Authors' responses

Manuscript egusphere-2023-2212
"A comparison of two causal methods in the context of climate analyses"
D. Docquier, G. Di Capua, R. V. Donner, C. A. L. Pires, A. Simon, S. Vannitsem

We would like to thank the Editor Stefano Pierini for handling our manuscript and the two anonymous reviewers for their helpful comments. Below we list the reviewer's comments (in black) **and our point-by-point replies (including changes in the paper) (in bold blue). We provide the line numbers of both revised and track-change versions, starting with "L" and "LT", respectively.**

1. Reply to Anonymous Referee #1

The authors apply two causal inference methods to a variety of test models as well as a climate example. The paper is interesting and worthy of publication pending some revisions for clarity.

We thank the reviewer for the encouraging words.

Comments:
– L25–7: Although instantaneous correlations cannot determine the causal direction, lagged regressions are often used for this purpose, despite their drawbacks compared to causal inference methods. It may be worth citing McGraw and Barnes (2018) here, since they support your point about the need for causality analysis rather than lagged regressions for climate studies.

We thank the reviewer for the suggestion. We have now included some text related to lagged regression and added the proposed reference (L27-29; LT28-30).

– L92: The notation dw_k ∼ sqrt(dt) N(0, 1) is unusual. Can you please use the typical notation in which this is stated using non-infinitesimal increments, i.e., something like w_k(t+t_0) - w_k(t_0) ∼ sqrt(t) N(0, 1)?

We have re-written this expression following the reviewer's comment (L94; LT97).

– Eq. 5: In Liang and Kleeman (2005), they derive a different expression for discrete and continuous time. Do you need to use different expressions for LKIF here in either case?

Liang (2021) only considers the continuous flow case in his multivariate version, which is the one we use here.

– L211–2: There is not enough information about how the significance was computed. What is the null hypothesis, and how exactly was the bootstrap distribution used? Please add details. Also, the significance test for PCMCI is not described.

We have added this information at the end of Section 3.1 (L217-220; LT222-224) for LKIF and at the end of Section 3.2 (L259-261; LT266-268) for PCMCI.

– L222–4: The conditions here are described with too little detail to be useful to the reader who is not already familiar. I would suggest either taking them out and just referring to Runge (2018) or describing each one briefly.

As suggested by the reviewer, we have removed these conditions from the text and refer to Runge (2018) for more details (L228; LT232-234).

– Table 1: What does "Use of iterative conditioning" mean? This should be clarified in the text.

We explain the meaning of "iterative conditioning" in Section 3.2 (L232-244; LT238-250). We have added a sentence in that section to make it clear (L239; LT245). We also provide some explanation about iterative conditioning in Section 3.3 (L265-267; LT272-274).

– L262–263: How are lags incorporated into LKIF? This needs to be described in greater detail since it is used in the experiments.

We have added this information in the text (L275-276; LT282-283).

– Figure 1e: the analytical values for x1->x1 and x2->x2 are excluded here. Why is this?

We have now provided these analytical values in Fig. 1e.

– L295–296: In almost all cases, LKIF seems to suggest a strong self-influence even when this is not present. Do you have any explanation of this effect? Perhaps it would be good to recommend against using LKIF for detecting self-influences.

We thank the reviewer for this comment. We acknowledge that LKIF incorrectly detects significant self-influences for the 6D model (Fig. 2b), which we state in Section 4.2 (L327; LT334). We have also added a sentence following the reviewer's comment: "The latter result indicates that the LKIF method may fail in representing the correct self-influences, while PCMCI does not" (L327-328; LT334-335). However, LKIF correctly identifies all self-influences as significant for the 2D (Fig. 1b) and 9D (Fig. 3b) models. Thus, we think that our study is not conclusive in this respect and we would need to make further analysis to recommend one or the other method in terms of self-influences.

– Figure 2d: Can you please add the self-influences to this graph (as arrows from a node to itself), as well as all the other graphs?

This is now done (Figs. 1f, 2d and 4d).

– Figure 3: It would be easy to add the false negatives to these plots by having white squares with a blue (or black) rectangle for those links that exist but were not detected by a given method. Can you please do this for all the plots?

This is now done, but please note that this is only the case for $\tau_{5\rightarrow4}$ in Fig. 3b.

– Table 2: It would be nice to add something like the phi coefficient (https://en.wikipedia.org/wiki/Phi_coefficient) for each case and method, to summarize the performance with a single number.

We thank the reviewer for the excellent suggestion, which we have now incorporated in Table 2. We have also included in this table the true-negative rates to be complete. We have switched the methodological explanation of all these diagnostics from the caption of Table 2 to the new Section 3.4 (‘Comparison diagnostics’). We now discuss these results in more details in Section 5.1 (L477-492; LT485-502).

– Discussion: This paper does not test the robustness of the methods to noise, which is a crucial consideration in practice. This could be stated as another area for future work.

We thank the reviewer for this interesting suggestion, which we have added at the end of our paper (L557; LT568).

Minor:

– L36: "Taken's theorem" -> "Takens's theorem" or "Takens' theorem" (the namesake is Takens, not Taken)

Corrected (L37-38; LT39).

– L222: "such has" -> "such as"

This sentence has been re-written and does not contain “such as” anymore (L228; LT232).

References

– Liang, X. S., & Kleeman, R. (2005). Information Transfer between Dynamical System Components. Physical Review Letters, 95(24), 244101. https://doi.org/10.1103/PhysRevLett.95.244101

– McGraw, M. C., & Barnes, E. A. (2018). Memory Matters: A Case for Granger Causality in Climate Variability Studies. Journal of Climate, 31(8), 3289–3300. https://doi.org/10.1175/JCLI-D-17-0334.1

This reference has been added and discussed in the text (L27-29; LT28-30).

2. Reply to Anonymous Referee #2

The manuscript presents an interesting comparison of the output of two causality-detection methods, LKIF and PCMCI, on synthetic data generated by models of different characteristics, and real data (a set of relevant climatic indices). The manuscript is clearly written and the results are sound. Therefore, I am happy to recommend the acceptance of this manuscript, with a minor optional suggestion.

We thank the reviewer for the encouraging words.

In the abstract, the authors say "Detecting causal links from the fourth model is more challenging as the system is nonlinear and chaotic." Model 4 is the well-known 3D Lorenz (1963) model, while Model 2 is a linear 6D model, and Model 3 is a 9D nonlinear model. Can the authors comment on the role of the model's dimensionality? How the dimensionality of the model affects the performance? Which method can be expected to provide more accurate results, in the case of high dimensional systems?

We thank the reviewer for this comment. We think we partly answered this question in our study. This is summarized in our abstract: "LKIF and PCMCI display some strengths and weaknesses for the three simplest models, with LKIF performing better with a smaller number of variables, and PCMCI being best with a larger number of variables". Indeed, we show that overall LKIF is better for the 2D model, both methods are fine for the 6D model (but PCMCI performs better for self-influences) and PCMCI is better for the 9D model. However, if we think of high dimensionality including a very large number of atmospheric and oceanic fields, then further analysis would need to be done to answer this question. We have added some text related to dimensionality at the end of Section 6 (L557-558; LT568-569).