Responses to the Referees’ Comments to Quantifying Causal Contributions in Earth Systems by Normalized Information Flow

We appreciate all the comments from referees. Before we proceed to detailed responses, we would like to highlight that our manuscript focuses more on the practicality of the method and hence lacks detailed theoretical background. Nevertheless, we would like to briefly share some further points that may help clarify some concerns and explain the results we have obtained:

When $R^2$ in regression approaches 1 with very strong causal strength, $|nIF|$ tends to approach 0.5 instead of 1. This differs from our earlier understanding. By carefully looking into the contributions between the three terms of $Z$ (equation 4): $|IF_{X\rightarrow Y}|$ and $|dH^*_{Y}/dt|$ tend to become similar to each other and the $|dH^{\text{noise}}_{Y}/dt|$ term approaches zero, resulting the $|nIF|$ approaching 0.5. Furthermore, small change in mock-up data value may often lead to contrasting values of $|nIF|$ at $\sim 0$ and $\sim 1$. This partly explains some of the sharp fluctuations of estimates by $|nIF|$ in our results. Nevertheless, an approximately proportional relationship between the causal sensitivity and $|nIF|$ still holds. This is because most significant causal contributions occur when $|nIF|$ is around 0.5. Comparing these outcomes to regression, in regression the range of $|m|/	ext{maximal}|m|$ lies from 0 to 1 while the corresponding range is 0 to 0.5 for $|nIF|$. The calibration factor for $|nIF|$, alpha, should hence be approximately equivalent to the 2x maximal $|m|$ when $R^2 \rightarrow 1$. We have also explored the results by removing the $|dH^*_{Y}/dt|$ term from $Z$. Not surprisingly, this modified $|nIF|$ (abbreviated as md.$|nIF|$) now approaches 1 when there is strong causal influence without any other causal driver (See Fig below). This also slightly improves the accuracy of estimated causal contributions, as compared to the original $|nIF|$. To summarize the pros and cons of various methods: both $|IF|$ and $|nIF|$ are reflective of (and approximately proportional to) causal sensitivity (equation 7) to certain extent. However, $|IF|$ is not normalized and problems surface when the ranges of variability differ; $|nIF|$ helps (imperfectly) to normalize the causal signal, but the issues with non-specified causal terms $|dH^*_{Y}/dt|$ and $|dH^{\text{noise}}_{Y}/dt|$ do lead to other effects. On the other hand, regressions reflect causality poorly (a high $R^2$ may still occur in a
situation of nil-causality), and a rather low $R^2$ also weakens the estimates of causality due to several effects, including the presence of extreme values, the occurrence of alternating sign within the analyzed period, and significant lead-lag times between cause and effects, alongside the problems associated with excluding strong noise contribution. Figure R1 (below) shows a case with a strong single-directional causal relationship (Y influences X but not the other way round). In the revised manuscript, we highlight the practicability of the (modified) normalized information flow for quantifying causal contributions, but we will also call for further study to improve methods to normalize the information flow.

Figure R1. The first column shows designed and estimated causal contributions from Y1 to X1, X2, and X3. Note that $\frac{\partial X2}{\partial t} = 2\frac{\partial X1}{\partial t}$, $\frac{\partial X3}{\partial t} = 3\frac{\partial X1}{\partial t}$, and this 1:2:3 ratio is applied to the calibration factors for $IF$ and $nIF$. Due to the influence of another constant positive contribution to X1 (the difference between $c$ and $a$), those negative causal contributions are not detected by regressions, while estimates by $IF_a$ (subscript “a” denotes the adjusted sign by correlation), $nIF_a$, and modified $nIF_a$ still show those negative contributions to some extent, if the negative correlation sign is detected. In addition, the 1:2:3 ratio is captured by estimates based on $IF_a$, $nIF_a$, and $md.nIF_a$ as this ratio is accounted for in the calibration factor. However, the gradient $m$ in regression is corrupted by a further constant positive contribution, and hence the 1:2:3 ratio is lost. Instead, for negative contributions with -1:-2:-3 ratio, it becomes more similar to 3:2:1 (sub-Figs. a,e,i). The second column shows the designed and estimated causal contributions from X1 to Y1, Y2, and Y3, which is zero (nil-causality). The nil-causality is not captured by regressions with $R^2 \rightarrow 1$ (sub-Fig f,h,j,l). It is best captured by $|IF|$. However, due to the extremely low $|IF|$ (maximal $|IF(X1\rightarrow Y1)| \sim 3.4e-15$), if there are even smaller signals from $|dH^*_{Y}/dt|$ or $|dH_{\text{noise}}/dt|$ detected, the $|nIF|$ or $|md.nIF|$ will then be non-zero or even approaches 0.5 or 1 (sub-Figs t, x). Also note sub-Fig. s and w. The values of $|nIF|$ tend to show a sharp change between 0 and 1. This may be associated with progress of wave. For some circumstances, it is impossible to differentiate the cause and effect, e.g. between two time-series sin(x) and sin(n-x) (also refer to Liang’s 2021 paper on Entropy: A Note on Causation versus Correlation in an Extreme Situation). For this reason, we remove the $|dH^*/dt|$ term for the modified $nIF$ and the $|md.nIF|$ now also approaches 1 when $R_2 \square 1$ with real causality (compare boxes g, k, and x). In this case, the calibration factor should then be approximately equal to the maximal $|m|$ when $R_2 \square 1$, since $|m|\max(|m|)$ between 0-1 will correspond to the $|md.nIF|$ if $R_2 \square 1$ and real causality is present (not in shown in this example).

With these new observations, we will further revise the manuscript with major revision