

Biogeosciences Discuss., referee comment RC1
<https://doi.org/10.5194/bg-2022-191-RC1>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on bg-2022-191

Anonymous Referee #1

Referee comment on "Revisiting and attributing the global controls over terrestrial ecosystem functions of climate and plant traits at FLUXNET sites via causal graphical models" by Haiyang Shi et al., Biogeosciences Discuss.,
<https://doi.org/10.5194/bg-2022-191-RC1>, 2022

Review: Revisiting and attributing the global controls on terrestrial ecosystem functions of climate and plant traits at FLUXNET sites with causal networks

1. General comments

The authors attempt to build causal links between plant traits, climate and ecosystem functions by constructing a Bayesian Network (BN), where links between traits and functions are based on expert knowledge, while the climatic variables are informed by the model. The authors then reevaluate the relative importance of plant traits and climate in determining ecosystem functions through a sensitivity analysis based mainly on FLUXNET data. Building on this they argue that climate indirectly affects ecosystem functions via its control on plant traits. We agree that, from an ecological perspective and considering the increasing availability of data, exploring different methods to analyze the interactions climate-vegetation involved in the ecosystem functions is a relevant and meaningful research topic. However, the paper is missing an appropriate model validation (making it difficult to evaluate the robustness of the results) and suffers from reproducibility issues as some important methodological points and choices require clarification or additional information. We start by providing some major concerns followed by minor comments in order of appearance.

2. Specific comments

- The data used in this study relies on a database collected by Magliavacca et al. (2021). In the cited paper, there are three complementary variables to quantify water use efficiency (G1, WUEt and uWUE); Rb was calculated in terms of both mean and max (95th percentile) values; and aCUE data is also available. The authors do not specify the criteria they followed when choosing which variables to include in the BN among the available

variables in Magliavacca et al. (2021). Why was uWUE used and not the other water use efficiency metrics? How sensitive are the results to this choice? Why was the calculation of Rb unique in using the mean – compared to the other ecosystem function variables? If the network was based on expert knowledge, why to exclude aCUE which is explicitly included in the expert framework (Fig 1)? What would be the effect of adding CUE as an extra constrain? How sensitive is the network to its inclusion?

- In the BN (Figure 1), the causal relationships plant traits – ecosystem functions were assigned based on expert knowledge (Reichstein et al, 2014). Then, the climatic variables and the respective causal relationships were added: how were the links climate-plant traits and climate-ecosystem functions determined? Considering that BN based on expert knowledge rely heavily on the prior understanding of the processes, the approach used to assign these links should be clearly stated in the methods.

- Each plant-trait and ecosystem-function variable used in this study has a clear equivalent in the expert knowledge frame (upper panel in Figure 1) (Reichstein et al., 2014), except for AGB – which is unique in the authors model. Please provide more information about the assignment of links to this variable. How did you link G_{max} with AGB? Why not to link LAI_{max} to AGB? AGB is confounded with wood density, height and other size metrics as plant diameter. Which variable on Reichstein's frame is being represented with AGB? How was the confounding controlled for? Can the authors ensure no circularity was added to the framework? This is because AGB from Globbiomass is inferred from models and algorithms that use as input remote indicator variables that are correlated with many of the other variables in the authors network. Finally, AGB from Globbiomass is subject to large error – how was this error controlled for?

- More information is required regarding the specific criteria taken into account when defining the discretization thresholds. It is mentioned in the text (Lines 113-114) that the "meanings of the thresholds" were considered, but it is not completely clear how the thresholds were chosen nor what their meanings are. Ideally, BN models developed using different discretization methods should be considered and compared. If the results of such models are different, the choice of one method over the other should be justifiable (see e.g. Nojavan et al., 2017 - Comparative analysis of discretization methods in Bayesian networks).

- Not all FLUXNET stations used in Magliavacca et al. (2021) have data available for all the variables. Only 94 out of 203 sites have data regarding vegetation structure, N%, LAI_{max}, H_c and AGB. This means that there is a substantial amount of missing data in the model. Please report the missing fractions in the manuscript, or another indicator of the amount of missing data treated with the Expectation-Maximization method that is mentioned in the text (Line 131). Also, how robust are the results to the imputation methods used? It is critical to show that the results are not dependent on data imputation when a large amount of data is missing.

- One core question arising from this study is: how to show that the artificial-rules-based model can reveal the real rules compared with a data-driven model? It is, therefore,

necessary to show validation results in the paper. What validation methods did the authors use? how does this validation result compare to typical standard seen in similar studies? If a validation (i.e. k-fold), and robustness checks are done and reported, then the model results can be interpreted with more certainty. However, with the available data in the current version of the manuscript, the model performance cannot be assessed. We believe, that even though the BN was calculated for categorical data instead of continuous data, there is always a need to show that the model predicted well through model validation techniques. Validation results are also important for comparison/verification with future ecological-knowledge-based models.

- Finally, to prevent confusion and over interpretation of the result, the authors should acknowledge that BN are not necessarily causal networks, they are essentially a set of conditional (in)dependencies that factorize the joint probability distribution of all the variables. Causal deductions hence may not be made (Ramazi et al., 2020 - Exploiting the full potential of Bayesian networks in predictive ecology).

3. Technical corrections

3.1. Introduction

- Since the main focus of this study are ecosystem functions, a clear and concise definition of this term is needed in the Introduction section. Also, it would be useful adding some supporting references regarding the theory linking the functional traits included in the paper and Reichstein et al.'s frame.

- There are some terms used along the paper referring to climate variables, vegetation structure variables, or ecosystem functions. These terms change along the manuscript. Try to use a consistent terminology. Examples of these terms are:

Line 27: "complex ecosystems", "environmental systems".

Line 37: "environmental conditions"

Line 60: "environments"

Line 158: "ecosystem service functions"

Line 271: "ecosystem systems"

- Lines 66-67: The paper by Gregorutti et al. (2017) is used to support the statement that IMP-based attribution can be unreliable when the aim is explaining systematic causality. The cited paper does not discuss systematic causality. Please support this idea with appropriate references.

3.2. Methodology

- Line 95: "Climatic variables:..."

- Table 1: Though the detailed methods for the ecosystem functions' calculation is in Magliavacca et al. (2021), provide a complete summary for each variable in the column "Approach". E.g. if the percentile used in the calculations will be reported, report it for all the variables consistently (GSmax – 90th percentile). The use of medians instead of means for some of the variables may also be important information.

- Line 119: Figure 1 is presented as a column with three panels. It is not clear what the author refers to when pointing to the lower left panel.

- Lines 142-147: This sentence is quite long. Try to split it or make it shorter.

3.3. Results

- Line 154: Incomplete sentence at the end of this line: "... SWin, VPD, and showed..."

- Figure 3: In the text, some information is extracted from this figure regarding the correlations climate vs. ecosystem functions and ecosystem functions vs. ecosystem functions. More information can be extracted from this figure regarding the correlations ecosystem function vs. plant traits and plant traits vs. climate.

- Figure 4: Can the display of this figure be improved? E.g. locating the tables in a more equidistant layout.

- Lines 195-196: When looking at Figure 1, it is not clear what the author is referring to with the loop of Tair controlling LAImax. What variables are included in this loop? This word should be used with caution since it could be taken as equivalent to a feedback.

3.4. Discussion

- The interactions among plant traits, climate and ecosystem function variables may be

complex when high-order effects are considered. Does the author think these effects play an important role when trying to explain the causal links to ecosystem functions? If these effects are not considered in this study, this limitation should be stated in the Discussion section.