

Referee #3

This manuscript applies the multivariate, time-varying LKIF framework to compare bivariate and multivariate information-flow estimates via ΔIF and related diagnostics, focusing on SM/VPD $\rightarrow$ LAI/GPP couplings in Southeast China and the Amazon Basin. The topic remains relevant and the paper is generally readable. Relative to the previous round, the authors have now provided acceptable responses to most of the comments raised earlier. However, one central issue remains unresolved, and it concerns not a peripheral sensitivity test but one of the paper's most important empirical findings. For Southeast China SM $\rightarrow$ GPP (Fig. 1c), the bivariate IF is mostly insignificant except around 2011, whereas the conditioned estimate SM $\rightarrow$ GPP|VPD,T,SSR is significant over multiple periods. Conditioning therefore changes both magnitude and significance. This is precisely the kind of confounding-/conditioner-related behaviour that the manuscript claims to diagnose. Yet the revision still does not provide a clear account of why this happens—neither from the observational causal analysis itself, nor from an idealized dynamical model, nor from a physically grounded mechanistic interpretation. Without such an account, the credibility of the main conclusion is substantially weakened. Major revision is still required on this point.

Comments on title/scope wording, Appendix A's limited structural coverage, lag memory, Fig. 2 year selection, IF-sign interpretation, and Fig. 3 missing $|rIF|$ cells have been answered at an acceptable level for this round (through clarification, limitation statements, and/or textual correction). I do not request further major work on those items beyond ordinary consistency checks.

We thank the reviewer for their thorough and constructive comments.

Below, we provide a point-by-point response to each comment. Reviewer comments are shown in black, our responses in blue, and the corresponding revisions in the manuscript are highlighted in yellow.

However, previous Major Comments 2–3 noted that Appendix A emphasizes directional fidelity and near-overlapping bivariate/multivariate curves for $x_1 \rightarrow x_2$ and therefore does not support the Fig. 1c pattern. In the response, the authors agree that Appendix A does not reproduce conditioning-induced significance transitions and add limiting text. That is acknowledgement, not explanation.

What is still missing is a coherent treatment of this finding itself: (i) a focused observational diagnosis of why SM $\rightarrow$ GPP is weak/insignificant in the bivariate form but significant after conditioning on VPD, T, and SSR—Section 3.1 still speaks of “little adjustments” for SM-driven GPP couplings, which is inconsistent with Fig. 1c and the later diagnostics; (ii) a controlled dynamical example in which a strong direct pathway is masked bivariate and recovered under multivariate conditioning (referring to a different synthetic pathway with only modest amplitude change does not address Fig. 1c); and (iii) a clear land–atmosphere mechanistic interpretation of why the SM $\rightarrow$ GPP signal should be suppressed in the bivariate projection and revealed after conditioning on atmospheric demand/energy variables. Absent such an account, the result risks being read as an estimator or significance-threshold artefact rather than a physically meaningful confounding/conditioning effect.

Because this pattern is one of the manuscript's most consequential empirical claims, leaving it unexplained greatly reduces the paper's credibility. A satisfactory revision must therefore provide a clear exposition along at least one of the three lines above (data diagnosis, idealized model, and mechanistic interpretation), and must correct the inconsistent wording in Section 3.1.

Response: We thank the reviewer for this comment. We agree that the Southeast China SM→GPP result required a clearer explanation. We have revised the manuscript accordingly by updating Section 3.1, adding supporting analysis in Section 3.5, adding a synthetic experiment to the Supplementary Material and referencing it in Appendix A, expanding the Discussion, and revising the Conclusion to avoid overinterpretation.

Revised text in Section 3.1:

Figures 1c–d show IFs of SM/VPD to GPP, and their multivariate forms for both regions. For the SM→GPP coupling, conditioning produces a pronounced change in Southeast China (Fig. 1c, top panel). The bivariate SM→GPP IF is statistically significant in only 13 out of the 180 months (7.2%), whereas conditioning on VPD increases the number of significant months to 116 (64.4%). Conditioning on VPD and T, or VPD and SSR, or VPD, T and SSR also provides a large number of significant months (105, 115 and 105 months, respectively; Supplementary Table S1). Thus, conditioning substantially modifies the inferred SM→GPP coupling in Southeast China, particularly the frequency with which the estimated IF is statistically significant (Supplementary Note S1). In contrast, the Amazon Basin (Fig. 1c, bottom panel) shows a decrease in the number of significant SM→GPP months, from 73 out of 180 months (40.6%) in the bivariate form to 41 months (22.8%) in the fully conditioned form (Supplementary Table S2). This contrasting behavior indicates that the influence of conditioners on the SM→GPP coupling is region-dependent, even when the same set of variables is considered. Pronounced differences between the bivariate and multivariate estimates are also observed for the VPD→GPP couplings (Fig. 1d), showing their greater sensitivity to the selected conditioning variables. These results reveal that dependency on conditioners is both spatially and temporally sensitive, given the same variables. The changes towards positive or negative IF are strongly impacted by ecosystem stress, here more consistent with drought occurrences in the region (Sun and Yang, 2012; Feldpausch et al., 2016). These results reveal that dependency on conditioners is both spatially and temporally sensitive, given the same variables. Thus, the influence of conditioning variables should not be assumed to remain stationary across time or space.

Supplementary Tables S1 and S2:

Supplementary Table S1. Statistical significance of the SM→GPP IF in Southeast China under progressive conditioning. Significant months are reported as counts and percentages out of the 180-month analysis period. Mean IF (nats/month) is calculated only over months in which the corresponding IF estimate is statistically significant.

SM→GPP IF	Number of significant months	Percentage of significant months (%)	Mean IF over significant months (nats/month)
Bivariate	13/180	7.2%	0.0352
Conditioned on VPD	116/180	64.4%	0.0461
Conditioned on VPD + T	105/180	58.3%	0.0472
Conditioned on VPD + SSR	115/180	63.9%	0.0456
Conditioned on VPD + T + SSR	105/180	58.3%	0.0466

Supplementary Table S2. Statistical significance of the SM→GPP IF in the Amazon Basin under progressive conditioning. Significant months are reported as counts and percentages out of the 180-month analysis period. Mean IF (nats/month) is calculated only over months in which the corresponding IF estimate is statistically significant.

SM→GPP IF	Number of significant months	Percentage of significant months (%)	Mean IF over significant months (nats/month)
Bivariate	73/180	40.6%	-0.0225
Conditioned on VPD	50/180	27.8%	0.0175
Conditioned on VPD + T	43/180	23.9%	0.0224
Conditioned on VPD + SSR	54/180	30.0%	0.0124
Conditioned on VPD + T + SSR	41/180	22.8%	0.0205

Addition to Section 3.5:

This dominant VPD contribution is also consistent with the change in statistical significance described in Section 3.1, where adding VPD as a conditioner produces the largest increase in the occurrence of statistically significant SM→GPP IF (Supplementary Table S1).

Paragraph added to the Discussion:

This state dependence is particularly evident for the SM→GPP coupling in Southeast China, where conditioning on VPD substantially increases the temporal persistence of statistically significant IF compared to the bivariate form (Section 3.1; Supplementary Table S1). This behavior is consistent with the joint regulation of vegetation productivity by soil water supply and atmospheric demand; low SM restricts plant water supply, whereas high VPD increases atmospheric water demand and can induce stomatal closure, thereby reducing photosynthetic carbon uptake (Novick et al., 2016; Yuan et al., 2019). Because SM and VPD covary through land–atmosphere interactions, distinguishing their individual influences on GPP in a bivariate analysis can be challenging. In our analysis, the bivariate SM→GPP pathway is significant primarily around the 2011 drought (Sun and Yang, 2012). Under such dry conditions, increasing soil water limitation can constrain stomatal conductance and photosynthetic carbon uptake, thereby increasing the sensitivity of GPP to SM (Fu et al., 2022). This may help explain why the SM→GPP relationship becomes detectable even without conditioning. Conversely, the VPD-conditioned coupling remains significant over a much larger fraction of the record. This contrast is consistent with partial masking of the SM-related contribution by shared SM-VPD variability in the bivariate representation, whereas accounting for VPD

allows the SM-related contribution to be detected more consistently. Temperature and SSR may further modulate this separation through their influence on atmospheric demand and photosynthetic energy availability, consistent with the higher-order hydroclimatic interactions identified in Section 3.4. The Amazon Basin shows the opposite significance response (Section 3.1; Supplementary Table S2), indicating that the effect of conditioning on the statistical significance of SM→GPP is region-dependent and can operate in either direction.

Addition to Appendix A:

Because conditioning can affect not only IF magnitude but also its statistical significance, we performed an additional 3-variable controlled experiment specifically designed to examine this transition. With a fixed direct pathway, progressively strengthening an opposing common-driver pathway causes the bivariate estimate to become statistically non-significant, while the corresponding conditioned estimate remains significant (Supplementary Note S1; Supplementary Fig. S13).

Addition to Supplementary Materials:

Supplementary Note S1. Synthetic data experiment

To complement the hidden-confounder experiment presented in Appendix A, we perform an additional controlled synthetic experiment designed specifically to determine whether a prescribed direct source-to-target coupling can appear weak or statistically non-significant in a bivariate formulation but remain detectable when an observed common driver is included as a conditioner, as in our real-world case study (Fig. 1c). The experiment uses a three-dimensional, linear, stable first-order Vector Autoregressive (VAR1) process defined as (Eq. S1);

$$\mathbf{x}_t = A\mathbf{x}_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \Sigma), \quad (\text{S1})$$

where $\mathbf{x}_t = (x_{1,t}, x_{2,t}, x_{3,t})^\top$ are the three different time series and $\Sigma = \text{diag}(b_1^2, b_2^2, b_3^2)$, with b_i being the stochastic perturbation amplitudes. The coefficient matrix A is (Eq. S2);

$$A = \begin{pmatrix} \phi & c_{21} & c_{31} \\ c_{12} & \phi & c_{32} \\ 0 & 0 & \phi \end{pmatrix} \quad (\text{S2})$$

where $\phi = 0.44$, $c_{21} = 0.15$, $c_{31} = 0.25$, and $c_{12} = 0.30$ are held fixed, whereas c_{32} is varied. The innovation scales are $(b_1, b_2, b_3) = (0.35, 1.15, 0.53)$. The fixed coefficients are assigned different magnitudes because the pathways play different dynamical roles. The direct $x_1 \rightarrow x_2$ pathway is the strongest fixed cross-variable coupling the $x_3 \rightarrow x_1$ common-driver pathway is intermediate, and the $x_2 \rightarrow x_1$ feedback is weaker. These parameter values were chosen to define a stable asymmetric synthetic system for the controlled experiment and were not estimated from the observational data. Expanding the (VAR1) system by rows gives (Eq. S3);

$$x_{1,t} = \phi x_{1,t-1} + c_{21} x_{2,t-1} + c_{31} x_{3,t-1} + \varepsilon_{1,t}$$

$$x_{2,t} = \phi x_{2,t-1} + c_{12} x_{1,t-1} + c_{32} x_{3,t-1} + \varepsilon_{2,t} \quad (S3)$$

$$x_{3,t} = \phi x_{3,t-1} + \varepsilon_{3,t}$$

The causal structure contains a bidirectional interaction between x_1 and x_2 together with unidirectional influences $x_3 \rightarrow x_1$ and $x_3 \rightarrow x_2$. As c_{32} departs from zero, x_3 acts as an observed common driver because it influences both the source x_1 and the target x_2 . In contrast, c_{32} is assigned the opposite sign to the $x_1 \rightarrow x_2$ pathway and is progressively strengthened. To sample this transition systematically, $|c_{32}|$ is varied uniformly from 0.0 to 2.3 in increments of 0.1 (equivalently, $c_{32} = 0, -0.1, \dots, -2.3$). For presentation, the horizontal axis is expressed as the magnitude of $x_3 \rightarrow x_2$ coupling, $|c_{32}|$. Importantly, the prescribed direct $x_1 \rightarrow x_2$ coefficient remains unchanged throughout the experiment.

For every $|c_{32}|$ value, 80 independent realizations of length $N = 3000$ are generated after discarding a burn-in of 200 time steps. We compute the bivariate information flow $x_1 \rightarrow x_2$ using only x_1 and x_2 , and the multivariate information flow $x_1 \rightarrow x_{2|3}$ after including x_3 as the conditioning variable. To evaluate how conditioning alters the estimated causal strength, the differential information flow is calculated using the same convention as in the main manuscript (Eq. S4),

$$\Delta IF_{1 \rightarrow 2} = T_{1 \rightarrow 2}^{(bi)} - T_{1 \rightarrow 2|3}^{(mv)} \quad (S4)$$

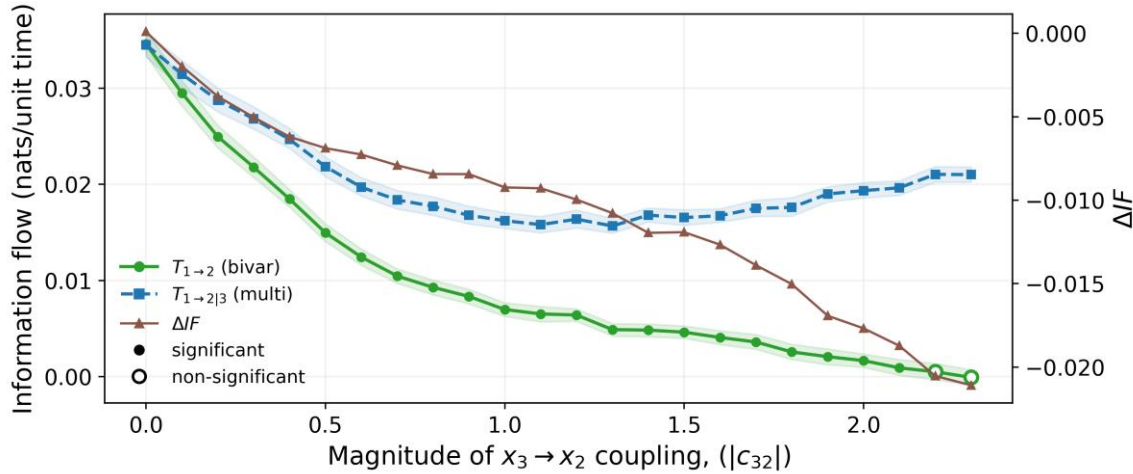
Across the 80 realizations, the ensemble-mean information flow and its 95% percentile bootstrap confidence interval are estimated using $B=1000$ bootstrap samples. For this synthetic experiment, an ensemble-mean IF is classified as statistically significant when its 95% bootstrap confidence interval excludes zero.

The results are summarized in Supplementary Fig. S13. When $|c_{32}| = 0$, the bivariate and conditioned estimates are almost identical, indicating no effect of conditioning. As the $x_3 \rightarrow x_2$ influence becomes stronger, however, the bivariate $T_{1 \rightarrow 2}$ progressively decreases even though the prescribed direct $x_1 \rightarrow x_2$ coefficient remains fixed. $T_{1 \rightarrow 2|3}$ also decreases as $|c_{32}|$ increases until $|c_{32}| = 1$ and then it increases, but it always remains larger than the bivariate $T_{1 \rightarrow 2}$ and statistically significant throughout the tested range.

Correspondingly, ΔIF becomes increasingly negative because the conditioned estimate exceeds the bivariate estimate. The negative ΔIF therefore does not indicate a negative conditioned causal pathway. Instead, it indicates that accounting for x_3 yields a larger estimate of the $x_1 \rightarrow x_2$ information flow than is apparent in the bivariate projection.

Overall, this controlled experiment demonstrates that a fixed direct source-to-target coupling can be masked in a bivariate representation when an observed common driver influences the source and target through opposing contributions, as in our real-world case study (Fig. 1c). In this configuration, omission of x_3 progressively suppresses the estimated $x_1 \rightarrow x_2$ information flow until the bivariate estimate becomes

statistically indistinguishable from zero, whereas inclusion of x_3 as a conditioner retains a significantly positive $x_1 \rightarrow x_2$ pathway. The experiment therefore provides a controlled proof of principle for a conditioning-induced transition in statistical significance. It is intended as a generic dynamical experiment rather than a representation of individual land–atmosphere variables.



Supplementary Fig. S13. Controlled synthetic experiment illustrating bivariate masking and conditioned detection of LKIF under an observed common-driver structure. The direct $x_1 \rightarrow x_2$ coupling ($c_{12} = 0.30$), the $x_3 \rightarrow x_1$ coupling ($c_{31} = 0.25$), and the $x_2 \rightarrow x_1$ feedback ($c_{21} = 0.15$) remain fixed, while the magnitude of the opposing $x_3 \rightarrow x_2$ coupling is varied uniformly from 0.0 to 2.3 in increments of 0.1. The solid green and dashed blue curves show the ensemble-mean bivariate $T_{1 \rightarrow 2}$ and conditioned $T_{1 \rightarrow 2|3}$ estimates, respectively, and the shaded regions show their 95% bootstrap intervals. Filled markers indicate statistically significant estimates, whereas empty markers indicate estimates whose 95% bootstrap intervals include zero. The brown line shows ΔIF on the right axis. At $|c_{32}| = 2.2$ and 2.3 , the bivariate estimate is statistically non-significant, while the conditioned estimate remains positive and significant.

Finally, the Conclusion still states too strongly that when driver–response relationships are “predominantly more direct, conditioners contribute little”. This overreaches relative to the revised Methods/Discussion, which now acknowledge that small $|\Delta IF|$ does not establish causal sufficiency, and relative to Fig. 1c, whose bivariate-insignificant / multivariate-significant pattern remains unexplained. The concluding claim should be toned down accordingly.

Response: We agree with the reviewer and have revised the Conclusion to clarify that similarity between bivariate and conditioned estimates does not imply causal sufficiency. We also clarify that conditioning can affect both the magnitude and statistical significance of the inferred coupling.

Revised text in the Conclusion:

Importantly, similarity between bivariate and conditioned estimates should not be interpreted as evidence of causal sufficiency, because relevant conditioner effects may offset one another or important variables may

remain unobserved. In contrast, under more complex or compound-stress conditions, conditioning can substantially modify the inferred coupling, including its magnitude and statistical significance.

Minor comments / language

Please also harmonize abbreviations between the main text and figure captions: GPP is expanded inconsistently as “Gross Primary Productivity” and “Gross Primary Production”, while and SSR is inconsistently as “Solar Surface Radiation” and “Surface Solar Radiation”.

Response: Thank you. We have harmonized the abbreviations and their expansions throughout the manuscript and figure captions.

L93: and more importantly, the physical limitations leading to their differences. → and more importantly, did not clarify the physical limitations leading to their differences.

L316: Fig 1a → Fig. 1a

L322: SM_driven → SM-driven

L326: results in little adjustments → results in few adjustments

L363: southeast China → Southeast China

L420: “Fig. b-d” → “Fig. 3 b-d”

L424: “SM-based couplings (a-c) ... VPD-based couplings (b-d)” → “SM-based couplings (a, c) ... VPD-based couplings (b, d)”

Response: Thank you. We have made all the suggested language and formatting corrections and checked the manuscript for consistency.

There are also multiple subject–verb agreement and article/count-noun issues throughout; the following are only examples, and similar problems occur elsewhere in the text: L355: “how these increases or decrease T”→“how these increase or decrease T”; L547: “attempts to understand how conditioners may influence its causal structures is still limited” → “are still limited”

Response: Thank you. We have corrected the indicated grammatical issues and carefully checked the entire manuscript for similar inconsistencies. Corresponding corrections have been made throughout the main text and appendices.

Referee #4

The authors have adequately addressed all the questions raised in the report.