Biogeosciences Discuss.,
https://doi.org/10.5194/bg-2020-374-AC1, 2020

Interactive comment on "Functional convergence of biosphere–atmosphere interactions in response to meteorology" *by* Christopher Krich et al.

Christopher Krich et al.

ckrich@bgc-jena.mpg.de

Received and published: 18 November 2020

The manuscript, "Functional convergence of biosphere-atmosphere interactions in response to meteorology," investigates a number of variables and their connections from publicly-available FLUXNET datasets using a relatively novel causal analysis method called "Peter Clark Momentary Conditional Independence" (PCMCI), in conjunction with a dimensionality reduction technique called "t-distributed stochastic neighbor embedding" (t-SNE) and a subsequent clustering algorithm called "Ordering Points To Identify the Clustering Structure" (OPTICS). The specific research questions motivating the study are not clearly stated; the general motivation provided is, "to investigate how biosphere–atmosphere interactions vary across vegetation types and climate

zones." This the manuscript accomplishes through a notion of linkages between biosphere and atmospheric variables, with the primary units of analysis being 1) network representations of those variables and their causal interactions over three-month windows at daily scale, 2) a two-dimensional representation of the structure of those high-dimensional networks, and 3) clusters formed within that 2d space of high dimensional networks. The methods will likely be unfamiliar to most readers, and the units of analysis are quite abstract and require considerable explanation for readers to fully grasp the results being presented: this explanation is not currently sufficient in the manuscript. Broad discussion focuses on some very interesting topics, such as 1) the universality or functional convergence of biosphere/atmosphere processes, 2) trajectories of ecosystems through a 2d space of land surface "network" states, including seasonal cycles and deviations due to extreme events, and 3) linkages between biosphere and atmospheric variables, and how their causal relationships could be represented as clinal processes along some continuum from linked to unlinked. Ultimately though, this discussion turns back to separating water/energy/radiation/temperature limitations on ecosystem productivity from land-atmosphere feedbacks (both of which are areas of deep physical research), which leads the reader to ask what the analysis gains from combining them in the first place. While providing an interesting lens for looking at highly complex interactions between the biosphere and atmosphere across time and space, I found that this study failed to specify its intents and rather motivated too much using the tools (which instead should be motivated as useful for answering the question at hand). This led to results for which I am hard-pressed to find applications. I am not convinced that sufficiently substantial conclusions have been reached. I recommend a major structural overhaul of the paper, driven by specific, answerable scientific questions. At the same time – and this is the difficulty in a study with such "boutique" methodology (with no judgement passed on that label)– the readers will still need *more* description of what is being shown in

the analysis, all leading back to the primary research questions. I have tried to provide as specific guidance as I can in the following comments.

We thank the reviewer for this comprehensive review and the critical view on our manuscript. We recognise the potential to improve the accessibility of the manuscript. In the following we will discuss the reviewer's comments and explain how we will deal with them. We as well try to give answers to the many questions posed and take those question as a guidance to improve the manuscript.

Specific comments:

1. **What is/are the primary research question(s) being asked here? What is the knowledge gap?**

   We indeed did not formulate a specific research question as we conducted a rather exploratory study which was motivated in line 37ff. The world is attempted to be categorised: ecosystems by their appearance, or climate regions by their temperature and precipitation. While discrete categories contradict the natural continuum, they hold certain benefits. Here we wanted to examine, whether ecosystems show distinct functional states and how these functional states can be characterised. Do those states form a continuum or separate classes? Are ecosystems limited within the accessible states? The special quality of our approach is that the states characterise interactions and only those, i.e they are not build by the mean of a certain variable but are an aggregation of many. To lie the foundation to straighten our story line, we will add a hypothesis to our introduction. It might read the following: It is known that ecosystems can enter various functional states, i.e. interactions with the atmosphere. We hypothesise first that the manifold of existing states is bound by few characteristic functional states despite the complexity and differences among ecosystems. Second, attributing to an ecosystems adaptation an ecosystem's accessible functional states are limited to a certain range.

2. **The fundamental units of analysis need to be very clearly specified, as readers will be unfamiliar. Each point in Figure 1 is a network of connections between a bunch of variables at daily scale, but each representing three months of data, including some lagged effects. This network is the primary unit of analysis. It would really help to show an example of one of these networks at one location for three months before jumping into Figure 1, even though the authors have written papers on these networks before.**

   Thanks for reminding us that our daily work isn't that of others and that we need to guide readers more carefully. We regard it as sensible to address this comment together with comment 3. We will add a schematic/a flow chart as a first figure which will explain our work flow including which kind of data was used, which method was employed with which purpose and the meaning of a network and its constituents.

3. **The authors need to discuss seasonality at some point earlier in the analysis to let readers know that all seasons will be studied, and that points in Figure 1 will represent different locations and different times of year.**

   see comment 2.

4. **Line 80: (Probably an easy, but major point) "A comprehensive description from theoretical assumptions..." These assumptions should be stated clearly here, as the method is not well-known. As with any paper using basic regression analysis, a statement of the ways in which the analysis meets basic methodological assumptions is necessary. Rationale/justification for using the method when assumptions are not met are necessary as well. Some of this discussion can happen in supplementary materials if it is particularly involved, but the assumptions and their validity should probably be stated in the main text.**

   We agree, 'easy but major'. Thus we will add a paragraph within the method section summarising the assumptions and to which degree they are fulfilled.

5. **Line 96: "Unobserved common drivers can still render links as spurious." How do the critical *non-stationary* variables (at the time scale of your analyses) of biomass and phenostage influence 1) the validity of the estimation of your networks, and 2) the structure of the 2d space in Figs 1 and 2?**

   The reviewer points to one assumption out of the set of assumptions requested in comment 4. Causal stationarity is an assumption of PCMCI. This does not mean that a causal dependence may not change in strength but that it persists over the time period of interest. Causal stationarity would not be given for many ecosystems when estimating networks over a whole year. We attempt to increase causal stationarity by chunking the time series into 3 month periods. Especially in autumn or spring one can still argue causal stationarity to be violated but as we are talking of a gradual shift, the network representation remains a valid representation of the functional state. An example can be given by rather consistent trajectories in Figure 4 and 5. This comment will be addressed by listing the set of assumptions (comment 4).

6. **Line 105: "subtracted a smoothed seasonal mean from each variable..." I agree that this needs to be done to remove non-stationarity which can cause spurious correlations. At the same time, subtraction of a Fourier series from a time series could either solve that problem or partially solve the problem while introducing new non-stationarities. How robust is the de-seasoning technique? Do the results change if you use other filtering methods?**

   In a previous paper (Krich et al. 2020) we studied the performance of PCMCI on an artificial dataset. Here we could show, that the subtraction of the seasonal mean by a Fourier Series does decrease the false positive rate but leaves the true positive rate mostly unaffected.

7. **Use of PCMCI, t-SNE, and OPTICS really makes this difficult for readers to follow the methodology. I would guess almost no one (particularly outside the author's list) is familiar with all of these. The authors need to motivate why they are using these methods with respect to some research question, and not just because causal tools exist.**

   We agree, that the combination of methods is heavy but it is indeed mandatory to address our research question (see answer to comment 1). For example, in Krich et al. 2020 we could show, that PCMCI enables to focus on a few but relevant links compared to correlation. Due to the quantity of observed ecosystems and the high dimensionality we require further methods for the analysis. We will motivate their use more strongly and clarify their purpose. It is actually not needed to understand each method to its details.

8. **Line 156: "The strongest gradients measured via distance correlations..." As the manuscript stands, I don't think even the most careful methodologically-focused readers are going to know how to interpret these results. It took me a lot of rereading to get the idea that the distance correlations represent the spatial (in this 2d space) coherence of the link strengths. The link strengths themselves should be more clearly explained and motivated, probably in a preliminary figure showing an example network. The meaning of the link strength should be clarified (does link strength 1 mean fully causal? Completely dependent? One-to-one?)**

   We will give a short explanation of distance correlation and its intended use in a new paragraph of the methods section. We try to explain the link strength more clearly in the methods and maybe even in the schematic figure mentioned in comment 1.

9. **Figure 1: As the manuscript stands, I do not think readers are equipped to understand what is being shown in this figure, which needs to change.**

While some methods maybe dense and opaque, results need to be comprehensible to readers in the field, even if they are not close enough to the sub-field method specifics.

Thanks for pointing out this issue. Based on your previous comments and our suggestions to address them, we expect that to change, i.e. we will revise the motivation for using the methods and clarify the workflow with help of a new figure.

10. **While Szekely et al. 2007 is highly-cited in the statistics literature, it is unlikely that many of your readers in biogeosciences will be familiar. How do we interpret these distance correlations? Having referred to Szekely myself, I can see that the correlations are metrics of dependence between random vectors, but can you clarify what are the vectors in question here (say for NEE-LE)? What is their dimensionality, what are the constituent dimensional components? Are they across space and time (I think so) and season of the year (I don't think so), and if so, how do these constituent components combine to give a single number (rho=0.75)? Does this represent something like a fraction of explained variance, and if so across what conditions? Can I compare the rho for NEE-LE and the rho for T-H to infer something about bivariate coupling? What time-scale should I think about these metrics representing? Mostly daily? Does Rg→ H mean that Rg almost always causes H (with positive partial correlation)? Does T-VPD being red mean that T causes VPD with positive correlation or that VPD causes T with negative correlation? Your readers need their hands held through all of this to interpret your results and see the patterns you are seeing in your analysis.**

    Thanks for pointing out all these questions that still remain open. As they still refer to the metrics distance correlation and causal networks we aim to address them in the sections added to the methods and additional schematics.

11. **Figure 1 caption: "As Rg can only be a cause..." Is this true? I'd imagine that LE →Rg often if Rg is measured at the tower (as opposed to top of atmosphere). There are certainly LE → humidity → cloud formation processes at the local scale in many locations, aren't there?**

    We have as well considered the possibility for such processes. We have come to the conclusion, that LE can affect Rg but likely does so at an other location (due to lateral transport). Thus we decided to set Rg as the main driver of our system.

12. **Line 156: "The colouring reveals that the link strengths are ordered along gradients." This sounds like A Finding, but of course the world is spatially autocorrelated and neighborhoods are similar. What is it explicitly that is interesting about this? Is it expected or unexpected (I would expect it) and why?**

    We indeed see this as a minor (not major!) result/finding even though we as well expected it (as we wrote in line 168). Yet, we do not see spatial autocorrelation as the reason. First of all, we are dealing with data on ecosystem level and on this spatial scale FLUXNET towers can be regarded as rather sparsely distributed (in contrast to e.g. satellite data). Second, Figure 1 displays the distribution of link strengths, i.e. the strength of interaction between a set of variables resembling atmosphere and biosphere. No information of location enters here and as can be seen in Fig. 1, the information of location exhibit a lower distance correlation with the tSNE axis as any link.

    That Figure 1 of the manuscript is presented as a finding builds upon following: By attempting to preserve local neighbourhoods, tSNE also preserves gradients within the data. The stronger the gradient the more likely it is to observe it in the low dimensional embedding. Within our approach, only link strengths are handed to tSNE. If the networks are not random (which we expect), we will see gradients of link strength . Further, even though gradients in link strength are expected, we did not know which are dominating before doing the analysis. Figure 1 presents the emerging gradients ranked via distance correlation.

13. **Figure 2: This figure could in theory be used to add interpretability to Figure 1, but otherwise the main take-away is that GPP, NEE, and LE are fairly correlated, as are Rg and T. I enjoy looking at this for patterns, and I can imagine spending time in a study looking for structure and emergent relationships in this data projection, but I am not sure what "results" it represents. How could I as a biogeoscientist make use of the information in this Figure? What question is this helping to answer?**

    As stated already in the answer to comment 12, only link strength values enter the dimensionality reduction process. Thus, any gradient that emerges besides those of link strength values is unexpected. Figure 2 in the manuscript shows 3 month mean values. As the value range of variables is not per se affecting any link strength (see Fig. 2, the finding that certain links correlate with a mean variable value is interesting. For example, such correlation could help to tailor the dependence structure in model parametrisations.

14. **Line 165: "The results show that a high dimensional space encompassing more than 10000 ecosystem networks representing the states of biosphere–atmosphere interactions from ecosystems of various geographic origins can be reduced to a compact two-dimensional manifold characterized by four edges and gradients of biosphere and atmosphere conditions." Maybe I'm missing a key piece of nuance here. It is by definition true that applying a dimensionality-reduction algorithm to high-dimensional data will yield a lower dimensional representation.**

    The reviewer is right, a dimensionality-reduction algorithm will project any high dimensional data onto a low dimensional space, i.e. 2 dimensions. However, it is not granted, that the projection yields any meaningful insight. PCA for instance

projects the high dimensional network space onto the axes with the highest explained variances. However, as Figure A1 shows, points which have been far away in the high dimensional space are now lying close to each other. t-SNE on the contrary manages to unwrap the intrinsically high dimensional network space. Further, as gradients are preserved according to their strength, the embedding also reveals which links dominate transitions between the networks (Figure 1 of the manuscript). Further, the space has a rather quadrilateral shape (if we would not use any significance threshold, that shape would be more prominent). Having four corners is neither anything that is known beforehand.

Are you claiming the positive (and sufficiently large absolute values) of the distance correlation metric imply something more significant about biosphere-atmosphere interactions and coupling? no **Isn't the t-SNE method designed to do something *close* to maximizing these distance correlations?** yes **And didn't you select your dimensionality reduction method to basically do that (maybe not with the explicit cost function of maximal distance correlations, but with local and global neighborhood coherence maximization)?** yes **I'm either 1) not seeing how this is a finding rather than the necessary outcome of your approach, or 2) not seeing how significant these specific metrics are relative to what I should be expecting (maybe the rho values would for some reason be expected to all be less than 0.1 for some reason?)**.

We hope with the above given explanation it becomes clear, why the result is not trivial.

I think it is well-known that there are continua of all of these variables (ranges of GPP, ranges of LE, etc.) and that stepping from one location to another nearby location (in space,time, or say VPD space) will lead to small changes in the biospheric and atmospheric states. This is not surprising. If you can quantify or qualify something ABOUT those gradients, it would

be very interesting because that science is wide-open, but I don't see how Figure 2 is doing that. I see that you are suggesting that this is not obvious in the statement, "While gradients in MCI partial correlation strength are expected as they were used as features in the dimensionality reduction, gradients in climatic and biospheric conditions were not." But I don't see that as actually surprising as there are entire disciplines focused on the biogeographical structure of ecosystems and their gradients. How could we not expect a clinal change in LE and GPP to be related to a clinal change in LE-GPP coupling?

Whether there are continua or not and independent of how variable values among ecosystems change, a link strength estimated via PCMCI is not dependent on the mean value of its variables. To visualise that, we make use of an artificial dataset created from random variables with preset dependencies (Fig. 2 left column). Scaling this dataset by a factor (we chose 10) (Fig. 2 right column) PCMCI detects the same network (including link strength). This is due to the fact that PCMCI, before assessing any dependence, standardises the data.

Additionally, the contemporaneous link NEE–LE, for example, changes not necessarily according to changes in position (latitude, longitude) as Fig. 3 demonstrates.

15. **Line 183: "LE and NEE are weakly, not, or even negatively connected" This is interesting because it is commonly thought that arid/semi-arid locations have the highest coupling between LE (or Bowen ratio) and NEE because of omnipresent water limitation (e.g., in references below). Are these networks so arid as to not have vegetation? (; ; ;)**

    I guess here we have a misunderstanding due to some imprecise wording. The sentence will be changed to: LE and NEE are weakly, not, or even negatively connected to the atmosphere.

But still, the statement is true also for the connection between NEE and LE. The ecosystems showing such states are vegetated. Yet, depending on the state, this vegetation might be dormant or even dead (grass cover). We will have a closer look at the suggested literature and try to discuss it in more detail.

16. **Figure 3: How were these archetypal clusters determined, by eye? It's a little weird to define 17 clusters algorithmically and then define 4 clusters of clusters by hand. Do the 4 types fall out on their own if restricted to 4 clusters? Type 2 looks like barren,arid landscapes. Type 3 is the mid-latitudes growing season. Type 1 is winter. Type4 is interesting in that I wouldn't expect strong coupling in the tropics between T and anything or NEE and anything, since coupling (and even causality) is generally thought of as related to bottleneck variables. What leads to this full connectedness in physical terms do you think?**

    The low dimensional embedding takes the form of quadrilateral shape (the lower the applied significance threshold, the more prominent the four corners, the less prominent the clusters). The four archetypes are the average networks found in the clusters at each corner of the low dimensional embedding (line 288). Defining such archetypes is based on the concept of endmember states. We try to clarify the choice of the archetypes. The full connectedness can be explained the following way. We focus on the example of T and NEE: The optimum temperature range for photosynthesis is between 10 and 20 °C. This temperature range is given in archetype 3. As any fluctuation in this temperature range barely affects photosynthesis (it remains close to optimum), T and NEE are unconnected. In archetype 4 the temperature is above 20°C and thus affecting photosynthesis which links T and NEE. Similarly, radiation changes can be detected in photosynthetic activity linking Rg and NEE. High water availability and energy input allows for large latent heat fluxes and stomata to remain open linking Rg and LE as well as NEE and LE.

17. **Line 149: "The monthly median network is the *average* of the networks..." The mixing of average and median here is complicating an already complex processing step. Are these averages being taken in the 2d reduced space axes?**

    Indeed, using both terms can be confusing but their use is intended. The 'median networks' are calculated using a concept of a median calculation in 2d, which is why we call them 'median' instead of 'mean' networks. Unfortunately, their calculation involves the calculation of an average (similar to the median calculation in 1d with an even number of observations). In the current state we regard the choice of words justified to maintain mathematical correctness but we will attempt to improve the description when revising the manuscript.

18. **Line 216: "for a given month" One month or three month window, shifting by one month at a time? This is confusing throughout. If using overlapping data (single-month network definition, but using sliding three-month windows), I think that will cause some real problems in discussions of the inter-connectedness of your neighborhoods. Your rho values in Figs 1 and 2 will be artificially high by triply counting your data. I am sure you don't want this, but I think you need to either switch to 1-month windows or remove any networks with overlapping data windows (which will reduce your data points by a factor of 3 if I am understanding the method correctly). You can't talk about how nice and smooth the 2d space is when the analysis units are all 2/3 the same data as other neighbouring units**

    Each month of each year is attributed a network. This network is calculated from a time period of three month (centred three month window). Therefore, the reviewer is right when saying that (almost) each datapoint is used three fold. This might also increase the distance correlations calculated for Figure 1 and 2 compared to taking only every third month (every third network). However, this does not affect the result. As the distance correlation is only used to rank the

    links. If 'triply counting' increases the distance correlation value, it is done for all links alike. We do not calculate networks on one month time windows as this would be too few data points. Further, for example the median networks of the month February and May of the towers DE-Hai and FI-Hyy lie pretty far apart. Including March and April thus appears mandatory to create the trajectories.

19. **Figure 4: Why aren't Bowen ratios defined for the whole year for any of the sites except US-SRM? Aren't those variable observed? Don't you need them to map the seasonal trajectories for the plot on the left? Maybe not, and a lot of the network points are fit with partial data? Is that a problem in terms of robustness of the 2d space, the clusters, or the link strengths? You need to clarify how you deal with missing data throughout.**

    Thanks for pointing out this inaccuracy. Bowen ratios are defined for each month but since we use a log scale, we can not plot them when negative. Setting a log scale also might not be necessary. Thus the graph could be changed to Fig. 3. Nevertheless, neither of the values on the right of Fig **??** is needed to map/plot the trajectories on the left. The trajectories are based on the network structure (link strengths) only. Missing datapoints are flagged and not included in any calculation.

20. **Lines 229-230: The fact that you are post-hoc trying to talk about this in terms of water/energy/temperature limitation on ecosystem productivity, but then calling out a separate, loosely connected concept of atmospheric interaction covariance highlights a general weakness in your storyline. There are physical concepts that are well-understood here: water, energy, and radiation (and temperature) can all act as limitation on photosynthetic activity in ways we largely understand (at the plot scale). At the same time, the land surface and atmosphere feed back on one another. I respect and am intrigued by the way in which you are attempting to link those two, but the question remains: what are we learning about how the**

land surface works by doing so? This is a major issue to be resolved for this manuscript.

We hope to understand this comment correctly. We are not using our analysis to post-hoc identify limitations of productivity. Instead, we are trying to understand how biosphere–atmosphere interactions vary: Which states of interactions exist? Which interaction states are dominating? When are the different states reached?. This is very different from: What are the limitations of GPP? Yet, the way we refer to Kraemer et al. 2020 might cause the impression. When rewriting the manuscript, we will more clearly link the finding to the actual research question and clarify how findings of Kraemer et al. 2020 support our findings.

21. **Figure 5: I like the idea of this figure, and think it is a compelling way to look at extreme events. At the same time, it is worth asking whether this 2d space is good at representing extreme events. Does it make sense to think that a tropical rainforest undergoing an extreme drought is really just suddenly (and temporarily) turning into a system akin to a woody savanna, with all of the accompanying *causal* land-atmosphere feedbacks and carbon-energy-water coupling? I wouldn't assume so. That doesn't mean that this isn't a fine first-order way to think about extreme events from a new analytical framework, but I would not a priori think that this is physically representative in any way beyond very coarse correlational descriptions. Presumably extreme events are another suite of dimensions that could be characterized, except for their lack of statistical representation in any data set (by definition). This warrants an explanation and some discussion of limitations.**

    A tropical rainforest will not turn suddenly into a woody savanna (structure wise). The processes we capture in our interaction networks are certainly not covering all processes that characterise ecosystem and distinguish them from each other. However, our analysis shows, that the biosphere-atmosphere interactions (limited

to the chosen set of variables within the analysis) can become very similar.

22. **Technical Corrections:**

23. **Line 29: "only consider two variables..." Granger Causality and Transfer Entropy at this point are only reasonably considered bivariate if you state "bivariate Granger Causality" as the authors do. This bivariance by necessity stance is a false position to take here. I don't know as much about CCM (although a quick search turns up a few recent multivariate extensions), but no econometricians think of VAR-based GC as bivariate, and there have been wikipedia articles about multivariate mutual information for more than a decade. I don't know that you need to have this discussion, depending on how you reframe your research questions and motivation, but you can't publish this sort of claim.**

    The reviewer is right in stating that multivariate extensions exists. Though their practicality for large datasets (especially transfer entropy) is low, as of high computational complexity and data size requirements. Yet we will rethink our argumentation when revising the masnuscript to avoid any misjudgement.

24. **Line 38-42: A nice synthesizing motivation. The motivation for the tie-in to *extremes* is not very clear though. Are you going to be looking at just biosphere–atmosphere interactions under meteorologically-extreme conditions? Or across the whole range of observed conditions?**

    We agree with the reviewer. Linking the motivation to extremes in not needed.

25. **Line 59: Strange citation format for Nelson.** Thanks, we changed the format.

26. **Line 87-95: In the partial correlations, are the correlations controlling for (multiple) lags of the X and Y variables as well, or just other variables?**

    Before calculating partial correlations between Y at time t and X at time t-tau (tau between 0 and 5) other correlating variables, i.e. drivers, are removed via regression. Those drivers can be third variables Z(t-tau) with tau from 1 to 5 and the past of Y(t), Y(t-tau) with tau from 1 to 5 as well as the past of X(t-tau) which is given for tau+1 to 5.

27. **Line 116: And SWC?** The issue with SWC is its lower availability and for those sites that have such measurements it might be applied at differing depth. The depth that is mostly present is at shallows depth of 5 or 10 cm. The upper soil layer, however, dries out quickly and can explain only little of the latent heat flux.

28. **Line 147: "non-intercepting convex hulls..." Even as a very methodological reader I am completely lost here. Is this a typo? What does non-intercepting mean? Intersecting maybe?**

    The reviewer is right. We will change the phrase to 'non-intersecting convex hulls'.

29. **Line 180: "Leave" → leaf**

30. **Line 226: month→ months** We will perform the suggested changes.

References

W. T. Crow, C. A. Gomez, J. M. Sabater, T. Holmes, C. R. Hain, F. Lei, J. Dong, J. G. Alfieri, and M. C. Anderson. Soil moisture–evapotranspiration overcoupling and l-band brightness temperature assimilation: Sources and forecast implications. *Journal of Hydrometeorology*, 21(10):2359–2374, 2020.

P. A. Dirmeyer, F. J. Zeng, A. Ducharne, J. C. Morrill, and R. D. Koster. The sensitivity of surface fluxes to soil water content in three land surface schemes. *Journal of Hydrometeorology*, 1(2):121–134, 2000.

D. J. Short Gianotti, A. J. Rigden, G. D. Salvucci, and D. Entekhabi. Satellite and station observations demonstrate water availability's effect on continental-scale evaporative and photosynthetic land surface dynamics. *Water Resources Research*, 55(1):540–554, 2019.

I. N. Williams, Y. Lu, L. M. Kueppers, W. J. Riley, S. C. Biraud, J. E. Bagley, and M. S. Torn. Land-atmosphere coupling and climate prediction over the us southern great plains. *Journal of Geophysical Research: Atmospheres*, 121(20):12–125, 2016.

Interactive comment on Biogeosciences Discuss., https://doi.org/10.5194/bg-2020-374, 2020.

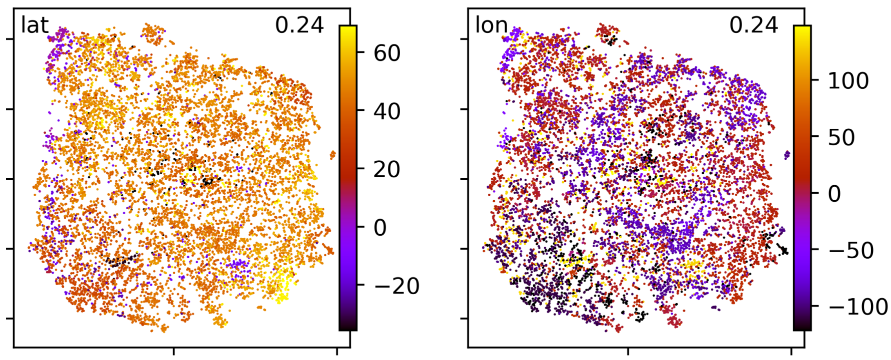

Fig. 1. Same as Figure 2 of the manuscript but colored by latitude and longitude. The distance correlation value of (upper right of each inset) is much lower than that of any link.

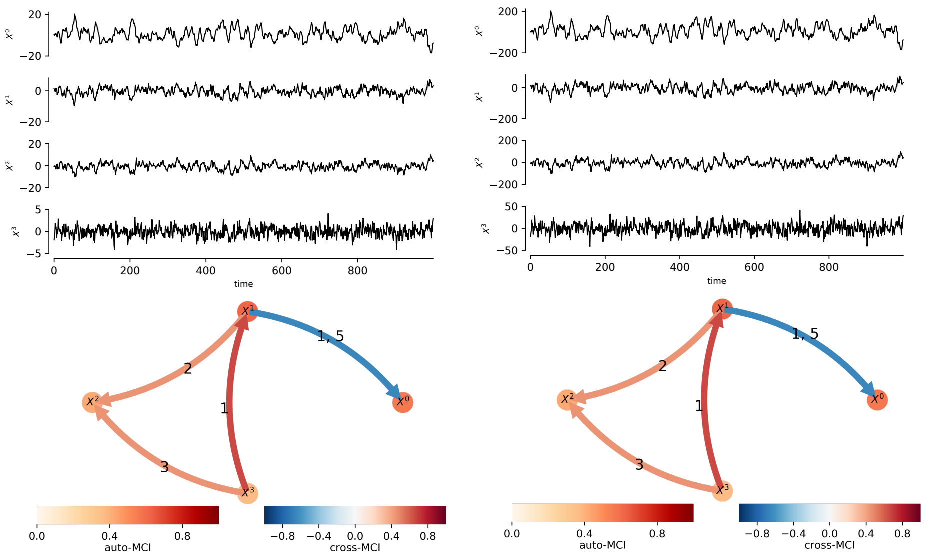

Fig. 2. We created a dataset with given dependencies from random variables. The difference between the left and the right column is simply that the dataset in multiplied by 10 (see y axis).

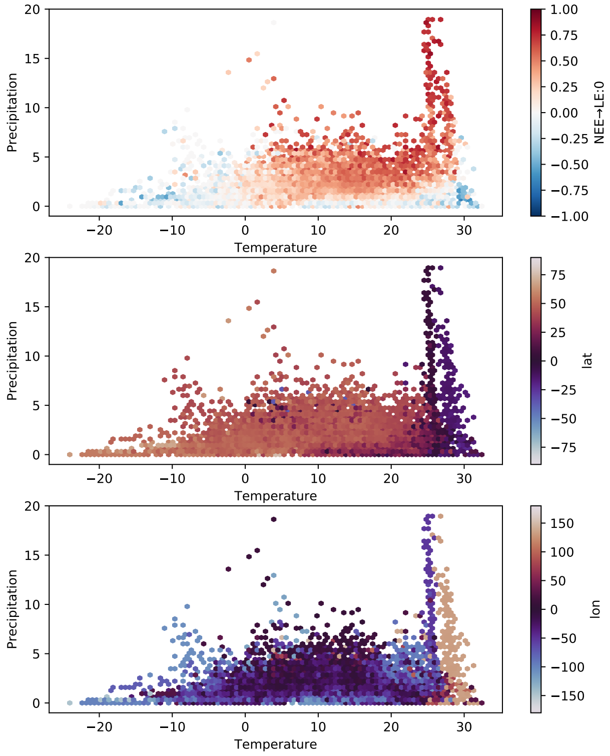

Fig. 3. Behaviour of link strength NEE–LE in the climate space precipitation - temperature.

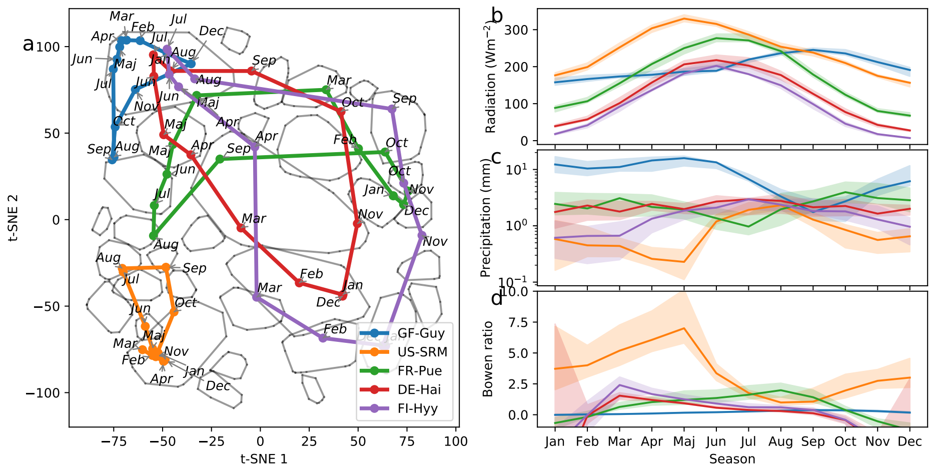

Fig. 4.