

RC2:

This manuscript presents a benchmarking framework designed to evaluate the influence of conditioning variables, specifically mediators and confounders, within the Liang-Kleeman Information Flow (LKIF) framework. Building upon the non-stationary, multivariate extension of LKIF, the authors investigate the systematic divergence between bivariate and multivariate information flow estimates. To characterize these causal shifts, the study introduces four novel diagnostic indices: the Mediator Dominance Index, Moderation Gain, Confounding Pressure, and Convergence Rate. The framework is then applied to land-atmosphere interactions in Southeast China and the Amazon Basin using soil moisture, vapor pressure deficit, LAI and GPP.

The manuscript is generally well written and addresses an important problem in Earth system analysis, namely the conditions under which adding conditioning variables alters inferred causal structure and how to quantify this systematically. However, there are several methodological and interpretative aspects that should be revised before the manuscript can be accepted. I thus recommend a major revision of the manuscript.

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.

Major Comments

1. Ordering dependence in sequential mediator decomposition

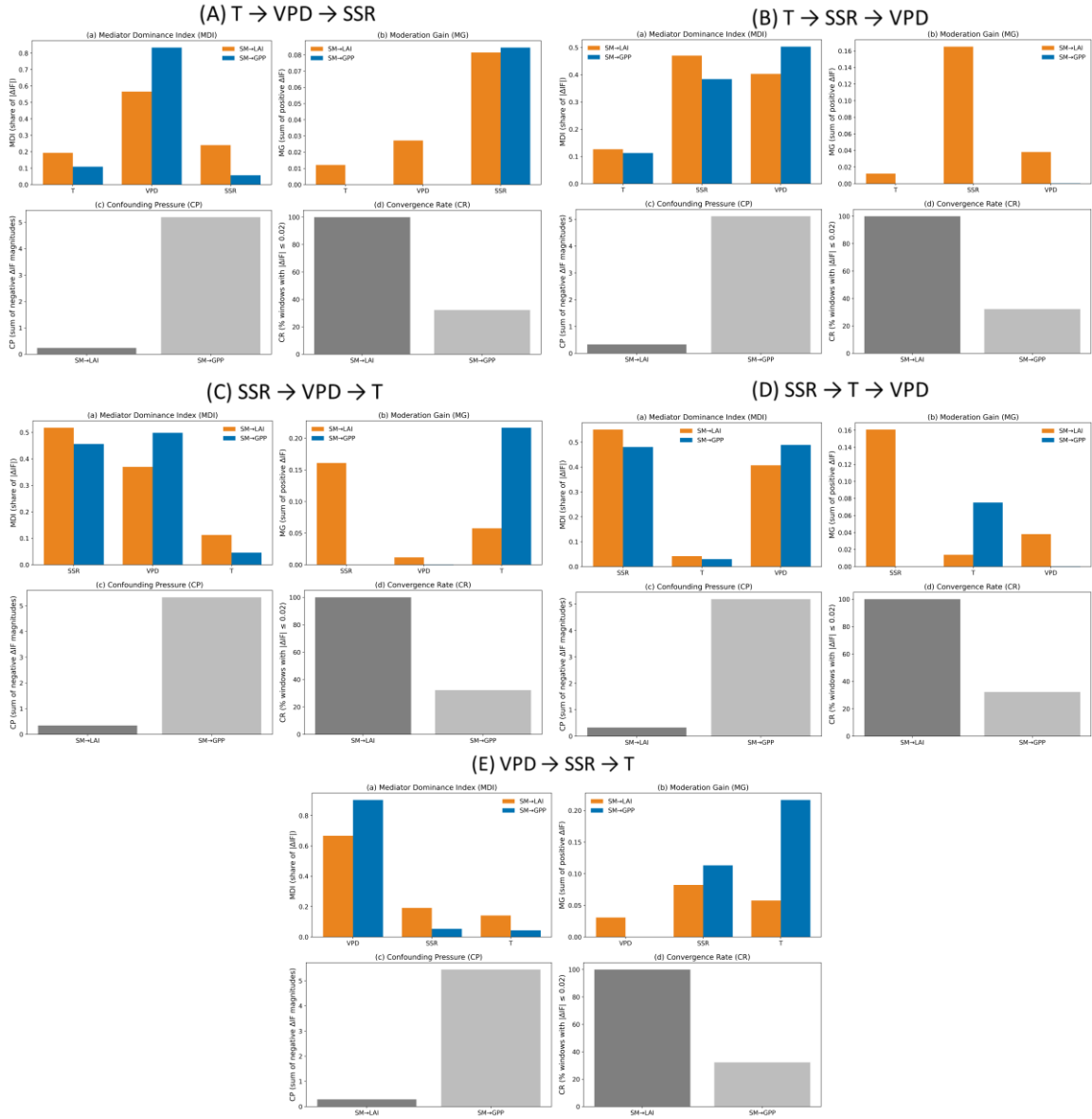
The sequential decomposition in Equations (7)-(9) - computing ΔIF contributions for each conditioner by progressively adding them to the conditioning set - is path-dependent. It seems that the contribution of each mediator depends on the order in which variables are added. The current choice places VPD first (m1), followed by T (m2) and SS3 (m3) for the SM-driver cause, and SM first for the VPD-driver case. It remains unclear whether some of the results, like the dominance of VPD in the MDI (Fig. 5a), is a robust physical finding or potentially related to the artifact of this ordering. Would the contribution of e.g. T to MDI increase if it would be placed first in this conditioning sequence? I recommend that the authors (1) provide a sensitivity analysis demonstrating how MDI, MG, CP and CR change when the ordering of the conditioners is changed, (2) that potential limitations of ordering are discussed in the method section and (3) that the conclusions from this analysis (e.g. that VPD emerges as a dominant mediator) are checked based on a sensitivity analysis of the ordering.

Response: We thank the reviewer for this important comment. We agree that the sequential decomposition in Equations (7)-(9) is order-dependent by construction, because each conditioner is quantified relative to the conditioning set that precedes it. To address this, we repeated the decomposition for all permutations of the

three conditioners (VPD, T, SSR) for the SM→LAI and SM→GPP pathways, and we now report these results in the [Supplementary Fig. S7](#).

The permutation analysis confirms that MDI and MG are order-sensitive, and therefore these indices should be interpreted as sequential attribution measures, not as permutation-invariant measures of intrinsic mediator importance. In particular, the contribution of T does increase when it is introduced earlier in the conditioning sequence, but it does not become the dominant conditioner in our application. For SM→LAI, the relative dominance between VPD and SSR varies with ordering. For SM→GPP, however, VPD remains the strongest across permutations, indicating that this conclusion is comparatively robust, even though its exact MDI and MG values vary with ordering.

We also note that CR is invariant to the ordering because it depends only on the difference between the bivariate and fully conditioned IF estimates. In our application, CP remained unchanged across all tested permutations. We have revised the Methods to clarify the order-dependence of the sequential decomposition and have updated the Results/Discussion to qualify the mediator-dominance statements accordingly.



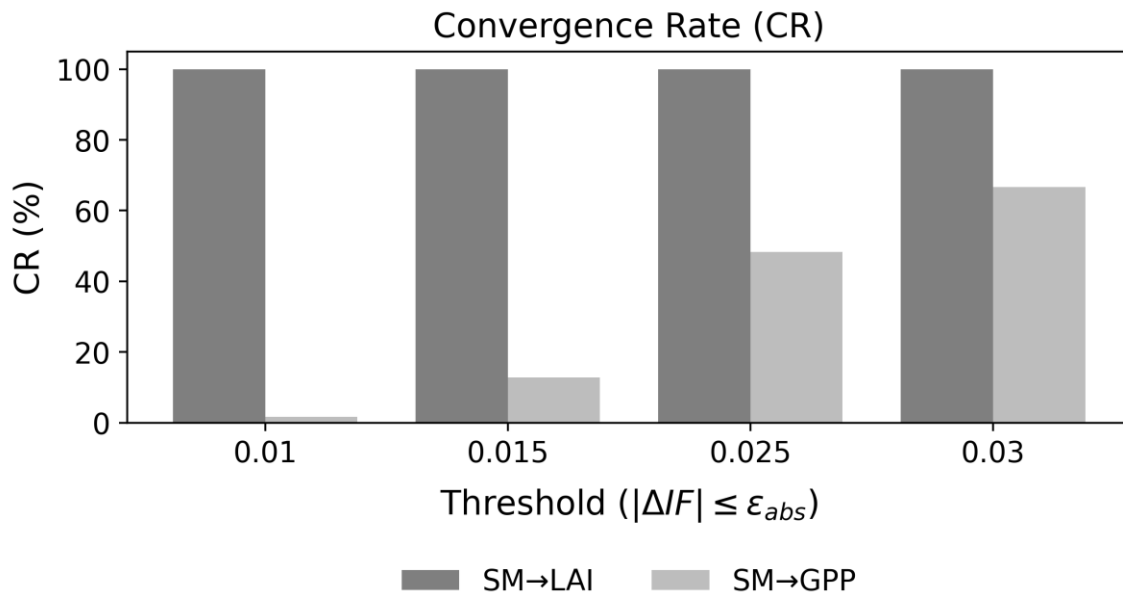
Supplementary Fig. S7. Sensitivity of decomposition indices to conditioner ordering. MDI, MG, CP, and CR are computed for different permutations of the conditioning variables (VPD, T, SSR). Panels (A–E) represent alternative ordering sequences: (A) $T \rightarrow VPD \rightarrow SSR$, (B) $T \rightarrow SSR \rightarrow VPD$, (C) $SSR \rightarrow VPD \rightarrow T$, (D) $SSR \rightarrow T \rightarrow VPD$, and (E) $VPD \rightarrow SSR \rightarrow T$. MDI and MG vary with permutation order, indicating order dependence, whereas CR remains invariant and the aggregated CP remains unchanged in this application. Results are shown for both SM–LAI and SM–GPP pathways.

2. Absolute convergence rate threshold

The four new metrics introduced are very interesting and could be quite relevant for the field. However, the conference rate (CR, Eq. 13) is defined based on an absolute threshold of ϵabs which is stated to be a “fixed tolerance reflecting a physically meaningful scale of information f low variation (e.g. the 10th percentile of

ΔIF across all couplings)”. However, as far as I could find the actual value used for ϵ_{abs} is not defined in the text and the sensitivity of CR to this choice of fixed threshold is not tested. Since CR is used as an important diagnostic, it seems important to test whether the choice of threshold can qualitatively change the interpretation. It would be helpful if the authors would report the exact numerical values of ϵ_{abs} used in their analysis and provide a sensitivity analysis showing how CR changes as a function of the fixed ϵ_{abs} chosen.

Response: We thank the reviewer for this important comment, which was also pointed out by the other reviewer. We agree that the Convergence Rate (CR) is threshold-dependent by construction and that the choice of ϵ_{abs} should therefore be reported explicitly and supported by a sensitivity analysis. In the revised manuscript, we now state the exact threshold used in the main analysis, $\epsilon_{abs}=0.02$, in Sect. 2.4.4, rather than describing it only through the earlier percentile-based example. We also added a supplementary sensitivity analysis in which CR was recomputed for $\epsilon_{abs}= 0.01, 0.015, 0.025$, and 0.03 (Supplementary Fig. S6). The results show that, although the absolute CR value for SM→GPP varies substantially with the threshold, the main qualitative interpretation remains unchanged: SM→LAI remains consistently close to the fully conditioned estimate across all tested thresholds, whereas SM→GPP remains much more threshold-sensitive and less convergent overall. We have revised the Methods and Discussion accordingly.



Supplementary Fig. S6. Convergence Rate (CR) to the absolute divergence threshold ϵ_{abs} for the Southeast China SM→LAI and SM→GPP couplings. CR was recalculated using $\epsilon_{abs} = 0.01, 0.015, 0.025$, and 0.03 . SM→LAI remains consistently near full convergence across all tested thresholds, whereas SM→GPP shows strong threshold dependence, although it remains substantially less convergent than SM→LAI.

3. Linearity assumptions in MTvLK

The MtvLK framework (based on Zhou et al., 2024) utilizes a Kalman filter to estimate the time varying covariance matrix. This approach inherently assumes that the underlying dynamics of variables are treated as approximately linear at the temporal scale of estimation. In land atmosphere interactions, the response of vegetation (GPP, LAI) to environmental drivers is likely not linear. For instance, the authors find that the VPD-LAI coupling shows opposite signs between bivariate and multivariate forms (e.g. Fig 1b, SE China). While the authors interpret this as a result of mediation/confounding, I am wondering whether this might be an artifact of applying a linear estimator to a system that might be undergoing a nonlinear shift. I recommend that the authors add a paragraph in the methods discussing the linearity assumption of the MtvLK framework and also mention specific environmental conditions where this assumption likely does not hold. Additionally, it would be helpful to add in the results section a cautionary note that some of the sign reversals may, as I understand it, reflect a real mediated causal effect or a potential artifact stemming from the linear approximation made in a likely nonlinear system.

Response: We thank the reviewer for this important comment. We agree that land–atmosphere interactions, particularly vegetation responses to hydroclimatic stress, can exhibit substantial nonlinearity. In the present study, we indeed apply the MtvLK implementation of Zhou et al. (2024), in which the resulting inference is based on a linear approximation over short temporal intervals. We acknowledge that under strongly nonlinear or threshold-like environmental conditions, part of the observed divergence or sign reversal may also be influenced by the use of a linear approximation in a system with more complex dynamics.

To address this point, we have added the following clarification to the Methods section:

In the present implementation, the Z24 framework tracks evolving covariance structure through a Kalman-filter-based formulation and therefore relies on a linear approximation over short temporal intervals. This approximation is useful for resolving time-varying behavior in non-stationary systems, but it may not fully represent strongly nonlinear responses. In land–atmosphere systems, such behavior may arise under strongly varying or extreme environmental conditions. Accordingly, the inferred time-varying information flow (IF) should be interpreted in light of this approximation.

In addition, we have added the following statement to the Results section to provide a cautious interpretation of sign reversals:

“Some of these sign differences may reflect conditioning effects, although under strongly nonlinear environmental transitions, they may also be influenced by the linear approximation underlying the present Z24 implementation.”

We have also added the limitation related to the linearity assumption in the new Discussion section, see more details below Point 4.

This clarification is also consistent with our conclusion that future work should explore nonlinear extensions to better represent complex feedback during extreme events.

4. Comparison to previous literature and discussion of limitations

The introduction of the paper is very well written and introduces the scientific question in an interesting and engaging way. However, I am missing a contextualisation of the results with previous literature. In particular, a comparison of the main findings regarding the causal pathways found in the application to the land-atmosphere context is largely absent. The authors could compare their results to the growing literature about causal discovery in land-atmosphere science. It might be furthermore interesting to compare some of these observation-derived results with findings from CMIP6 climate or other land-atmosphere models or at least mention this as an option for future work. Additionally, the discussion section could further illustrate in which applications the new introduced metrics might be helpful for the community. Lastly, a discussion of limitations of the presented methods, new metrics introduced and interpretation in the land-atmosphere context would be helpful to place these methods in the context and limitations of other methods and findings.

Response: We thank the reviewer for this suggestion. We have revised the Introduction to strengthen the contextualization within the existing literature on land–atmosphere interactions by adding a paragraph synthesizing prior studies on causal inference approaches applied in this context. This addition is highlighted in the revised manuscript (see paragraph below). A detailed comparison with previous findings is presented in the Discussion section (below).

Most existing studies have investigated the relationships among SM, VPD, and vegetation using conventional statistical techniques such as correlation and regression (Yuan et al., 2019; Zhong et al., 2023). While these approaches can reveal co-variability, they identify symmetrical relationships that do not imply causation or directionality (Stips et al., 2016; Docquier et al., 2023). As a result, utilizing basic statistical methods makes it challenging to differentiate cause from effect. To address these limitations, several causal inference techniques, such as Granger Causality (GC) and Peter and Clark Momentary Conditional Independence (PCMCI), have been suggested to investigate complex land–atmosphere interactions (Papagiannopoulou et al., 2017; Krich et al., 2020; Dubey et al., 2023). In land–atmosphere studies, these methods have been used to infer lagged directional links among hydroclimatic and vegetation variables, to reconstruct process or causal networks, and to disentangle more direct dependencies from indirect pathways and observed common drivers; for example, information-theoretic process-network approaches have been applied to ecohydrological systems, and PCMCI has also been used to estimate biosphere–atmosphere causal networks (Ruddell and Kumar, 2009; Krich et al., 2020). Although these methods allow to retrieve causal relationships, they continue to encounter challenges such as inconsistent or method-dependent causal estimates, strong sensitivity to assumptions (e.g., stationarity), difficulty in representing multiscale feedbacks, and limited robustness in the presence of hidden or time-varying confounders (Docquier et al., 2024; Shao et al., 2024). Recently, the Liang–Kleeman information flow (LKIF) has surfaced as an encouraging formalism to quantify cause–effect relationships (Liang, 2013, 2014, 2016, 2021). In land–atmosphere applications, LKIF has been used to quantify directional SM influences on precipitation and temperature, and to diagnose evolving coupling strength in nonstationary systems (Hagan et al., 2019; Liu and Pu, 2019; Sun et al., 2025; Takaya

et al., 2025). This formalism provides a quantitative measure of causation, directly derived from first principles in contrast to information flow being axiomatically proposed as an ansatz, making it well-suited for understanding directional influences in environmental systems (Liang 2016).

The Discussion reads as follows:

Understanding cause–effect relationships in land–atmosphere systems remains challenging because interacting drivers and shared variability obscure direct and indirect influences. Several causal methods have been used to identify directional relationships, although they differ in how they handle confounding. GC infers lagged directional dependence through autoregressive predictability and can be extended to conditional or multivariate forms, yet its ability to control confounding depends on explicitly including additional variables and can weaken as dimensionality increases (Papagiannopoulou et al., 2017; Krich et al., 2020). Information-theoretic approaches, such as transfer entropy and process-network frameworks, can capture nonlinear dependencies and feedback structures, but their multivariate application is limited by data demands and the curse of dimensionality (Ruddell and Kumar, 2009; Krich et al., 2020). PCMCI addresses this more explicitly by using conditional independence tests to remove indirect pathways and the influence of observed common drivers, although it still relies on assumptions such as causal sufficiency and stationarity, leaving potential sensitivity to hidden or time-varying confounders (Krich et al., 2020; Docquier et al., 2024). LKIF provides a physically grounded asymmetric measure of directed influence and has been applied in land–atmosphere studies of SM impacts on temperature and precipitation (Liu and Pu, 2019; Hagan et al., 2019; Takaya et al., 2025; Sun et al., 2025), with recent multivariate extensions allowing explicit conditioning on additional variables (Liang, 2021; Zhou et al., 2024). However, attempts to understand how conditioners may influence its causal structures is still limited. Here, using the differences between bivariate and multivariate LKIF (Δ IF), we explore the impact of conditioners on its causal estimations. We find that this difference is physically informative, such that the divergence between the two identifies when apparent driver–response relationships are robust to conditioning and when they depend on additional variables.

Building on this framework, Δ IF provides a direct measure of how strongly conditioning alters the inferred pathway. In the present application, Δ IF is not uniformly large or small, but varies systematically across couplings and environmental regimes. The influence of SM on LAI shows persistently low Δ IF and strong agreement between bivariate and multivariate estimates, whereas the influence of SM on GPP and the VPD-driven pathways exhibits much larger and more variable divergence, particularly under dry, high VPD, hot, and high radiation conditions. This indicates that the role of conditioning is state dependent, such that under some conditions, the bivariate estimate captures the dominant pathway, whereas under others, the inferred relationship changes substantially once additional variables are included. Such state dependence impact is particularly relevant in land–atmosphere systems, where interactions among moisture, energy, and atmospheric demand intensify under compound stress.

The four proposed indices (MDI, MG, CP, and CR) are important because they make this divergence diagnostically interpretable and comparable across pathways. Δ IF indicates whether conditioning matters, but not whether its effect is pathway-specific, concentrated under particular environmental conditions, or associated with convergence between simpler and more complete formulations. The indices provide a compact framework for resolving these dimensions systematically. They allow the analysis to move beyond detecting divergence to identifying when a simpler representation is adequate, when additional variables materially alter inference, and where causal refinement is most needed. In this way, they strengthen the methodological contribution of the LKIF framework by making pathway sensitivity to conditioning explicit and comparable across couplings.

These findings also point to a broader application of the framework in model evaluation. Because the framework identifies when inferred pathways are stable and when they are strongly conditioning dependent, it offers a means to assess how land–atmosphere interactions are represented in Earth system models. Applied to climate model simulations, including CMIP6, these diagnostics could test whether models reproduce not only the magnitude of vegetation–climate coupling, but also its sensitivity to co-varying drivers and its reorganization under compound stress. This would provide a more process-based benchmark than evaluations based only on mean states or correlations.

Several limitations should be considered when interpreting these results. First, the analysis is based on a time-varying linear formulation of the Liang–Kleeman framework, which may not fully capture nonlinear feedbacks and state-dependent shifts under extreme conditions. Second, the inferred causal structure depends on the selected conditioning variables. Although SM, VPD, T, and SSR represent dominant environmental controls, omitted variables may still influence the estimated relationships. Third, the proposed indices depend on the conditioning design and should therefore be interpreted as diagnostic summaries of pathway sensitivity, rather than as universal causal measures. Fourth, uncertainties in observational and reanalysis datasets, together with the aggregation to monthly temporal scales, may affect both the magnitude and variability of the estimated information flows. Finally, the real-world application is limited to two humid forest biomes and therefore does not capture the broader range of hydroclimatic variability across global ecosystems. Vegetation responses in drylands are more directly constrained by SM deficits, whereas high-latitude systems are more strongly influenced by temperature and radiation limitations (D’Orangeville, Liu et al., 2020; Xie et al., 2024). These differences may alter both the magnitude and structure of information flow, as well as the relative importance of conditioners. Extending the framework across a broader range of biomes, incorporating additional conditioning variables, and exploring nonlinear extensions of the method, therefore, represent important directions for future work.

5. Representativeness of study regions

The study focuses on two humid forest biomes, Southeast China and the Amazon basin, to apply the presented methods. While this choice allows for a deeper analysis of these ecosystems, it likely limits the

generalisability of the conclusions. However, the abstract and conclusions make quite general statements (e.g. “this highlights the need for multivariate conditioning when assessing vegetation responses”), although these conclusions are solely based on humid forests which will be among the least moisture-stressed ecosystems. Drylands or high-latitude boreal forests would likely show quite different conditioner importance. It would be highly beneficial to add additional representative biome regions to the analysis or if that is not feasible, clearly state in which biome region these analyses were performed.

Response: We thank the reviewer for this important comment. We agree that the real-world application in this study is limited to two humid forest regions (Southeast China and the Amazon Basin), and that the conclusions should be interpreted within this specific ecological context.

We intended to highlight the methodological implications of multivariate conditioning within the systems analyzed here, rather than to imply universal applicability across all biome types. To address this, we have revised the manuscript to explicitly constrain the scope of our conclusions to humid forest ecosystems.

Specifically, we have (i) refined the wording in the abstract to reflect the ecological context of the analysis, (ii) revised the Conclusions to clarify the scope of inference and acknowledge that conditioner importance may differ in other biome types, such as drylands or high-latitude forests. We agree that extending the analysis to a broader range of biomes would be valuable and now highlight this as an important direction for future work.

6. Definition of proposed indices

The four indices (MDI, MG, CP, CR) are an interesting contribution of this manuscript. However, it should be stated more explicitly in the method section that these indices are introduced the first time in this study. Additionally, it would be helpful to relate them where possible to indices in existing frameworks (such as the Shapley-value framework or similar). Additionally, a bit of additional information in the method section would be helpful regarding how these four indices relate to each other mathematically and physically to better understand their unique information they provide.

Response: We thank the reviewer for this constructive comment. We agree that the role, novelty, and interpretation of the four indices (MDI, MG, CP, and CR) should be stated more explicitly. In the revised manuscript, we now clarify in Section 2.4 that these indices are introduced here as diagnostic measures derived from the sequential decomposition of ΔIF . We have also added a short paragraph to explain how these indices relate to each other mathematically and physically, and the complementary information they provide.

The added paragraph reads as follows:

In this study, we introduce four diagnostic indices: Mediator Dominance Index (MDI), Moderation Gain (MG), Confounding Pressure (CP), and Convergence Rate (CR), to characterize complementary aspects of divergence between bivariate and multivariate IF. These indices are derived from the sequential decomposition of ΔIF and together provide a unified description of how conditioning variables modify the inferred coupling. Specifically, MG and CP represent the positive and negative components of the decomposition, capturing enhancement and suppression effects, respectively, while MDI quantifies the relative contribution of each conditioner to the total magnitude of restructuring. CR, in contrast, summarizes the overall degree of agreement between the bivariate and fully conditioned estimates.

We also acknowledge the reviewer's suggestion regarding connections to existing attribution frameworks. While the proposed indices are conceptually related in that they summarize the contribution of individual conditioning variables, they are introduced here as diagnostic measures tailored to the LKIF framework.

Minor comments

1. L. 82: "This begs a spatio-temporal quality of the role" This could be rephrased as it sounds unnecessarily complex.

Response: Thank you for this suggestion. We have revised the sentence for clarity and readability. The revised sentence now reads:

This raises the question of how the influence of mediators/confounders (hereafter referred to as "conditioners") varies in space and time in the causal estimation based on LKIF, especially under a nonstationary assumption.

2. The acronym 'Z24' is defined at line 63 but is used inconsistently - sometimes 'Z24 framework,' sometimes 'MtvLK method,' sometimes 'Zhou et al. (2024).' Pick one convention and use it throughout.

Response: Thank you for this helpful comment. We have revised the manuscript to ensure consistent terminology. We now use "Z24" consistently for Zhou et al. (2024), introduced at its first occurrence, and avoid mixing it with alternative expressions. The terminology has been standardized throughout the manuscript.

3. L. 78: Please add an example of interactions in the land-atmosphere continuum being tightly coupled with overlapping drivers. This sounds very vague.

Response: Thank you for this helpful suggestion. We have revised the sentence to include a concrete example of overlapping drivers in the land-atmosphere continuum and added supporting references. The revised sentence now reads:

Furthermore, interactions in, say, the land–atmosphere continuum, are tightly coupled with overlapping drivers; for example, vegetation responses may be shaped simultaneously by SM, VPD, and temperature (Novick et al., 2016; Green et al., 2017; Liu et al., 2020).

4. L. 88: It was mentioned in L85 that the authors refer to conditioners as the umbrella term of mediators and/or confounders so please remove this here.

Response: Thank you for pointing this out. We have removed the redundant definition at L88 and now use “conditioners” consistently, following its earlier definition.

5. L. 91-93: Can be removed since it is the standard paper format.

Response: Thank you for this suggestion. We have removed this part as recommended.

6. L. 98: The primary timeseries and the influenced timeseries are here introduced as X_2 and X_1 , but in the introduction X , Y (and Z) was used. Please make this consistent. Also, maybe I am missing something but would not a notation of the information flow from X_1 to X_2 make intuitively more sense?

Response: Thank you for this helpful comment. We have revised the manuscript to ensure consistent notation throughout, and now use X_i, X_j across both the Introduction and Methods sections. Regarding the direction of the information flow, we follow the standard LKIF formulation, where $T_{2 \rightarrow 1}$ denotes the influence of X_2 on X_1 .

7. L. 100: The d should be in italics.

Response: Thank you for pointing this out. We have corrected the notation and now use italic formatting for d and ensured consistent formatting of mathematical variables throughout the manuscript.

8. L. 178: m_1, m_2, m_3 should be in italics with subscripts.

Response: Thank you for pointing this out. We have corrected the notation and now use italic variables with subscripts (e.g., m_1, m_2, m_3) throughout the manuscript.

9. In L. 221 following the authors describe how the monthly anomalies are computed. On top of that is any seasonal cycle detrending applied?

Response: Thank you for this question. The monthly anomalies were computed by removing the climatological seasonal cycle from each variable. No additional detrending (e.g., linear trend removal) was applied.

10. L. 468-470: The mention of “potentially important applications for feature selection for machine learning development” is very vague. Please provide a bit more detail of how this could be helpful in this context.

Response: Thank you for this helpful comment. We agree that the original statement was too vague. We have revised the sentence to clarify how the framework could support machine learning feature selection by identifying drivers with a direct causal influence and revealing when inferred relationships are strongly affected by confounding or mediation, while retaining a forward-looking perspective. The revised sentence now reads:

Furthermore, this framework may also support feature selection in machine learning by identifying drivers with a direct causal influence and revealing when inferred relationships are strongly affected by confounding or mediation, an application that remains to be explored.

11. The authors do not mention which Kalman filtering parameters (like filter gain and noise parameters) were used which is important since these parameters can affect the bandwidth of temporal smoothing and so also the timescale of detected causal variations.

Response: We thank the reviewer for this important comment. We have now revised Section 2.1 to report the Kalman filter settings used in our implementation. The added paragraph reads:

In this study, the time-varying covariance matrix was estimated using a square-root Kalman filter within the Z24 framework. The Kalman gain was computed recursively within the filter from the evolving covariance estimates. The process-noise covariance matrix Q and measurement-noise covariance matrix R were estimated from the input series using a moving-window procedure. Specifically, the input vector was first smoothed using an unweighted moving average (UWMA) with a window length of 24 timesteps, and covariance estimation was then performed over a lookback window of 90 timesteps. The filter was initialized using the smoothed signal as the initial a posteriori state estimate and identity matrices as the initial covariance factors.

12. The paper defines 'conditioners' as the umbrella term for both mediators and confounders, but the physical implications of a variable being a mediator versus a confounder are very different. Is there a way within the LKIF framework to distinguish which role each conditioning variable primarily plays? This would be a valuable methodological addition.

Response: We thank the reviewer for this important and insightful comment. We agree that the physical implications of mediators and confounders are different, even though they are grouped under the umbrella term "conditioners" in this study. In the present LKIF-based framework, conditioning allows us to quantify how an additional variable modifies the inferred driver–response coupling, but it does not by itself identify whether that variable acts primarily as a mediator or as a confounder. Our analysis therefore focuses on the influence of conditioning variables on the estimated coupling, rather than on a formal classification of their causal role. We agree that such a distinction would be a valuable methodological extension, but it is beyond the scope of the present study.

13. The paper uses lots of abbreviations and sometimes it is not clear whether the abbreviations are consistent or not. Additionally, some abbreviations like IF are used before they are introduced. Also, Δ IF is used once for differential information flow and once for time-varying LKIF, please stick to one description.

Response: Thank you for this important comment. We have carefully revised the manuscript to improve clarity and consistency in the use of abbreviations. All abbreviations (e.g., Information Flow (IF), differential Information Flow (Δ IF), and correlation of IF (rIF)) are now defined at their first occurrence and used consistently throughout the manuscript. In particular, Δ IF is used exclusively to denote the difference between bivariate and multivariate IF estimates, and no longer used to describe time-varying IF.

14. Fig 2: It would help the reader to digest the many causal graphs if the years in (b), (c) etc if they are annotated like yearX (large/small IF divergence)

Response: Thank you for this helpful suggestion. We have revised Fig. 2 by annotating each selected year with the corresponding magnitude of divergence between the bivariate and multivariate IF estimates (e.g., “2011 (large IF divergence)”), which clarifies the rationale for selecting these representative years. The figure caption has also been updated accordingly. The revised version of Fig. 2 is provided below for reference.

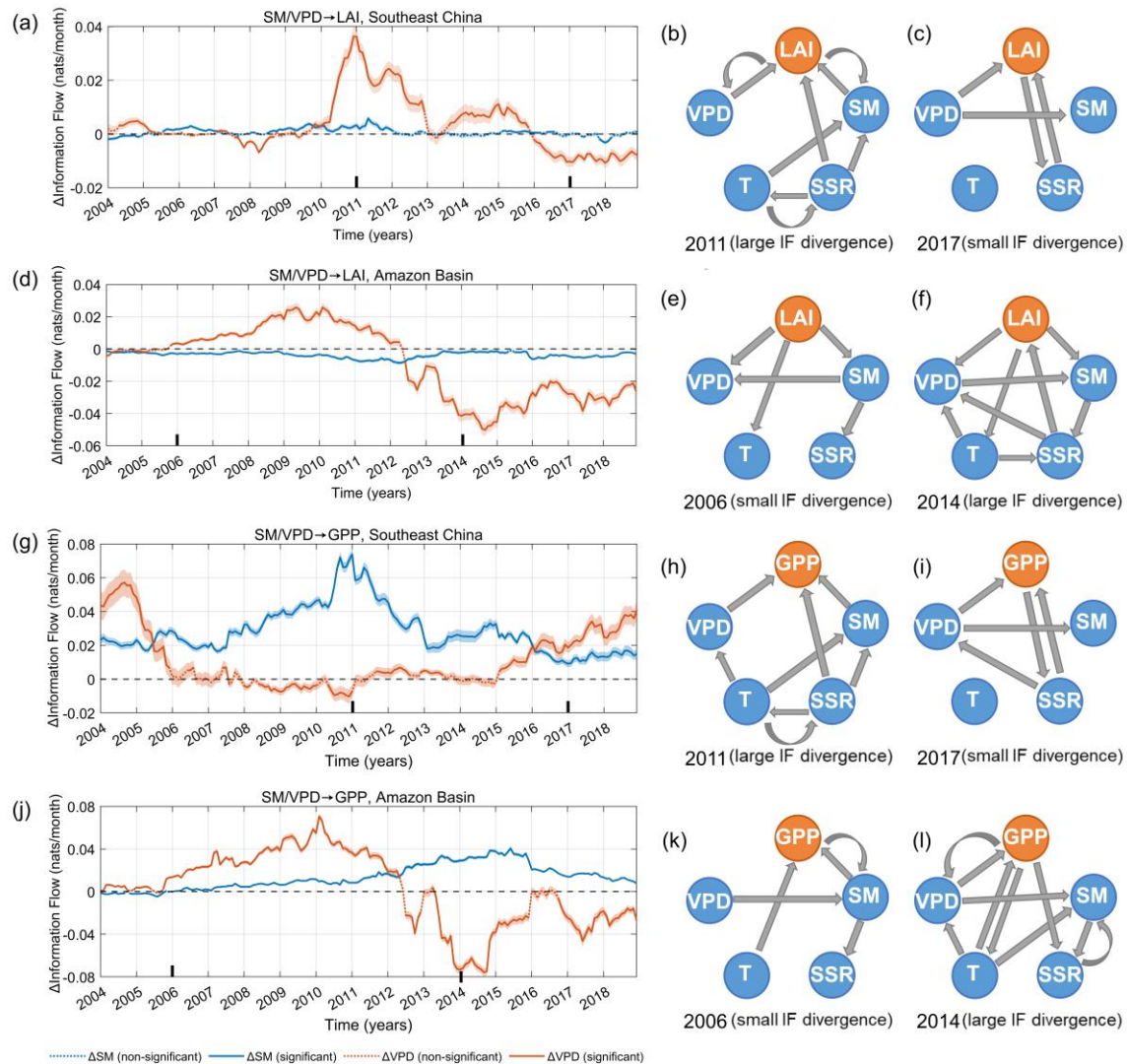


Figure 2: Time series of differential Information Flow (ΔIF) for Southeast China and the Amazon Basin (2004–2018). (a,d) for SM/VPD to LAI, and (g,j) from SM/VPD to GPP. Solid lines denote periods where ΔIF is statistically significant ($p < 0.05$); dotted lines indicate non-significant values. Shaded bands represent the 95% confidence interval. ($\Delta SM = |TSM \rightarrow LAI/GPP - TSM \rightarrow LAI/GPP|_{VPD, T, SSR}$ and $\Delta VPD = |TVPD \rightarrow LAI/GPP - TVPD \rightarrow LAI/GPP|_{SM, T, SSR}$). Vertical tick marks denote the time points selected for causal graph construction. Summary of the causal graphs for years exhibiting large and small IF divergence for LAI (b–c, e–f), and for GPP (h–i, k–l). The selected years are annotated in each panel according to the magnitude of IF divergence (ΔIF), indicating whether the bivariate and multivariate estimates strongly diverge or converge. Stronger ΔIF corresponds to years with richer coupling and feedback structure, whereas smaller ΔIF reflects periods with fewer couplings. (LAI= Leaf Area Index, GPP= Gross Primary Production, SM= Soil Moisture, VPD= Vapor Pressure Deficit, T= Temperature, SSR= Surface Solar Radiation).

15. Fig 3: This graph would benefit from some additional explanation in the figure caption along the lines of “If both triangles are dark coloured and of opposing signs ...,”

Response: Thank you for this helpful suggestion. We have revised the Fig. 3 caption to provide clearer guidance on how to interpret the triangular encoding. Specifically, we now explain how different combinations of upper-triangle ($|rIF|$) and lower-triangle ($|\Delta IF|$) shading reflect differences in temporal behavior and magnitude between the bivariate and multivariate IF estimates.

Added to caption:

Dark upper and lower triangles indicate similar temporal evolution ($|rIF|$) but large magnitude ($|\Delta IF|$) differences between bivariate and multivariate IF. Light upper and lower triangles indicate small differences in both temporal behavior and magnitude. Dark upper triangles with light lower triangles indicate similar temporal evolution with small magnitude differences, whereas light upper triangles with dark lower triangles indicate differences in both temporal behavior and magnitude under conditioning.

16. The section labelled discussion is not a real discussion and rather introduces the new metrics and thus should be part of the results section. As mentioned in the major comments, please include a “common” discussion comparing the main results and methods to other literature and mentioning relevant limitations of the study.

Response: We thank the reviewer for this suggestion. We have added a common Discussion section to the manuscript. The revised discussion text is provided in our response to Major Comment 4 (Comparison to previous literature and discussion of limitations).