

Interactive comment on “Global hydro-climatic biomes identified via multi-task learning” by Christina Papagiannopoulou et al.

Christina Papagiannopoulou et al.

christina.papagiannopoulou@ugent.be

Received and published: 18 June 2018

Response to the review comments of Stephanie Horion

The work presented by Papagiannopoulou et al. in this manuscript is of interest for the reader of GMD and is also very relevant for the ecosystem and climate research community. Overall the manuscript is well structured and the methodology section generally well documented. Knowing that the focus of GMD is on the progress and novelty in computation and model development, I support the need for in-depth description of the MTL model and its performances (e.g. STL vs MTL, capability to detect Granger causality, etc.). However I believe that the manuscript would be strengthened and results better supported if the authors could really demonstrate that

C1

the new product (i.e. map of hydro-climatic biomes) is outperforming other bio-climatic maps that did not consider in their design the vegetation response to climate variability. This is still lacking in the current manuscript. In addition some methodological aspects that led to the final design of the MTL and clustering should also be improved to backup the authors' statement on the performances of the final models and derived product. Based on these observations and on the detailed comments provided below I recommend the paper for major revision.

We would like to thank the reviewer for her appreciation of our manuscript, the constructive feedback, and thorough assessment. Below we provide a point-by-point response to each comment.

In general, we would like to clarify that the goal of our study is to provide a new methodology that can identify coherent regions in which vegetation responds to climate in a similar way. We model our problem with a multi-task learning approach that considers the different locations as different tasks and learns the relationship between the tasks during the learning process. Hence, the climate–vegetation interaction is simultaneously learned for all locations. The similarity between the learned relationships (between the tasks) is also discovered during the process. This is the first time (to the best of our knowledge) that an approach of this kind, which discovers the structure of the relationships between the different locations, is applied on this setting. As such, we try to avoid the claim that our *hydro-climatic biomes* ‘outperform’ other schemes, which rely on climate and/or vegetation data and not on the modeled interaction between climate and vegetation. It is not really our intent to outperform these land cover classifications, and the comparison that is provided against them is to assure that – despite the fact that our approach does not prescribe any explicit information on land cover types – comparable regions arise from our data-guided appraisal.

C2

Specific comments

Introduction

- Studying vegetation response to climate variability is and has been the focus of numerous researches. I know the objective of the authors is to create a new bio-climatic map, however I can imagine that their work build up on recent developments in science regarding ecosystem response to climate variability. This is not well reflected in the introduction. Please add some references to key papers, studies in the matter. Some suggestions below:

Liu L, Zhang Y, Wu S, Li S, Qin D (2018) Water memory effects and their impacts on global vegetation productivity and resilience. *Sci Rep*, 8, 2962.

Seddon AW, Macias-Fauria M, Long PR, Benz D, Willis KJ (2016) Sensitivity of global terrestrial ecosystems to climate variability. *Nature*, 531, 229-232.

De Keersmaecker W, Lhermitte S, Tits L, Honnay O, Somers B, Coppin P (2015) A model quantifying global vegetation resistance and resilience to short-term climate anomalies and their relationship with vegetation cover. *Global Ecology and Biogeography*, 24, 539-548.

Nemani RR, Keeling CD, Hashimoto H et al. (2003) Climate-driven increases in global terrestrial net primary production from 1982 to 1999. *Science*, 300, 1560-1563

True. We will include the suggested literature in the revised manuscript.

- The authors claim (p2, l23) that it is the first time that ecoregions are being defined based on the analysis of vegetation response to climate variability. I agree that the idea is relatively novel and definitely relevant. Yet previous attempts have been made, notably by combining PCA and clustering techniques applied to climate and vegetation dataset. See the following reference as an example:

Ivits E, Horion S, Fensholt R, Cherlet M (2014) Global Ecosystem Response Types

C3

Derived from the Standardized Precipitation Evapotranspiration Index and FPAR3g Series. *Remote Sensing*, 6, 4266-4288.

Thanks for pointing us to this paper. The suggested literature is relevant to our study and will also be referred to in the revised version. In addition, the differences compared to this and other studies will also be highlighted in the revised version.

Methodology

- Sect. 2.4. The authors mentioned that the ASO method used here should not be confused with PCA. It would be useful to develop this statement. Indeed for both techniques orthonormal vectors are derived from the high dimensional feature space, creating a new 'optimized' low-dimensional feature space. The authors mentioned that the goal of the ASO method is to detect the PC of the predictive structure. Knowing that PCA can be performed in two ways (t-mode and s-mode), the t-mode being the most frequently used by climatologist to identify recurrent spatial patterns over time, whereas the S-mode allows for identifying recurrent temporal patterns over space. How would the current method differ from an extended PCA in S-mode? I can imagine that using EPCA over a dataset as large as the one used here could be a real challenge for example. But I would like the authors to elaborate on the pros and cons of the new method as compared to already established techniques in the climate research such as PCA/EPCA for example.

Thanks for this comment. In the last paragraph of Sect. 2.4, we mention the main difference between the commonly-used PCA approaches and the proposed method. However, we will elaborate on the differences and potential advantages of our approach in the revised manuscript.

To give an example, in the work of Ivits et al. (2014), PCA is performed on the data

C4

matrix of drought anomalies (measured by Standard Precipitation Evapotranspiration Index data, SPEI) and vegetation state (measured by Fraction of Photosynthetically Active Radiation data, FPAR3g), while the clustering is applied to the correlation coefficients based on the spatio-temporal patterns obtained by PCA.

Our approach is based on different principles, and as such it is expected to yield different results. We explicitly consider the climatic variables as predictors and the vegetation variable as target variable, and we learn the relationship between them in a supervised setting. As such, the regions that we define rely on the relationship between climate and vegetation in a prediction setting, and the clustering is calculated based on similarity of this relationship (i.e. the model coefficients for different locations). As such, we learn relationships between climate and vegetation in a supervised setting, whereas PCA-based methods are fully unsupervised. In our study the SVD decomposition is used as part of the optimization algorithm, thus in a supervised setting. In this setting, the model weights are optimized based on a given training set. Therefore, the discovered structures are obtained during the training process. This novel part of our methodology will be stated more clear in the revised version.

- Sect. 2.5. The authors do not give any name or reference for the clustering technique used here. Please clarify if a new algorithm has been developed for the study or if an already developed clustering technique was applied.

In the manuscript, it is mentioned that the clustering technique that we use is the agglomerative hierarchical clustering (with Euclidean distance measure) which is a well-known clustering method in Statistics (see Sect. 2.5 and 3.2 of the manuscript). To make it more clear to the broad audience of GMD, we will mention in the revised manuscript that we use the hierarchical clustering python implementation of scikit-learn, and add a specific reference.

C5

- General comment on the use of R^2 for assessing the model performance: at several occasions (in the manuscript and in the supplementary material), the authors used R^2 to quantify the performance of different models (MTL vs. STL, models with and without Granger causality, inclusion of higher-level features in the input dataset, final decision on the number of clusters). They generally conclude that the best model is the one with the highest R^2 . I agree on the principle, however looking at the differences between R^2 (e.g. figures 3b and 3d, large areas present difference in R^2 below 0.1), I wonder whether all these differences are statistically significant. As based on the analysis of R^2 , the authors are deciding on the final set of input data, the final design of the MTL model, and the final number of clusters, I would really urge the need for further statistical assessment of the model performances. One first analysis could simply be to estimate the percentage area of pixels with statistically significant increase in R^2 .

The distributions depicted in Figs. 3c and 3f of the manuscript show that the results of the MTL and STL methods are substantially different. Specifically, the distributions of the MTL results are shifted to the right, meaning that STL is outperformed by MTL at global scale. This result can be confirmed by any paired statistical test (Demšar, 2006). The same comparison can be applied for the performance comparison of the full and the baseline MTL models.

However, we agree that a significance test of this difference was not included in the original manuscript. At pixel level, traditional statistical tests usually have too many assumptions for our purposes. Alternatively, non-parametric tests based on resampling, such as permutation or bootstrap tests, cannot really be applied due to the size of our data set. A proposed solution is to use the Diebold-Mariano statistical test (Diebold, 2015). This test can be used here to compare the MTL and the STL approaches and will be used in the revised version of the manuscript.

C6

For the final decision about the number of clusters, see our answer below.

Results

- General comment on the final number of clusters: the fact that the majority of the Iberian Peninsula is included in the transitional energy driven cluster together with Ireland, an important part of SE Asia, part of Brasil and Venezuela – Colombia makes me wonder if a higher number of clusters would not be more appropriate. The authors mentioned already in Figure S2 that the differences in the predictive performance for $h = 6 - 15$ are marginal. Further assessments should therefore be performed in order to identify the optimal number of hydro-climatic biomes. Part of this assessment should be dedicated to the understanding of the actual drivers (main predictors) for each biome. I believe providing a solid justification for the naming of the different biomes (by referring back to the main predictors) would be beneficial for the paper.

We agree that the differences in predictive performance for $h = 6-15$ are marginal. However, the proposed method is a fully data-driven approach that is not fine-tuned based on any kind of prior knowledge. Therefore, the selection of the final value of the h parameter is based on an objective criterion, i.e. the model performance. As for the resulting map (Fig. 4a), although we are aware that this map may not fully reflect all particular expectations, we do believe that the spatial distribution broadly captures the expected regimes of climate–vegetation interactions, as described in the results section. Note as well that in our early experiments we ran our approach with a different number of clusters to visually inspect the resulting regions. The regions formed with h values close to 11 are similar to the reported ones (Fig. 4a of the manuscript). This result proves the robustness of the proposed method to detect the basic vegetation response types with respect to climate. These results (for $h = 8-12$) will be included as supplementary figures in the revised manuscript.

C7

Concerning the labels scheme, we should stress that the names of the biomes are inspired by the main predictors based on Papagiannopoulou et al. (2017). We are afraid that making the labels reflect these predictors more accurately would make them extremely complex and rather impractical.

- In relation to the previous comment, how does the new global map of hydro-climatic biomes perform as compared to previous ones (not including information of vegetation condition and response to climate)? It would be really interesting if the authors could showcase for one (or more) bio-climatic zone how the new bio-climatic zone provide a finer, more accurate picture of global terrestrial biomes by analysis the specific (/sub-local) ecosystem response to climate variability. To this regard, the bioclimatic map produced by Metzger et al. (see reference below) could also be of interest for comparison.

Metzger MJ, Bunce RGH, Jongman RHG et al. (2012) A high-resolution bioclimate map of the world: a unifying framework for global biodiversity research and monitoring. *Global Ecology and Biogeography*, 22, 630-638.

Thanks for the relevant reference, it will definitely be cited in the revised manuscript. However, as we mentioned above, by using our approach we really aim for detecting regions of consistent behavior in response to climate (based on the learned weights). That is what we should evaluate. As such, we cannot really aim for ‘accurate’ biomes. This is the reason why we do not compare our result to other data-driven approaches that rely on climate and/or vegetation data (as Metzger et al. (2012)), since our study tries to detect regions based on different criteria (based on the interaction between climate–vegetation and not on the data). This point will also be stressed in the revised manuscript. We also note again that the comparison that is provided against traditional land classification schemes is to assure that comparable regions arise from our data-guided approach, despite these land cover types not being exppecificaly

C8

prescribed.

- Figure 4. (c) The Köppen classification divides the world into 5 main classes and 29 sub-classes. The authors should justify the use of 10 classes in the figure. This can be very misleading when looking and interpreting the results. An example: I do not think that the statement p14, l21-23 ‘...the region of North Asia is coherent in terms of climate, but quite diverse in terms of vegetation types; the hydro-climatic biomes show a clear distinction from shrublands (...) to coniferous ...’ holds entirely when looking at the high level details (29 classes) of the Köppen classification. Please justify your choice here.

It is true that the Köppen climate classification scheme consists of divisions and sub-divisions of the five main climate types. We could choose to use the divisions of the Köppen classification, which are basically 12 (if we also divide the tropics further) and not 10 as in Fig. 4. However, the use of 10 instead of 12 classes will not make the map look much different. Moreover, from the color scheme used in Fig. 4, it is clear that there are five main classes. In Fig. 4, we aim for comparing the regions detected by the proposed method to the regions based on the Köppen climate classification scheme. Since the division of 10 climate classes is closer to the number of regions detected by our approach, we choose this number of regions (10) on Köppen’s map. Nonetheless, we agree that the statements mentioned in the comment sound a bit strong, so in our discussion we should take into account also the sub-division (29 classes) of Köppen classification. Again, the comparison to the Köppen and IGBP maps serves only as a general evaluation or proof of concept for our hydro-climatic biomes map, since in the end they are based on a different rationale. Thus, we will clarify in the revised manuscript that we do not claim that our map is capable of ‘outperforming’ these classification schemes.

C9

- Supplementary material S4. The authors indicate that the best-formed clusters are depicted in FigS4a (hence by the hydro-climatic biomes). I find very difficult to make any final judgment of the best “depiction” (/detection) of biomes based on the 2-dimensional graphs provided here.

Yes, true. Another dimensionality reduction technique, such as the t-sne, might give a visually better result. We are exploring the potential of other methods in order to improve these figures in the revised version.

Technical comments

- P5, l14: please add a reference for the statement: ‘...this kind of modelling is becoming more common in climate science...’

This sentence refers to the previously mentioned studies, which are described in the same paragraph, and serves as a conclusion that MTL approaches are used more common recently than in the past in climate science. However, we will repeat the references at this point as well in the revised manuscript.

- P10, l10: please clarify what you mean by multi-month vegetation dynamics. Is it seasonal, subseasonal, yearly?

We mean monthly vegetation. It will be corrected.

- P12, l5: please correct ‘Geanger’ with ‘Granger’

True, thanks.

- Figure 4. (a) the color code for the clusters sub-tropical energy driven and

C10

mid-latitude temperature driven are too similar. It is difficult to differentiate them. Please adjust the color scheme of the legend.

Indeed. We will adjust the color scheme in the revised manuscript.

- p15, l22: The term 'turning point' has only been introduced recently in ecosystem and climate science so for clarity, you can refer to:

Horion S, Prishchepov AV, Verbesselt J, De Beurs K, Tagesson T, Fensholt R (2016) Revealing turning points in ecosystem functioning over the Northern Eurasian agricultural frontier. *Glob Chang Biol*, 22, 2801-2817.

We will include this relevant reference in the revised manuscript.

References

Demšar, J. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine learning research*. 2006, 7(Jan), 1-30.

Diebold, F. X. Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of Diebold-Mariano tests. 2015, *J. Bus. Econ. Stat.*, 33.

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, *Environ. Res. Lett.*, 2017, doi:10.1088/1748-9326/aa7145.

Ivits, E., Horion, S., Fensholt, R., Cherlet, M. Global Ecosystem Response Types Derived from the Standardized Precipitation Evapotranspiration Index and FPAR3g

C11

Series. *Remote Sensing*, 2014, 6, 4266-4288.

Metzger, M. J., Bunce, R. G. H., Jongman, R. H. G. et al. A high-resolution bioclimate map of the world: a unifying framework for global biodiversity research and monitoring. *Global Ecology and Biogeography*, 2012, 22, 630-638.

C12