Authors' response to the Referees

For clarifying our answers to the reviewers' comments, the following scheme is used: comments of the reviewers are denoted in **bold** font, our answers are denoted in plain font and quotes from the manuscript are denoted in *italic*. We also number the different comments of the reviewers as: I.J, where I is the number of the referee and J the number of the remark.

Anonymous Referee #1

The manuscript introduces a Granger causal inference approach to investigate climate-vegetation dynamics. A great effort in collecting a representative enough dataset has been pursued to study such dependencies. The authors put emphasis on the non-linearity of the approach since the VAR method typically used in the canonical Granger approach is here replaced by a non-linear regression tool, the random forests method. Authors claim that the causal patterns are more clearly identifiable than with traditional linear models. Overall, I think this is a very nice piece of work that is worth publishing after some clarifications and addressing some problems.

We would like to thank the reviewer for the appreciation of the manuscript, for the constructive feedback, and for the thorough assessment. Below we provide a point-to-point response to each comment and we present the changes in the revised version of the manuscript.

Below authors will find a long list of minor and major comments that I hope they can address.
- 1.1) abstract: 3: unravel the influence... : this looks like an ambitious goal that I'm not sure authors finally managed to address

We agree with the reviewer and changed the sentence by stating that the technique "allows to further unravel the influence". The sentence now reads *'Data of this kind [Earth observations] provide new means to further unravel the influence of climate on vegetation dynamics'.*

- 1.2) 4: existing statistical methods: do authors refer to linear ones only, right?

'Existing' has been replaced by 'commonly-used' in the revised manuscript.

Abstract: 3: *However, as advocated in this article, commonly-used statistical methods that assume linearity are often too simplistic to represent complex climate-vegetation relationships.*

- 1.3) 8: (also in the title) the word 'framework? looks too ambitious.

In the end, authors only proposed to follow the Granger approach with a different feature selection and regression method. Does this qualify to call it framework?

We understand the concern, yet 'framework' can be defined as 'a basic conceptional structure (as of ideas)' (Merriam-Webster dictionary). We chose to use the word 'framework', because we believe it reflects well the conceptional structure followed here, which goes beyond adopting the Granger approach with a different feature selection and regression method, and consists of several sequential steps. First, multiple datasets of the most important climatic variables have been collected and converted into a common temporal and spatial resolution (a multidimensional data-cube). Second, by applying feature extraction techniques and domain knowledge, predictor variables have been constructed. Third, a non-linear machine learning algorithm has been designed and applied. Fourth, causality has been assessed based on Granger causality. As such, we think that we propose a complete framework that can be used for knowledge discovery in climate sciences, and which in this paper is applied to unveil climate-vegetation dynamics, but that could be used to detect other causal patterns in the climate system. Moreover, by using non-linear models and feature extraction techniques, we argue that this framework substantially differs from methodologies that are common practice in the field. For these reasons, we opted to keep the word 'framework' in the revised manuscript.

1.4) p2.29: y alludes to the NDVI time series: shouldn't be the IAV of NDVI thereof?

It is true that we finally model the IAV of NDVI, this is just a starting point of explaining the basic model. We have added the word 'anomalies' in order to clarify which is the target variable.
p2.31: *In this work* $\boldsymbol{y}$ *alludes to the NDVI anomalies time series at a given pixel, whereas* $\boldsymbol{x}$ *can represent the time series of any climatic variable at that pixel (e.g. temperature, precipitation or radiation).*

1.5) p3.7: for me, describing the R^2 is too verbose and useless in a scientific journal nowadays

The R^2 is indeed a well-known performance measure, we wrote down the formula just to avoid confusion: for linear models one often computes the R^2 as a correlation coefficient, while for non-linear models one has to compute the R^2 in a different way, using the formula in the manuscript. This is perhaps obvious, but it might be useful to have the formula in the paper for readers that are less familiar with the different definitions of R^2. We decided to keep this formula in the revised manuscript. We also stress that the R^2 is computed on data that was not used for model training, so there is no need for adjustments to account for the degrees of freedom of the models, like is commonly done in statistics.

1.6) p3.eq2-3: the $\approx$ symbol is meaningless here. I'd suggest to include the signal model here ($y = \hat{y} + e$), and describe the assumptions about the noise model (Gaussian, uncorrelated?). Also, I don't find natural that both eqs. have the same model coefficients β_{11p}.

We agree. We have incorporated these changes in the revised manuscript.
p3.21:

$$y_t = \hat{y}_t + \epsilon_1 = \beta_{01} + \sum_{p=1}^{P} \left(\beta_{11p} y_{t-p} + \beta_{12p} x_{t-p} \right) + \epsilon_1 \quad (1)$$

$$y_t = \hat{y}_t + \epsilon_1 = \beta_{01} + \sum_{p=1}^{P} \beta_{11p} y_{t-p} + \epsilon_1 \quad (2)$$

with β_{ij} being parameters that need to be estimated and ϵ_1 and ϵ_2 referring to two white noise error terms.

1.7) p3.27: authors should clarify the sentence "neither variables nor observational ... and errors are ...". Independent of what? each other? independent noise? Please be explicit and consistent in the use of the terms 'error', 'noise', 'residuals'.

We meant independent from each other. We added the phrase "from each other" to the revised manuscript.
p3.27: *However, our above definition differs from the perspective in research papers that develop statistical tests for Granger causality (Hacker and Hatemi-J, 2006), because we intend to move away from statistical hypothesis testing, since the assumptions behind such testing are typically violated when working with climate data where neither variables nor observational techniques are fully independent from each other in most cases, and errors are not normally distributed (see Sect. 2.4 for a further discussion).*

1.8) p4.10, eq4: describe the meaning of β_{13} and all terms involved in the equation

We have included a more extensive explanation of the tri-variate extension of Granger causality in the revised manuscript.
p4.15: *As previously mentioned, the time series $\boldsymbol{w}$ may also have a causal effect on $\boldsymbol{y}$ and be correlated with $\boldsymbol{x}$. For this reason, $\boldsymbol{w}$ should be included in both models (baseline and full), so that the method can cope with cross-correlations between climatic drivers of vegetation anomalies.*

1.9) p4.26: Maybe I'm missing something but if you split the data this way, aren't you discarding long-term correlations. Also, by simple xval, results depend to a large extent of the selected data splits. To avoid this, why not LOO?

Our motivation for doing 5-fold cross-validation instead of leave-one-out (LOO) was mainly motivated by computational reasons. LOO takes a long time to compute and is generally not the recommended method when analyzing large datasets (Elisseeff and Pontil, 2003). As we are working with an extremely large dataset here, computational efficiency is always the first criterion to look for. For this reason, we keep the same evaluation procedure in the revised manuscript.

Elisseeff, A., & Pontil, M. (2003). Leave-one-out error and stability of learning algorithms with applications. NATO science series sub series iii computer and systems sciences, 190, 111-130.

1.10) p5.10: the same comment about the $\approx$ symbol before: please include the signal model equations here too.

We have changed it in the revised manuscript.
p5.14:

$$y_t = \hat{y}_t + \epsilon_1 = \beta_{01} + \sum_{p=1}^{P} \left(\beta_{11p} y_{t-p} + \beta_{12p} x_{t-p} + \beta_{13p} w_{t-p} \right) + \epsilon_1 \quad (3)$$

$$x_t = \hat{x}_t + \epsilon_2 = \beta_{02} + \sum_{p=1}^{P} \left(\beta_{21p} y_{t-p} + \beta_{22p} x_{t-p} + \beta_{23p} w_{t-p} \right) + \epsilon_2 \quad (4)$$

$$w_t = \hat{w}_t + \epsilon_3 = \beta_{03} + \sum_{p=1}^{P} \left(\beta_{31p} y_{t-p} + \beta_{32p} x_{t-p} + \beta_{33p} w_{t-p} \right) + \epsilon_3 \quad (5)$$

1.11) p5.15: formally it is straightforward, but not computationally or for decision making which may be an infeasible problem.

Indeed. We only want to convey here that the formal definition of Granger causality does not change in the case of more than three time series.

1.12) p6.1-3: if you want to keep this statement, please discuss about the theoretical implications, and cite other non-linear Granger causality methods (a simple search in Google will return you several dozens of works in machine learning, kernel methods, time series forecasting, econometrics and finance).

We agree that there is previous work on non-linear Granger causality. These methods typically assume stationarity in the time series, and they are hence not immediately applicable for climatic time series. We have extended the paragraph with a more thorough discussion on related work and new references to these articles. We also clarify that we have not introduced non-linear Granger causality for the first time, yet, to our knowledge, more complex methods that use

Granger causality have not been widely applied in the field of climate sciences. We refer to the literature cited in Sect. 2.3.
p6.2: *In other fields, such as in neurosciences, kernel methods or other non-linear models have been used for the investigation of non-linear Granger causality relationships between time series (Marinazzo et al., 2008; Ancona et al., 2004). In our analysis, we stick to simple non-linear methods that are applicable to large datasets. More sophisticated approaches typically do not scale well enough in global climate-vegetation datasets.*

1.13) p6.1-14: verbose, remove or summarize a lot.

This might be obvious for a well-informed reader, but we believe that an explanation of that kind is needed for readers that are less familiar with Granger causality and time series forecasting. We therefore decided to keep this paragraph as it is.

1.14) p9.eq: the upperscript T may confuse as in standard algebra that symbol stands for transpose.

We have replaced the symbol T with T_r in order to make it clearer in the revised manuscript.
p10.20:

$$y_t \approx y_t^{T_r} = \alpha_0 + \alpha_1 \tag{6}$$

1.15) p9.3: obvious non-stationary: sometimes it is not that obvious.

We deleted the word 'obvious' in the revised manuscript.

1.16) p12.6: a sentence does not conform a paragraph.

We tend to disagree, a paragraph is 'a subdivision of a written composition that consists of one or more sentences' (again from the Merriam-Webster). We decided to keep it a stand-alone sentence in order to highlight it as a conclusion from the entire section

1.17) And by the way... is 1 degree enough resolution to claim something about causation? do the expected relations occur at such broad scale?

Most atmospheric variables change consistently at spatial resolutions that are even coarser than 1 degree; in fact most current climate models resolve the land-atmospheric interactions while working at coarser resolutions. We also note that there is a trade-off between spatial resolution and time period covered by the datasets. The 1-degree resolution is a characteristic of the datasets we are working with, and if we wanted to focus on finer resolutions, we would need to incorporate datasets from sensors covering more recent years only, thus multi-decadal analysis would not be possible. The 1-degree resolution, in addition, still also allows us to perform our calculations in a reasonable amount of time.

1.18) p12.13: please avoid overoptimistic phrases like "our non-linear random forestS".

It has been rephrased to "the non-linear random forest model" in the revised manuscript.
p14.9: *To analyze the effect of climate on vegetation more thoroughly, we substitute the linear ridge regression model (VAR) by the non-linear random forest model.*

1.19) p12.17: "simple correlations" should be "spurious correlations"? in any case this sentences deserves more clarification and be more explicit Fig4: some discussions and words of caution should be given about deriving conclusions out of $R^2 \sim 0.4$. By the way, why the maximum in the scale is not explicit for R^2 and you select that threshold in 0.4? Why not using the statistical significance of the correlation rather than the R^2 score? Can authors include and discuss the maps of R p-values?

As mentioned in the manuscript, the assumptions of common statistical tests are violated due to the non-stationarity of the data and the non-linearity of the proposed model. Developing a statistical test that is able to handle non-stationary time series and non-linear models is not a trivial task. As far as we know, no such test exists. Therefore, we decided to focus on expressing Granger causality in a quantitative way instead of a qualitative way, and stress the gained improvement with the use of a non-linear model. We have included the relevant references and a more thorough discussion about existing statistical tests in Section 2.4 of the revised manuscript.
p7.1: *See entire Section 2.4.*

1.20) Fig4 caption: 'with respect to a the' to be corrected

Done.
Fig4 caption: *Improvement in terms of R^2 by the full ridge regression model with respect to the baseline ridge regression model that uses only past values of NDVI anomalies as predictors;*

1.21) p13.3: 'our'?

We replaced "our" with "the" in the revised manuscript.

1.22) p14.3: what are these patterns of the explained variance? some clarification is needed here? I guess authors refer to spatial patterns of variation? If that is the case, it looks not really obvious to talk about spatial relations when no such relations are considered to build up the regression models.

In fact this section does not refer to spatial patterns, but to a general improvement of the full model versus the restricted model. We removed the word "patterns" from the referred statement in the revised manuscript.

1.23) p14.7: unambiguous? some more comments are needed, and if possible supported by numerical scores.

With unambiguous we just mean that the improvement is clearly visible here (in the order of 20 to 60%). This claim is supported by Fig. 5b. So we opted to keep this word.

1.24) p13-14: as a reader I'd prefer to have in the same figure panel the current figures 4 and 5 so I could directly compare results in one shot.

We understand this concern and understand that a 3×2 figure would appear somehow more convenient. However, we chose to have Fig. 5 in a separate panel because this is the main figure of the paper.

1.25) p15.3: what do authors mean by 'higher-lever variables'? are you thinking of higher-order statistical relations between variables? this is absolutely confusing.

We agree. With the term 'higher-level variables' we refer to the past cumulative climate, lagged variables and climate extreme indices that are considered as predictor variables. This has now been clarified in Section 3.3 of the revised manuscript.
p12.1: *We do not limit our approach to considering raw versus anomaly time series of the data sets in Table 1 as predictors, but also take into consideration different lag times, past-time cumulative values and extreme indices. These additional predictors, further referred to as higher-level variables, are calculated based on raw and anomaly time series.*

1.26) p15.5: please provide a copy of the (Papagiannopolou et al, in review) so reviewers can appreciate differences in approaches and results. Alternative, cite an accessible work to support the claims in this paper.

The referred article is enclosed. This should be made available to the editorial and reviewers only.

1.27) p16.3-18: please clarify these paragraphs in several ways: 1) the spatial encoding is not at all clear since typically the input (feature) space is augmented with the neighbors which are then used to predict on the central pixel (the length of the observation variable does not change), which seems not to be the case here. 2) it is weird that the spatial info didn't improve the results: I'd thank the authors to include such 'negative results' but then some comments and clarifications are needed (e.g. 1 degree is already integrating too much info, or spatial encoding was not taking into account pixel spatiotemporal variances?)

Yes, the approach we followed is as described by the reviewer. The feature space of one pixel is augmented with the features of the 8 neighboring pixels. We also expected to see a more substantial improvement using spatial information but this is not the case. We have extended the discussion in page 18 of the revised manuscript to hypothesize a reason for this limited improvement.
p18.20: *A possible explanation for this result is that the model without the spatial information cannot be outperformed because of the large dimensionality of the feature space, which may include redundant information, in combination with the low number of observations per pixel (Fig. 5a). Note that in this case the number of observations per pixel remains the same as in the original model (360 observations) while the number of predictor variables is 9 times larger.*

1.28) p17.9: as said before I feel claiming a 'novel framework? is far too much for this contribution.

See response 1.3.

1.29) p17.15-20: some claims are contained here without empirical justification. I think that authors lost a nice opportunity here to explain the causal relations. For example, to me it seems ad hoc to justify results with a simple 'the predictive power of the model is especially high in water-limited regions'. Probably this is true but some numbers are needed to support it. I suggest to include a summarizing feature ranking of the LR vs RFs (e.g. permutation analysis, and surrogate analysis). Also, summarize results per regions and biomes would help discussing the results more profoundly, elevating the debate. Of course, these two issues may require some more work, but I sincerely think they are mandatory to make a sound publication.

Actually, we have performed this kind of analyses, taking feature rankings using RFs. However, these rankings become unstable due to highly-correlated predictors. A specialized approach would be needed here, in which groups of features are ranked instead of individual features. This makes the rankings more stable

and improves the interpretability. It is exactly what we do in the complementary paper (Papagiannopoulou et al., under review). We also agree with the reviewer that a stratification of the results according to regions/biomes is a relevant addition to the paper. The revised version provides the results stratified according to IGBP land cover classes for both the baseline and the full random forest model. These new results are discussed in Section 4.2 and a new figure, Fig. 6 has been added.
p15.20: *For a better understanding of the results obtained by the two models, we average the performance of each model regionally. More specifically, we use the International Geosphere-Biosphere Program (IGBP) (Loveland and Belward, 1997) land cover classification to stratify the mean and variance of R^2 for both the baseline and the full model in Fig. 5 per IGBP land cover class. The barplot in Fig. 6 shows that the full model outperforms the baseline model in all IGBP land cover classes, i.e. that Granger causality exists for all these biomes. In the parentheses we note the number of pixels per region. The error bars indicate that the variances of the two models are analogous, i.e. they are low or high in both models in the same land cover class. For the Closed Shrublands region, one can observe the highest difference between the two models, yet only 19 pixels belong to this biome type. In savanna regions, the performance of the full model is high in comparison with other regions (see Fig. 5). On the other hand, the lowest performance improvement of the full model with respect to the baseline is observed for the regions of Deciduous Needleleaf Forests and Evergreen Broadleaf Forests. This shows that for these two regions climate is not identified as a major control over vegetation dynamics (see discussion in previous paragraph about tropical and boreal regions).*

1.30) p18.8: reproducibility is not possible as data is not available yet. do authors plan to make these data available to the community?

All codes are freely available and documented on GitHub and will comply with the Copernicus data policy. On the other hand, the full database is formed by a collection of datasets that are all publicly available and that, due to copyright conflicts, cannot be openly distributed. The relevant link is provided in Section 6.

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 detrended 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 is 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-up 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.
Below we provide our pointwise response to the review, as well as the changes in the revised manuscript.

Detailed comments:
2.1) 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 relation-*

ship 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. Therefore, we opted to keep the sentence as it is.

2.2) 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)

See response 1.25.

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

See response 1.25.

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

See response 1.5.

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

We have changed this phrase in the revised manuscript.
p4.4: *However, such an analysis might lead to incorrect conclusions, because additional (confounding) effects exerted by other climatic or environmental variables are not taken into account (Geiger et al, 2015).*

2.6) 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.
This is stated in Section 4.1.

2.7) 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.

2.8) 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 in these regions. Therefore, the results indicate that the Granger causal effects of climate on vegetation anomalies in these areas 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.

- 2.9) 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.

- 2.10) The citation references error for soil moisture.

Corrected.
Dorigo, W., Wagner, W., and et al.: ESA CCI Soil Moisture for improved Earth System understanding: state-of-the-art and future directions, Remote Sensing of Environment, in review.

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

See response 2.8.

2.12) 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. However, we understand the fact that the presentation of additional results in the "Discussion" section is a bit confusing. Therefore we decided to combine the two sections in one with the name "Results and discussion".

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

We have added the word 'anomalies' here in the revised manuscript.
p15.7: *(b) these are regions in which vegetation anomalies are not necessarily primarily controlled by climate, but may also be driven by phenological and bi-*

otic factors

2.14) 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. We kept the word in the revised manuscript.

2.15) Page 14 line 7-12: move to discussion.

See response 2.12.

2.16) 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.

2.17) 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.

See response 1.24.

2.18) 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 response 1.26.

2.19) 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. Results of isolating the impact of extremes are also presented in that same manuscript. Again, this paper is enclosed in the resubmission.

2.20) 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 response 1.26.

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

We have connected the two paragraphs in the revised manuscript by adding the following sentence:
p18.8: *In addition, it is quite likely that neighboring areas have similar climatic conditions, which in their turn affect vegetation dynamics in a similar manner.*

2.22) 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.

See response 1.3.

2.23) Page 17 line 11: explain "feature construction"

With the term 'feature construction' we refer to the process of extracting 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. The explanation is given in the corresponding section (Section 3.3).

2.24) Page 17 line 16: update word order to read "causality based approaches indicate"

Corrected.
p19.21: *Comparisons to traditional Granger causality based approaches indicate*

that the random forest framework can predict 14% more variability of vegetation anomalies on average globally.

A non-linear Granger causality framework to investigate climate–vegetation dynamics

Christina Papagiannopoulou[1], Diego G. Miralles[2,3], Stijn Decubber[1], Matthias Demuzere[2], Niko E. C. Verhoest[2], Wouter A. Dorigo[4], and Willem Waegeman[1]

[1]Depart. of Mathematical Modelling, Statistics and Bioinformatics, Ghent University, Belgium,
[2]Laboratory of Hydrology and Water Management, Ghent University, Belgium
[3]Depart. of Earth Sciences, VU University Amsterdam, the Netherlands
[4]Depart. of Geodesy and Geo-Information, Vienna University of Technology, Austria

Correspondence to: Christina Papagiannopoulou (christina.papagiannopoulou@ugent.be)

Abstract. Satellite Earth observation has led to the creation of global climate data records of many important environmental and climatic variables. These ~~take~~ come in the form of multivariate time series with different spatial and temporal resolutions. Data of this kind provide new means to further unravel the influence of climate on vegetation dynamics. However, as advocated in this article, ~~existing~~ commonly-used statistical methods are often too simplistic to represent complex climate-vegetation relationships due to ~~the assumption of linearity of these relationships~~linearity assumptions. Therefore, as an extension of linear Granger causality analysis, we present a novel non-linear framework consisting of several components, such as data collection from various databases, time series decomposition techniques, feature construction methods and predictive modelling by means of random ~~forests~~forest. Experimental results on global data sets indicate that, with this framework, it is possible to detect non-linear patterns that are much less visible with traditional Granger causality methods. In addition, we ~~also~~ discuss extensive experimental results that highlight the importance of considering ~~the~~ non-linear ~~aspect~~ aspects of climate–vegetation dynamics.

1 Introduction

Vegetation dynamics and ~~ecosystem distribution~~ the distribution of ecosystems are largely driven by the availability of light, temperature and water, thus they are mostly sensitive to climate conditions (Nemani et al., 2003; Seddon et al., 2016; Papagiannopoulou et al., *in review*). Meanwhile, vegetation also plays a crucial role in the global climate system. Plant life alters the characteristics of the atmosphere through the transfer of water vapour, exchange of carbon dioxide, partition of surface net radiation (e.g. albedo), and ~~impact of~~ impacts on wind speed and direction (~~McPherson et al., 2007; Bonan, 2008; Nemani et al., 2003; Seddon~~ Papagiannopoulou et al., *in review*). Because of the strong two-way relationship between terrestrial vegetation and climate variability, predictions of future climate can be improved through a better understanding of the vegetation response to past climate variability.

The current wealth of Earth ~~Observation (EO)~~ observation data can be used for this purpose. Nowadays, independent sensors on different platforms collect optical, thermal, microwave, altimetry and gravimetry information, and are used to monitor vegetation, soils, oceans and atmosphere (e.g. ~~Su et al., 2011; Lettenmaier et al., 2015~~ Su et al., 2011; Lettenmaier et al., 2015; McCabe et al., 2

The longest composite records of environmental and climatic variables already span up to 35 years, enabling the study of multi-decadal climate–biosphere interactions. Simple correlation statistics and multi-linear regressions using some of these data sets have led to important steps forward in understanding the links between vegetation and climate (e.g. ~~Nemani et al., 2003; Wu et al., 2015; Ba~~ However, ~~correlations~~ these methods in general are insufficient when it comes to assessing causality, particularly in systems like the land–atmosphere continuum in which complex feedback mechanisms are involved. A commonly used alternative consists of Granger causality modelling ~~(Attanasio et al., 2013)~~ (Granger, 1969) . Analyses of this kind have been ~~commonly~~ applied in climate attribution studies, to investigate the influence of one climatic variable on ~~an~~ another, e.g., the Granger ~~causality~~ causal effect of CO_2 on global temperature (Triacca, 2005; Kodra et al., 2011; Attanasio, 2012), of vegetation and snow coverage on temperature (Kaufmann et al., 2003), of sea surface temperatures on the North Atlantic Oscillation (Mosedale et al., 2006), or of the El Niño Southern Oscillation on the Indian monsoons (Mokhov et al., 2011). Nonetheless, Granger causality should ~~neither~~ not be interpreted as 'real causality'; ~~to simplify,~~ one assumes that a time series A Granger-causes a time series B if the past of A is helpful in predicting the future of B (see Sect. 2 for a more formal definition). ~~The~~ However, the underlying statistical model that is commonly considered in such a context is a linear vector autoregressive model, which is (again), by definition, linear – see e.g. Shahin et al. (2014); Chapman et al. (2015).

In this article we show new experimental evidence that advocates ~~for the use~~ the need of non-linear methods to study climate–vegetation dynamics, due to the non-linear nature of these interactions (Foley et al., 1998; Zeng et al., 2002; Verbesselt et al., 2016). To this end, we have assembled a ~~vast~~ large, comprehensive database, comprising various global data sets of temperature, radiation and precipitation, originating from multiple online resources. We use the Normalized Difference Vegetation Index (NDVI) to characterise vegetation, which is commonly used as a proxy of plant productivity ~~(Myneni et al., 1997)~~ (Myneni et al., 1997; Ner We followed an inclusive data collection approach, aiming to consider all ~~the~~ available data sets with a worldwide coverage, and at least a thirty-year time span and monthly temporal resolution (Sect. 3). Our novel non-linear Granger causality framework is used for finding climatic drivers of vegetation and consists of several steps (Sect. 2). In a first step, we apply time series decomposition techniques to the vegetation and the various climatic time series to isolate seasonal cycles, trends and anomalies. Subsequently, we explore various techniques for constructing ~~high-level~~ more complex features from the decomposed climatic time series. In a final step, we run a Granger causality analysis on the NDVI anomalies, while replacing traditional linear vector autoregressive models by random forests. This framework allows for modelling non-linear relationships ~~, incorporating higher-level climatic variables and avoiding~~ and prevents over-fitting. The results of the global application of our framework are discussed in Sect. 4.

2 A Granger causality framework for geosciences

2.1 Linear Granger causality revisited

We start with a formal introduction to Granger causality for the case of two times series, denoted as $\boldsymbol{x} = [x_1, x_2, ..., x_N]$ and $\boldsymbol{y} = [y_1, y_2, ..., y_N]$, with N being the length of the time series. In this work $\boldsymbol{y}$ alludes to the NDVI anomalies time series at a given pixel, whereas $\boldsymbol{x}$ can represent the time series of any climatic variable at that pixel (e.g. temperature, precipitation~~or~~

, radiation). Granger causality can be interpreted as predictive causality, for which one attempts to forecast y_t (at the specific time stamp t) given the values of $\boldsymbol{x}$ and $\boldsymbol{y}$ in previous time stamps~~of $\boldsymbol{y}$ and $\boldsymbol{x}$~~. Granger (1969) postulated that $\boldsymbol{x}$ causes $\boldsymbol{y}$ if the autoregressive forecast of $\boldsymbol{y}$ improves when information of $\boldsymbol{x}$ is taken into account. In order to make this definition more precise, it is important to introduce a performance measure to evaluate the forecast. Below we will work with the coefficient of determination R^2, which is here defined as follows:

$$R^2(\boldsymbol{y}, \hat{\boldsymbol{y}}) = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum_{i=P+1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=P+1}^{N}(y_i - \bar{y})^2} \tag{1}$$

where $\boldsymbol{y}$ represents the observed time series, $\bar{y}$ is the mean of this time series, $\hat{\boldsymbol{y}}$ is the predicted time series obtained from a given forecasting model, and P is the length of the lag-time moving window. Therefore, the R^2 can be interpreted as the fraction of explained variance by the forecasting model, and it increases when the performance of the model increases, reaching the theoretical optimum of 1 for an error-free forecast, and being negative when the predictions are less representative of the observations than the mean of the observations. Using R^2, one can now define Granger causality in a more formal way.

Definition 1. *We say that time series $\boldsymbol{x}$ Granger-causes $\boldsymbol{y}$ if $R^2(\boldsymbol{y}, \hat{\boldsymbol{y}})$ increases when $x_{t-1}, x_{t-2}, ..., x_{t-P}$ are ~~considered for predicting~~ included in the prediction of y_t, in contrast to considering $y_{t-1}, y_{t-2}, ..., y_{t-P}$ only, where P is the lag-time moving window.*

In climate sciences, linear vector autoregressive (VAR) models are often employed to make forecasts (Stock and Watson, 2001; Triacca, 2005; Kodra et al., 2011; Attanasio, 2012). A linear VAR model of order P boils down to the following representation:

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} \beta_{01} \\ \beta_{02} \end{bmatrix} + \sum_{p=1}^{P} \begin{bmatrix} \beta_{11p} & \beta_{12p} \\ \beta_{21p} & \beta_{22p} \end{bmatrix} \begin{bmatrix} y_{t-p} \\ x_{t-p} \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix}$$

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} \beta_{01} \\ \beta_{02} \end{bmatrix} + \sum_{p=1}^{P} \begin{bmatrix} \beta_{11p} & \beta_{12p} \\ \beta_{21p} & \beta_{22p} \end{bmatrix} \begin{bmatrix} y_{t-p} \\ x_{t-p} \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix} \tag{2}$$

with β_{ij} being parameters that need to be estimated and ϵ_1 and ϵ_2 referring to two white noise error terms. This model can be used to derive the predictions required to determine Granger causality. In that sense, time series $\boldsymbol{x}$ Granger-causes time series $\boldsymbol{y}$ if at least one of the parameters β_{12p} for any p significantly differs from zero. Specifically, and since we are focusing ~~only~~ on the vegetation time series as ~~target time series~~the only target, the following two models are compared:

$$y_t \cancel{\approx} = \hat{y}_t + \epsilon_1 = \beta_{01} + \sum_{p=1}^{P}\left(\beta_{11p} y_{t-p} + \beta_{12p} x_{t-p}\right) + \epsilon_1 \tag{3}$$

$$y_t \cancel{\approx} = \hat{y}_t + \epsilon_1 = \beta_{01} + \sum_{p=1}^{P} \beta_{11p} y_{t-p} + \epsilon_1 \tag{4}$$

We will refer to model (3) as the *full model* and to model (4) as the *baseline model*, since the former incorporates all available information and the latter only information of $\boldsymbol{y}$.

Comparing the above two models, $\boldsymbol{x}$ Granger-causes $\boldsymbol{y}$ if the full model manifests a ~~substantially-better predictive performance~~ substantially better predictive performance in terms of R^2 than the baseline model. To this end, statistical tests can be employed, for which one typically assumes that the errors in the model follow a Gaussian distribution (Maddala and Lahiri, 1992). However, our above definition differs from the perspective in research papers that develop statistical tests for Granger causality (Hacker and Hatemi-J, 2006), because we intend to move away from statistical hypothesis testing, since the assumptions behind such testing are typically violated when working with climate data where neither variables nor observational techniques are fully independent from each other in most cases, and errors are not normally distributed ~~.~~ (see Sect. 2.4 for a further discussion).

In climate studies, the Granger causal relationship between two time series $\boldsymbol{x}$ and $\boldsymbol{y}$ has often been investigated in the bivariate setting (Elsner, 2006, 2007; Kodra et al., 2011; Attanasio, 2012; Attanasio et al., 2012). However, such an analysis might lead to ~~wrong~~ incorrect conclusions, because additional (confounding) effects exerted by other climatic or environmental variables are not taken into account (Geiger et al., 2015). This problem can be ~~solved~~ mitigated by considering time series of additional variables. For example, let us assume one has observed a third variable $\boldsymbol{w}$, which might act as a confounder in deciding whether $\boldsymbol{x}$ Granger-causes $\boldsymbol{y}$. The above definition then naturally extends as follows.

Definition 2. *We say that time series $\boldsymbol{x}$ Granger-causes $\boldsymbol{y}$ conditioned on time series $\boldsymbol{w}$ if $R^2(\boldsymbol{y},\hat{\boldsymbol{y}})$ increases when $x_{t-1}, x_{t-2}, \ldots, x_{t-P}$ are ~~considered for predicting~~ included in the prediction of y_t, in contrast to considering $y_{t-1}, y_{t-2}, \ldots, y_{t-P}$ and $w_{t-1}, w_{t-2}, \ldots, w_{t-P}$ only, where P is the lag-time moving window.*

~~An extension of this definition for more than three times series is straightforward. In~~ Similarly as above, we refer to the two models as full and baseline model, respectively. Therefore, in the tri-variate setting, Granger causality might be tested using the following linear VAR model:

$$\begin{bmatrix} y_t \\ x_t \\ w_t \end{bmatrix} = \begin{bmatrix} \beta_{01} \\ \beta_{02} \\ \beta_{03} \end{bmatrix} + \sum_{p=1}^{P} \begin{bmatrix} \beta_{11p} & \beta_{12p} & \beta_{13p} \\ \beta_{21p} & \beta_{22p} & \beta_{23p} \\ \beta_{31p} & \beta_{32p} & \beta_{33p} \end{bmatrix} \begin{bmatrix} y_{t-p} \\ x_{t-p} \\ w_{t-p} \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{bmatrix}, \tag{5}$$

where a causal relationship between $\boldsymbol{x}$ and $\boldsymbol{y}$ exists if at least one β_{12p} significantly differs from zero. As previously mentioned, the time series $\boldsymbol{w}$ might also have a causal effect on $\boldsymbol{y}$ and be correlated with $\boldsymbol{x}$. For this reason $\boldsymbol{w}$, should be included in both models (baseline and full), so that the method can cope with cross-correlations between predictors, in our case between the climatic drivers of vegetation anomalies. An extension of this definition for more than three times series is straightforward.

2.2 Over-fitting and out-of-sample testing

It is well known in the statistical literature that predictions made on in-sample data, that is, ~~data~~ the same data that was used to fit ~~a~~ the statistical model, tend to be optimistic. This process is often referred to as over-fitting, i.e., by definition, the fitting process ~~chooses parameters that mimic the~~ leads to parameter values that cause the model to mimic the observed data as closely as possible (Friedman et al., 2001). In the context of Granger causality analysis, over-fitting will occur more prominently in the multivariate case, when the number of considered time series increases. ~~In the experimental analysis that will be presented~~ The results in Sect. 4 ~~, we will consider in total 21 time series, resulting in $21(P+1)$ different parameters for each equation of the VAR model. Simple models of that kind will as a result already be~~ are based on multivariate analysis, thus they are vulnerable to over-fitting~~. The~~ ; the situation further aggravates when switching from linear to non-linear models, because then the number of parameters typically increases ~~since the general form of the model is unknown~~to allow for a more flexible functional model form.

To prevent over-fitting, out-of-sample data should be used in evaluating the predictive performance in Granger causality studies (Gelper and Croux, 2007). The most straightforward procedure for creating out-of-sample data is to separate the time frame into two parts, a training set and a test set, which typically constitute the first and last half of the time frame. A few authors have adopted this approach for ~~climate~~ climatic attribution (Attanasio et al., 2012; Pasini et al., 2012); however, satellite ~~EO~~ Earth observation time series are usually too short to allow for train-test splitting in that fashion. An alternative approach, which uses the available data in an efficient manner, is cross-validation. To this end, the time frame is divided in a number of short intervals, typically a few years of data, in which one interval serves as a test set, while all remaining data are used for parameter fitting. This procedure is repeated until all intervals have served once as a test set, and the prediction errors obtained in each round are aggregated, so that one global performance measure can be computed. We direct the reader to ~~Von Storch and Zwiers (2001) and Michaelsen (1987)~~ Michaelsen (1987) and Von Storch and Zwiers (2001) for further discussion.

The inclusion of a regularization term in the fitting process of over-parameterized linear models will avoid over-fitting. Typical regularizers that shrink the parameter vectors of linear models towards zero are L2-norms as in ridge regression, L1-norms as in Least Absolute Shrinkage and Selection Operator (LASSO) models, or a combination of the two norms, as in elastic nets (Friedman et al., 2001). Translated to VAR models, this implies that one should impose restrictions on the parameter matrix of Eq. (5), as done in ~~recent theoretical papers (Gregorova et al., 2015)~~ the recent theoretical paper of Gregorova et al. (2015) . In this work, we want to identify causal relationships between a vegetation time series and various climatic time series~~, for which~~ . Hence, there is only one target variable of interest, and a simpler approach can be adopted~~, since the target variable, in our case is only one~~. Denoting the vegetation time series by $\boldsymbol{y}$, one can mimic in the tri-variate setting a VAR model by means of three autoregressive ridge regression models:

$$y_t \quad \cancel{\approx} = \quad \hat{y}_t + \epsilon_1 = \beta_{01} + \sum_{p=1}^{P} \left(\beta_{11p} y_{t-p} + \beta_{12p} x_{t-p} + \beta_{13p} w_{t-p} \right) + \epsilon_1 \tag{6}$$

$$x_t \quad \cancel{\approx} = \quad \hat{x}_t + \epsilon_2 = \beta_{02} + \sum_{p=1}^{P} \left(\beta_{21p} y_{t-p} + \beta_{22p} x_{t-p} + \beta_{23p} w_{t-p} \right) + \epsilon_2 \tag{7}$$

$$w_t \quad \cancel{\approx} = \quad \hat{w}_t + \epsilon_3 = \beta_{03} + \sum_{p=1}^{P} \left(\beta_{31p} y_{t-p} + \beta_{32p} x_{t-p} + \beta_{33p} w_{t-p} \right) + \epsilon_3 \tag{8}$$

In this article we aim to detect the climate drivers of vegetation, and not the feedback of vegetation on climate (see e.g., Green and et al., *in review*). Therefore, it suffices to retain Eq. (~~9~~6) in our analysis as is stated above for the ~~bivariate~~ tri-variate case (Eq. ~~4~~5). Concatenating all parameters of this model into a vector ~~$\boldsymbol{\beta} = [\beta_{01}, \beta_{111}, ..., \beta_{13p}]$~~ $\boldsymbol{\beta} = [\beta_{01}, \beta_{11p}, ..., \beta_{13p}]$, one fits in ridge regression the parameters by solving the following optimization problem:

$$\min_{\boldsymbol{\beta}} \sum_{P+1}^{N} (y_t - \hat{y}_t)^2 + \lambda ||\boldsymbol{\beta}||^2,$$

$$\min_{\boldsymbol{\beta}} \sum_{P+1}^{N} (y_t - \hat{y}_t)^2 + \lambda ||\boldsymbol{\beta}||^2 \tag{9}$$

with λ being a regularization parameter, that is tuned using a validation set or nested cross-validation and $||\boldsymbol{\beta}||^2$ being a penalty term, i.e. the squared ℓ_2 norm of the coefficient vector. The sum only starts at $P+1$ because a moving window of P lags is considered. For simplicity, we describe the above approach for the tri-variate setting, even though the total number of variables used in our study ~~amounts to 3,884~~is a lot larger (see Sect. 3); nonetheless, extensions ~~of this~~ to the multivariate setting are straightforward.

2.3 Non-linear Granger causality

The methodology that we develop in this paper is closely connected to the methods explained in the previous section. However, as we hypothesize that the relationships between climate and vegetation can be highly non-linear (Foley et al., 1998; Zeng et al., 2002; Verbesselt et al., 2016), we also replace the linear VAR-models in the Granger causality framework with non-linear machine learning models. ~~The~~ In other fields, such as in neurosciences, kernel methods or other non-linear models have been used for the investigation of non-linear Granger causality relationships between time series (Ancona et al., 2004; Marinazzo et al., 2008) . In our analysis, we stick to simple non-linear methods that are applicable to large datasets. More sophisticated approaches typically do not scale well enough in global climate-vegetation datasets. Therefore, in our work, the machine learning algorithm we choose is random forests ~~(Breiman, 2001) ,~~ due to its excellent computational scalability (Breiman, 2001) . Random forests are a well-known method that has shown its merits in diverse application domains, and that has successfully been applied to ~~EO~~ Earth observations in both classification and regression problems (Dorigo et al., 2012; Rodriguez-Galiano et al., 2012; Loosvelt et al., 2012a, b). Briefly summarized, the random forest algorithm forms a combination of multiple decision trees, where each tree contributes with a single vote to the final output, which is the most frequent class (for classification problems) or the average (for regression problems).

~~However, compared~~ Compared to most application domains where random forests are applied, we employ the algorithm in a slightly different way~~:~~ , as an autoregressive non-linear method for time series forecasting~~, similar to VAR models but non-linear~~. In practice, this means that we ~~simply~~ replace the full and baseline linear model of Sect. 2.1 by a random forest model. At each pixel, the vegetation time series ~~will still be~~ is still considered as response variable, and the various climate

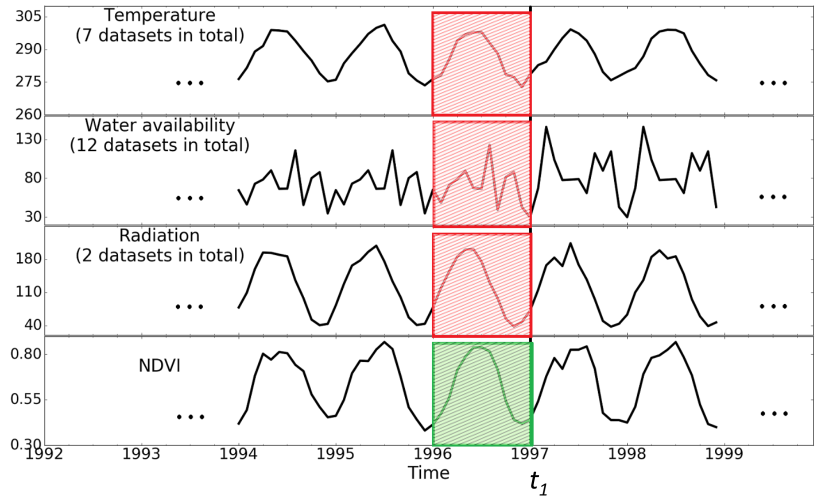

Figure 1. An illustrative example of the moving window approach considered in the analysis of vegetation drivers at a given time stamp t_1. NDVI takes here the role of the time series $\boldsymbol{y}$ in Eq. 3. In addition three climate predictor time series are shown. The baseline random forest model only considers the green moving window, whereas the full random forest model includes the red moving windows as well. The pixel corresponds to a location in North America (lat: 37.5, long: -87.5).

time series serve as predictor variables (see Sect. 3.1 for an overview of our database). For a given value of the NDVI time series $\boldsymbol{y}$ at time stamp t, we investigate properties of the different predictor time series – i.e., temperature, radiation, etc. – by considering a moving window including a number of previous months (Fig.1). In this way, the definition of Granger causality in Sect. 2.1 is adopted. Any climatic time series $\boldsymbol{x}$ Granger-causes vegetation time series $\boldsymbol{y}$ if the predictive performance in terms of R^2 improves when the moving window $x_{t-1}, x_{t-2}, ..., x_{t-P}$ is incorporated in the random forest, in contrast to considering $y_{t-1}, y_{t-2}, ..., y_{t-P}$ and $w_{t-1}, w_{t-2}, ..., w_{t-P}$ only. Analogous to the linear case, we will ~~refer to the~~ speak of a full random forest model when all variables are taken into account and ~~to the~~ of a baseline random forest model when only the moving window $y_{t-1}, y_{t-2}, ..., y_{t-P}$ of $\boldsymbol{y}$ is considered as predictor. In Fig. 1 this principle is extended to four time series. The baseline random forest predictions of NDVI at t_1 are based on the observations from the green moving window only, whereas the full random forest model includes the three red moving windows as well~~when predicting the NDVI at t_1~~.

In our experiments, we treat each continental pixel as a separate problem, and use the ~~implementation of~~ scikit-learn library (Pedregosa et al., 2011) for the random forest regressor ~~, setting~~ implementation, with the number of trees equal to 100 and the maximum number of predictor variables per node equal to the square root of the total number of predictor variables. Changes in these parameters ~~(or~~ or in the randomness of the algorithm ~~) did~~ do not cause substantial changes in the results (not shown). ~~The model was~~ Model performance is assessed by means of five-fold cross-validation. The window length ~~was~~ is fixed to twelve months because initial experimental results revealed that longer time windows did not lead to improvements in the predictions (results omitted). ~~We~~ Finally, we also experimented with techniques that exploit spatial correlations to improve the ~~predicted~~ predictive performance of the model (see Sect. 4.1).

2.4 Granger causal inference

Generally, the null hypothesis (H_0) of Granger causality is that the baseline model has equal prediction error as the full model. Alternatively, if the full model predicts the target variable $\boldsymbol{y}$ significantly better than the baseline model, H_0 is rejected. In some applications, inference is drawn in VAR by testing for significance of individual model parameters. Other studies have used likelihood-ratio tests, in which the full and baseline models are nested models (Mosedale et al., 2006) . However, in both cases, the models are trained and evaluated on the same in-sample data. As it has been discussed above, the performance of any Granger causal model should be validated on out-of-sample data to avoid overfitting (see Sect. 2.2). Therefore, the null hypothesis of non-causality in the formulation stated above should be tested for by comparing out-of-sample prediction errors. To this end, statistical tests have been proposed and applied both in the econometric literature as well as in Granger causality studies in the context of climate science. This kind of tests, which compare out-of-sample prediction errors, are available for models for which parameter estimation is done through ordinary least squares or maximum likelihood estimation (Attanasio et al., 2013) . Moreover, the asymptotic and finite-sample properties of a battery of tests for comparing forecasting accuracies of different models have been studied and more recently, further tests aiming specifically at nested models have been proposed (Clark and McCracken, 2001) .

Unfortunately, all the tests mentioned above were designed to compare the out-of-sample prediction errors of linear parametric models (McCracken, 2007) . In climate, relations between variables are highly non-linear and tend to become even more non-linear as the temporal resolution of the data becomes finer (Attanasio et al., 2013) . Therefore, it would be convenient to have at our disposal a statistical test to assess the significance of any quantitative evidence of climate Granger-causing vegetation anomalies that we can find. Ideally, the test would be model-independent so that any non-linear model could be used. One well-known model-independent test to compare the accuracy of two forecasts is the Diebold-Mariano test (DM-test) (Diebold, 2015) . Although its application to Granger causality is promising, the test does not hold for nested models, because under H_0, the prediction errors from two nested models are exactly the same and perfectly correlated (McCracken, 2007) . An alternative approach for comparing the predictive performance of different models is to use resampling methods such as the bootstrap or schemes such as 5×2 cross-validation (Dietterich, 1998) . Methods based on the bootstrap have been used before in Granger causality studies with climate data (Diks and Mudelsee, 2000; Attanasio et al., 2013) . However, these results need to be interpreted with care because, by increasing the number of bootstrap samples, the power of any paired test (such as the Wilcoxon signed rank test) to detect significant differences between the error distributions of both models (full and baseline) increases as well. For these reasons, we conclude that developing a statistical test that is able to handle non-stationary time series and non-linear models is not a trivial task. To the best of our knowledge, no such test exists in the current literature. In this paper, we focus on expressing Granger causality in a quantitative instead of a qualitative way, and stress the gained improvement with the use of a non-linear model.

3 Database creation and variable construction

3.1 Global data sets

Our non-linear Granger causality framework is used to disentangle the effect of past-time climate variability on global vegetation dynamics. To this end, climate data sets of observational nature – mostly based on satellite and *in situ* observations – have been assembled to construct time series (see Sect. 3.3) that are then used to predict NDVI anomalies following the linear and non-linear causality frameworks described in Sect. 2. Data sets have been selected from the current pool of satellite and *in situ* observations on the basis of meeting a series of spatio-temporal requirements: (a) expected relevance of the variable for driving vegetation dynamics, (b) multi-decadal record and global coverage available, and (c) adequate spatial and temporal resolution. The selected data sets can be classified into three different categories: water availability (including precipitation, snow water equivalent and soil moisture data sets), temperature (both for the land surface and the near-surface atmosphere), and radiation (considering different radiative fluxes independently). Rather than using a single data set for each variable, we have collected all data sets meeting the above requirements. This has led to a total of twenty-one different data sets which are listed in Table 1. They span the study period 1981–2010 at the global scale, and have been converted to a common monthly temporal resolution and $1^\circ \times 1^\circ$ latitude-longitude spatial resolution. To do so, we have used averages to re-sample original data sets found at finer native resolution, and linear interpolation to resample coarser-resolution ones.

For temperature we consider seven different products based on *in situ* and satellite data: Climate Research Unit (CRU-HR) ~~Harris et al. (2014)~~ (Harris et al., 2014) , University of Delaware (UDel) (Willmott et al., 2001), NASA Goddard Institute for Space Studies (GISS) (Hansen et al., 2010), Merged Land-Ocean Surface Temperature (MLOST) (Smith et al., 2008), International Satellite Cloud Climatology Project (ISCCP) (Rossow and Duenas, 2004), and Global Land Surface Temperature Data (LST) (Coccia et al., 2015). We also included one reanalysis data set, the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-Interim (Dee et al., 2011). In the case of precipitation, eight products have been collected. Four of them result from the merging of *in-situ* data only: Climate Research Unit (CRU-HR) (Harris et al., 2014), University of Delaware (UDel) (Willmott et al., 2001), Climate Prediction Center Unified analysis (CPC-U) (Xie et al., 2007), and the Global Precipitation Climatology Centre (GPCC) (Schneider et al., 2008). The rest result from a combination of *in-situ* and satellite data, and may include reanalysis: CPC Merged Analysis of Precipitation (CMAP) (Xie and Arkin, 1997), ERA-Interim (Dee et al., 2011), Global Precipitation Climatology Project (GPCP) (Adler et al., 2003), and Multi-Source Weighted-Ensemble Precipitation (MSWEP) (Beck et al., 2017). For radiation two different products have been collected (considering incoming shortwave/longwave and surface net radiation as different time series); first the NASA Global Energy and Water cycle Exchanges (GEWEX) Surface Radiation Budget (SRB) (Stackhouse et al., 2004) based on satellite data, and the second one the ERA-Interim reanalysis (Dee et al., 2011). For soil moisture we use the Global Land Evaporation Amsterdam Model (GLEAM) ~~(Miralles et al., 2011)~~ (Miralles et al., 2011; Martens et al., 2016) , and the Climate Change Initiative (CCI) product (Liu et al., 2012, 2011); two different soil moisture products by CCI are considered: the passive microwave dataset and the combined active/passive product (Dorigo et al., *in review*). Moreover, snow water ~~equivalents data come~~ equivalent data comes from the GlobSnow project (Luojus et al., 2010).

Table 1. Data sets used in our experiments. Basic data set characteristics are provided, including the native spatial and temporal resolutions~~and the primary data source~~.

Variable	Product Name	Spatial Resolution	Temporal Resolution	Primary data source	Reference
Temperature	CRU-HR	0.5°	monthly	*in situ*	Harris et al. (2014)
	UDel	0.5°	monthly	*in situ*	Willmott et al. (2001)
	ISCCP	1°	daily	satellite	Rossow and Duenas (2004)
	ERA-Interim	~~0.25°~~ 0.75°	~~daily~~3-hourly	reanalysis	Dee et al. (2011)
	GISS	2°	monthly	*in situ*	Hansen et al. (2010)
	MLOST	5°	monthly	*in situ*	Smith et al. (2008)
	LST	0.5°	daily	satellite	Coccia et al. (2015)
Water availability	CRU-HR	0.5°	monthly	*in situ*	Harris et al. (2014)
	MSWEP	0.25°	~~daily~~3-hourly	satellite/*in situ*	Beck et al. (2017)
	UDel	0.5°	monthly	*in situ*	Willmott et al. (2001)
	CMAP	2.5°	monthly	satellite/*in situ*	Xie and Arkin (1997)
	CPC-U	0.25°	daily	*in situ*	Xie et al. (2007)
	GPCC	0.5°	monthly	*in situ*	Schneider et al. (2008)
	GPCP	2.5°	monthly	satellite/*in situ*	Adler et al. (2003)
	ERA-Interim	~~1°~~0.75°	~~daily~~3-hourly	reanalysis	Dee et al. (2011)
	GLEAM	0.25°	daily	satellite	Miralles et al. (2011)
	ESA CCI-PASSIVE	0.25°	daily	satellite	~~Liu et al. (2012)~~ Dorigo et al. (in review)
	ESA CCI-COMBINED	0.25°	daily	satellite	Liu et al. (2012)
	~~GLOBSNOW~~GlobSnow	0.25°	daily	satellite	Luojus et al. (2010)
Radiation	SRB	1°	~~daily~~3-hourly	satellite	Stackhouse et al. (2004)
	ERA-Interim	~~0.25°~~0.75°	~~daily~~3-hourly	reanalysis	Dee et al. (2011)
Greenness (NDVI)	GIMMS	0.25°	monthly	satellite	Tucker et al. (2005)

To conclude, as a proxy for the state and activity of vegetation, we use the third generation (3G) Global Inventory Modeling and Mapping Studies (GIMMS) satellite-based NDVI (Tucker et al., 2005), a commonly used long-term global record of NDVI (Beck et al., 2011). We note that this dataset is used to derive the response variable in our approach (seasonal NDVI anomalies, see Sect. 3.2), while all other data sets are converted to predictor variables. The length of the NDVI record (1981–2010) sets the study period to an interval of 30 years.

3.2 Anomaly decomposition

In climate studies, ~~it is common to apply Granger causality analyses~~ Granger causality has already been applied on time series of seasonal anomalies (Attanasio, 2012; Tuttle and Salvucci, 2016) . The latter may be obtained in a two-step decomposition procedure, by first subtracting the seasonal cycle and then the long-term trend from the raw time series~~(Attanasio, 2012)~~ . Several competing decomposition methods have been proposed in the literature, including additive models, multiplicative models and more sophisticated methods based on break points (see e.g. ~~Cleveland et al., 1990; Verbesselt et al., 2010; Grieser et al., 2002~~ Clevelan
In our framework we used the following approach: ~~In~~ in a first step, at each given pixel, the 'raw' time series of the target variable $\boldsymbol{y_t}$ and the climate predictors ($\boldsymbol{x_t}$, $\boldsymbol{w_t}$,...) are de-trended linearly based on a simple linear regression with the time stamp t as predictor variable applied to the entire study period. For the case of the target variable this can be denoted as follows:

$$\underline{y_t \approx y_t^T = \alpha_0 + \alpha_1 t\,,}$$

$$y_t \approx y_t^{T_r} = \alpha_0 + \alpha_1 t \tag{10}$$

with α_0 and α_1 being the intersect and the slope of the linear regression, respectively. We obtain in this way the de-trended time series $\cancel{\boldsymbol{y_t^D} = \boldsymbol{y_t} - \boldsymbol{y_t^T}}$ $\boldsymbol{y_t^D} = \boldsymbol{y_t} - \boldsymbol{y_t^{T_r}}$. This de-trending is needed to remove ~~obvious~~ non-stationary signals in climatic time series, and allows us to draw the emphasis to the shorter-term multi-month dynamics. By de-trending one can assure that the mean of the probability distribution does not change over time; however, other moments of the probability distribution, such as the variance, might still be time-dependent. As classical statistical procedures for testing Granger causality (i.e. autoregressive model, statistical tests) are developed for stationary time series, those methods are in fact not applicable to non-stationary climate data. In a second step, after subtracting the trend from the raw time series, the seasonal cycle $\boldsymbol{y_t^S}$ is calculated. When the assumption is made that the seasonal cycle is annual and constant over time, one can simply estimate it as the monthly expectation. To this end, the multi-year average for each of the twelve months of the year $\cancel{(\boldsymbol{y_t^S})}$ is calculated. Finally, the anomalies $\boldsymbol{y_t^R}$ can then be computed by subtracting the corresponding monthly expectation from the de-trended time series: $\boldsymbol{y_t^R} = \boldsymbol{y_t^D} - \boldsymbol{y_t^S}$. This procedure is schematically represented in Fig. 2.

3.3 Predictor variable construction

We do not limit our approach to considering raw ~~versus~~ and anomaly time series of the data sets in Table 1 as predictors, but also take into consideration different lag times, past-time cumulative values and extreme indices (see next). These additional predictors, here referred to as 'higher-level variables', are calculated based on ~~these~~ raw and anomaly time series. Our application of Granger causality can be interpreted as a way to identify patterns in climate during past-time moving windows (see Fig. 1) that are predictive with respect to the anomalies of vegetation time series. Therefore, by feeding predictor variables from previous time stamps to a linear (or non-linear) predictive model, one can identify sub-sequences of interest in the moving window specified for time stamp t, a technique that is similar to so-called shapelets (Ye and Keogh, 2009). ~~Techniques for finding~~

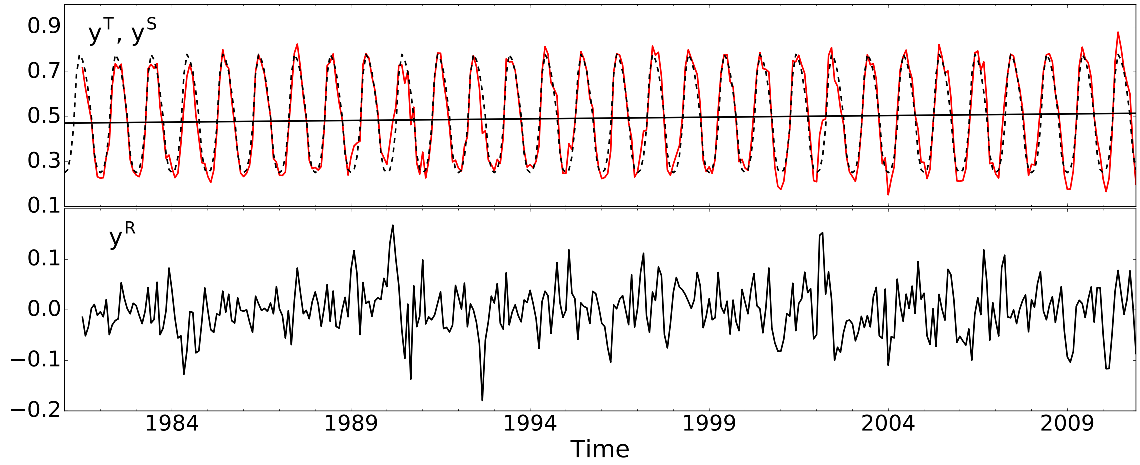

Figure 2. The three components of the NDVI time series decomposition of a specific pixel of the ~~northern~~ Northern hemisphere (lat: 53.5, long: 26.5). On top, the linear trend (black continuous line) and the seasonal cycle (dashed black line) fitted on the raw data ~~, on~~ (red). On the bottom the remaining anomalies. ~~All three components of the time series are shown in black, whereas the raw time series is shown in red.~~ See text for details.

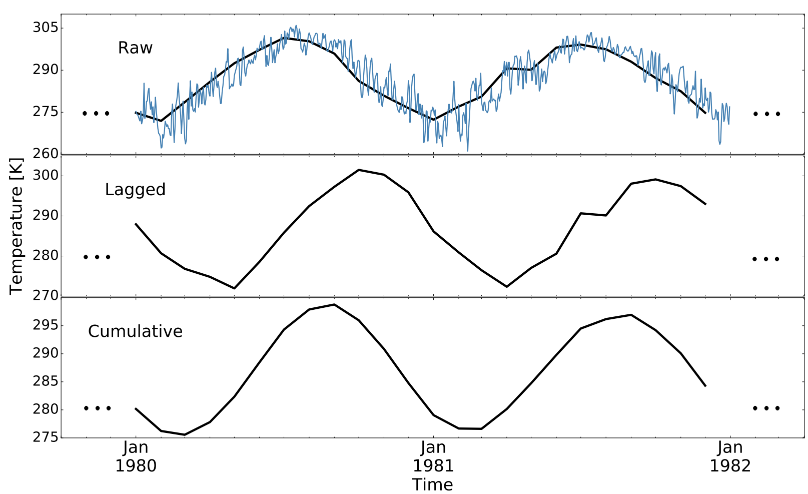

Figure 3. Example of lagged and cumulative variables extracted from a temperature time series. On top, part of a raw daily time series with its monthly aggregation. In the middle, the 4-month lag-time monthly time series~~is presented, while on~~. On the bottom, the corresponding 4-month cumulative variable. The pixel corresponds to a location in ~~North America~~ Kentucky US (lat: 37.5, long: -87.5).

~~shapelets have been mainly applied to the problem of time series classification, where they are used to extract meaningful information from raw time series (Grabocka et al., 2014; Rakthanmanon and Keogh, 2013; Zakaria et al., 2012; Ye and Keogh, 2009; Mue~~

~~In our case~~In addition, vegetation dynamics may not necessarily reflect the climatic conditions from (e.g.) three months ago, but the average of the (e.g.) three antecedent months. This integrated response to antecedent environmental and climatic conditions is referred here as a ~~'~~‘cumulative’ response. More formally, we construct a cumulative variable of k months as the sum of time series observations in the last k months:

$$\mathrm{Cum}[x_{t-1}, x_{t-2}, ..., x_{t-k}] = \sum_{p=1}^{k} x_{t-p}\,.$$

$$\mathrm{Cumul}[x_{t-1}, x_{t-2}, ..., x_{t-k}] = \sum_{p=1}^{k} x_{t-p} \tag{11}$$

Note that, unlike in the case of lagged variables, cumulative ones include always the period up to time t. Figure 3 illustrates an example of a 4-month cumulative variable. In our analysis we ~~have~~ experimented with time lags covering a wide range of time-lag values ~~, concluding~~ and concluded that including lags of more than six months did not yield substantial predictive power.

Another type of higher-level predictor ~~variables~~ variable that can be constructed from the data sets in Table 1 are extreme indices. Over the last few years, ~~many~~ several research studies have focused on ~~identifying~~ defining and indexing climate extremes (Nicholls and Alexander, 2007; Zwiers et al., 2013)~~, even though fewer have concentrated on the effect of climate extremes on vegetation (Reyer et al., 2013; Smith, 2011; Zscheischler et al., 2014a, b, c) . The~~ . As an example, the Expert Team on Climate Change Detection and Indices (ETCCDI) recommends the use of a range of extreme indices related to temperature and precipitation ~~(Donat et al., 2013; Zhang et al., 2011)~~ (Zhang et al., 2011; Donat et al., 2013) . Here we calculate a variety of analogous indices for the whole set of the collected climatic variables, based on both the raw data sets as well as on the seasonal anomalies (see Table 2). In addition, we derived lagged and cumulative predictor variables from these extremes indices ~~as described above; this is done~~ to incorporate the potential impact of climatic extremes occurring (e.g.) three months ago, or during the previous (e.g.) three months, respectively. All these resulting time series appear as additional predictor variables in our non-linear Granger causality framework (see Sect. 2.3).

Combining the different climate and environmental predictor variables described above, we obtain a database of ~~3,884~~ 4,571 predictor variables per 1° pixel, covering thirty years at a monthly temporal resolution.

4 Results and discussion

4.1 Detecting linear Granger-causal relationships

In a first experiment, we evaluate the extent to which climate variability Granger-causes the anomalies in vegetation using a standard Granger causality approach, in which only linear relationships between climate (predictors) and vegetation (target

Table 2. Extreme indices considered as predictive variables. These indices are derived from the raw (daily) data and the (daily) anomalies of the data sets in Table 1. We also calculate the lagged and cumulative variables from these extreme indices.

Name	Description
STD	Standard deviation of daily values per month
DIR	Difference between max and min daily value per month
Xx	Max daily value per month
Xn	Min daily value per month
Max5day	Max over 5 consecutive days per month
Min5day	Min over 5 consecutive days per month
X99p/X95p/X90p	Number of days per month over $99^{\mathrm{th}}/95^{\mathrm{th}}/90^{\mathrm{th}}$ percentile
X1p/X5p/X10p	Number of days per month under $1^{\mathrm{th}}/5^{\mathrm{th}}/10^{\mathrm{th}}$ percentile
T25C[a]	Number of days per month over 25°C
T0C[a]	Number of days per month below 0°C
R10mm/R20mm[b]	Number of days per month over 10/20 mm
CHD (Consecutive High value Days)	Number of consecutive days per month over 90^{th} percentile
CLD (Consecutive Low value Days)	Number of consecutive days per month below 10^{th} percentile
CDD (Consecutive Dry Days)[b]	Number of consecutive days per month when precipitation < 1 mm
CWD (Consecutive Wet Days)[b]	Number of consecutive days per month when precipitation $\geq$ 1 mm
Spatial Heterogeneity[c]	Difference between max and min values within 1 degree box

[a]Only for temperature data sets
[b]Only for precipitation data sets
[c]Only for data sets with native spatial resolution <1° lat-lon

variable) are considered. To this end, ridge regression is used as a linear vector autoregressive (VAR) model in the Granger causality approach (note this ridge regression will be substituted by ~~our~~ the non-linear random forest approach in Sect. 4.2). In the application of the ridge regression, we use all climatic and environmental predictor variables (Sect. 3.3), and adopt a nested 5-fold cross-validation to properly tune the hyper-parameter λ ~~.~~ (see Eq. 9). Figure 4a shows the predictive performance of the full ridge regression model. While the model explains more than 40% of the variability in NDVI anomalies in some regions (R^2 > 0.4), this is by itself not necessarily indicative of climate Granger-causing the vegetation anomalies, as it may reflect simple correlations. In order to test the latter, we compare the results of the full model to a baseline model, i.e., an autoregressive ridge regression model that only uses previous values of NDVI to predict the NDVI at time t (see Sect. 2.1). If climate Granger-caused the variability of NDVI at a given pixel, the full ridge regression model (Fig. 4a) would show an increase in the predictive power over the predictions based on the baseline ridge regression model. However, the results unequivocally show that – when only linear relationships between vegetation and climate are considered – the areas for which

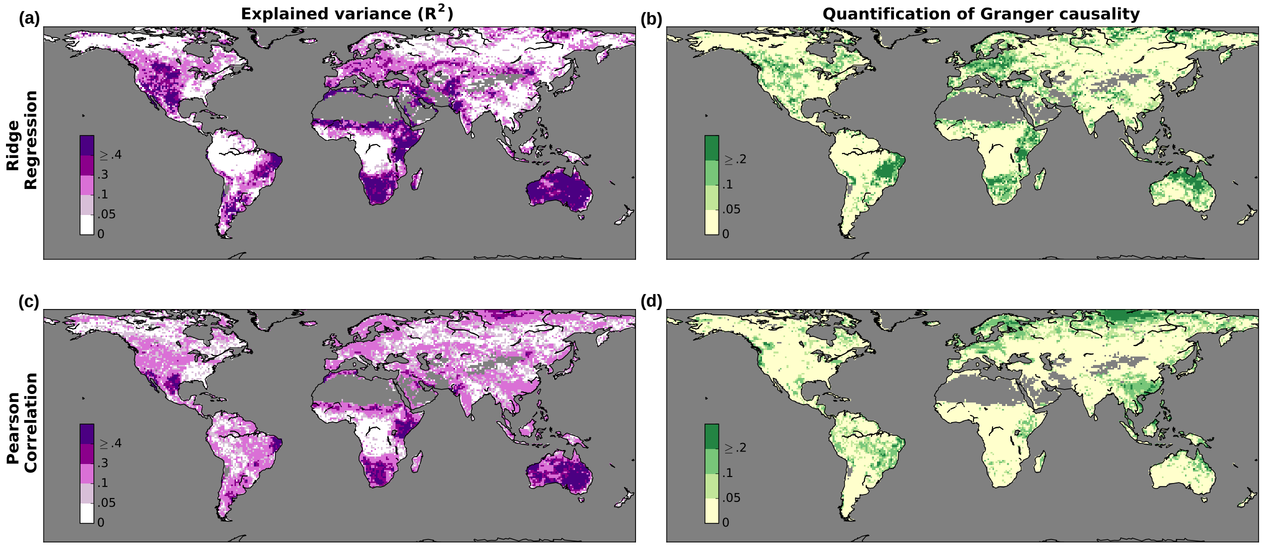

Figure 4. Linear Granger causality of climate on vegetation. **(a)** Explained variance (R^2) of NDVI anomalies based on a full ridge regression model in which all climatic variables are included as predictors. **(b)** Improvement in terms of R^2 by the full ridge regression model with respect to ~~a~~ the baseline ridge regression model that uses only past values of NDVI anomalies as predictors; positive values indicate (linear) Granger causality. **(c)** A filter approach in which the variable with the highest squared Pearson correlation against the NDVI anomalies is selected. **(d)** Improvement in terms of R^2 by the filter approach with respect to ~~a~~ the same baseline ridge regression model that uses only past values of NDVI anomalies~~only~~.

vegetation anomalies are Granger-caused by climate are very limited, ~~including~~ involving mainly semiarid regions and central Europe (Fig. 4b).

For further comparison, we analyze the predictive performance obtained when (linear) Pearson correlation coefficients are calculated on the training data sets, selecting the highest correlation to the target variable for any of the ~~3,884~~ 4,571 predictor variables at each pixel. Figure 4c shows that the explained variance is again rather low, and for most regions substantially lower than the R^2 of the baseline ridge regression model, here considered as the minimum to interpret this predictive power as Granger-causal. These results indicate that, despite being routinely used as a standard tool in ~~many~~ climate–biosphere studies (see e.g. ~~Nemani et al., 2003; Wu et al., 2015~~ Nemani et al., 2003), univariate correlation analyses are unable to extract the nuances of the relationships between climate and vegetation dynamics.

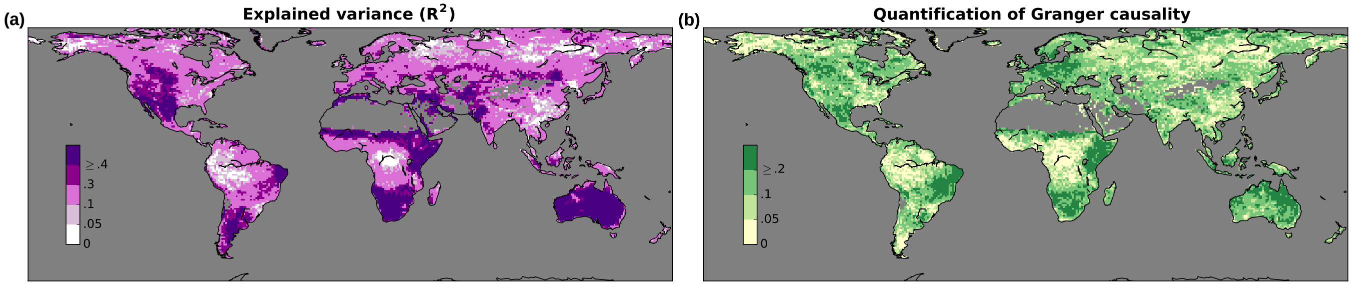

Figure 5. Non-linear Granger causality of climate on vegetation. **(a)** Explained variance (R^2) of NDVI anomalies based on a full random forest model in which all climatic variables are included as predictors. **(b)** Improvement in terms of R^2 by the full random forest model with respect to the baseline random forest model that uses only past values of NDVI anomalies as predictors; positive values indicate (non-linear) Granger causality.

4.2 Linear versus non-linear Granger causality

To analyze the effect of climate on vegetation more thoroughly, we substitute the linear ridge regression model (VAR) by ~~our~~ the non-linear random forest model. Results in Fig. 5 highlight the differences. Compared to the results in ~~the previous~~ Sect. 4.1, the predictive power substantially increases by considering non-linear relationships between vegetation and climate (Fig. 5a). This is the case for most land regions, but is especially remarkable in semiarid regions of Australia, Africa, Central and North America, which are frequently exposed to water limitations. In those regions, more that 40% of the variance of NDVI anomalies can be explained by antecedent climate variability. These results are further investigated by Papagiannopoulou et al. (*in review*), who highlight the crucial role of water supply for the anomalies in vegetation greenness in these and other regions. On the other hand, the variance of NDVI explained in other ~~regions~~ areas such as the Eurasian taiga, tropical rainforests or China is again below 10%. We hypothesize two potential reasons: (a) the uncertainty in the observations used as ~~predictor~~ target and predictors are typically larger in these regions (especially in tropical forests and at higher latitudes)~~(Dorigo et al., 2010)~~ , and (b) these are regions in which vegetation ~~is~~ anomalies are not necessarily primarily controlled by climate, but may ~~also be~~ be predominantly driven by phenological and biotic factors (Hutyra et al., 2007), occurrence of wild fires (Van der Werf et al., 2010), limitations imposed by the availability of soil nutrients (Fisher et al., 2012) or agricultural practices (Liu et al., 2015). Nonetheless, the ~~patterns of~~ explained variance shown in Fig. 5a ~~are~~ is again not necessarily indicative of Granger causality. As we did in Fig. 4b, in order to test whether the climatic and environmental controls do, in fact, Granger-cause the vegetation anomalies, we compare the results of our full random forest model to a baseline random forest model which only uses previous values of NDVI to predict the NDVI at time t~~(see~~ . As seen in Fig. 5b~~). In~~ , in this case, the improvement over the baseline is unambiguous. One can conclude that – while not bearing into consideration all potential control variables in our analysis – climate dynamics indeed Granger-cause vegetation anomalies in most of the continental land surface, ~~being larger their~~ with a larger impact on subtropical regions and mid-latitudes. Moreover, a comparison between Figs. 4b and 5b unveils that these causal relationships are highly non-linear~~as expected, given the progressive response of vegetation to environmental changes and the recovery time~~ , as expected given the distinct resistence and resilience of different ecosystems, which is reflected by a progressive response and recovery of vegetation to these perturbations (Foley et al., 1998; Zeng et al., 2002; Verbesselt et al., 2016).

For a better understanding of the results obtained by the two models, we average the performance of each model regionally. More specifically, we use the International Geosphere-Biosphere Program (IGBP) (Loveland and Belward, 1997) land cover classification to stratify the mean and variance of R^2 for both the baseline and the full model in Fig. 5 per IGBP land cover class. The barplot in Fig. 6 shows that the full model outperforms the baseline model in all IGBP land cover classes, i.e. that Granger causality exists for all these biomes. In the parentheses we note the number of pixels per region. The error bars indicate that the variances of the two models are analogous, i.e. they are low or high in both models in the same land cover class. For the Closed Shrublands region, one can observe the highest difference between the two models, yet only 19 pixels belong to this biome type. In savanna regions the performance of the full model is high in comparison with other regions (see Fig. 5). On the other hand, the lowest performance improvement of the full model with respect to the baseline is observed for the regions of Deciduous Needleleaf Forests and Evergreen Broadleaf Forests. This shows that for these two regions climate is not identified as a major control over vegetation dynamics (see discussion in previous paragraph about tropical and boreal regions).

5 ~~Discussion~~

4.1 Spatial and temporal aspects

Environmental dynamics reveal their effect on vegetation at different time scales. Since the adaptation of vegetation to environmental changes requires some time, and because soil and atmosphere have a memory, a necessary aspect to investigate is the potential lag-time response of vegetation to climate dynamics which relates to the ecosystem resistence and resilience properties. The idea of exploring lag times ~~has been~~ was introduced by several studies in the past (see e.g. Davis, 1984; Braswell et al., 1997), and it has been adopted in ~~many~~ various studies more recently (Anderson et al., 2010; Kuzyakov and Gavrichkova, 2010; Chen et al., 2014; Rammig et al., 2014). These studies indicate that lag times depend on both the specific climatic control variable and the characteristics of the ecosystem. As explained in Sect. 3.3, in our analysis shown in Fig. 4 and 5, we moved beyond traditional cross-correlations, and incorporate higher-lever variables in the form of cumulative and lagged responses to extreme climate. As mentioned in Sect. 3.3, our experiments indicated that lags of more than six months do not add extra predictive power (not shown), even though the effect of anomalies in water availability on vegetation can extend for several months (Papagiannopoulou et al., *in review*).

To disentangle the response of vegetation to past cumulative climate anomalies and climatic extremes, Fig. 7a visualizes the predictive performance when cumulative variables and extreme indices are not included as predictive variables in the random forest model. As shown in Fig. 7b, in almost all regions of the world the predictive performance decreases substantially compared to the full random forest model approach, i.e. using the full repository of predictors ~~,~~ (Fig. 5a), especially in regions

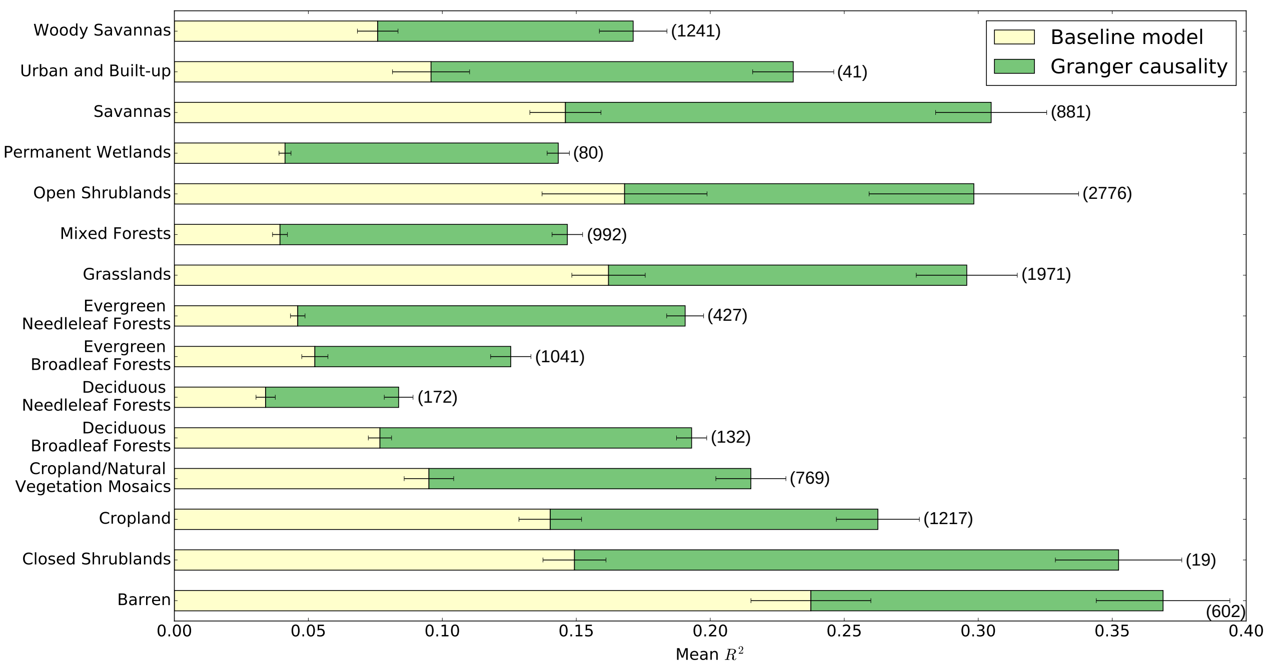

Figure 6. ~~Analysis of spatiotemporal aspects of our framework.~~ ~~**(a)** Explained variance (~~Mean R^2 ~~) of NDVI anomalies based on a~~ and variance per IGBP land cover class for both the baseline and full random forest model~~in which all climatic variables are included as predictors and in Fig~~. ~~5a, except for~~ The green part indicates the ~~past cumulative climate and climate extreme indices (see Sect. 3.3).~~ ~~**(b)** Difference~~ improvement in ~~terms~~ performance of ~~R^2 between~~ the ~~model without cumulative and extreme predictors and the~~ full ~~random forest~~ model ~~in Fig~~with respect to the baseline, i.~~5a~~e. ~~**(c)** Explained variance (R^2)~~ the quantification of ~~NDVI anomalies based on a full random forest model in which all climatic variables are included~~ Granger causality (as ~~predictors and~~ in Fig. 5~~a, but including also the predictors from the 8 nearest neighbors~~b). ~~**(d)** Difference in terms~~ The number of ~~R^2 between this full random forest model which includes spatial information from neighbouring~~ pixels ~~and the full random forest model~~ per IGBP class is noted in ~~Fig~~the parentheses.~~5a.~~

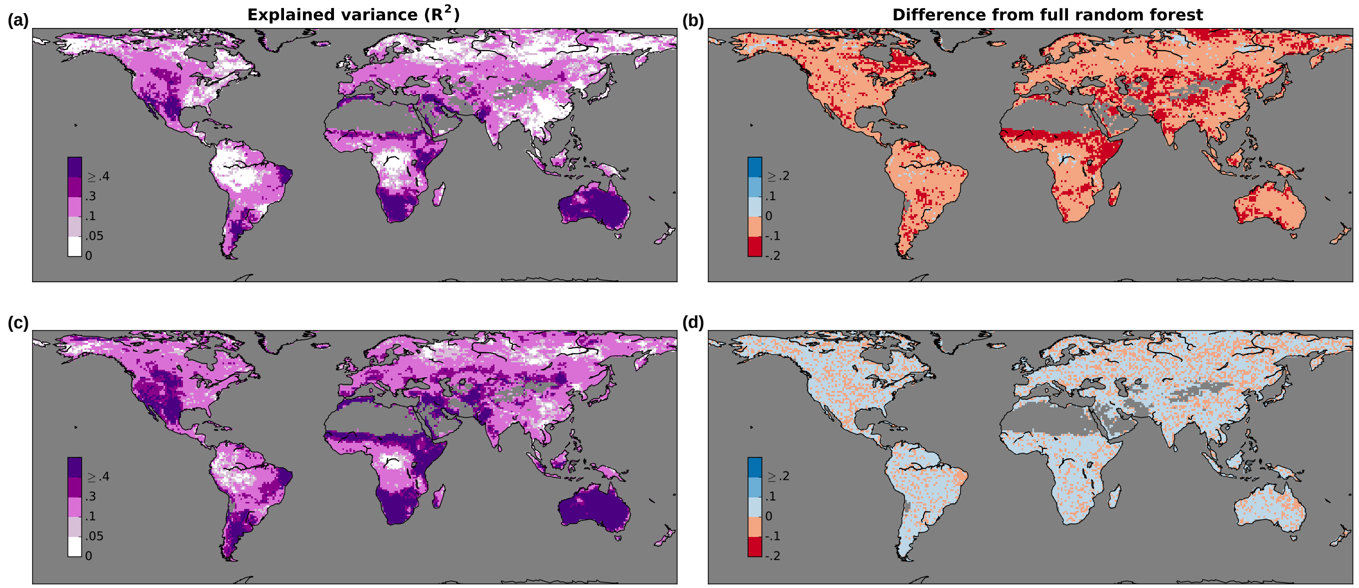

Figure 7. Analysis of spatiotemporal aspects of our framework. **(a)** Explained variance (R^2) of NDVI anomalies based on a full random forest model in which all climatic variables are included as predictors as in Fig. 5a, except for the cumulative variables and the extreme indices (see Sect. 3.3). **(b)** Difference in terms of R^2 between the model without cumulative and extreme predictors and the full random forest model in Fig. 5a. **(c)** Explained variance (R^2) of NDVI anomalies based on a full random forest model in which all climatic variables are included as predictors as in Fig. 5a, but including also the predictors from the 8 nearest neighbors. **(d)** Difference in terms of R^2 between this full random forest model which includes spatial information from neighbouring pixels and the full random forest model in Fig. 5a.

such as the Sahel, the Horn of Africa or North America. In those regions 10-20% of the variability in NDVI is explained by the occurrence of prolonged anomalies and/or extremes in climate, illustrating again the non-linear responses of vegetation ~~to climate~~. For more detailed results about lagged vegetation responses for specific climate drivers and the effect of climate extremes on vegetation, the reader is referred to Papagiannopoulou et al. (*in review*).

Because of uncertainties in the observational records used in our study to represent climate and predict vegetation dynamics, and given that ecosystems and regional climate conditions usually extend over areas that exceed the spatial resolution of these records, one may expect that the predictive performance of our models ~~become~~ becomes more robust when including climate information from neighboring pixels.

In addition, it is quite likely that neighboring areas have similar climatic conditions, which in their turn affect vegetation dynamics in a similar manner. We therefore also consider an extension of our framework to exploit spatial autocorrelations, inspired by Lozano et al. (2009), who achieved spatial smoothness via an additional penalty term that punishes dissimilarity between coefficients for spatial neighbors. In our analysis, we incorporate at a given pixel spatial autocorrelations by extending the predictor variables of our models with the predictor variables of the 8 neighboring pixels. We provide such an extension both for the full and the baseline random forest model. As such, for the full random forest model, a vector of ~~34,956 (3,884~~ 41,139 (4,571 $\times$ 9) predictor variables is formed for each pixel.

Figure 7c illustrates the performance of the full random forest model that includes the spatial information. As one can observe in Fig. 7d, the explained variance of NDVI anomalies remains similar to the original model ~~which~~ that depicts the same approach without spatial autocorrelation (Fig. 5a). While in most areas the performance slightly increases, the explained variance never improves by more than 10%; as a result, incorporating spatial autocorrelations in our framework does not seem to further improve the quantification of Granger causality and is not considered in further applications of the framework (see Papagiannopoulou et al., *in review*). A possible explanation for this result is that the model without the spatial information cannot be outperformed because of the large dimensionality of the feature space, which may include redundant information, in combination with the low number of observations per pixel (Fig. 5a). Note that in this case the number of observations per pixel remains the same as in the original model (360 observations) while the number of predictor variables is 9 times larger.

4.2 The importance of focusing on vegetation anomalies

In Sect. 3.2 we advocated that Granger causality analysis should target on NDVI anomalies, as opposed to raw NDVI values. There are several fundamental reasons for this. First, by applying a decomposition, one can subtract long-term trends from the NDVI time series, making the resulting time series more stationary. This is absolutely needed, as existing Granger causality tests cannot be applied for non-stationary time series. Secondly, by subtracting the seasonal cycle from the time series, one is ~~able not only~~ not only able to remove a confounding factor that may contribute predictive power without ~~having~~ bearing causality, but also able to remove a clear autoregressive component that can be well explained from the NDVI time series themselves. As vegetation has a strong seasonal cycle, it is not difficult to predict subsequent vegetation conditions by using the past observations of the seasonal cycle only. To corroborate this aspect, we repeat our analysis in Sect. 4.2, but this time considering the raw NDVI time series instead of the NDVI anomalies are considered as the target variable. We again compare the full and the baseline random forest models.

The results are visualized in Fig. 8a. As it can be observed, worldwide the R^2 is close to the optimum of one. However, due to the overwhelming domination of the seasonal cycle, it becomes very difficult, or even impossible, to unravel any potential Granger-causal relationships with climate time series in the ~~northern~~ Northern hemisphere – see Fig. 8b. The predictability of NDVI based on the seasonal NDVI cycle itself is already so high that nothing can be gained by adding additional climatic predictor variables; see also the large amplitude of the seasonal cycle of NDVI at those latitudes compared to the NDVI anomalies, as illustrated in Fig. 2. Therefore, a non-linear baseline autoregressive model is able to explain most of the variance in the time series. Moreover, as observed in Fig. 1, temperature and radiation also manifest strong seasonal cycles that often coincide with the NDVI cycle. For most regions on Earth, such a stationary seasonal cycle is less present for variables such as precipitation. This can potentially yield wrong conclusions~~. For instance,~~ , such as that temperature in the ~~northern hemisphere becomes really predictive for raw NDVI~~ Northern hemisphere is driving most NDVI variability, since the two seasonal cycles have the same pattern. However, based on the above discussion, it ~~is obvious~~ becomes clear that results of that kind should be

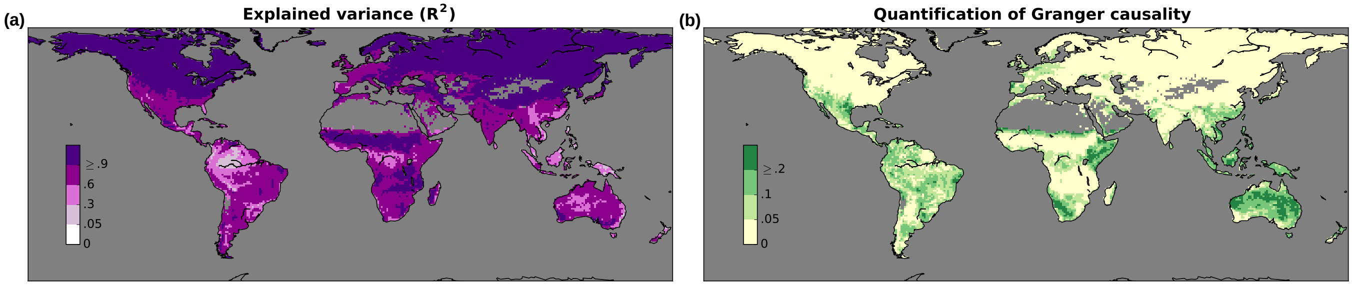

Figure 8. Comparison of model performance with R^2 as metric with the raw NDVI time series as target variable. **(a)** Full random forest model **(b)** Improvement in terms of R^2 of the full random forest model over the baseline random forest model.

treated with ~~great caution. The sudden rise in importance for temperature can simply be explained by the presence of a stronger seasonal cycle for this variable.~~ caution: ~~We therefore conclude that,~~ for climate data, a Granger causality analysis should be applied after decomposing time series into seasonal anomalies.

5 Conclusions

In this paper we introduced a novel framework for studying Granger causality in climate-vegetation dynamics. We compiled a global database of observational records spanning a thirty-year time frame, containing satellite, *in situ* and reanalysis-based datasets. Our approach consists of the combination of data fusion, feature construction and non-linear predictive modelling~~on climate and vegetation data~~. The choice of random forest as a non-linear algorithm has been motivated by its excellent computational scalability with regards to extremely large data sets, but could be easily replaced by any other non-linear machine learning technique, such as neural networks or kernel methods.

Our results highlight the non-linear nature of climate–vegetation interactions and the need to move beyond the traditional application of Granger causality within a linear framework. Comparisons to ~~traditional Granger causality~~ linear Granger causality based approaches indicate that ~~our~~ the random forest framework can predict 14% more variability of vegetation anomalies on average globally. The predictive power of the model is especially high in water-limited regions where a large part of the vegetation dynamics responds to the occurrence of antecedent rainfall. Moreover, our results indicate the need to consider multi-month antecedent periods to capture the effect of climate on vegetation, ~~and the need for considering the impact~~ in particular to account for the effects of climate extremes on vegetation ~~dynamics~~resilience. The reader is referred to Papagiannopoulou et al. (*in review*) for a detailed analysis of the effect of different climate predictors on the variability of global vegetation using the mathematical approach described here.

6 Code and data availability

~~We use the implementation of scikit-learn (Pedregosa et al., 2011) library in Python for the random forest regressor. Data~~ Our code and the data used in this manuscript can be accessed using http://www.SAT-EX.ugent.be as gateway.

Author contributions. Diego G. Miralles, Willem Waegemann and Niko E. C. Verhoest conceived the study. Christina Papagiannopoulou conducted the analysis. Willem Waegeman, Diego G. Miralles and Christina Papagiannopoulou led the writing. All co-authors contributed to the design of the experiments, discussion and interpretation of results, and editing of the manuscript.

Acknowledgements. This work is funded by the Belgian Science Policy Office (BELSPO) in the framework of the STEREO III programme, project SAT-EX (SR/00/306). D. G. Miralles acknowledges ~~financial support from The Netherlands Organization for Scientific Research through grant 863.14.004~~support from the European Research Council (ERC) under grant agreement n° 715254 (DRY-2-DRY). W. Dorigo is supported by the 'TU Wien Wissenschaftspreis 2015', a personal grant awarded by the Vienna University of Technology. The authors thank Mathieu Depoorter ~~, Stijn Decubber~~ and Julia Green for the fruitful discussions. Finally, the authors sincerely thank the individual developers of the wide range of global data sets used in this study.

References

Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P.-P., Janowiak, J., Rudolf, B., Schneider, U., Curtis, S., Bolvin, D., et al.: The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979-present), Journal of hydrometeorology, 4, 1147–1167, 2003.

Ancona, N., Marinazzo, D., and Stramaglia, S.: Radial basis function approach to nonlinear Granger causality of time series, Physical Review E, 70, 056 221, 2004.

Anderson, L. O., Malhi, Y., Aragão, L. E., Ladle, R., Arai, E., Barbier, N., and Phillips, O.: Remote sensing detection of droughts in Amazonian forest canopies, New Phytologist, 187, 733–750, 2010.

Attanasio, A.: Testing for linear Granger causality from natural/anthropogenic forcings to global temperature anomalies, Theoretical and Applied Climatology, 110, 281–289, 2012.

Attanasio, A., Pasini, A., and Triacca, U.: A contribution to attribution of recent global warming by out-of-sample Granger causality analysis, Atmospheric Science Letters, 13, 67–72, 2012.

Attanasio, A., Pasini, A., and Triacca, U.: Granger causality analyses for climatic attribution, Atmospheric and Climate Sciences, 2013, 2013.

Barichivich, J., Briffa, K. R., Myneni, R., van der Schrier, G., Dorigo, W., Tucker, C. J., Osborn, T. J., and Melvin, T. M.: Temperature and Snow-Mediated Moisture Controls of Summer Photosynthetic Activity in Northern Terrestrial Ecosystems between 1982 and 2011, Remote Sensing, 6, 1390, doi:10.3390/rs6021390, 2014.

Beck, H. E., McVicar, T. R., van Dijk, A. I., Schellekens, J., de Jeu, R. A., and Bruijnzeel, L. A.: Global evaluation of four AVHRR–NDVI data sets: Intercomparison and assessment against Landsat imagery, Remote Sensing of Environment, 115, 2547–2563, 2011.

Beck, H. E., van Dijk, A. I. J. M., Levizzani, V., Schellekens, J., Miralles, D. G., Martens, B., and de Roo, A.: MSWEP: 3-hourly 0.25° global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data, Hydrology and Earth System Sciences, 21, 589–615, doi:10.5194/hess-21-589-2017, http://www.hydrol-earth-syst-sci.net/21/589/2017/, 2017.

Bonan, G. B.: Forests and climate change: forcings, feedbacks, and the climate benefits of forests, science, 320, 1444–1449, 2008.

Braswell, B., Schimel, D., Linder, E., and Moore, B.: The response of global terrestrial ecosystems to interannual temperature variability, Science, 278, 870–873, 1997.

Breiman, L.: Random forests, Machine learning, 45, 5–32, 2001.

Chapman, D., Cane, M. A., Henderson, N., Lee, D. E., and Chen, C.: A vector autoregressive ENSO prediction model, Journal of Climate, 28, 8511–8520, 2015.

Chen, T., De Jeu, R., Liu, Y., Van der Werf, G., and Dolman, A.: Using satellite based soil moisture to quantify the water driven variability in NDVI: A case study over mainland Australia, Remote Sensing of Environment, 140, 330–338, 2014.

Clark, T. E. and McCracken, M. W.: Tests of equal forecast accuracy and encompassing for nested models, Journal of econometrics, 105, 85–110, 2001.

Cleveland, R. B., Cleveland, W. S., McRae, J. E., and Terpenning, I.: STL: A seasonal-trend decomposition procedure based on loess, Journal of Official Statistics, 6, 3–73, 1990.

Coccia, G., Siemann, A. L., Pan, M., and Wood, E. F.: Creating consistent datasets by combining remotely-sensed data and land surface model estimates through Bayesian uncertainty post-processing: The case of Land Surface Temperature from {HIRS}, Remote Sensing of Environment, 170, 290 – 305, doi:http://dx.doi.org/10.1016/j.rse.2015.09.010, 2015.

Davis, M. B.: Climatic instability, time, lags, and community disequilibrium, Harper & Row, 1984.

Dee, D., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, P., et al.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society, 137, 553–597, 2011.

Diebold, F. X.: Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of Diebold–Mariano tests, Journal of Business & Economic Statistics, 33, 1–1, 2015.

Dietterich, T. G.: Approximate statistical tests for comparing supervised classification learning algorithms, Neural computation, 10, 1895–1923, 1998.

Diks, C. and Mudelsee, M.: Redundancies in the Earth's climatological time series, Physics Letters A, 275, 407–414, 2000.

Donat, M., Alexander, L., Yang, H., Durre, I., Vose, R., Dunn, R., Willett, K., Aguilar, E., Brunet, M., Caesar, J., et al.: Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: The HadEX2 dataset, Journal of Geophysical Research: Atmospheres, 118, 2098–2118, 2013.

Dorigo, W., Wagner, W., and et al.: ESA CCI Soil Moisture for improved Earth System understanding: state-of-the-art and future directions, Remote Sensing of Environment, *in review*.

Dorigo, W., Lucieer, A., Podobnikar, T., and Čarni, A.: Mapping invasive Fallopia japonica by combined spectral, spatial, and temporal analysis of digital orthophotos, International Journal of Applied Earth Observation and Geoinformation, 19, 185–195, 2012.

Dorigo, W. A., Scipal, K., Parinussa, R. M., Liu, Y., Wagner, W., De Jeu, R. A., and Naeimi, V.: Error characterisation of global active and passive microwave soil moisture datasets, Hydrology and Earth System Sciences, 14, 2605–2616, 2010.

Elsner, J. B.: Evidence in support of the climate change–Atlantic hurricane hypothesis, Geophysical Research Letters, 33, 2006.

Elsner, J. B.: Granger causality and Atlantic hurricanes, Tellus A, 59, 476–485, 2007.

Fisher, J. B., Badgley, G., and Blyth, E.: Global nutrient limitation in terrestrial vegetation, Global Biogeochemical Cycles, 26, 2012.

Foley, J. A., Levis, S., Prentice, I. C., Pollard, D., and Thompson, S. L.: Coupling dynamic models of climate and vegetation, Global change biology, 4, 561–579, 1998.

Friedman, J., Hastie, T., and Tibshirani, R.: The elements of statistical learning, vol. 1, Springer series in statistics Springer, Berlin, 2001.

Geiger, P., Zhang, K., Gong, M., Janzing, D., and Schölkopf, B.: Causal inference by identification of vector autoregressive processes with hidden components, in: Proceedings of 32th International Conference on Machine Learning (ICML 2015), 2015.

Gelper, S. and Croux, C.: Multivariate out-of-sample tests for Granger causality, Computational statistics & data analysis, 51, 3319–3329, 2007.

Grabocka, J., Schilling, N., Wistuba, M., and Schmidt-Thieme, L.: Learning time-series shapelets, in: Proc. of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 392–401, ACM, 2014.

Granger, C. W.: Investigating causal relations by econometric models and cross-spectral methods, Econometrica: Journal of the Econometric Society, pp. 424–438, 1969.

Green, J. K. and et al.: Hotspots of terrestrial biosphere-atmosphere feedbacks, Manuscript submitted for publication.

Gregorova, M., Kalousis, A., Marchand-Maillet, S., and Wang, J.: Learning vector autoregressive models with focalised Granger-causality graphs, arXiv preprint arXiv:1507.01978, 2015.

Grieser, J., Trömel, S., and Schönwiese, C.-D.: Statistical time series decomposition into significant components and application to European temperature, Theoretical and applied climatology, 71, 171–183, 2002.

Hacker, R. S. and Hatemi-J, A.: Tests for causality between integrated variables using asymptotic and bootstrap distributions: theory and application, Applied Economics, 38, 1489–1500, doi:10.1080/00036840500405763, 2006.

Hansen, J., Ruedy, R., Sato, M., and Lo, K.: Global surface temperature change, Reviews of Geophysics, 48, 2010.

Harris, I., Jones, P., Osborn, T., and Lister, D.: Updated high-resolution grids of monthly climatic observations–the CRU TS3. 10 Dataset, International Journal of Climatology, 34, 623–642, 2014.

Hills, J., Lines, J., Baranauskas, E., Mapp, J., and Bagnall, A.: Classification of time series by shapelet transformation, Data Mining and Knowledge Discovery, 28, 851–881, doi:10.1007/s10618-013-0322-1, 2013.

Hutyra, L. R., Munger, J. W., Saleska, S. R., Gottlieb, E., Daube, B. C., Dunn, A. L., Amaral, D. F., De Camargo, P. B., and Wofsy, S. C.: Seasonal controls on the exchange of carbon and water in an Amazonian rain forest, Journal of Geophysical Research: Biogeosciences, 112, 2007.

Kaufmann, R., Zhou, L., Myneni, R., Tucker, C., Slayback, D., Shabanov, N., and Pinzon, J.: The effect of vegetation on surface temperature: A statistical analysis of NDVI and climate data, Geophysical Research Letters, 30, 2003.

Kodra, E., Chatterjee, S., and Ganguly, A. R.: Exploring Granger causality between global average observed time series of carbon dioxide and temperature, Theoretical and applied climatology, 104, 325–335, 2011.

Kuzyakov, Y. and Gavrichkova, O.: REVIEW: Time lag between photosynthesis and carbon dioxide efflux from soil: a review of mechanisms and controls, Global Change Biology, 16, 3386–3406, 2010.

Lettenmaier, D. P., Alsdorf, D., Dozier, J., Huffman, G. J., Pan, M., and Wood, E. F.: Inroads of remote sensing into hydrologic science during the WRR era, Water Resources Research, 51, 7309–7342, doi:10.1002/2015WR017616, 2015.

Lines, J., Davis, L. M., Hills, J., and Bagnall, A.: A shapelet transform for time series classification, in: Proc. of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '12, p. 289, ACM Press, New York, New York, USA, doi:10.1145/2339530.2339579, 2012.

Liu, Y., Parinussa, R., Dorigo, W., De Jeu, R., Wagner, W., Van Dijk, A., McCabe, M., and Evans, J.: Developing an improved soil moisture dataset by blending passive and active microwave satellite-based retrievals, Hydrology and Earth System Sciences, 15, 425–436, 2011.

Liu, Y., Dorigo, W., Parinussa, R., De Jeu, R., Wagner, W., McCabe, M., Evans, J., and Van Dijk, A.: Trend-preserving blending of passive and active microwave soil moisture retrievals, Remote Sensing of Environment, 123, 280–297, 2012.

Liu, Y., Pan, Z., Zhuang, Q., Miralles, D. G., Teuling, A. J., Zhang, T., An, P., Dong, Z., Zhang, J., He, D., et al.: Agriculture intensifies soil moisture decline in Northern China, Scientific reports, 5, 2015.

Loosvelt, L., Peters, J., Skriver, H., De Baets, B., and Verhoest, N. E.: Impact of reducing polarimetric SAR input on the uncertainty of crop classifications based on the random forests algorithm, IEEE Transactions on Geoscience and Remote Sensing, 50, 4185–4200, 2012a.

Loosvelt, L., Peters, J., Skriver, H., Lievens, H., Van Coillie, F. M., De Baets, B., and Verhoest, N. E. C.: Random Forests as a tool for estimating uncertainty at pixel-level in SAR image classification, International Journal of Applied Earth Observation and Geoinformation, 19, 173–184, 2012b.

Loveland, T. and Belward, A.: The IGBP-DIS global 1km land cover data set, DISCover: first results, International Journal of Remote Sensing, 18, 3289–3295, 1997.

Lozano, A. C., Li, H., Niculescu-Mizil, A., Liu, Y., Perlich, C., Hosking, J., and Abe, N.: Spatial-temporal causal modeling for climate change attribution, in: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 587–596, ACM, 2009.

Luojus, K., Pulliainen, J., Takala, M., Derksen, C., Rott, H., Nagler, T., Solberg, R., Wiesmann, A., Metsamaki, S., Malnes, E., et al.: Investigating the feasibility of the GlobSnow snow water equivalent data for climate research purposes, in: Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International, IEEE, 2010.

Maddala, G. S. and Lahiri, K.: Introduction to econometrics, vol. 2, Macmillan New York, 1992.

Marinazzo, D., Pellicoro, M., and Stramaglia, S.: Kernel method for nonlinear Granger causality, Physical Review Letters, 100, 144 103, 2008.

Martens, B., Miralles, D. G., Lievens, H., van der Schalie, R., de Jeu, R. A. M., Férnandez-Prieto, D., Beck, H. E., Dorigo, W. A., and Verhoest, N. E. C.: GLEAM v3: satellite-based land evaporation and root-zone soil moisture, Geoscientific Model Development Discussions, 2016, 1–36, doi:10.5194/gmd-2016-162, http://www.geosci-model-dev-discuss.net/gmd-2016-162/, 2016.

McCabe, M. F., Rodell, M., Alsdorf, D. E., Miralles, D. G., Uijlenhoet, R., Wagner, W., Lucieer, A., Houborg, R., Verhoest, N. E. C., Franz, T. E., Shi, J., Gao, H., and Wood, E. F.: The Future of Earth Observation in Hydrology, Hydrology and Earth System Sciences Discussions, 2017, 1–55, doi:10.5194/hess-2017-54, http://www.hydrol-earth-syst-sci-discuss.net/hess-2017-54/, 2017.

McCracken, M. W.: Asymptotics for out of sample tests of Granger causality, Journal of Econometrics, 140, 719–752, 2007.

McPherson, R. A., Fiebrich, C. A., Crawford, K. C., Elliott, R. L., Kilby, J. R., Grimsley, D. L., Martinez, J. E., Basara, J. B., Illston, B. G., Morris, D. A., Kloesel, K. A., Stadler, S. J., Melvin, A. D., Sutherland, A. J., Shrivastava, H., Carlson, J. D., Wolfinbarger, J. M., Bostic, J. P., and Demko, D. B.: Statewide monitoring of the mesoscale environment: A technical update on the Oklahoma Mesonet, 24, 301–321, doi:10.1175/JTECH1976.1, 2007.

Michaelsen, J.: Cross-validation in statistical climate forecast models, Journal of Climate and Applied Meteorology, 26, 1589–1600, 1987.

Miralles, D., Holmes, T., De Jeu, R., Gash, J., Meesters, A., and Dolman, A.: Global land-surface evaporation estimated from satellite-based observations, Hydrology and Earth System Sciences, 15, 453–469, 2011.

Mokhov, I. I., Smirnov, D. A., Nakonechny, P. I., Kozlenko, S. S., Seleznev, E. P., and Kurths, J.: Alternating mutual influence of El-Niño/Southern Oscillation and Indian monsoon, Geophysical Research Letters, 38, 2011.

Mosedale, T. J., Stephenson, D. B., Collins, M., and Mills, T. C.: Granger causality of coupled climate processes: Ocean feedback on the North Atlantic Oscillation, Journal of climate, 19, 1182–1194, 2006.

Mueen, A., Keogh, E., and Young, N.: Logical-shapelets: an expressive primitive for time series classification, in: Proc. of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '11, p. 1154, ACM Press, New York, New York, USA, doi:10.1145/2020408.2020587, 2011.

Myneni, R. B., Keeling, C. D., Tucker, C. J., Asrar, G., and Nemani, R. R.: Increased plant growth in the northern high latitudes from 1981 to 1991, Nature, 386, 698–702, 1997.

Nemani, R. R., Keeling, C. D., Hashimoto, H., Jolly, W. M., Piper, S. C., Tucker, C. J., Myneni, R. B., and Running, S. W.: Climate-Driven Increases in Global Terrestrial Net Primary Production from 1982 to 1999, Science (New York, N.Y.), 300, 1560–3, doi:10.1126/science.1082750, 2003.

Nicholls, N. and Alexander, L.: Has the climate become more variable or extreme? Progress 1992-2006, Progress in Physical Geography, 31, 77–87, 2007.

Papagiannopoulou, C., Miralles, D. G., Dorigo, W. A., Verhoest, N. E. C., Depoorter, M., and Waegeman, W.: Vegetation anomalies caused by antecedent precipitation in most of the world, Environmental Research Letters, *in review*.

Pasini, A., Triacca, U., and Attanasio, A.: Evidence of recent causal decoupling between solar radiation and global temperature, Environmental Research Letters, 7, 034 020, 2012.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E.: Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research, 12, 2825–2830, 2011.

Rakthanmanon, T. and Keogh, E.: Fast shapelets: A scalable algorithm for discovering time series shapelets, in: Proc. of the thirteenth SIAM conference on data mining (SDM), pp. 668–676, SIAM, 2013.

Rammig, A., Wiedermann, M., Donges, J., Babst, F., von Bloh, W., Frank, D., Thonicke, K., and Mahecha, M.: Tree-ring responses to extreme climate events as benchmarks for terrestrial dynamic vegetation models, Biogeosciences Discussions, 11, 2537–2568, 2014.

Reyer, C. P., Leuzinger, S., Rammig, A., Wolf, A., Bartholomeus, R. P., Bonfante, A., de Lorenzi, F., Dury, M., Gloning, P., Abou Jaoudé, R., et al.: A plant's perspective of extremes: terrestrial plant responses to changing climatic variability, Global change biology, 19, 75–89, 2013.

Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., and Rigol-Sanchez, J. P.: An assessment of the effectiveness of a random forest classifier for land-cover classification, ISPRS Journal of Photogrammetry and Remote Sensing, 67, 93–104, 2012.

Rossow, W. and Duenas, E.: The international satellite cloud climatology project (ISCCP) web site: An online resource for research, Bulletin of the American Meteorological Society, 85, 167–172, 2004.

Schneider, U., Fuchs, T., Meyer-Christoffer, A., and Rudolf, B.: Global precipitation analysis products of the GPCC, Global Precipitation Climatology Centre (GPCC), DWD, Internet Publikation, 112, 2008.

Seddon, A. W., Macias-Fauria, M., Long, P. R., Benz, D., and Willis, K. J.: Sensitivity of global terrestrial ecosystems to climate variability, Nature, 531, 229–232, 2016.

Shahin, M. A., Ali, M. A., and Ali, A. S.: Vector Autoregression (VAR) Modeling and Forecasting of Temperature, Humidity, and Cloud Coverage, in: Computational Intelligence Techniques in Earth and Environmental Sciences, pp. 29–51, Springer, 2014.

Smith, M. D.: An ecological perspective on extreme climatic events: a synthetic definition and framework to guide future research, Journal of Ecology, 99, 656–663, 2011.

Smith, T. M., Reynolds, R. W., Peterson, T. C., and Lawrimore, J.: Improvements to NOAA's historical merged land-ocean surface temperature analysis (1880-2006), Journal of Climate, 21, 2283–2296, 2008.

Stackhouse, Jr. Paul, W., Gupta, Shashi, K., Cox, Stephen, J., Mikovitz, C., Zhang, T., and Chiacchio, M.: 12-Year Surface Radiation Budget Dataset, GEWEX News, 14, 10–12, 2004.

Stock, J. H. and Watson, M. W.: Vector autoregressions, The Journal of Economic Perspectives, 15, 101–115, 2001.

Su, L., Jia, W., Hou, C., and Lei, Y.: Microbial biosensors: a review, Biosensors and Bioelectronics, 26, 1788–1799, 2011.

Triacca, U.: Is Granger causality analysis appropriate to investigate the relationship between atmospheric concentration of carbon dioxide and global surface air temperature?, Theoretical and applied climatology, 81, 133–135, 2005.

Tucker, C. J., Pinzon, J. E., Brown, M. E., Slayback, D. A., Pak, E. W., Mahoney, R., Vermote, E. F., and El Saleous, N.: An extended AVHRR 8-km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data, International Journal of Remote Sensing, 26, 4485–4498, 2005.

Tuttle, S. and Salvucci, G.: Empirical evidence of contrasting soil moisture–precipitation feedbacks across the United States, Science, 352, 825–828, 2016.

Van der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G., Mu, M., Kasibhatla, P. S., Morton, D. C., DeFries, R., Jin, Y. v., and van Leeuwen, T. T.: Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997–2009), Atmospheric Chemistry and Physics, 10, 11 707–11 735, 2010.

Verbesselt, J., Hyndman, R., Newnham, G., and Culvenor, D.: Detecting trend and seasonal changes in satellite image time series, Remote sensing of Environment, 114, 106–115, 2010.

Verbesselt, J., Umlauf, N., Hirota, M., Holmgren, M., Van Nes, E. H., Herold, M., Zeileis, A., and Scheffer, M.: Remotely sensed resilience of tropical forests, Nature Climate Change, 2016.

Von Storch, H. and Zwiers, F. W.: Statistical analysis in climate research, Cambridge university press, 2001.

Willmott, C. J., Matsuura, K., and Legates, D.: Terrestrial air temperature and precipitation: Monthly and annual time series (1950-1999), Center for climate research version, 1, 2001.

Wu, D., Zhao, X., Liang, S., Zhou, T., Huang, K., Tang, B., and Zhao, W.: Time-lag effects of global vegetation responses to climate change, Global change biology, 2015.

Xie, P. and Arkin, P. A.: Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs, Bulletin of the American Meteorological Society, 78, 2539–2558, 1997.

Xie, P., Chen, M., Yang, S., Yatagai, A., Hayasaka, T., Fukushima, Y., and Liu, C.: A gauge-based analysis of daily precipitation over East Asia, Journal of Hydrometeorology, 8, 607–626, 2007.

Ye, L. and Keogh, E.: Time series shapelets: a new primitive for data mining, in: Proc. of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD ’09, p. 947, ACM Press, New York, USA, doi:10.1145/1557019.1557122, 2009.

Zakaria, J., Mueen, A., and Keogh, E.: Clustering Time Series Using Unsupervised-Shapelets, in: 2012 IEEE 12th International Conference on Data Mining, pp. 785–794, IEEE, doi:10.1109/ICDM.2012.26, 2012.

Zeng, N., Hales, K., and Neelin, J. D.: Nonlinear dynamics in a coupled vegetation-atmosphere system and implications for desert-forest gradient, Journal of Climate, 15, 3474–3487, 2002.

Zhang, X., Alexander, L., Hegerl, G. C., Jones, P., Tank, A. K., Peterson, T. C., Trewin, B., and Zwiers, F. W.: Indices for monitoring changes in extremes based on daily temperature and precipitation data, Wiley Interdisciplinary Reviews: Climate Change, 2, 851–870, 2011.

Zscheischler, J., Mahecha, M. D., von Buttlar, J., Harmeling, S., Jung, M., Rammig, A., Randerson, J. T., Schölkopf, B., Seneviratne, S. I., Tomelleri, E., et al.: A few extreme events dominate global interannual variability in gross primary production, Environmental Research Letters, 9, 035 001, 2014a.

Zscheischler, J., Michalak, A. M., Schwalm, C., Mahecha, M. D., Huntzinger, D. N., Reichstein, M., Berthier, G., Ciais, P., Cook, R. B., El-Masri, B., et al.: Impact of large-scale climate extremes on biospheric carbon fluxes: An intercomparison based on MsTMIP data, Global Biogeochemical Cycles, 28, 585–600, 2014b.

Zscheischler, J., Reichstein, M., Harmeling, S., Rammig, A., Tomelleri, E., and Mahecha, M. D.: Extreme events in gross primary production: a characterization across continents, Biogeosciences, 11, 2909–2924, 2014c.

Zwiers, F. W., Alexander, L. V., Hegerl, G. C., Knutson, T. R., Kossin, J. P., Naveau, P., Nicholls, N., Schär, C., Seneviratne, S. I., and Zhang, X.: Climate extremes: challenges in estimating and understanding recent changes in the frequency and intensity of extreme climate and weather events, in: Climate Science for Serving Society, pp. 339–389, Springer, 2013.