EGUsphere, referee comment RC2
https://doi.org/10.5194/egusphere-2022-580-RC2, 2022

Comment on egusphere-2022-580

Anonymous Referee #2

Referee comment on "What determines peat swamp vegetation type in the Central Congo Basin?" by Selena Georgiou et al., EGUsphere,
https://doi.org/10.5194/egusphere-2022-580-RC2, 2022

General comments:

The manuscript by Georgiou et al addresses the question what controls the distribution of vegetation types in the Congo Basin peat swamp. The peat swamp has only been recently mapped based on field and remote sensing data. I found it very interesting to read the manuscript and think about what the distribution of hardwood trees and palms could tell us about hydrological processes and climatological boundary conditions. The topic is highly relevant given the threats the Congo peatland is currently facing due to regional anthropogenic alterations of the carbon and water cycle and global climate change. However, I do see two fundamental problems in the methodological approach of the manuscript. Addressing those problems may fundamentally change the findings of the manuscript.

(1) Ground truth data

The manuscript uses as ground truth data the mapping product of Crezee et al. 2022. The supplementary Figure 1 of the paper by Crezee et al. shows the nine remote-sensing products that were used to map peat-associated vegetation, i.e. the ground truth data used here in the work of Georgiou et al. Three of the nine input variables were based on elevation data. The fact that detailed elevation data was already used in the generation of the ground truth data conceptually prohibits that in Georgiou et al. elevation data is again used to build a regression model. In Georgiou et al., it is found that peat swamp vegetation is mainly a function of elevation. Knowing that the ground truth data was already created with elevation data makes this a trivial finding. Any discussion in Georgiou et al. on the influence of elevation-based variables is far-fetched given this fundamental problem of the ground truth data. To analyze the influence of elevation, the authors would need to work with ground truth data that is e.g. solely based on optical and microwave satellite signatures, but not on elevation.

(2) Division into sub-basins and random cross validation

The distribution of hardwood trees and palm shows patterns with clear spatial autocorrelation structure. The authors ignored this structure in their 'random' cross-validation approach at sub-basin scale, and thus seriously underestimated predictive error and likely have built overfitted models with non-causal predictors. For details I refer to the highly cited methodological paper of Roberts et al. 2017 on data structure and cross validation (see below). The derived models at sub-basin scale that use, apart from elevation, many different types of climatological-based variables are therefore highly questionable. The authors would need to show that the proposed climatological variables are reliable in a stratified cross-validation that acknowledges the spatial auto-correlation of the data. I believe that this would require an aggregation of sub-basins into larger regions. Perhaps one model for RoC and one for DRC in which one perhaps e.g. stratify the cross-validation by sub-basins (= not building a model for each sub-basin but building a model for four sub-basins and cross-validate against the fifth). Only variables that survive as reliable predictors in such a stratified cross-validation could be used as basis for an interpretation of optimal vegetation conditions

Detailed comments:

Line 35:

Harmonize use of Pg C and Gt C in the paper.

Line 98:

A useful variable might be the 'topographic wetness index' that combines subbasin area and local slope to estimate ground- and surface water impacts on soil wetness (e.g. Kopecky et al. 2021).

Line 163-164:

Sentence unclear

Line 210:

It's not 'train-test' since “test” data needs to be independent. With a random sampling, test data points are spatially auto-correlated with training points, thus they are not independent.

Line 261:

Also for RoC sub-basins not all show a positive correlation b/w palm fraction and annual rainfall (Roc5 show negative correlation)

Figure 6:

spatial variation of precipiation in RoC is only 100 mm, ~ 6-7%. In this example, it's quite likely that this trend will prove unreliable in a stratified cross validation.

Line 455:

Is there any physiological indication why palms should be less able to tolerate wetness than hardwood trees? Based on the methodological problems of the study, I found the discussion on the optimal water amounts for palms based on the negative correlation of palms with rainfall far-fetched.

Roberts et al. 2017. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography 40: 913–929. doi: 10.1111/ecog.02881

Kopecky et al. 2021. Topographic Wetness Index calculation guidelines based on measured soil moisture and plant species composition. Science of the Total Environment. doi.org/10.1016/j.scitotenv.2020.143785