Geosci. Model Dev. Discuss.,
doi:10.5194/gmd-2016-266-AC2, 2017


Geoscientific Model Development Discussions

# *Interactive comment on* "A non-linear Granger causality framework to investigate climate–vegetation dynamics" *by* Christina Papagiannopoulou et al.

**Christina Papagiannopoulou et al.**

christina.papagiannopoulou@ugent.be

Received and published: 28 February 2017

**Response to Anonymous Referee #2**

**General Comments:**
**Reviewer summary: The manuscript presents a non-linear Granger causality analysis to investigate climate-vegetation interactions. Anomalies of the normalized vegetation index (NDVI) are analyzed in conjunction with a full set of climate variables taken from re-analysis, in situ, and satellite observations. The data provide multi-decadal global coverage for water availability (precipitation, snow water equivalent and soil moisture data), temperature, and radiation. All**



**data spans the period 1981-2010 at the global scale and has been converted to a common monthly temporal resolution and 1×1 degree spatial resolution. At each pixel the NDVI data is considered the response and the climate data the predictor variables. A moving window of twelve months is used to determine if the climate data Granger-causes the NDVI value. Analysis is performed on NDVI anomalies computed by subtracting the corresponding monthly expectation from the de-trended time series. The climate data as well as cumulative values and extreme indices calculated from the climate data were included as predictor variables. The non-linear Granger causality uses a non-linear random forest model, and is shown to explain more of the variance than the linear Granger analysis.**
**Article contribution and overall impact: This study makes an effort to use multiple climate data sources to tease out predictability for vegetation anomalies. The authors highlight improvements with the non-linear method compared to traditional Granger causality, as well as the importance of using extreme events. The discussion would benefit from a more explicit discussion of the uncertainty associated with the climate datasets used as predictors. Given that this study precedes or supports Papagiannopoulou et al (in review), more discussion of those results and their importance would be useful as that study is not available to the reader. Specifically, the follow-on study highlights the importance of specific climate predictors for particular regions. It is not clear how those variables are chosen from the many climate predictors, and it would be useful to provide an example in this manuscript to highlight the strength of this method with a clear detailed regional example.**

We would like to thank the reviewer for the feedback, and the thorough assessment of the manuscript.
We agree that the study Papagiannopoulou et al. (in review), in which we apply the method to discern the importance of different climatic drivers, may be useful for the

referees to assess the potential of our framework. As we mentioned in our response to the Anonymous Referee #1, that article will be enclosed in the resubmission of the revised paper, so it can be available to the editor and reviewers. As the referred article is a follow on from this GMD paper, as the referee states, we do not see the need to provide details about its specific results within the GMD paper.
Regarding the above comment on input uncertainty, the revised article will include a few statements discussing the impact of these uncertainties.
Below we provide our pointwise response to the reviewer.

**Detailed comments:**
**Page 1 line 17-18: Should this read "predictions of vegetation in response to future climate can be improved through a better understanding..." ? as you are looking for climate drivers of vegetation.**

We think that the initial sentence "Because of the strong two-way relationship between terrestrial vegetation and climate variability, predictions of future climate can be improved through a better understanding..." is in fact correct. In this paragraph, we discuss the complex two-way interactive relationship between vegetation and climate in order to state the importance of understanding climate dynamics to predict climate accurately. Therefore, a better understanding of the vegetation response to past climate variability, brings us one step further in understanding future climate, since the latter will also be affected by the fate of vegetation.

**Page 2 line 22: define "higher-level features" here and throughout manuscript. It is not clear what these are. (Pg 11 line 4, pg.15 line 2)**

With the terms 'higher-level features' or 'higher-level climate variables' we refer to the 'cumulative' and 'lagged variables' as well the climate extreme indices, which have been calculated on the raw time series and serve as additional predictor variables



in the dataset. We will clarify this in the revised manuscript.

**Page 2 line 24: define "higher-level climate variables" not clear what this is.**

True. See above response.

**Page 3 line 2-7: May not be necessary to include full definition of $R^2$.**

This comment has been addressed in our response to Anonymous Referee #1. We acknowledge the reviewers' claim is correct, but we should clarify the different definitions of $R^2$ in linear and non-linear models. For non-linear models, $R^2$ is calculated using the formula in the manuscript, while for linear models it can be calculated as a correlation coefficient. We believe that since the audience of GMD is rather multidisciplinary, including this definition may be helpful to some readers.

**Page 3 line 30: update "might lead to wrong" to "might lead to incorrect"**

Yes, we will change this phrase in the revised manuscript.

**Page 12 line 15-23: Are the results for all variables, or the most predictive variable, or a set of variables at each pixel?**

We use all the variables at each pixel in order to obtain the results presented in this section. We will clarify it in the revised manuscript.

**Page 12 line 26-27: Why is this chosen as the minimum? Please explain or provide citation.**

This statement comes from the definition of Granger causality. The minimum explained variance can be achieved by using the history of the target variable only, and this is basically the model referred to as 'baseline model' in the manuscript. We will make this more explicit in the revised version.

**Page 13 line 10: by what margin is the uncertainty larger in these regions, and for what reasons?**

As one can notice from the map in Figure 5b, the improvement of the full model in terms of $R^2$ compared to the baseline is low. Therefore, the results indicate that the Granger causal effects of climate on vegetation anomalies in these regions are not obvious. This is why we enumerate a set of studies which explore the main drivers of vegetation in these regions, explaining the poor predictive performance of the full model with respect to the baseline model. We will extend the discussion of this part in the revised manuscript to clarify this point.

**- Are you referring to all the climate variables, if not please qualify.**

Yes, we are referring to all the climate variables included in the dataset. We will clarify it in the revised manuscript.

**- The citation references error for soil moisture.**

Thanks, we will update.

**- Add citations, which support the amount of uncertainty in these regions for the remaining data types.**

See above response.



**Page 13 line 7 to bottom and page 14 line 1-4: Move this to discussion.**

The subject of this section is to present the results obtained from the proposed methodology. The part the reviewer refers to discusses the performance of the proposed methodology in comparison to other models. We therefore believe that this is the appropriate location to place this discussion, and we would prefer to keep it this way, unless the reviewer is insisting on this issue.

**Page 14 line 1: Update to "vegetation anomalies are not necessarily"**

We will add the word 'anomalies' in the revised manuscript.

**Page 14 line 7: Use different phrasing for "unambiguous"**

By "unambiguous" we just mean that the improvement is clearly visible here, which we believe reflects correctly the meaning of this word.

**Page 14 line 7-12: move to discussion.**

See the above response.

**Page 14 line 8-10: Recommend re-wording this. The limit for figure 5 and the presentation of the non-linear analysis is still to a limit of $R^2$ = 0.4 as in figure 4? An $R^2$ of 0.4 does not seem like a strong correlation. Though figure 5 is improved from figure 4 there are large portions that show no improvement, and the overall explained variance is below 40% in most regions.**

We kept the same colorbar scale in both figures for a better figure-to-figure comparison. We note that $R^2$ is not to be mistaken by 'correlation' in the non-linear models; it is a performance measure that indicates how close the model predictions are to the real values of the target variable. In our study, we remove the seasonal cycle from the NDVI time series and we target the NDVI anomalies, making the task more difficult than predicting the raw NDVI time series, since the autocorrelation in the NDVI anomalies time series is much lower. Note that if we target the raw NDVI time series (which includes the seasonal component), the $R^2$ is close to 1 in most of the regions (see Fig. 7 in the manuscript). In addition, it is worth noting that there are other factors such as fires or harvest that affect vegetation dynamics but are not included in the dataset, as mentioned in the discussions. Therefore, we should be aware that we focus on explaining the variance of the NDVI anomalies, taking into account only climatic variables, and focusing on the part of their explanatory power that truly reflects Granger-causal relationships.

**Page 14 line 10: "comparison between figs 4b and 5b" explain in more detail. It would be easier for the reader to compare these if they were in one figure block, or on the same page.**

Yes, it is true that a 3 $\times$ 2 figure would be more convenient for the reader; it was also a comment from the Anonymous Referee #1. However, Figure 5 is the main figure of the paper and we decided to make it in a separate block to highlight it.

**Page 15 line 5: Please provide more detail about this study. It comes up frequently in the manuscript, and a larger summary with details (supportive numbers or examples from regions) would be helpful since we do not have access to the manuscript.**

See first response.



**Page 15 line 11: Has a test been run with only the anomalies and extremes? Would that sub-set of predictors provide strong predictive performance?**

We agree with the reviewer. We can include more experimental results at this point in order to figure out the importance of anomalies and extremes in particular. In the paper Papagiannopoulou et al. (in review), there is a thorough discussion about the different groups of variables that have been tested for Granger causality, and provides the results of isolating the impact of extremes. Again this paper will be enclosed in the resubmission.

**Page 16 line 1-2: Provide more detail from supporting manuscript for current manuscript. It is necessary to support this analysis that you can separate specific drivers.**

See above response.

**Page 16 line 3-6: Connect this sentence to the following paragraph.**

Thanks. We will connect the two paragraphs in the revised manuscript.

**Page 16 line 17: Is the "framework" the non-linear component? Maybe just call it that ? non-linear, rather than a framework. This implies a more complex process.**

As we have explained in our response to Referee #1, it is not only the non-linear component that has been applied in this study. Our approach consists of several components including data collection, feature construction, non-linear machine learning algorithms and Granger causality analyses. We will try to make it more clear in the revised manuscript.

**Page 17 line 11: explain “feature construction"**

With the term ‘feature construction’ we mean the process that is followed in order to extract informative predictors for a model. In our case, predictors have been extracted from the raw time series using domain knowledge and include climatic cumulative/lagged variables and climate extreme indices. This term is explained in detail in Sect. 3.3 of the manuscript.

**Page 17 line 16: update word order to read “causality based approaches indicate"**

Yes, thanks.