Biogeosciences Discuss.,
https://doi.org/10.5194/bg-2019-297-RC1, 2019

Biogeosciences Discussions
Open Access
EGU

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

Anonymous Referee #1

Received and published: 4 September 2019

The manuscript presents the application of the newly developed PCMCI algorithm for the detection of causal links in geophysical data, focusing on biosphere/atmosphere interactions. Applications of the algorithm is done using flux tower eddy covariance data, for fine temporal scale analysis, and satellite derived NDVI along with climate data are used for global scale analysis at coarser scales. The topic is important and clearly within the scope of Biogeosciences. Identifying and quantifying causation in geophysical data is crucial for understanding the interplay of the processes involved and developing models. The present manuscript is intended to my understanding to be primarily a proof of concept of the applicability of the PCMCI algorithm. While the paper is well written several parts need to be better clarified.

Specific comments:

Printer-friendly version

Discussion paper

o I find the description of the algorithm on the paper slightly confusing (in particular for a Biogeosciences audience). I believe that the reader must refer to (Runge et al., 2018) to understand the basic principles behind the algorithm. I strongly suggest the authors to restructure and clarify this section. Simplifying the description and reporting the algorithm details as supplementary information could benefit the fluency of the manuscript.

o The synthetic test developed to quantify the skill of the algorithm, when its assumptions are not valid (e.g. seasonality, heteroscedacity) is clearly important. However, it is not currently clear how the results derived from this analysis can be generalized beyond the Hainich site. Emphasis should be given on the transferability of the magnitude of expected biases at a global scale.

o In a broader sense, a key question is why would the authors use a procedure for the identification of causal links, when the basic assumptions of the procedure are violated by the data?

o I believe the authors should better explain why the proposed algorithm in more efficient than e.g. the spectral Granger causality algorithm proposed earlier by Detto et al. (2012), which, to my understanding, is inherently non-parametric and not sensitive to periodicities.

o In Fig2 it is shown that with an increasing sample size, the fraction of falsely identified causal links increases, when the algorithm's assumptions are not valid. This can be a significant drawback as the best datasets (i.e. with long records), are more prone to errors. The authors should better discuss this.

o If I understood correctly, it is shown that for the baseline case, the algorithm cannot robustly identify the true causal non-linear links (as expected). However, the identification rate increases incorporating seasonality (which I presume also violates the stationarity assumption of the algorithm), which is counter intuitive. The authors in their discussion attribute this behaviour to the variance of the parent variable. Can the authors discuss

Printer-friendly version

Discussion paper

how this artefact can limit the range of applicability of the procedure for global scale applications (i.e. differences in regions with distinct seasonality or not)?

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

Printer-friendly version

Discussion paper