445-AC3>, 2021  
© Author(s) 2021. This work is distributed under  
the Creative Commons Attribution 4.0 License.



## Reply on RC4

Damien Delforge et al.

---

Author comment on "Detecting hydrological connectivity using causal inference from time-series: synthetic and real karstic study cases" by Damien Delforge et al., Hydrol. Earth Syst. Sci. Discuss., <https://doi.org/10.5194/hess-2021-445-AC3>, 2021

---

### 1. General Response

Dear reviewer, we would like to thank you for your attention to the preprint manuscript and the helpful comments you made. We understood that your major concerns are related to the virtual experiment about (1) the fair comparison of CCM when it comes to noisy signals and (2) the robustness and performance of the PCMCI-CMI method under limited data points. For both points, as explained here below, we do not think that extending our synthetic case study can reply to the very valid points you raised. We rather choose to better explain the results we get and the limits of the conclusions that can be drawn out of them. To this end, we clarify our point of view in a revised discussion section, which we hope will meet your concerns and that of the readers.

We also improve references to the existing literature on these concerns within a revised discussion structured as follows:

- a summary and appreciation of our results (improved for CCM as well);
- a particular focus on the robust estimation of Conditional Mutual Information (CMI) concerning missing values, record length, dimensionality, the nature of the dependencies, or noise;
- and practical recommendations for the uses of causal inference methods and future research perspectives.

We hope you will find below satisfactory answers to your observations.

Thank you again for your contribution to this discussion,

On behalf of co-authors,

Damien Delforge

### 2. Major Comments

#### 2.1 Major Comment 1

**Comparison between CCM and CMI-based PCMC. The current comparison based on noisy data is unfair to CCM, because CCM is more suitable for deterministic dynamics and does not work well in a stochastic system. Also, the authors used different levels of noises in the synthetic study. Still, only the averaged results are reported in Figure 3, and the noise impact on the performances of the four methods remains unknown. Therefore, I suggest, at least in the synthetic case study, performing the comparison based on a noise-free/deterministic system and a thorough evaluation of the noise impact.**

Because CCM is based on the theory of nonlinear deterministic systems (or chaos theory), it is generally considered a tool for deterministic signals. Therefore, CCM is also considered as a method that would not work well for a stochastic system, which is your legitimate point and potentially the one shared by many readers familiar and unfamiliar with the CCM method. Despite this widely held view, we consider that noise or stochasticity should not be a significant issue in our case and, thus, that the problem should not be emphasized in the manuscript. Noise is not the main reason why CCM could be deceiving. In our experiment, we face many False Positive and believe that the False Positive Rate should be the major concern of hydrologists, with confounding being the primary reason for False Positive misclassifications.

We consider that CCM's inability to deal with confounding is the critical message to illustrate and convey, as this limit is not apparent in the current literature. Because of this, researchers may have high and misguided expectations about the capabilities of CCM (e.g., see Sugihara, 2017), just as us when we first turned to this method. These expectations are all the more pronounced since the CCM method refers directly to the concept of causality. This is not the case for the correlation or the cross-correlation function. Researchers use them implicitly to make causal inferences while being aware of the limit, as the maxim "correlation is not causality" says. Many papers on CCM, including the original (Sugihara et al. 2012) or some in hydrology (e.g., Ombadi et al., 2020), refer to this maxim to motivate the use of more advanced methods such as CCM. Yet, when one types the maxim into an internet search engine, most of the illustrations refer precisely to the problem of confounding. Therefore, it seems crucial to dispel the myth that CCM could do better than correlation or CCF by escaping the confounding problems. Our interpretation of our results is that CCM differs from CCF and correlation in its ability to reveal nonlinear associations, as well as linear associations, not in any capability of solving confounding problems because it is a bivariate method. Since most hydrological signals already exhibit linear correlations, we believe that the usefulness of CCM for causal inference is quite limited in hydrology. However, the method is interesting for studying the dynamic and nonlinear properties of signals (e.g., see Delforge et al., 2020).

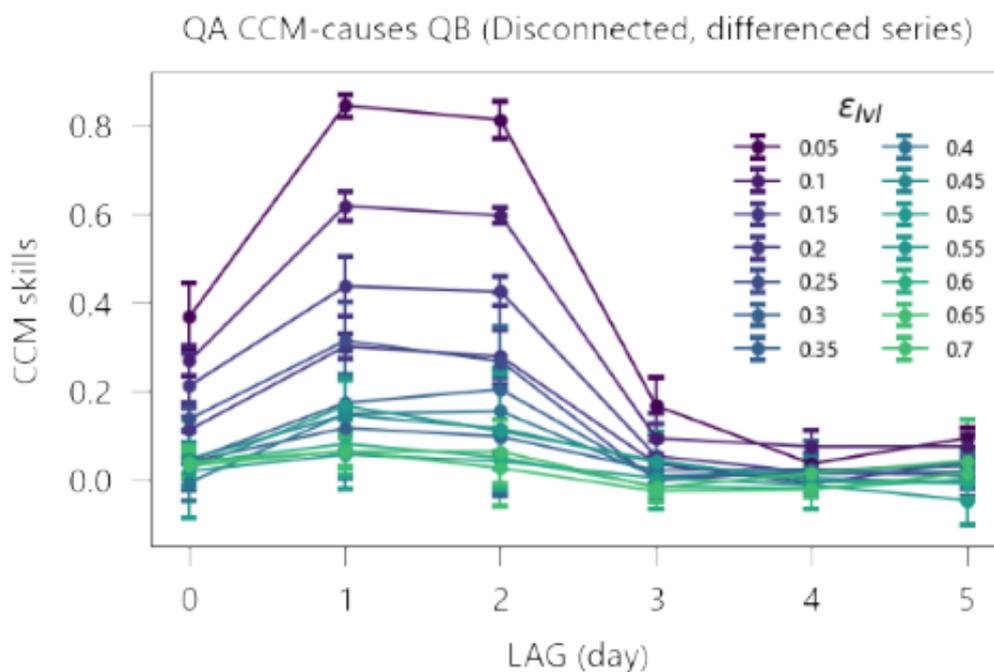
In conclusion, we understand that you have not made any remark relative to the main message we are trying to convey, and we consider your remark about the noise relevant. Yet, we have chosen not to modify the manuscript to introduce the noise issues because 1) we consider its implications to be secondary in our case, (2) exposing the issue could outshine the main message concerning CCM, and (3) we also need to discuss three other methods in our comparative study.

In response to your comment, you will still find below specific elements answering your concerns on the effect of noise, together with other elements motivating our opinion regarding CCM.

---

*About the effect of noise in our virtual experiment*

Regarding the virtual experiment, Figure 3 (in the preprint) displays the average dependencies, as you said, but also the interquartile range. Hence, Fig. 3 shows some dispersion possibly imputed to noise or model configurations. To illustrate further the effect of noise, the Figure below reports the CCM results for the different noise level parameters  $\epsilon_{|v|}$  for the case where Q\_A and Q\_B are disconnected. Error bars are the remaining interquartile ranges. In this case, the figure shows that injecting noise decreases the mapping skills. Beware that we do not consider that noise decreasing the mapping skills would be systematic, for instance, if we inject auto-correlated noise and have relatively short time-series. However, in our case, we used a Theiler window (Theiler, 1986)  $tw$  of 10 days, which should limit the potentiality that CCM performance is related to the autocorrelation (See supplementary materials for details).



If we define a case without noise, as you suggest, the result would still be deceiving by reporting a significant link between Q\_A and Q\_B, most likely above the case with the minimal noise in the Figure ( $\epsilon_{|v|}=0.05$ ). When noise is injected, CCM improves its causality detection skills for the wrong reason. CCM mapping skills are simply reduced because of noise. Similar conclusions are reported in the literature. Reviewer 1 pointed to the study of Ombadi et al. 2020, which also shows that CCM has a high False Positive Rate based on another virtual experiment and that the False Positive Rate, alongside the True Positive Rate, decreases when noise is injected. Accordingly, without adding this additional figure to the manuscript, we propose a short modification to mention that we observed that noise reduced the mapping skills, as in Ombadi et al. (2020).

Still, even if the mapping skills are noise-sensitive, CCM will always attempt to report such a link because we concluded that CCM could not deal with confounding given its bivariate nature. Hence, as long as the virtual experiment tends to suggest a dependency between Q\_A and Q\_B at lag 1, we consider that our application is fair and not compromised by noise. More generally, we consider that CCM could be applied to noisy systems with awareness of its limits and the necessary cautions (e.g., a Theiler window). After all, CCM relies on algorithms developed to be more robust to short and noisy time series (Sugihara and May, 1990; Sugihara, 1994). If convergence there is, CCM succeeds in revealing an association, either linear or not. If not, either there is no association, or it is drowned in

the noise.

### *Notes about CCM, synchrony, and confounding*

Initially, Sugihara et al. (2012) do not refer to confounding but to the concept of synchrony, such as the famous synchronization discovered by Huygens for its pendulum clocks. Even if there may be substantial differences between the concepts, we consider that synchrony refers to confounding as synchrony could occur in dynamical systems subject to strong forcing (See the SM of Sugihara 2012 for a discussion on synchrony, and Sugihara et al. 2017). Since we fall in this case, we do as suggested in Ye et al. (2015) and varied the predictive horizon to infer causality from the principle of priority when synchrony occurs. Yet, in this case, we show that it is still deceiving, reinforcing our point that CCM cannot deal with confounding and the common cause problem. To make sense of our results, we had to conclude that if both  $Q_A$  and  $Q_B$  are deterministically mapping to  $P_{eff}$ , it would seem logical that a deterministic map exists between the two, just as when one rearranges deterministic equations, and that they present some isomorphism in their reconstructed attractor (what is tested with CCM). Referring to Sugihara et al. (2012), "*time-series variables are causally linked if they are from the same dynamic system that is, they share a common attractor manifold  $M$ . This means that each variable can identify the state of the other*". As a matter of hydrological connectivity, given reasonably low noise, it would always be possible to identify the state of a hydrological reservoir from another parallel, yet disconnected reservoir, as two buckets sitting next to each other.

## **2.2 Major Comment 2**

**Limited data points for computing CMI. In the real case study, the inferred causality from CMI-based PCMCI is much less trustable, given only 465 datapoints of 7 variables (what is the maximum allowed number of conditioned variables set in PCMCI by the way?). In fact, it is somehow expected that the CMI-based PCMCI does not work well using this limited dataset (even for a three-dimensional CMI estimation, several hundred data points might not be sufficient). Although the authors acknowledged this limitation, I strongly recommend a corresponding synthetic study to evaluate the impact of dataset size and the number of variables in CMI-based PCMCI, which is very critical to guide the current and future causality analysis in earth science inferred by the PCMCI algorithm.**

Your present concern meets those of the other reviewers as well, in particular, reviewer 2. For clarification, the virtual experiment is a three-variable case that spans 365 days, while the real study case is an 11 (All) or 9 (P1, P2, P3) variable case over a restricted time domain of 48 (All), and 184 (P1), 62 (P2), and 218 (P3) timestamps. This time-domain is very short because the PCMCI algorithm dismisses all timestamps where the time-series and their lags up to  $2d_{max}$  exhibits at least one missing value. Therefore, missing values are a problem, mainly when they are unevenly distributed over time, because the analysis falls down to a very small sample size compared to the length of the individual time series. There is no maximum number of conditioned variables. This number would depend on the parameter  $\alpha_{PC}$  and the preselection step with the PC algorithm (see supplementary materials for details). Our philosophy (better exposed in the revised manuscript) is to use the PCMCI without adding constraints on which variable is supposed to influence another or not. Hence, this attitude is very empirical as what we are testing is the ability of causality methods, such as the PCCMI, to screen and detect the existing links between the datasets. We assume that many PCMCI users will also have the same initial mindset of letting the method work by itself. Our revised discussion points out the resulting problems, as you do in your comment, and evaluate that mindset.

We share your view that the small sample size is problematic and relative to the number

of variables. Our strategy was to use the CMI estimators and test that is the most forgiving (Runge, 2018). The nearest-neighbor estimator and the shuffle test of Runge (2018) are better suited than kernel-based approaches for short record length ( $< 1000$ ), based on numerical experiments covering sample sizes from 50 to 2,000 and dimensions up to 10. Yet, despite the use of method recommended for small records, the real study case of the manuscript is concerned by the pitfall of estimating CMI with short record length and high dimensions resulting in non-robust test results, as you also interpret. We showed that the robustness could be increased by performing an ensemble of tests. Of course, the robustness of a numerical result is one issue; its reliability in terms of connectivity is another.

So, we agree with you that this is a critical point to discuss. However, extending toy model-based sensitivity analysis will not provide sufficiently robust results to guide the scientific community properly for other applications. The reasons go beyond that. First, our study remains a comparative study. Such a focus on the PCMCI-CMI method would rather deserve a separate issue. Also, the conclusion of such an extended virtual experiment is reasonably known a priori: the results become more robust with increasing sample length or decreasing number of variables. We see no reason why a non-trivial conclusion, such as recommendations of sample length as a function of the number of variables, would be transposable to a problem with different characteristics, such as different noise levels, the number of variables, the maximum delay  $d_{\max}$ , different model coupling patterns, signal behavior, or representative scales. In addition to the sample size and dimensionality, the CMI also depends on the nature of the CMI dependencies, smooth or not smooth as it could be expected in systems with highly dynamic connectivity, and the magnitude and the characteristic of noise. This CMI dependency and noise vary across spatial and time scales. The results also depend on the methods, for instance, kernel-based or nearest neighbors estimators and their hyperparameters. Altogether, we believe that proper recommendations and fit to one's specific case would require an extended and computationally intensive sensitivity analysis with another model, which is not the scope of our study.

Our point, with the synthetic studies, was to show the divergence of the methods on the same – simplistic - case study, not as an answer to the question "what should we do", but rather as an exploration of the behavior of the tested methodology in a case where we can give meaningful interpretations of the results. As a recommendation for each problem on which those methods are used, we consider that a good strategy would be to test the issues met and the insights gained by using fit-for-purpose models mimicking the property of signals they want to study. This is suggested in the revised perspective.

Another revision is to discuss the question "should we condition on all variables?". As mentioned above, with missing values, the small size of the time domain is because we test and condition each variable with respect to each variable. This is why focusing on P1, P2, P3 only increases the size of the cross-domain while removing the possibility that P1, P2, P3 influence each other. Perhaps, when it comes to testing the connectivity between two points A or B within a system, conditioning on their past and the common driver, e.g., evaporation and rainfall or effective precipitation alone, is a dimensionally adequate and contained representation of the problem, on which CMI can be more reliably estimated.

### **3. Minor Comments**

#### **3.1 Minor Comment 1**

**Lines 67 and 68: Please spell out ParCorr and CMI.**

Corrected

### **3.2 Minor Comment 2**

**Line 144: "entropy transfer" -> "transfer entropy"**

Corrected

### **3.3 Minor Comment 3**

**Lines 147 and 148: "The nearest-neighbor estimator is recommended for time-series below 1000 samples"... under what dimensionality?**

As mentioned in our response to your major comment 2, this heuristic value is based on Runge, 2018 for a varying dimensionality up to 10. This is clarified in the revised manuscript.

### **3.4 Minor Comment 4**

**Line 238: Qb -> QB**

Corrected. B has been capitalized.

### **3.5 Minor Comment 5**

**Line 356: "constraint causal inference" -> "constrain causal inference"**

Corrected.

## **4. Cited References**

Theiler, J.: Spurious dimension from correlation algorithms applied to limited time-series data, *Phys Rev A Gen Phys*, 34, 2427–2432, 1986.

Sugihara, G. and May, R. M.: Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series, 344, 734–741, <https://doi.org/10.1038/344734a0>, 1990.

Sugihara, G.: Nonlinear forecasting for the classification of natural time series, *Phil. Trans. R. Soc. Lond. A*, 348, 477–495, <https://doi.org/10.1098/rsta.1994.0106>, 1994.

Ye, H., Deyle, E. R., Gilarranz, L. J., and Sugihara, G.: Distinguishing time-delayed causal interactions using convergent cross mapping, 5, 14750, <https://doi.org/10.1038/srep14750>, 2015.

Sugihara, G., Deyle, E. R., and Ye, H.: Reply to Baskerville and Cobey: Misconceptions about causation with synchrony and seasonal drivers, *PNAS*, 114, E2272–E2274, <https://doi.org/10.1073/pnas.1700998114>, 2017.

Runge, J.: Conditional independence testing based on a nearest-neighbor estimator of conditional mutual information, in: *International Conference on Artificial Intelligence and Statistics, International Conference on Artificial Intelligence and Statistics*, 938–947, 2018.

Delforge, D., Muñoz-Carpena, R., Van Camp, M., and Vanclooster, M.: A Parsimonious Empirical Approach to Streamflow Recession Analysis and Forecasting, 56, e2019WR025771, <https://doi.org/10.1029/2019WR025771>, 2020.

Ombadi, M., Nguyen, P., Sorooshian, S., and Hsu, K.: Evaluation of Methods for Causal Discovery in Hydrometeorological Systems, 56, e2020WR027251, <https://doi.org/10.1029/2020WR027251>, 2020.