SOIL Discuss., referee comment RC2
https://doi.org/10.5194/soil-2021-107-RC2, 2022

Comment on soil-2021-107

Anonymous Referee #2

Referee comment on "Spatial prediction of organic carbon in German agricultural topsoil using machine learning algorithms" by Ali Sakhaee et al., SOIL Discuss., https://doi.org/10.5194/soil-2021-107-RC2, 2022

In general, a well-written paper that presents soil organic carbon modelling at a national scale (Germany). The author focus on three aspects, namely the comparison of three machine learning models, expanding the national dataset with samples from a continental scale survey (LUCAS dataset) and how generating two separate models for mineral and organic soils affects the performance of such models.

Since this is a national scale digital soil mapping (DSM) study, I think that a major revision is required.

General comments

- I have a problem with the way maps are presented. As far as I understand, the paper is a digital soil mapping (DSM) study but I do not see any maps with continuous predictions but just some points on a map. Or are those the areas corresponding to croplands? Please clarify. Second, you use a discrete colour map to show the results which do not allow the reader to see the spatial pattern of the predicted maps. You discuss the distribution of the residuals but a more detailed visual inspection of maps could be useful (which is common in DSM). For instance, Boosted Regression Tree (BRT) seems to mostly use categorical covariates (except for total nitrogen). How does that map look like?

- The largest difference can be seen when you split the dataset in mineral/organic. There is no doubt that the difference is significant. What about the rest of the comparisons? You use a Kruskal-Wallis to show that extrapolation in depth of the LUCAS dataset is valid but it is not clear if the main comparison (between three models according to the title) is significant.

- Perhaps the paper is focussing too much on the differences between models which is not very interesting. We have seen hundreds of papers comparing different models just to confirm that the "best model" depends on many factors. However, your results on modelling mineral and organic models separately seem interesting and perhaps focussing on that could benefit the community and the readers.

- How do you actually use two separate models (mineral/organic) in practice? In this approach, to make a SOC prediction you first need to decide which model to use. But to make that decision, you need to know the SOC concentration. This is an important point that should be discussed. For instance, how do we generate a national map in this particular study? Is your potential solution applicable to other countries?

- I think a bit more discussion about the covariates could be useful. Many of the soil covariates used correspond to continental scale predictions (with significant uncertainty) which usually perform poorly at other scales (national). In addition to that, is interesting to see how just a few covariates are actually used by the models. Are we using too many useless covariates in DSM (studies with dozens of covariates)?

Specific comments

- Section 2.6.1: I think the way parameter tuning is described is not correct. First, you mention that grid search parameters need to be discrete or discretised, which is not true. You can use continuous parameters without problem (e.g. [1.0, 0.1, 0.01]). Second, you used a DE algorithm for BRT since the parameters are continuous but `number of trees` and `interaction depth` are discrete. Based on your criteria, you couldn't use any of the strategies for BRT. A clarification is required.

- Figure 2: The limits of the whiskers and boxes sometimes represent different things depending on the library. Please add what they represent in the caption.