Interactive comment on "Predicting the spatial distribution of soil organic carbon stock in Swedish forests using remotely sensed and site-specific variables" *by* Kpade O. L. Hounkpatin et al.

Anonymous Referee #1

Received and published: 13 January 2021

The manuscript attempted to compare the model performance of global and local models in predicting humus layer, mineral soil and total SOC stock and to identify the controlling factors for SOC stock prediction. Besides, this study also investigated the effect different combinations of data from site characteristics and remotely sensed variables on model performance. The results from independent dataset indicated that the local models generally had better model performance than the global models. The only use of remotely sensed variables had limited predictive ability while site characteristics had better explanatory strength in estimating SOC stocks. The authors suggest that further

work can focus on mapping these influential site covariates. The manuscript is overall well-written with clear objectives and reasonable methodology. However, two major limitation of this manuscript are: (1) in comparing global and local models, I can tell from Table 4 and Table 5 that some local models had higher R2 than that of global models while the remaining local ones had lower R2. Therefore, the conclusion that local models have a comparative advantage over global models is not convincing for me. Again, it is not fair to compare the performance indicators as mentioned in this study for local and global models. Instead, for the global model, authors should also calculate indicators for three regions separately to make them comparable for these in local models; (2) As mentioned by authors, the site characteristics are only available at the visited plots, therefore the digital maps can not be produced by the models built with site characteristics, which certainly limits the usage of site characteristics. I am afraid that the importance of these site characteristics in this study have been overlooked as the observed site characteristics are directly used in independent validation which certainly ignores the inaccuracy of these site characteristics if they are mapped by certain algorithms. I mean, the site characteristics used in a fair independent validation should come from the predicted maps of these relevant site characteristics, not from observed data. Therefore, I suggest a major revision before it can be published. Specific comments are listed below: Line 49: which are nonspatial environmental covariates? If they are not spatial, how they can be used in DSM? Line 58: Authors miss at least two papers on comparing local and global models in DSM. Piikki, K., & Söderström, M. (2019). Digital soil mapping of arable land in Sweden–Validation of performance at multiple scales. Geoderma, 352, 342-350. Song, X. D., Wu, H. Y., Ju, B., Liu, F., Yang, F., Li, D. C., ... & Zhang, G. L. (2020). Pedoclimatic zone-based three-dimensional soil organic carbon mapping in China. Geoderma, 363, 114145. Line 66: SCORPAN also includes soil information compared to Jenny's soil forming theory. Lines 89-90: It seems to me that this manuscript does not take the SOC changes during three periods into account. I would add relevant statements with supporting references. Lines 103-104: Since mineral soil is sampled at 0-10, 10-20 and 55-65 cm depth, which kind

of interpolation is used to harmonize them at 0-50 cm? Please provide more details. Line 158: RF is commonly used instead of RFR. Line 216: Which method is used in recursive feature elimination in caret package? Random forest? Line 247: What does p mentioned here?

Interactive comment

Printer-friendly version

Discussion paper