

Geosci. Model Dev. Discuss., author comment AC4
<https://doi.org/10.5194/gmd-2021-196-AC4>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Reply on RC2

Chin-Hsien Cheng and Simon A. T. Redfern

Author comment on "Quantifying Causal Contributions in Earth Systems by Normalized Information Flow" by Chin-Hsien Cheng and Simon A. T. Redfern, Geosci. Model Dev. Discuss., <https://doi.org/10.5194/gmd-2021-196-AC4>, 2021

Responses to referee 2:

The paper uses the methodology developed of Liang over the last years for the information flow in coupled systems. It advocates the use of normalized information flow, or nIF , to (i) characterize causal relationships of the coupled variables and (ii) to assess the quality of model simulations from a causality perspective.

The paper is timely in the sense that causal methods are increasingly popular, with very different approaches nicely summarized e.g. by the Runge et al. (2019) paper cited in this manuscript where, as far as the reviewer remembers, the information flow of Liang is missing. So this could be a welcome contribution to causal approaches in the geosciences.

Unfortunately, the paper is excessively difficult to read. It starts with the equation (2) - which is eq. (5) of Liang (2021) in Entropy, which should be referenced to, but this has been mentioned already in the comments - which to the uninitiated comes out of nowhere, and is not motivated but only described. Why should this particular algebraic combination of covariances be called "information flow"? In the same spirit, how should a reader easily grasp the meaning of the normalization factor Z from eq. (4)? How do you separate the changes in marginal entropy into the self-dependent and the noise term for observed (as opposed to generated) time series; in other words, how do you calculate Z and nIF in this situation? There is no hint given in ch. 2.1.

The key to this paper is exploration of the hypothesis that a proportional relationship exists between causal sensitivity between two coupled variables and the IF , nIF or $md.nIF$, (i.e. equation 7), and thus its useful but conditional applications. We feel that further detailed theoretical discussion of Liang's earlier papers will not help clarify this, especially since the normalizing method is still being researched and has scope for potential future development. However, we are very grateful for the comments on different ways of thinking of Z which has driven us to attempt to answer these questions by developing the use of $md.nIF$ as outlined above.

Next, extensive use is made of an artificial example (called mock-up data), but should the reader be interested in this, she has to refer to the Supplementary Material. Here, she finds a system of two coupled first-order equations for the variables $X1$ and $Y1$ - written in a manner more complicated than necessary, since the overall term $dY1/dt$ is not taken out but repeated three times, and with seemingly arbitrary numerical constants (1.1, 1.5,

300, 1.8, 0.000005 and so on) - which is not motivated by any means. The reviewer also wonders in which sense the noise terms on the right side of the Table deserve that name - being a sum of a constant, a deterministic trigonometric function of time, and the function itself? How could that be noisy, where is a stochastic process involved?

The Table is placed in the Supplementary because it does not provide much information beyond which equations were used. The Figures in the main text present the results that explain the situations where the proposed proportional relationship works better than regressions. We will improve the clarity of these Figures and the Tables. Regarding the "arbitrary numerical constants", these refer to the calibration factor alpha in equation 1 which can be removed from the legends and labels in the revision.

In the main text, there is talk about X and Y (not X1 and Y1) and in Figure 2, we suddenly have d(partial) X1, dX2 and dX3 - what are these? In the heading of the Figure, it says that dX2=dX1 and dX3=dX1, so it is only one quantity after all? The graphs show three curves, so they are again different? The reader should look for red boxes according to the legend - there are no red boxes in Fig. 2. And what do we see on the y axis, actually? The variables themselves? Their partial derivatives? The information flow? What does the axis legend "by mR^2" even mean?

We thank the referee for pointing out this error in the original submission. The legend should be " $\frac{\partial X_2}{\partial t} = 2 \frac{\partial X_1}{\partial t}$ ", " $\frac{\partial X_3}{\partial t} = 3 \frac{\partial X_1}{\partial t}$ ". We missed out the 2 and 3 in front of $\frac{\partial X_1}{\partial t}$. The Figure captions will be corrected to eliminate this omission. The red boxes highlight trends with negative values of the ordinate, which represent the designed and estimated causal contributions (i.e. the partial derivatives) using the variety of methods explored. Thus, when we state "by mR²" we refer to estimates based on mR² as the multiplier in equation 1.

At the beginning of ch. 3.1, there is talk of "where causality becomes more important" - as opposed to what? And how do you know that, given observations and measurements (only)?

This word "important" refers to relative conditions when such causality is not captured or misinterpreted by regressions. We will rephrase this sentence.

The reviewer didn't get the concept of the "1:2:3 ratio" either nor how this could convert (apparently at high noise level) to a "-1:-2:-3 ratio" - isn't that the same since pairwise signs would cancel out?

Noise operates differently from regressions since both the sign and magnitude of the gradient is influenced by the noise. Hence, when the sign is incorrect, the cancelling effect for the 1:2:3 ratio may be result in a set of ratios in the order -3:-2:-1. This is seen in sub-Figs a,e,i of the new Figure R1 described above. However, this does not greatly affect the magnitude of IF, nIF, and md.nIF, especially the IF. Hence, 1:2:3 may turn into -1:-2:-3 when the noise is too large, resulting in the correlation sign misrepresenting the direction.

For the comparison of observations and model runs, here for CH4 growth rates, the reviewer has a hard time to discern the upper panels of Fig. 6 and 7. They look exactly identical, apart from the fact that the time axis for Fig. 7 is shorter (up to 2012). CESM2 can't be that "perfect"? Also, the observations/simulations give a rather blurred image along the latitudes and the time axis, whereas the estimates have a finer resolution. How is that possible, and how do the authors come to the conclusion that nIF is doing best, and that CESM2 fails to reproduce the spatial pattern?

Yes, the first rows are the same, because we were testing the observed and modelled temperatures and precipitations against the observed methane concentrations. A failure to

capture the causal trend based on either modelled C_{CH_4} or observed C_{CH_4} implies improper underlying model processes. In fact, a typical research method for understanding the methane-climate feedback processes is to tune various parameters in process-based models to explore how to best fit the observed concentration and isotope-ratio trends. It is hence justifiable to show the inadequacy of CESM2 based on modelled climatic variables and observed C_{CH_4} .

The relatively 'blurred' image from the observation is due to smoothing in reconstructing the observed trends. The apparently sharper resolution of estimated contribution trends is not a result of image resolution, but has two possible origins: i) the estimated contributions better reflect the location of the causes of net-emissions, while the observations are affected by atmospheric mixing; ii) the nIF method tends to accentuate sharp fluctuations due to the difficulty to differentiate whether $|nIF|$ is approaching 0 or 1.

We will also revise the examples to show estimates based on zonal C_{CH_4} instead of (or in addition to) estimates based on global C_{CH_4} . Also, we will include estimates based on $md.nIF$.

The paper raises a lot of questions. It requires a substantial revision (major) before having the chance to come close to be readable. The potential of the method is present, but it has to be motivated much more explicitly and the examples have to be explained, shown in the main paper (a time series graph of the very target variables X and Y would be handy) and then a clearer demonstration and indication why nIF and its variants are superior to a regression approach, or in other words, why correlation and causality are different concepts, to the extent that you can have two causally connected variables with a Pearson correlaton coefficient of zero.

We are grateful to the reviewer for their comments and will indeed carry out major revision to improve readability and clarity.

A promising application of the framework seems to be to determine the effective lag between cause and effect by lagging one of the two such that the nIF value (or one of its variants) is maximized, and compare this to a conventional cross correlation analysis. This is merely a suggestion since these lags are to be expected, in particular in the context of teleconnected variables.

This will require further analysis. For $|nIF|$, we wonder if this lead-lag effect is more accurately captured when the $|nIF|$ is around 0.5, which is often the case when there are significant causal contributions. However, some sharp spikes due to the $|dH_Y^*/dt|$ term may mislead the lead-lag information. We will explore if this problem can be mitigated by $md.nIF$.

Citation: <https://doi.org/10.5194/gmd-2021-196-RC2>