Biogeosciences Discuss.,
https://doi.org/10.5194/bg-2020-374-AC3, 2020

Biogeosciences Discussions
Open Access

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

This is a very interesting paper on biosphere-atmosphere interaction, but it is also a bit difficult for me to understand. I have some background in causality detection (GrangerCausality, CCM), but still find the paper hard to follow when I read it for the first time. This is mostly due to the incomplete description of the methods the authors used. When I went back and read the paper the same lead author published earlier this year (Kirch et al 2020, BG), this paper becomes clearer. The author used a causal relationship detection method, PCMCI, and quantify the interactions between biosphere and atmosphere (represented by the energy, water and carbon fluxes measured by eddy covariance flux towers). With the resultant 10038 networks obtained from PCMCI, the authors applied a dimen-

Printer-friendly version

Discussion paper

sion reduction algorithm and visualize and analyze these networks along two dimensions. By quantifying the 2-dimension space into four regions, the authors can show the trajectory of biosphere–atmosphere interactions changes through time and across different sites. The authors lastly claim that environment are the major factors that regulates the biosphere-atmosphere interactions, effect from the vegetation type is small. This paper tackles a very important question, and use rather novel method. And also because of this, the presentation is not very clear. The final conclusion is drawn based on qualitative evidence which weakens the importance of this study. I have several comments below for the authors to consider.

We thank the reviewer for the effort undertaken to give us valuable feedback to improve the manuscript. We will attempt to improve mentioned weaknesses and consider suggested improvements.

1. **From a reader perspective, I find this paper difficult to read. Clearly, there is a large gap between "what the readers know" and "what the authors assume the readers know". For example, in the introduction (P2 L33) when the author first mention PCMCI, I would expect further explanation on this new method because it is clearly not known by most or at least some readers. Instead, the author did not give any explanation on this but just mentioned one paper. I think this is a good place where the authors can briefly explain what is the basic ideas behind this method and what kind of information it can tell us. With this in mind, the readers can better follow the research question. Another example is in section 2.5, the author mentioned a method called OPTICS, but also did not provide enough explanations. There are also some good practices to improve the readability. For example, in the last paragraph of the introduction where the authors describe the structure of the paper. The authors can define the aim of each section first before directly stating what they did in each section. This also apply for the**

method description which is very dense and filled with lots of jargons and acronyms.

We can comprehend the reviewers critique on sparsely given information regarding the applied methods (PCMCI P2L33, Optics Sect. 2.5). The brevity though is not due to some sort of sloppiness. We tried to provide as much information as necessary to gain an intuition of the methods while reading the main text. For example, based on the paragraph (L27:L36) the reader could gain following understanding about PCMCI: PCMCI is a causal discovery method, that allows to create causal networks based on a set of assumptions. It is a multivariate approach attempting to remove the influence of third variables when evaluation the dependence among two variables. The method was already tested on datasets similar to the ones considered in the present study. In short, what is the method supposed to do, how is it achieved and does it work for the intended use? Yet, we fully respect the reviewers opinion, are grateful that it is stated and acknowledge the effect of "what the readers know" and "what the authors assume the readers know". When revising the manuscript we will focus on increasing the level of information. This will go in hand with addressing suggestions of reviewer 1. We further appreciate the suggestion how to improve the description of the structure of the paper.

2. **The major conclusion of this study, to me, is not well supported. The authors claim that the biosphere-atmosphere interactions are determined mostly by the environment, with limited effect from vegetation types, etc. This is supported by the similar causal network shown in the 2-D space of the t-SNE. However, there are two problems with this. First, when looking at this network, much of the linkage are physically based, with limited effect of vegetation, for example, the relationship between T-VPD, T-H, Rg-T, etc. Vegetation would have limited effect on these relationships and for the other carbon fluxes related relationships, different biome types may have**

different linkage strength. For example, in Figure 1, when looking at link strength related to NEE and other variable, the patterns become more sporadic, especially at bottom left corner, these differences in responses may be caused by the vegetation types or differences in months, but may have limited contribution to the overall network. That is to say, the vegetation types can have effect on the biosphere-atmosphere interaction, but only contribute limited to the network evaluated, therefore, their effects are ignored. ...

The reviewer is right with the statement 'much of the linkage are physically based, with limited effect of vegetation, for example, the relationship between T-VPD, T-H, Rg-T, etc'. However, it seems to be overlooked that we addressed this in L255:259. Four (if including H) variables are physical/atmospheric variables, two are biospheric variables. The most dominant links (Figure 1) though include the biospheric variables. Further Figure 4 shows, that the transitions between the archetypes are dominated by changes of biosphere variable dependencies. In addition, we show the distribution of IGBP classes in Figure D2 which shows a much lower distance correlation with the axes than any three month mean value of a physical variable. The reason for overlooking this argumentation might be the fact that the role of the distance correlation as well as the information revealed by the dimensionality reduction step are not clearly posed (as reviewer 1 suggests). We therefore currently see this comment addressed with measures undertaken to address comments of reviewer 1.

3. **... Another problem related to this is that the authors showed several cases of the change in climate can cause a shift in interaction in the t-SNE space. However, these effects are not quantitively analyzed, how much of this change in interaction strength are caused by climate and how much is caused by differences in ecosystems. are they both significant enough? Without these information, the conclusion is draw without solid support.**

Printer-friendly version

Discussion paper

The reviewer is presumably talking about Figure 4 and 5. Figure 4 shows median trajectories (kind of mean seasonal cycle) in the t-SNE space for five ecosystems. In addition to the median trajectories, we show the corresponding mean values of radiation, precipitation and Bowen ratio. The choice of these variables is made because Figure 2 shows that transitions between archetype one and four covary with energy availability (Rg, T) and transitions between archetype two and three covary with water availability (P, SWC, Bowenratio). Figure 5 shows strong deviations (due to changes in interaction strength) from such median trajectories for certain ecosystems. The strong deviations in the trajectory coincide with strong anomalies in precipitation (also shown).

The suggested comparison of 'how much of this change in interaction strength are caused by climate and how much is caused by differences in ecosystems' thus proves difficult or trivial, as we look at trajectory changes within one ecosystem (Figure 5). Thus the ecosystem stays constant resulting in a covariance of zero with any given variable. Other ecosystem variables like phenological changes within that ecosystem are again driven by climate. Figure 4 indeed allows a comparison of ecosystems and actually is intended to do so. Figure 2 shows, that the low dimensional space covaries with atmospheric variables. This would cause ecosystems with prevailing atmospheric conditions to populate certain regions in that low dimensional space. Figure 4 shows that this is the case. We choose the ecosystems due to their contrasting climate. Further ecosystem effect quantification again appears trivial. The only ecosystem effect quantification that appeared sensible to us within the scope of the manuscript is to quantify the covariance of the low dimensional embedding with the IGBP class, i.e. the covariance of IGBP class underlying each network with its location in the low dimensional space. This is done using the metric distance correlation and the result is shown in Figure D2.

As we discussed at the end of the manuscript (L260ff), we as well regard it important to quantify the effect of biotic factors on the network structure. However, this would include the use of further datasets as for instance standage, vegetation coverage, soil properties and species diversity. Additionally, the framework would need to be developed further to quantify and compare changes in the t-SNE space. We regard it to be out of scope of this paper.

4. **Some detailed comments: P2L49, "one high dimension observation" is not clear. Is it better to say "one facet in the high dimension space"** Thanks for pointing out a lack of clarity. Yet, we do not see that the suggestion improves the situation into the correct direction. We suggest: 'Each of the estimated networks constitutes one observation, i.e. measurement, in a high-dimensional space. This space is spanned by the network's links.'

5. **P4L92, based on the information theory, the causal relationship that happens within the smallest time step of observation cannot be detected. For example, although we know that Rg has a causal relationship with NEE, but this happens in seconds or minutes considering the lags in measurements, this causal relationship cannot be detected by the algorithm and will be regarded as unidirectional. This need to be mentioned or discussed as a limitation for interpretation of the results.**

   We agree with the reviewer. NEE, i.e. carbon uptake into the biosphere due to photosynthesis, responses much quicker to Rg than the time resolution of 30 minutes would allow to detect. This leads to non-directed (not unidirectional) links. Such links are called contemporaneous links in PCMCI (no direction of influence can be inferred). This can be indeed more clearly stated. Actually, this is going to happen when stating the assumptions underlying PCMCI, as reviewer 1 suggested (see comment 4).

6. **P5L129, This is not clear enough, is computation efficiency the only difference? Why would it generate different results as compared to t-SNE?**

Indeed, the given differences between t-SNE and UMAP are a bit shallow. We add a bit more description.

7. **P7 Fig 1, Rg can also be affected by the cloud which can be affected by ET, H and other factors. See Green et al. 2017 Nature Geoscience.** Also reviewer one pointed to this possibility. We acknowledge the possibility of Rg being affected by sensible (H) or latent (LE) heat fluxes which is investigated in Green et al. 2017 Nature Geoscience. However, here satellite data is used, not eddy covariance data. The larger the area over which the variables are aggregated, the higher the possibility to detect any effect of LE or H on Rg. Thus, we regard the possibility of Le and H influencing Rg and its effect as rather small as we look on ecosystem level. We therefore decided to do without detecting this influence in favour of setting Rg as a main driver.

   Setting Rg as a main driver will avoid the estimation of any driver of Rg an thus any regression on Rg. This has the benefit, that any seasonal changes in the variables (i.e. remaining non-stationarities) caused by seasonal changes in radiation are attributed to radiation.

8. **P11 Fig 4, Maj→May** Thanks.

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