Response to Reviewer 2

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.

In this paper, the authors apply a method developed in previous works to the particular case of different climate indices. The work is an addition to the literature on non-linear methods reporting complex, significant, non-linear interactions between teleconnections and large-scale patterns of the climate system.

Thank you very much for your evaluation of our manuscript and your constructive suggestions.

Some comments below:

Sec.3.1

- *It should be stressed that the SSA is applied to a single time series, or the comparison with EOFs is very misleading.*

Thank you very much for pointing out this aspect. We have slightly modified the text to stress the similarities/differences of the method:

> The Singular Spectrum Analysis (SSA) shows similarities with the principal component analysis where a covariance matrix is diagonalized. In the SSA, the lag-covariance matrix of a single time series is diagonalized where the eigenvectors or Temporal-Empirical Orthogonal Functions (T-EOFs) are finite time sequences providing the more frequent and higher amplitude finite time-spells of that variable.

- *The authors say that the modes of Table 1 have been chosen arbitrarily. Still, as a justification is not sufficient, they should explain the logic that allowed them to perform the reduction. Furthermore, there should be a minimum study of what would be the impact of another choice (for example, random).*

We first rewrote the part of the text as follows, hoping that now the logic is clearer than before:

> For each index, we compute the SSA spectrum and evaluate visually each of the 40 SSA modes corresponding to the 40 dominant eigenvalues. These different modes have a time length of 240 months (the M-window mentioned earlier). If the dominant period in each mode evolution is shorter than a year, the mode is discarded, the idea being to keep the low-frequency variability large than a year, only. After filtering out the modes displaying high frequencies, we end up with new low-frequency variability series of the original monthly anomalies of the climate indices (the monthly mean has been removed before the application of the SSA). The modes that are kept in the low-frequency signal are listed in Table 1. There is here a certain degree of arbitrariness as we discard sometimes modes that display a mix of low-frequency and high-

> frequency variabilities. We do believe however that the essence of the LFV dynamics is well captured by our selection. Note also that the LFV of most of the oceanic modes are essentially concentrated in the dominant SSA modes of variability.

The suggestion is very interesting. We have done similar analyses by selecting arbitrarily the modes in two ways, first only selecting the even modes from 2 to 40, and second selecting the odd modes from 1 to 39 of all the indices. In that case, low-frequency variability and high-frequency variability may mix up. The analyses have been performed using the whole set of linear and nonlinear predictors. For the odd selection of modes, the influences which are significant at the 1% level are: AO ➔ NAO; NAO and QBO^2 ➔ AO; PNA, AO PDO, TNA QBO and AMO^2 ➔ El Niño; NAO, PNA, El Niño ➔ PDO; NAO, QBO^2 and AMO ➔ TNA. For the even modes: AO ➔ NAO; NAO, PDO ➔ AO; El Niño ➔ PNA; TNA, El Niño ➔ AMO; PNA, El Niño ➔ PDO; PDO and TNA ➔ El Niño; NAO, AMO, El Niño, AMO^2 and PDO^2 ➔TNA. Note that all the significant influences associated with the nonlinearities are very small (around 1%). Moreover,

A general remark is now that only a few nonlinearities (5 instances for the odd modes, and 2 for the even modes) are involved, suggesting that the visual selection made above provides a good choice to isolate the impact of nonlinearities.

We will add following the text above:

> 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.

Sec. 3.2

- *Throughout the section, terms like "false positive" or "true negative" are frequently used in singular form when they should be plural.*

Thank you very much for pointing this out. This will be corrected.

- *Lines146-154: The discussion in this part is not sound. By definition, causality requires distinguishing between past and present. A more sound interpretation would be that the method can detect influences propagating at the time-step scale (likely due to the finite derivative used) but is inherently unsuitable for lagged interactions since it does not account for any lag by construction.*

Thank you very much for pointing this out. The text has been reformulated by removing the question of the convergence of the lag to zero as follows:

The approach proposed by Liang allows for constructing a network of directional connections between observables that are measured concomitantly. This approach is distinct to techniques that assume that causation should be based on a time lag between events like the classical network approach (Runge, 2019; DiCapua et al, 2020a). If real processes indeed display a lag -- like for instance in the propagation of a wave --, the information will propagate from one point to another, and this will be isolated in the Liang's method through a specific path through the network. As in reality we usually do not have all variables (at all grid points for instance), this could not show up, but filtration through (spatial or temporal) averaging or frequencies selections should help in disentangling the impact of one distant observable to another, as for instance in Vannitsem et al (2022).

- *The dependence of the method's results on its hyperparameters, such as time series length, should be discussed and reported.*

This question has already been addressed in other papers. For instance, 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. 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.

We will add the following paragraphs on that aspect in the current 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 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.

Sec. 4.

- *The quality of the figures is rather poor. The choice of "**2" for the squaring operator is unfortunate and could be improved.*

We modify the figures accordingly.

- *The reasoning behind choosing a quadratic non-linearity requires further elaboration. The authors mention its presence in "many dynamical systems" but they should tailor the choice depending on what is currently known regarding the teleconnection they are using and their interactions. If no specific knowledge is available, the reasoning should be physically sound. The authors also mention the quadratic terms as second-order terms in Taylor expansion, which makes sense only if these quadratic terms a smaller than the linear counterparts, which apparently is not checked or discussed afterwards.*

Thank you very much for pointing out this weakness is the justification. Stopping at the second order of the Taylor expansion could indeed point out to the smallness of these terms. They are probably small as these are not detected in the original series, but we agree that the justification is weak. In fact, in the dynamical equations issued from the classical conservation laws, such quadratic terms arise naturally, associated with the nonlinear advection terms in the equations. So, we rewrite the text as follows:

> To disentangle the role of nonlinearities in the context of our 8 climate indices, all combinations of quadratic terms are constructed. This choice is made since in many dynamical systems such nonlinearities are present. These quadratic nonlinearities are typically associated with the presence of nonlinear advection terms in the classical conservation laws (momentum equation, thermodynamic equation, …), as for instance illustrated in the work done recently in the context of the Charney – Straus model (Vannitsem et al, 2024). These however could be viewed as restrictive and tests should be done in the future to evaluate the impact of higher order or more complicate nonlinearities.

Sec. 5.

- *The first sentence of the conclusions is not clear.*

Thank you for pointing this out. We rewrite it as follows:

> This work investigates the quadratic nonlinear influences on a set of climate indices. This is done by considering, as predictors in the analysis, products of indices. This extension to nonlinear predictors has proven to be successful in the idealized context of a reduced order model (Vannitsem et al, 2024). A few key conclusions may be drawn from this analysis:…

The sentence "This would also allow to build a reduced-order data-driven climate model that could potentially help in understanding the global climate evolution." is not supported in both its claims.

Thank you for pointing this out. We will modify the sentence as follows:

> This knowledge could also provide hints on the nonlinearities that would be useful to build a data driven model of the large scale climate indices.