Second Reply to Referee #2 (Milan Paluš)

Manuscript egusphere-2022-942
"The rate of information transfer as a measure of ocean-atmosphere interactions"
D. Docquier, S. Vannitsem, A. Bellucci

We would like to thank the reviewer M. Paluš for his follow-up comments regarding our analysis, which allowed to considerably improve our manuscript. Below we present the reviewers' comments (in black) **and our point-by-point replies (including changes in the paper) (in bold blue).**

Comments

First, I have to apologize for mixing bivariate and multivariate results of applications of the Liang information flow (LIF) to coupled Roessler systems in my first review, and thank the authors for recomputing the various set-ups of this example. Still we see that the results depend on sampling frequency, and none case reproduced the result I have presented. The latter was my reproduction of original results of Liang. The reason - according to my experience - is that the multivar LIF estimate depends also on the arrangement of variables in the matrix used in the formula.

We agree with Dr. M. Paluš that results of information transfer (multivariate case) depend on the sampling frequency in the case of the Rössler systems, as shown in <u>his previous comment</u> and in <u>our previous reply</u> to his comment. We now emphasize this aspect in the final paragraph of Section 3.5 (which has been renamed from "Linearity assumption" into "Limitations of the method") and we added results carried out with the Rössler systems in a new Appendix C to illustrate that.

I appreciate the work the authors have dome by computing the conditional mutual information (CMI). The bivariate case I(THF(t);SST(t+1)|SST(t)) and the opposite showed that the causality between the analysed variables is not symmetric. This formulation of CMI is a causal measure, equivalent to transfer entropy (TE). Unfortunately, the trivariate case that the authors present, is not a causality measure. To get causal CMI/TE, one must condition on the presence/history of the effect variable. Putting a third variable in the condition one gets just CMI - a generalization of partial correlation. At the beginning I believed that the authors used lagged variables and that there is some symmetry in dependence, though not in causality. But in the revised MS the authors claim that the trivariate CMI was computed without lagging the variables - then the symmetry obtained is just trivial result due to basic properties:
I(X;Y)=I(Y;X)
I(X;Y|Z)=I(Y;X|Z)
I believe that the authors gladly accept my proposal for another revision of the manuscript in order to avoid publishing this incorrect interpretation.

We thank Dr. M. Paluš for this clarification related to CMI. Following his suggestion, we removed the results carried out with CMI in 3D. We decided not to include any result using the CMI method in our manuscript as a much deeper analysis, which is beyond the scope of the present study, would be needed to compare both methods adequately.

But back to the topic - we do not see any support from CMI for the symmetric causality in 2D LIF. Still I believe it is the same artefact as in the case of 2D estimation from Roessler systems. The authors should consider excluding the 2D case. ALso their explanation of differences between 2D and 3D case is not very understandable for me, maybe this could be avoided by removal of 2D case.

We agree with Dr. M. Paluš that the 2D case should be taken with caution due to the apparent symmetry found between the two directions when looking at causal interactions between SST and THF (Fig. 2), and between SST tendency and THF (Fig. 3). This symmetry is indeed present with the unidirectionally coupled Rössler systems when using the 2D Liang formula (see Fig. 2 in M. Palus' previous comment and Figs. C-D in our previous reply). However, we think that showing the 2D case in the paper is useful for two main reasons.

First, this symmetry is not present everywhere in the 2D case, especially for the causal interactions between SST tendency and THF (Fig. 3), where many regions present a significant positive or negative rate of information transfer τ in one direction and no significant τ in the other direction. For example, in the North Atlantic, there is a large band of positive τ from SST tendency to THF (Fig. 3a) and no significant τ from THF to SST tendency (Fig. 3b). Another example: south of the Gulf Stream, both directions present a region of significant positive τ . Many other exceptions are present.

Second, this is also a way to show the superiority of the 3D case compared to the 2D case. The multivariate approach from Liang (2021) is very recent and has not been used a lot compared to the original 2D case, thus we think it is useful to show differences arising when using the bivariate or multivariate approach. As we already wrote in the previous version of our paper, "the analysis done so far with two variables may provide a false impression of a two-way influence emerging due to the absence of a set of hidden variables" and "the 3D case provides additional sources of information compared to the 2D case and should thus be preferred in terms of result interpretation". So we think that showing the 2D case, which may be affected by the absence of an important variable in the system and a spurious symmetry (as said above), is still very interesting from a didactic point of view as a detailed comparison to the 3D case shows the superiority of the latter in the case of our study.

In our revised manuscript, we are much more careful about the use of results in the 2D case and we emphasize more the superiority of the 3D case compared to the 2D case. We added a discussion about the caution needed to interpret results from the 2D case in Section 3.5 (second limitation). We hope the manuscript is now clearer in this respect.

4D case: could the authors state explicitly the variables included and give also their order in each estimate?

We mean the 3 variables used throughout our study (SST, SST tendency and THF), plus THF leading SST by one month (THF(-1)). This information clearly appears in Fig. 6. We have added a sentence at the beginning of Section 3.3 to remove any ambiguity. The order of variables does not have any influence on the computation of the multivariate rate of information transfer, whatever the number of variables considered. Some tests in which we switched the order of variables were carried out with the Rössler systems to confirm this and results are identical to the ones shown in our previous reply and in Appendix C. This can been easily demonstrated by looking at equation (1) in our paper: if one or more variables are switched in the matrix of variables, this does not change anything to the different terms of the

equation (same determinant, same covariance terms, same sum of cofactors and covariance between variables and variable tendency).

Long scales- were the data resampled, e.g. each 12 samples were substituted by their mean, or the same number of samples remain, just moving average smoothing was done? Cannot the increase of causality strength be just a trivial consequence of smoothing the data, not the effect of the scale?

The same number of samples remains (second approach) when we compute the 12-month and 120-month running mean (moving average). We clarified this point in Section 3.4. The first approach (data resampling) would have resulted in a too low number of data points for computing the rate of information transfer, thus we preferred using the second approach (moving average). The smoothing resulting from the computation of the moving average does not necessarily increase the rate of information as shown in Vannitsem & Liang (2022), where there is no evidence of an increase in τ when using a sliding window.

Btw. LIF is not a time-lag free method, as claimed in the MS, the lag 1 is applied in computing the differenced series which is included in the formula.

The method does not explicitly take lags into account but rather indirectly through the computation of the Euler forward difference approximation dX/dt. So yes, we agree with the reviewer that the method is not entirely time-lag free. We added a sentence about this at the beginning of Section 3.3.