Review of "Benchmarking Conditioners in Liang–Kleeman Information Flow: Application to Land–Atmosphere Interactions"

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 Eart 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.

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 (m_1), followed by T (m_2) and SS3 (m_3) 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.

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 ϵ_{abs} which is stated to be a "fixed tolerance reflecting a physically meaningful scale of information-flow 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.

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.

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.

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.

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 farmwork 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.

Minor comments

1. L. 82: “This begs a spatio–temporal quality of the role” This could be rephrased as it sounds unnecessarily complex.
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.
3. L. 78: Please add an example of interactions in the land-atmosphere continuum being tightly coupled with overlapping drivers. This sounds very vague.
4. L. 88:Iit was mentioned in L85 that the authors refer to conditioners as the umbrella term of mediators and/or confounders so please remove this here.
5. L. 91-93: Can be removed since it is the standard paper format.
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?
7. L 100: The d should be in italics.
8. L. 178: m1, m2, m3 should be in italics with subscripts.
9. In L. 221 following the authors describe how the monthly anomalies are computed. On top of that is any seasonal cycle detrending 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.
11. The authors do not mention which Kalman filtering parameters (like filter gain and noise parameters) where used which is important since these parameters can affect the bandwidth of temporal smoothing and so also the timescale of detected causal variations.
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.
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, ΔIF is used once for differential information flow and once for time-varying LKIF, please stick to one description.
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)

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 …, “
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.