Response to Reviewer 1

Thank you very much for your review. Your suggestions are really appreciated, which helped us improving the manuscript. The answers to the review are provided below. The comments are italicized. Our responses are non-italicized and the paragraphs/sentences introduced in the manuscript are in red.

The manuscript by Stéphane Vannitsem et al. provides a significant contribution to our understanding of complex nonlinear interactions among climate modes, employing the Liang-Kleeman information flow technique. Through the examination of eight key climate indices, the authors offer valuable insights into low-frequency climate variability and nonlinear causal dependencies that shape the collective dynamics of the climate system. The study is distinguished by its rigorous methodology and its implications for advancing climate science. Overall, the paper is well-structured, clearly written, and holds considerable scientific merit. I recommend it for publication following minor revisions to address a few specific comments and clarifications.

Thank you very much for your very positive feedback on the work.

Specific Comments

The authors have made an impressive extension of Liang's method by incorporating nonlinear predictors, which allows for isolating specific quadratic nonlinearities among climate indices. This thoughtful approach offers a fresh perspective on capturing joint influences across multiple indices, greatly enriching our understanding of complex climate interactions. Additionally, the use of Singular Spectrum Analysis (SSA) to separate low- and high-frequency components for each climate mode adds valuable clarity to the dynamic relationships within the data. To further enhance the paper, a brief discussion on the effects of factors such as data length and predictor count on the accuracy and robustness of the results would be highly informative. A qualitative exploration of these potential limitations would add context to the findings and guide future applications of this methodology.

Thank you very much for your suggestion. Indeed, this is an important question that we already addressed in two previous works. In Vannitsem et al (2019), an analysis in the context of a 2 dimensional system has been made, showing that the analytical rate of information transfer can be estimated with some uncertainty which depends on the length of the series. And obviously, the shorter is the time series, the larger is the uncertainty. This implies that some dependencies may not show up.

The question of the number of bootstraps needed to estimate the uncertainty was also discussed in the same paper, and in Docquier et al (2024). In the latter the convergence was ensured for the different models investigated. We conservatively used 1000 bootstraps here for the real world application as in Docquier et al (2024). We cannot easily test the impact of the length of the series here as the series are already relatively short when interested in the low-frequency variability.

The impact of the number of predictors has also been tested in Docquier et al (2024) who suggested that the detection is better with a small number of predictors of maximum 10. This could indeed affect our results in a way that a lower number of nonlinearities would have been detected. It was, however, shown in the more theoretical study of Vannitsem et al (2024) that using a combination of linear and nonlinear terms up to 44 terms, still allow us to get meaningful results. The different conclusions reached are probably associated with the different setups used in both studies, and this question should be further addressed in the future.

We also propose to add the following text:

> Note that the time series used are relatively short. This implies that some links could not be detected as the uncertainty around the value of the rate of information transfer may be large. Given that caveat, some dependences of the method on the bootstrap sample have been explored in Docquier et al (2024). They concluded that 1000 bootstrap samples are a good choice to detect causal links on short climate time series.
>
> Another potential difficulty is the number of predictors. In Docquier et al (2024), it was indicated that to get a good detection, a small number of predictors should be used (of the order of 10). It was however shown in the more theoretical study of Vannitsem et al (2024) that using a combination of linear and nonlinear terms up to 44 predictors, still allow us to get meaningful results. The different conclusions reached are probably associated with the different setups used in both studies, and this question should be further addressed in the future.

The distinctions between low- and high-frequency influences are skillfully captured, particularly in the analysis of the NAO and El Niño indices. Interestingly, while low-frequency components reveal inter-dependencies among indices not evident in the original data, this phenomenon might stem from high-frequency components masking or "overwriting" lower-frequency signals. A brief clarification on this point would deepen readers' appreciation of the novelty and importance of the frequency filtering method used here, highlighting its role in uncovering hidden relationships otherwise obscured in unfiltered data.

Thank you very much for the suggestion. As discussed in the answer to reviewer 2, a more arbitrary choice of modes could lead to different detection. We have performed the same analysis using a set of even and odd modes among the 40 first SSA modes, and this ends up with a much lower detection of nonlinearities, suggesting, as you said, that the mixing of high and low frequencies are hindering the presence of nonlinear influences.

Moreover, more generic behavior emerges from the low-frequency filtering analysis. For most of the cases, the correlations between target and predictors (linear and nonlinear ones) grow in amplitude (three times in certain cases) when one passes from original signals to low-frequency data (see e.g. Fig. 2a,b and Fig. 3a,b). This means that most of the linear and nonlinear covariability and causality ranges from the annual to multiannual time scales for the chosen set of indices. This time-scale restriction is not very surprising since seasonal to decadal turbulent eddy activity exchanges between the ocean and the atmosphere can drive slow variability.

We added a sentence on that point, together with the answer of one point of the second reviewer concerning the sensitivity of the analysis to the number of SSA modes kept, as follows:

> In order to evaluate the impact of choosing specific SSA modes rather than others on the causality analysis, we have also considered arbitrary choices of modes, namely the even modes 2 k, and the odd modes 2 (k-1) +1, for k=1, …, 20. The analysis reveals that considerably less significant influences from nonlinearities are detected (5 instances for the odd modes and 2 instances for the even modes), indicating that these arbitrary choices are not providing optimal results. With such choices, high and low frequencies are again mixed up, leading to a rather suboptimal result. The presence of high frequencies is indeed hindering the proper detection of dependencies given the short time series.

The figures are generally clear and highly informative. To further support the study's central conclusion—that "nonlinear causal dependencies are a hallmark of the complexity of climate dynamics"—a comprehensive visualization would be a valuable addition. For instance, a general illustration showing the overall inter-dependencies between each mode could provide a clearer comparison of how nonlinear dependencies statistically differ from those in a linear or null model. This would enhance the manuscript's narrative, offering readers an intuitive grasp of the paper's findings and reinforcing the significance of the nonlinear dependencies identified.

Thank you very much for the suggestion. This is indeed a good idea, although it is a quite complicate situation. We have tried to do this and we end up with a set of visualizations, that could now be introduced in the manuscript and in the supplementary material. An example for the NAO is shown in Figure 1.

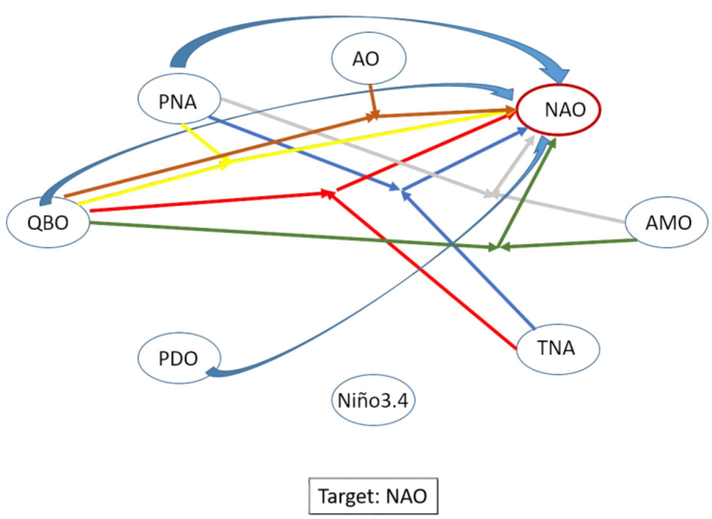

Figure 1: A visual representation of the linear and nonlinear influences from the set of indices to the NAO for the low-frequency variability dataset. the curved plain arrows refer to the linear influence, the triple straight arrows with the same colors refer to the influence of the quadratic nonlinearities: two straight arrows are emanating from two indices, joining somewhere in between the indices, and from there the third straight arrow indicates the target.

Overall, this manuscript is a thoughtful and impactful piece of research, with a few refinements adding further depth and clarity.

Thank you very much for your positive evaluation of our research and your suggestions.