Editor's Comments

MINOR REVISIONS

The four referees suggest that the submission may be publishable, but only after many minor revisions have been made to your manuscript. Therefore, I invite you to respond to their comments and revise your manuscript.

RE: We are grateful to the four referees for their kind words and fast and to the point review, which has helped us filter out some smaller uncertainties and improve the manuscript. Our response to specific questions you can follow below.

We sincerely apologize for some delay in our response, which was due to numerous conferences in May / June and some project obligations. If we still need to modify or adjust any section of our article we would be happy to do so.

FYI we have been tracking download statistics of the PNV maps (http://dx.doi.org/10.7910/DVN/QQHCIK and (https://www.arcgis.com/apps/MapJournal/index.html?appid=1856322400844a7cab348bccfa4bee76) and responding to the questions and comments by users. So far there were no serious objections to the spatial patterns in the maps so there seems to be no need for updating the maps / predictions. Our plan is to try to at least recompute all maps at finer resolution (250m) towards the end of 2018 most probably and release them under unique DOI. Note that update of the maps would not impact the main discoveries reported in the paper, a copy of all R code and maps will remain permanently accessible via DOI above.

Reviewer 1 (Anonymous)

Basic reporting

no comment

Experimental design

no comment

Validity of the findings

no comment

Comments for the Author

This paper is well structured and written. It compares four machine learning methods for potential natural vegetation (PNV) mapping from climate and environmental covariates. The results show that the random forest performs the best, which is coherent to many other findings solving classification/regression issues. I have a minor comment on this paper.

Many of MODIS FPAR is from the vegetation that has been affected by human activity. In the training stage, the author applied the WDPA and IFL mask to select samples with minimum affection by human activity. These two masks make the training pool data spatially unevenly distributed. Please clarify whether random sampling or stratified sampling is applied and discuss its possible effect on the results.

> RE: Random sampling has been used to select training points. This is a rather common sampling design which also enables us also to run Cross-validation without posing any special statistical assumptions.

Reviewer 2 (Jingzhe Wang)

Basic reporting

This manuscript is aimed to produce PNV maps that are both more detailed, richer in information, based on multiple machine learning algorithms.
The writing is clear enough, however the manuscript contains a number of spelling and grammatical errors which should be carefully revised before acceptance.
The literature is suitable for the purpose of the article.

Experimental design

no comments

Validity of the findings

The data are robust and results compatible with conclusions.
It is necessary to present evaluation of TPR outliers in the Table 2. (e.g., Carpinus orientalis and Cupressus sempervirens).

> RE: We have added some text explaining TPR outliers on P28L570-574.

In the revision round, it would be interesting to compare the performances of multiple MLA.

> RE: We already present extensive comparison of performances of different MLA's in Fig. 5, 8 and 11. Certainly even more MLA's even more comparison could be added, but this would maybe not impact the main results / discoveries of this paper.

Comments for the Author

1. As the title of this manuscript listed “Global Mapping of Potential Natural Vegetation: An Assessment of Machine Learning Algorithms for Estimating Land Potential”. Content of PNV and FAPAR is too much in the manuscript. Authors should focus on the effect on wind erosion potential by land use/cover changes and land management practices.

RE: This is certainly an interesting area of application as PNV maps could show to be especially useful in areas of high potential erosion. We encourage the reviewer to download our maps and use them for suggested purpose (the maps are available publicly at: http://dx.doi.org/10.7910/DVN/QQHCIK).

2. The text is extensive and repetitive in many parts. Thus, I recommend rewrite abstract and introduction. The conclusions can be improved.

RE: We have shortened the abstract and introduction, without omitting any important information. Abstract probably cannot be shortened more without taking out important information and potentially losing details about methods, main results, implications and limitations of the study.

3. There are some sentences which are difficult to understand, for instance:
Page 4/lines 105-106: For efforts aimed at land restoration ....
Page 9/lines 239-240: We use about 3× more training points ....
Page 12/lines 344-346: TPR values range from 0 to 1 where 1 ....
Page 16/lines 394- : Results of the spatial Cross-Validation, as implemented ....

RE: We have rewritten sentences from above to try to better clarify what we mean. Indeed the original text has not been clear enough.

4. The figures should be renumbered accordingly such that they are introduced in the correct sequence (e.g., Fig 1). You did not mention Fig 1a in the manuscript.

RE: Mention of Fig 1a and b is now added on L88 and L92.

5. The time stamps in Figure 9 are hard to read, as are the lines in Figure 10, which makes them difficult to interpret. There seems to be no explanation in the Figure legend or any values which describe the distribution of the tree species, it may be helpful to put these directly in the Figure.

RE: Explanation of reference years/months has been now added in Figures 9 and 10 captions. Legend has been added to Figure 9.

6. The reference citation in the title of Table 2 is inappropriate. I recommend deleting the citation and placing the correlated content information in the text. The Species name of forest tree taxa should be placed in italic type.

RE: DONE.

7. The spatial resolution of Input data (environmental covariates) should be illustrated in detail.

RE: Does reviewer maybe implies that we should provide visualizations of environmental covariates at native resolution? Most of environmental covariates we used have been described in detail in the original literature. In addition, we provide WMS that allows anyone to zoom in into maps / area and explore maps / specifically the effects of resolution (https://www.arcgis.com/apps/MapJournal/index.html?appid=1856322400844a7cab348bccfa4bee76). We hope this is satisfactory.

8. Line 287 The resampling methods of nearest neighbor interpolation, bilinear interpolation, and cubic convolution interpolation were commonly used. In this study, which one were used?

RE: We normally use "bilinear" resampling, unless downscaling covariates, in which case we use the "bicubic splines". All these are standard GDAL resampling techniques. Now added text on P11L295.

9. The model inputs seem to be complicated. It would be better to have a schematic picture that shows the model framework.

RE: Even better - we have documented all modeling steps at: https://github.com/Envirometrix/PNVmaps (publicly available repository) through R code available via the script "PNV_mapping.R".

10. The abbreviation of Scaled Shannon Entropy Index has two forms (EIs, SSEI, NDVI and some other technical terms) should be uniformed, and the corresponding full name should be given. Work on text clarification and formatting, improve figures, list of the abbreviations would be useful in this case to easily understand the text.

RE: That is a good point. The correct abbreviation is SSEI. This has now be harmonized across the document.

11. The quantitative comparison of results of modeling potential distribution of tree species in Europe should be illustrated, not only the visual comparison.

RE: If the reviewer will give us more ideas about what type of illustration does he has on mind. At the moment we provide three figures 8, 9 and 10 that illustrate predictive performance results and distinctness of the spatial patterns.

12. The calculating equations of RMSE and ME should be listed as standard equations with number (Page 13/line 366). The variables in the equations of them are not explained.

RE: Correct. Now added on P13.

13. More discussions on the ecological gradients should be added to the manuscript.

RE: This comment is unfortunately not detailed enough. If the reviewer could provide more detail where specifically in text should we add more discussion we would be happy to do so. The P4L109-130 already contains a comprehensive explanation of ecological gradients and how to represent them.

14. The detection and assessment of PNV is critical for the prevention of environmental deterioration especially in arid and semi-arid areas. Hence, I propose to consider the special region.

RE: Indeed, our PNV maps could be especially useful for land restoration projects in arid and semi-arid regions. Our maps are publicly available and we also leave it to other teams / groups to analyse usability of our maps.

15. The altitude, topography and soil types might have a certain impact on the PNV. Due to the lack of information it is not clear the need of pooling all data together. I am wondering if by comparing different climate (soil/topographical) regions and analyze the response of the methodology under these scenarios would not produce better results. In any case, this discussion was missing and need to be clarified.

RE: That is an interesting issue. We have decided not to use any soil layers as predictors as these are also based on same climatic and remote sensing images, but instead to only use global map of lithological units (http://dx.doi.org/10.1029/2012GC004370). The analysis the reviewer suggests is probably beyond this work, but we have added discussion on this topic on P25L485-490.

16. The challenges and limitation in mapping PNV using MLA should be further highlighted in the discussion.

RE: We have extended the existing discussion on P26L507-511 and L525-532.

17. Line 416 “Correlation analysis (predicted distribution maps) indicates that…” is not clear enough. Which fig (did you mean fig. 6)? There should be more details here.

RE: This paragraphs has now been improved; see P19L430. FYI we have implemented cross-correlation analysis of all produced maps as show in https://github.com/Envirometrix/PNVmaps/blob/master/R_code/PNV_mapping.R#L890).

18. In general, the machine learning algorithm with more parameters or hyper-parameters could lead to more complicated training, though it has better accuracy. A good algorithm should have good accuracy and meanwhile, include less trained parameters. It looks that RF has more complicated hyper-parameters. More discussions on this point should be added to the paper.

RE: This is an interesting point. Number of covariates and their repetition is not a big problem for RF since it is an ensemble model which on the end has low sensitivity to noise and over-parameterisation (https://stats.stackexchange.com/questions/47457/pca-on-high-dimensional-text-data-before-random-forest-classification). But yes it results in unnecessary high computing times if the number of covariates becomes excessive.

Reviewer 3 (Anonymous)

Basic reporting

This paper is well written and fits into the scope of this journal. This study contributes to the current knowledge of natural vegetation mapping using data-driven approaches. The methodology presented in this study would help other researchers to replicate the methods and apply to other parts of the world, provided all the data used in this study are available.

Experimental design

Modeling of natural vegetation across the globe using statistical and machine learning models is very useful for biodiversity management, and help quantify the effect of climate change and human activities on existing vegetation. Empirical/data-driven models can serve as surrogates to computationally intensive physical/numerical vegetation models. The authors did a commendable job in collecting the huge amount of data and developing models that bridge the gap between existing models and present knowledge of vegetation mapping.

RE: We thank the reviewer for his/her kind words.

Validity of the findings

Conclusions are well presented and answer the research questions.

Comments for the Author

1. Line 23: PNV is already spelled out in line 20, so just say PNV here.

RE: DONE.

2. Could you please elaborate the 'time' label on x-axis?

RE: You mean in the Figure 1b? PNV is a dynamic feature. It changes through time and time-reference is important as the differences can be large (especially with future global warming for example PNV might be significantly different than pre-industrial revolution). Additional explanation now added on P3L89-92.

3. Fig. 12: Bottom figure: It should be predicted FAPAR.

RE: DONE.

4. Please include a section on assumptions that are considered in this study or as a section describing limitations of the modeling framework. I would keep all the assumptions in one place to get a complete picture of what data are available and what are the limitations.

RE: Yes that is exactly the idea of the section "PNV mapping and species distribution modeling" on P4.

5. Fig. 14: Please elaborate 'To convert to percent divide by 253'.

RE: The original Copernicus FAPAR products are distributed in scale 0–235 (https://land.copernicus.eu/global/products/fapar). For consistency we stick to using the same scale as Copernicus (see also L231).

6. Did you run your models on GPU, if yes give the specifications.

RE: No, we have only used CPU's. Now revised on P12L340-341.

7. How long does it take to run each model (random forest, neural network etc.)? Please provide a comparison of computation time, performance, parameter sensitivity etc. maybe in the form of table for better understanding.

RE: This is a rather technical detail and is dependent on the OS, chipset characteristics etc, so we advise that the users refer to our R code repository for such issues (https://github.com/Envirometrix/PNVmaps/).

8. Could you please provide a comparison of the best-predicting model results to ground-truth observations?

RE: This is exactly what the Figs. 5, 8 and 11 show.

9. What are the percentage of relevance of the most important predictors like precipitation, temperatures etc.? Can you calculate those from connection weights of the neural network?

RE: In principle, we do not zoom in to NNs because these have also not performed so well in this comparison. We do however provide lists of most important covariates (https://github.com/Envirometrix/PNVmaps/blob/master/R_code/PNV_mapping.R#L149). Because these can be rather long tables, we have instead decided to only put general info about variable importance e.g. on P14L394-398.

10. It would be nice to refer some of the recent machine learning papers in related field.

For example:

Deo, R.C. and Şahin, M., 2015. Application of the extreme learning machine algorithm for the prediction of monthly Effective Drought Index in eastern Australia. Atmospheric Research, 153, pp.512-525.

Sahoo, S., T. A. Russo, J. Elliott, and I. Foster (2017), Machine learning algorithms for modeling groundwater level changes in agricultural regions of the U.S.,Water Resour. Res.,53, 3878–3895, doi:10.1002/2016WR019933.

Okujeni, A., van der Linden, S. and Hostert, P., 2015. Extending the vegetation–impervious–soil model using simulated EnMAP data and machine learning. Remote Sensing of Environment, 158, pp.69-80.

Li, X., Chen, W., Cheng, X. and Wang, L., 2016. A comparison of machine learning algorithms for mapping of complex surface-mined and agricultural landscapes using ZiYuan-3 stereo satellite imagery. Remote sensing, 8(6), p.514.

RE: Indeed the references by Li et al. (2016) and Deo et al 2015, are highly relevant and have been added now to the list of references.

PS: I've just first time heard about the Extreme Learning Machine algorithm (and of course it is available in R and well documented -> https://github.com/mlampros/elmNNRcpp). Indeed looks interesting for future testing.

Reviewer 4 (Jorge Maestre Vidal)

Basic reporting

The work presents a comprehensive study of the effectiveness of different machine learning approaches (in particular, Artificial Neural Networks, Random Forest, Generalized Boosted Regression and K-NN) applied to the global mapping of Potential Natural Vegetation (PNV). In general terms, it is an interesting contribution. It is well structured, follows a clear methodology and provides a wide experimentation. The writing is good, which is accompanied by visual content that facilitates its understanding. The paper meets all the requirements to be accepted. However, there are minor modifications that may greatly improve the dissemination of the performed research

RE: We thank reviewer for his kind words.

Experimental design

Although the experimentation is appropriate, the paper would gain in interest if it would provide a more in-depth discussion about the motivation that led to select the implemented algorithms, as well as their hyphotetized benefits/drawbacks when applied to PNV mapping.

RE: Our motivation is mainly to provide a more objective basis for land restoration studies and to emphasize how would the world look like without people (so possibly useful to e.g. the authors of the book "The world without us"). This is all now mentioned on P4L101-108.

Validity of the findings

The extensive and methodical experimentation proves the effectiveness of the selected methods. However, the paper would gain in clarity if the achieved results were correlated with the characteristics of the algorithms (from the ML and pattern recognition point of view).
On the other hand, I suggest to enhance the criteria considered for estimating the best TPR, since the ROC curve covers the spectrum of all possible values based on calibration, and a simple average may mislead the results interpretation. Maybe the best Youden index could fit with the experimental purposes.

RE: Thank you for this information. Indeed the Youden index could also be used to give insight into the mapping accuracy, but there are also many more indices that could be used. We use TPR mainly to emphasize which classes are mapped most efficiently and which are critically poor. Average TPR is not reported in the main results.

Since all our inputs and outputs are available publicly via https://github.com/Envirometrix/PNVmaps/ we hope that other users will consider further comparing and analysing our results. If the reviewer things that such additional analysis is crucial, however, we would be happy to add it.

Comments for the Author

The paper provides an interesting contribution, which a well-described research methodology and an extensive experimentation. Hence it meets all the requirements to be accepted

RE: Thank you.
