Responses to comments from RC1

Manuscript: ”Influence of plant traits on water cycle processes in the Amazon Basin”

Kien Nguyen and Maria J. Santos

Dear editor,

This letter serves to reply to to the comments provided by RC1 to our submitted manuscript. We appreciate the time to review our work, and have seriously considered and appreciated all the comments by the reviewer. We detail below on how we plan to address them in a revised version.

General Comments

The aim of the manuscript by Nguyen and Santos was to ask the highly relevant and interesting question, “Do plant traits affect the water cycle in the Amazon Basin?”. The manuscript is well-written; I enjoyed reading it. I appreciate the authors' attempt to address this question and sincerely hope my review is not too disheartening.

Response: We appreciate your assessment of our work and the encouragement regarding the topic we chose to study.

However, and unfortunately, there are statistical methodological violations in the analyses presented. More crucially, the authors use predicted global trait data derived using (in part) climate data and vegetation indices. The authors then use this climate-derived trait data and vegetation indices to predict climate variables. These climate variables used by Nguyen and Santos are likely highly correlated with the climate data used to predict the original trait data.

Response: Thank you for your concern, which we agree with. Indeed there could be a circularity problem in the use of climate data to predict trait values that were in turn generated by predictions that potentially used the same climate variables. The most problematic variable would have been LST which would have been the input to the bioclimatic variables that (Moreno-Martínez et al., 2018) used to develop the random forest algorithm. (Moreno-Martínez et al., 2018) showed that only 3 of the bioclimatic variables related to temperature were among those have considerable power in trait prediction (Maximum Annual Temperature, Isothermality and Mean temperature of warmest quarter), while LST (named LSTmed in their research) does not contribute significantly to gap-filling and prediction of traits.

The vegetation indices used are already documented components of the calculations for evapotranspiration. The circularity of the analysis strongly calls the validity of the presented results into *question. Correlations between variables where their relationships are already clearly documented (ET, LAI, NDVI) do not represent a novel scientific result*

Response: Thank you for your concern, and indeed this was an oversight from our side and therefore we will remove these analyses from the manuscript.

Specific Comments

The authors used multivariate and quantile regressions to examine relationships. Thus, the validity of the statistical methods is crucial to the validity of the results presented and conclusions drawn.

Response: Thank you for your statement. We fully agree.

RC1.1. *Independence of independent variables In the abstract, the authors state that the four plant traits (1-4 above) are derived from remotely sensed data. The authors tested for multicollinearity and, thus, the independence of independent variables by calculating the VIF. However, an examination of the documented methods used to derive the trait data produced by Moreno-Martinez et al. (2018) reveals that some of the "most relevant variables for the gapfilling" for a target trait were covarying traits (Table C.1). For example, LDMC was highly relevant to derive SLA values and LPC was highly relevant to derive LNC values. Clearly, SLA and LDMC are not independent. Likewise, LNC and LPC are not independent.*

ENV_PARAM	LPC	SLA	RSQUARED
ET_MEAN	-0.019	-0.470	0.225
PET_MEAN	-0.381	0.456	0.290
SM_MEAN	0.164	-0.236	0.068
VPD_MEAN	-0.324	0.367	0.196
LSTDAY_MEAN	-0.187	0.347	0.131
LSTNIGHT_MEAN	-0.260	-0.085	0.083
ET_STD	-0.201	0.301	0.109
PET_STD	-0.156	0.097	0.028
SM_STD	-0.244	0.199	0.081
VPD_STD	-0.449	0.479	0.353
LSTDAY_STD	0.095	0.512	0.289
LSTNIGHT_STD	-0.006	0.212	0.044

Figure 1: Revised results for the multivariate regression

Response: Thank you for pointing out the problems in our assumption of independence. Table C.1 in (Moreno-Martínez et al., 2018) shows the five most relevant variables for gap-filling of trait predictions for the generation of the global map. LDMC is the third most relevant variables used in the gap filling approach to fill SLA, and LPC is the second most relevant variable to fill LNC. Indeed this is an issue for our analyses and therefore will revise the results considering only two traits: LPC and SLA. These were chosen because of these dependencies, so SLA was chosen instead of LDMC and LPC was chosen because of the dependency to LNP and because phosphorous is a limiting factor to most tropical forests. Yet, the VIF test did report much larger values and met the criteria to not consider these variables so multicollinear that they would explain the same range of variability thus making our model overfit. Given these two potentially contradictory criteria, we opted to remove the analyses of LNC and LDMC to make our analyses more robust based on the rational explained above. We have revised the analysis while including only SLA and LPC (Figure 1). We find that the influence of traits on the water cycle parameters was lower, i.e. weaker associations (lower R2 values). However, the direction of these associations remains consistent with was reported in our previous manuscript version, enhancing the credibility of the results.

For the revision of the manuscript, we will rewrite the subsection related to the multivariate linear regression, having only SLA and LPC as independent variables. Further, and in relation to the comment numbered RC2.2 by RC2, we will also examine non-linear relationships in the new version of the manuscript.

RC1.2. *Additionally, components of the enhanced vegetation index EVI, which is essentially an improved version of NDVI, is highly relevant for prediction of SLA, LDMC, LNC, and LPC (see Tabs. 2, A.1, C.1, D.1 in Moreno-Martinez et al. (2018)). These are clear violations of the statistical assumption of independence. However, this does not necessarily disqualify the validity of the analysis.*

Response: We agree with this point and thank you for pointing out our oversight. In the revised version of the manuscript we removed these analyses, i.e. only focus on the relationships by LPC and SLA with water cycle parameters.

RC1.3. *Circularity and the generation of trait data used as independent variables.*

In this manuscript, the authors ask whether trait values correlate with climatic variables. They clearly do; climate data was used to generate them. The authors use trait data derived from statistical (machine learning) models, including climate data. Table C.1 and D.1 in Moreno Martinez et al. (2018) list the most influential. These include temperature and precipitation related climatological variables. To generate the trait data used, Moreno-Martinez et al. (2018) collated TRY plant trait data, satellite-derived vegetation index data, and climatological variable data and used a machine learning framework (see Tabs. 2, A.1, C.1, D.1 in Moreno-Martinez et al. (2018) to generate global predictions. The authors of this manuscript are essentially taking trait data derived using, in part, climate data, turning it around, and then asking whether this trait data correlates to (somewhat different?) climate data. The correlations found are not surprising, given the existing relationships between variables. How do the authors justify the validity of their approach where they use data derived from climate data to then predict climate data? How strongly do the temperature data used by Moreno-Martinez to derive trait predictions correlate with the LST data you are using?

Response: Thank you for questioning the validity of the approach due to the potential circularity in the use of climate data. Although LST was listed as an input variable for the downscaling of the MODIS cover type product (in Table 2 in (Moreno-Martínez et al., 2018)), it does not contribute much to the mapping of the traits (Table D1 in (Moreno-Martínez et al., 2018)). The most influential climatological variables used for trait prediction do not intersect with the ones that we used in our study. Notably, the influential variables identified by (Moreno-Martínez et al., 2018) predominantly relate to precipitation (BIO12-17 which correspond to Annual precipitation, Precipitation of driest month, Precipitation seasonality, Precipitation of wettest quarter, Precipitation of driest quarter in Table A.1), with temperature-related variables such as maximum annual temperature (BIO1), isothermality (BIO3) and Mean temperature of warmest quarter (BIO10) playing a lesser role. As such we think that the potential circularity is not fully justified.

RC1.4. *Relationships between independent (LAI, NDVI) and dependent (ET) variables.*

Before examining potential correlations between variables, examining the methods used to generate these respective variables is advisable. The methodology described by Mu et al. (2013) (https://modis-land.gsfc.nasa.gov/pdf/MOD16ATBD.pdf) to derive ET already includes LAI (see Fig. 2). The ET calculation also includes the fraction of photosynthetically active radiation (FPAR, Eq. 3 in Mu et al. (2013)) which is highly correlated with NDVI (see Myneni et al. (2002)). Other equations, e.g. for the soil heat flux, include NDVI (Eq. 12 in Mu et al. (2013). No statistical analyses are necessary to examine correlations between ET, PET, LAI, and NDVI. If one wishes to reproduce these values of ET and PET perfectly, simply using the LAI, NDVI, FPAR, Biome classification, climate data, etc., used by these authors and the methodology described in their publications is all that is required.

Response: We greatly appreciate the comments from the reviewer and acknowledge the oversight on our part. We have removed these analyses from the manuscript.

RC1.5. *4. Technical Corrections*

Comment: Line 128 and other instances: The authors associate SLA with larger leaves. SLA and leaf size are different traits. Myneni, R. B.; Hoffman, S.; Knyazikhin, Y.; Privette, J. L.; Glassy, J.; Tian, Y.; Wang, Y.; Song, X.; Zhang, Y.; Smith, G. R.; Lotsch, A.; Friedl, M.; Morisette, J. T.; Votava, P .; Nemani, R. R.; and Running, S. W., "Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data" (2002). NASA Publications. 39

Response: We apologise for the confusion. We understand that SLA and leaf area are different traits and that we may have communicated that in a misleading way. In our updated manuscript, we will correct the text where needed to ensure to accurately reflect this difference between the two traits.

References

Moreno-Martínez, , Camps-Valls, G., Kattge, J., Robinson, N., Reichstein, M., van Bodegom, P., Kramer, K., Cornelissen, J. H. C., Reich, P., Bahn, M., Niinemets, , Peñuelas, J., Craine, J. M., Cerabolini, B. E., Minden, V., Laughlin, D. C., Sack, L., Allred, B., Baraloto, C., Byun, C., Soudzilovskaia, N. A., & Running, S. W. (2018, December). A methodology to derive global maps of leaf traits using remote sensing and climate data. *Remote Sensing of Environment*, *218*, 69–88. (Publisher: Elsevier Inc.) doi: 10.1016/j.rse.2018.09.006