Biogeosciences Discuss.,
https://doi.org/10.5194/bg-2019-297-AC1, 2019

Biogeosciences
Discussions

Interactive comment on "Causal networks of biosphere–atmosphere interactions" *by* Christopher Krich et al.

Christopher Krich et al.

ckrich@bgc-jena.mpg.de

Received and published: 16 October 2019

Response to Ben Ruddell

This is Ben Ruddell writing; I waive anonymity for this review. When I saw this paper come across my desk it caught my attention, because I have been working on similar topics and also following the authors' work for several years. In general, I like this paper and after reading it I would like to see it published in this journal, with some changes. This general area of work needs a lot more attention because of the promise of the general approach and the urgency of getting our inference and modeling right for this kind of complex and coupled system. Thank you for this effort! After working on this kind of paper for more

than a decade (and contemplating many reviews of my own work) I've come to the opinion that we need to move past a focus on innovating methods and toward the challenge of showing how the methods can be used to produce actionable and fundamentally novel insights- or to test process theories in science. If we cannot use advanced inference techniques to learn about these systems or critique previously inaccessible scientific ideas, these methods will continue to fall on deaf ears, so to speak. So, I challenge the authors and anyone else listening to move forward aggressively with the intent to apply causal networks (Process Networks) and advanced inference techniques to interrogate scientific hypothesis and learn about systems. The current paper could do more along these lines, with added investment, by(for instance) comparing its statistical results with expectations from climate or ecological models, etc.

Dear Dr. Ruddell, thank you very much for your support and helpful advice. We fully agree that it is not enough to test and advocate for a new method only. However, in order to 'learn about these systems or critique previously inaccessible scientific ideas' the method of use has to be understood in its behaviour first. As PCMCI has not been tested or applied within the context of biosphere–atmosphere interactions to date, this was a necessary step to take before addressing specific scientific questions. The latter will be the aim of following studies, which will build upon the identified strengths of PCMCI. Further, testing or comparing the network structure between models requires non deterministic dependencies which, however, is typically not given.

In the following we try to respond accurately on your questions and comments and will try to integrate those as far as possible.

Before beginning the review, based purely on the expectations raised by the very broad title of the paper, I already had a few questions about the paper. I will pose and then evaluate those questions before moving on to line by line comments.

1. **Is the now-substantial body of literature on this topic adequately summarized and cited, giving credit where credit is due? Papagiannopoulou et al is cited twice, but the similar paper Seddon et al. 2016 is not cited; please cite appropriately. Please review GeoInfoTheory.org, which has a nice list of publications on related topics (https://geoinfotheory.org/reference-list/). In particular,there are a few recent papers that should really be cited appropriately in your paper,because they are recent and narrowly within the scope of your literature review; these treat global land-atmosphere interactions and feedbacks using similar methods to your own. Please describe in your introduction, methods, and/or results as relevant, how does your work relate to these? Yu et al. 2019 in GCB is especially important. A list of references that would seem to be highly relevant follows.** [please refer to original comment to see reference list]

   Thank you for this supportive comment. We will improve the discussion of related literature. Please refer to the answer of comment 3 for further details.

2. **Is the concept and any methods used for "causality" adequately posed and defended? You can expect such a strong claim and wording as "causal graph" to be aggressively challenged by readers and reviewers in this paper and any others using the term, fora long time to come. Renaming "correlation" or "information flow" to "causation" is a major and very aggressive departure from our disciplines' wording and conceptualization during the long and mature history of statistical inference, and requires very strong justification. Granger causality has never really been "causality"; it's a type of conditional time-lagged cross correlation. Please understand my point here; I'm not asking for you to give up on the use of "causal" language, but I am strongly requesting that you spend at least a paragraph in the introduction or methods section of this paper (and others, for the foreseeable future) to argue and explain to the reader exactly what is,and is not,**

   meant by "causal" in this context. It is otherwise too strong a term to be using. As a more general comment, it's extremely important for us to reach a consensus about what to call things. This is an iterative community process of communication that works through conversation and engagement, and through clarification about what is the same and what is different. It's not my place to decide whether we should be calling something a "causal network" or a "Process Network", but I do insist that we have the conversation. For the purposes of this paper, this means citing my recent & prior work,and that of others, and trying to explain exactly how your terms relate to our terms for similar things, and proposing what you understand the similarities and differences to be; this is particularly important when writing a methods paper such as this one under review here.

   See also answer to comment 3. No general standard notion of causality exists so far. Using PCMCI we naturally base our choice of vocabulary on the causal notion PCMCI is based on (cf. Runge2018a). Under fulfilled assumptions and in the limit of infinite time series length PCMCI converges to the 'true causal graph', which is why we use the term 'causal'. As we deal with finite sample length and partially unfulfilled assumptions, as we mention several times, spurious links can appear (see FPR), thus each detected link has to be interpreted carefully.

3. **Is there anything new here, and is that made clear? Yes! PCMCI is put through some rigorous tests for both satellite and 30m flux data and appears to hold up well; this is novel and interesting as a methodological development. However, in my opinion, it is important when describing this method in the methods section that you distinguish it precisely and detail from other similar methods,explaining its relative advantages and disadvantages. There are lots of other methods out there that have used Granger-adjacent directed coupling statistics in various application contexts. In this precise context, my 2009 papers you cited (and several**

since)used 30 minute flux tower time series data to determine atmo-bio networks, identifying ranges of statistically significant time lagged couplings, and also studies periodicity and noise in the method, calling these resulting patterns "Process Networks", and distinguishing the most "causally" relevant couplings using a Tz metric that compares directed vs correlative information flows. This is a well worn topic in 2019, so it's not sufficient in a methods paper to contrast your method with correlations anymore. Contrast your method precisely against others that claim similar goals and results, please. Why should we use PCMCI instead of one of several other existing similar methods?How would the results differ in theory and in practice? Should we use PCMCI in this case, and use Ruddell et al. 2009a "Tz" in another case? What are the pros and cons? Because this paper focuses on methods, it needs to be much more specific about how these methods relate to other adjacent methods and conflicting/overlapping terminologies already in use; this engagement is how we will build our community's knowledge and practice. (your treatment of the underlying assumptions is a strength of the paper and should help make these distinctions clear; thank you for this attention to detail here.)

Thank you for the above comments. All three comments refer to a stronger comparison to related literature and methods. We will address these comments by a more extensive literature review and discussion. While the first comment might be best addressed in the introduction, comment 2 and 3 belong to the method section. Here they can also help improve the accessibility of the method. However, an in-depth numerical comparison of the available methods is beyond the scope of the manuscript, and partially already done in Runge et al. 2018 and 2019. There are several causal inference methods available, with multiple additional modifications. Picking only one or two of them (e.g., Tz statistic of Ruddell et al. 2009) would unavoidably be rather arbitrary. We agree that this comparison is important but might be best tackled in a separate study, maybe even in

a combined effort. Such a comparison study would help users choose the most suitable method for each specific application, rather than addressing any specific question with the method at hand, as it is common practice.

4. **Is the very broad title justified, or is the paper actually about something much more narrow and specific?By the end of the abstract, I decided "negative" on 4. because this paper appears to be not a review or synthesis of the broad topic of causal networks in the bio-atmo-geo-sphere as implied by the title, but instead a methods case study establishing the robustness of a proposed method MCMCI in two land-atmosphere contexts. I suggest a much narrower title, like "PCMCI robustly identifies biosphere-atmosphere interdependencies", or some such. It is very important to use an accurate title that is not over-broad. The title directly summarizes the question and/or findings, in a nutshell.An overbroad or inaccurate title is grounds for rejection in my view.**

   Thank you for this advice. We will take your suggestions into consideration and will adjust the title so that it more precisely relates to the addressed questions.

Line by Line Comments

Sec. 2.1 **I've followed the derivations in Runge et al. (various, 2014-2018) and I don't have a problem with the methods. However, I have not seen here or in Runge et al. (various) an explicit comparison of the MCI approach with Ruddell and Kumar's(2009a) "Tz" or zero-lag ratio method for the disambiguation of "strongly causal" versus "common-source causal" indicated couplings. There appears to be a lot of shared intent and intuition here, and possibly some very similar (but differently named) mathematics and assumptions. Please explain what is similar or different.**

We will investigate the similarities of the methods for the disambiguation of of "strongly causal" versus "common-source causal" indicated couplings. In case of a similar intent and if a comparison can benefit the understanding, we will refer and compare to the existing literature.

Pg.20-10 **This discussion on "causal stationarity" and limitation of study to one season or system state appears to be treated in Ruddell and Kumar 2009(b) (second half of the paper you cited) under the terms "local" and "global" stationarity. What's the relationship here, please?**

We studied the paper "Ecohydrologic process networks: 2. Analysis and characterization" by Ruddell and Kumar. We could not identify a definition of local or global stationarity. To our understanding the terms 'local' and 'global' are used in the context of choosing the bounds for the binning intervals in the estimation of the conditional probability densities. A local scheme refers to a binning interval chosen by the minimum and maximum values of the month. A global scheme refers to the binning interval that is chosen by the minimum and maximum values of the whole time series or dataset. The global scheme is chosen if a comparison between process networks is intended.

Causal stationarity means: A process $\mathcal{X}$ with graph $\mathcal{G}$ is called causally stationary over a time index $\mathcal{T}$, iff for all links $X^i_{t-\tau} \to X^j$ in the graph $X^i_{t-\tau} \perp X^j \mid \mathbf{X}^-_t \setminus \{X^i_{t-\tau}\}$ holds for all $t \in \mathcal{T}$. An example: The influence from radiation to temperature exists in both summer and winter, it might weaken or strengthen but as the physical mechanisms remain active, the link satisfies causal stationarity through out the year. In contrast, the influence of radiation on photosynthesis in a deciduous forest exist in summer but can not exist in winter if no photoactive plant material is present. Thus causal stationarity is violated if the whole year is included in the analysis. Limiting the analysis to specific periods in time, e.g. summer, leads to causal stationarity. This masking in time in PCMCI could be done manually/fixed time intervals, e.g. monthly, or by choosing the mask for a specific value range of one specific variable, i.e. GPP or Rg. The latter might remind of the above mentioned local and global binning but still only marginally, from our perspective.

Pg.21-20 **Although it isn't the focus of your paper, Kumar and Ruddell (2010, Entropy) and some of my more recent papers (Yu et al., Gerken et al.) have shown very strong changes in coupling strength across space, as well as across time. I wouldn't make the claim that "the interaction between biosphere and atmosphere is expected to change only marginally across space" in the absence of strong arguments supporting this. I've argued the opposite in several recent papers- I've argued that the Process Network characterizing these systems and their states changes dramatically between places and times, and that this represents a qualitative shift in how the systems are functioning. (note that I'm not arguing that physics changes, only that its structure and expression in a complex system changes dramatically)...please engage with this argument,or remove the claim.**

The claim "the interaction between biosphere and atmosphere is expected to change only marginally across space" was used only in the context of the Majadas ecosystem and within this context we regard it justified and well supported: This ecosystem is a rather homogeneous Savanna. Within this ecosystem three eddy-covariance towers are situated within a distance of up to one kilometer (app.). Within this spatial scale, climatic conditions are very similar. Due to the homogeneity of the ecosystem "interaction between biosphere and atmosphere is expected to change only marginally across space" (for the Majadas ecosystem). We will clarify that this statement is meant for the Majadas ecosystem only.

Pg.21-25 **Most of my papers have focused their analysis and presentation of results on a single "most significant" time lag (usually chosen as the first/shortest peak lag in mypapers, called the "characteristic time lag" in my papers), or an average across a range of time lags (usually subdaily <18hrs) because of the extreme challenge of interpretation posed by a large number of statistically significant coupling links. Separating out every conditionally "momentary" coupling is not hard to do mechanically, but interpretation and**

communication is devilish. I think you're running into this problem here. Once we move past conditioning couplings on zero-lag correlations, it's not clear where to stop or how to interpret the results. I'd hope that PCMCI could add some clarity, but I'm not convinced based on this discussion that it is helping. Please comment and clarify if possible, or at least explain how what you're doing is different here from what Ruddelland Kumar 2009 did with T, I, Tz, canonical coupling types, and characteristic timelags. If possible, also engage with Goodwell and Kumar (recent) who have attempted to split out redundant, synergistic, and independent couplings in the land-atmosphere coupling context.

We agree that interpreting a process network incorporating many lags for one dependence can pose difficulties. That is why we omitted a detailed analysis/study of the monthly Majadas networks. Yet, we also did not want to aggregate or focus on one lag as this would have reduced the information content of the analysis. An averaging of lagged links, for example, would have caused a strong deviation in link strength for the dependence H→VPD in Fig. 4 (August) among the towers. I(T→VPD)$_{LMa}$ would be nearly 0 while I(T→VPD)$_{LM1}$ and I(T→VPD)$_{LM2}$ would be around 0.25. This is due to the possibility of negative coupling strengths using ParCorr. The dependence H→VPD appears at lag 1 and 3. The confidence intervals of the strength values from the three towers are overlapping for both lags, but as the link H→VPD at lag 3 is rather week, only one crosses the significance threshold.

Defining the maximum lag might be indeed difficult from a physiological/physical perspective. In Runge2018a following is suggested: "Choice of τ_{max}: The maximum time delay depends on the application and should be chosen according to the maximum physical time lag expected in the complex system. In practice we recommend a rather large choice that includes peaks in the lagged cross-correlation function (or a more general measure corresponding to the chosen

independence test), because a too large choice of τ_{max} merely leads to longer runtimes of PCMCI, but not to an increased estimation dimension as for FullCI."

Pg.22-15 **I am not convinced by biweekly or monthly scale correlation analysis in satellite or climate data represents causation in any real or approximate sense. There are several problems here. First, these data are modeled and abstracted several levels beyond primary observations, so patterns cannot be relied upon to strongly represent causal realities as well as insitu flux data. Second, once we move past subdaily timelags, we are well into the scales dominated by diurnal cycles, synoptic weather cy-cles and by seasonal rhythms, so it is hard to distinguish signal from noise when the"noise" is an overwhelmingly energetic diurnal, seasonal, or synoptic cycle. Third, we already have strong reason to believe that the main process timescales are subdaily, due to e.g. our flux tower analyses, so we must presume that super daily or monthly timescales indicated by the methods are merely echoes and confounding correlates of shorter timescale processes unless we can prove otherwise (e.g. through robust conditioning against shorter lags)- and that proof is not possible using coarse time resolution data. This is a basic problem with attempts to use satellite and coarse time resolution gridded data to establish "causal" relationships, and I haven't seen it adequately addressed in this paper or prior papers attempting similar. What am I missing here, please? Please explain how your method addresses these three problems. This gridded/monthly analysis may be a "bridge too far", so to speak, for this paper; it's different from and a weaker argument than your eddy covariance analysis, with several layers of practical problems weakening the conclusions.**

This comment might be addressed by addressing comment 2. Causal relationships are best examined by perturbing the system at a specific time and state (variable) (do calculus of Pearl). Though such experiments are usually not feasible in a controlled manner within Earth system sciences. Therefore, we (as a community) rely on the estimation of causal dependencies from time series and can only detect the signal imprinted in the time series. The signal of interactions depends on properties of the interaction itself, i.e. the strength,type and pattern, but also the signal-to-noise ratio, i.e. measurement noise, time sampling intervals and time aggregation. Therefore, the signal of interactions detectable within the time series (i.e. dependence within the conditional probability distributions, which using Markov condition determines connectedness in graph) might not correspond to the actual physical interactions anymore, but might very well allow valuable insight. Especially when trying to evaluate and compare dependence structures within model time series. Further, under aggregation information from fast interactions will be lost (and maybe visible as contemporaneous interactions in our networks) but processes which are dominant on larger time scales might appear as their signal is improved due to aggregation.

Further, Ruddell and Kumar 2009 and Krich 2019 find links on timescales below 30 min on 30 min time resolution fluxdata. Having the above in mind, i.e. keeping in mind that the time sampling interval determines the appearance of the causal graph, one can not even speak of the 'true causal relationships' using 30min resolution data. If links appear that happen on faster time scales than the time resolution, they will be shown as contemporaneous links (undirected) in our networks. In the method section, we state, that spurious links, both contemporaneous and lagged, can appear. This will be further elaborated in a revised manuscript.

Fig. 6,7 **These results are begging for a detailed comparison with Yu and Ruddell etal., published earlier this year in Global Change Biology, which attempts a very similar analysis but uses an extrapolation of 30m flux data derived couplings to the global terrasphere rather than monthly gridded data. Please provide this comparison.**

With all respect, we do not fully agree on the level of similarity between these two studies. Without a doubt the performed study "Anticipating global terrestrial ecosystem state change using FLUXNET" by Yu and Ruddell et al. 2019 is very interesting and we actually had similar ideas for another study. To explain why we prefer to omit a comparison with this study, we briefly summarize the method and subsequently give the reasoning.

Yu and Ruddell (2019) calculated two bivariate transfer entropy couplings (Temp-NEE, Precip-NEE) on monthly time periods of the available time series data of 204 Fluxnet towers. Thus they obtain a network per month which can be translated to monthly time series of couplings. These couplings are fitted with a specific model (using monthly averages or sums of Rg, Temp, Precip, EVI) to estimate an elasticity of that coupling to each 'driver'. Those elasticities are upscaled to global scale using an artificial neural network. Those maps of upscaled elasticities shall be compared to PCMCI strength values.

The choice of variables for the coupling calculations are based upon: (quote from the paper) "an eddy covariance tower's process network can be approximated using three functional subsystems: Synoptic, Atmospheric boundary layer (ABL), and Turbulent. We choose an essential timeseries from each of those three subsystems: for the Synoptic subsystem, air temperature; for the ABL subsystem, precipitation; and for the Turbulent subsystem, net ecosystem exchange of carbon".

We believe a comparison is not straightforward because of two reasons: The quantity we plot in Fig. 6 and 7 is a conditional independence measure, i.e. partial correlation coefficient between time series residuals at monthly resolution. The plotted elasticities in Fig. 3 of Yu and Ruddell et al. 2019 represent an upscale of a specific co variation (an exponential model) of a conditional independence measure, i.e. transfer entropy between time series at 30 min resolution, to monthly aggregates of climate and phenology variables. We have difficulties to relate these two quantities with each other. Furthermore, We want to validate the outcome of PCMCI. Thus we preferably compare our results to studies which calculate a dependence measure on approximately the same data as we used.

Second, Fig. 6 and 7 of our study show the dependence of phenology (NDVI) on climatic drivers. Fig. 3 of Yu and Ruddell et al. 2019 shows the elasticities of NEE to both climatic and phenological drivers. Fluctuations and responses of NEE and NDVI to climatic factors happen on very different time scales. Further, NEE and NDVI are difficult to compare in the first place.

We hope that we could convince you that the comparison of our global case study to Wu et al. (2015) and Papagiannopoulou et al. (2017b) is better suited for verification purposes than a comparison to Yu and Ruddell et al. 2019.

Interactive comment on Biogeosciences Discuss., https://doi.org/10.5194/bg-2019-297, 2019.