Review - Influence of plant traits on water cycle processes in the Amazon Basin

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.

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. 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.

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.

The dependent variables considered were:

a) Evapotranspiration (ET)
b) Potential evapotranspiration (PET)
c) Vapour pressure deficit (VPD)
d) Land surface temperature during the day and night (LST)
e) Soil moisture (SM)

The independent variables considered were:

1. Specific leaf area (SLA)
2. Leaf dry matter content (LDMC)
3. Leaf phosphorus content (LPC)
4. Leaf nitrogen content (LNC)
5. Normalised difference vegetation index (NDVI)
6. Leaf area index (LAI)

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. 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.

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?

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.

Technical Corrections

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). N ASA Publications. 39.