SOIL Discuss.,
https://doi.org/10.5194/soil-2017-30-AC4, 2018

Interactive comment on "Hot regions of labile and stable soil organic carbon in Germany – Spatial variability and driving factors" *by* Cora Vos et al.

Cora Vos et al.

axel.don@thuenen.de

Received and published: 1 March 2018

Dear Dr. Viscarra Rossel,

Thank you for your short comment regarding our manuscript. We very much appreciate your input that helps to improve our paper and to make it more clear and easy to read. Please find our answers to your suggestions below.

I thank the authors for their paper and I hope that my discussion helps. My comments here relate primarily to the lack of clarity in the description of the methods used for the spectroscopic modelling, and to missing quantification of robustness and uncertainty in the spectroscopic model predictions of the carbon fractions. I believe these to be crucially important because their further analyses and interpretation of the variability

Printer-friendly version

Discussion paper

and driving factors relies heavily on the spectroscopic model predictions. First, the description of the spectroscopic modelling is inadequate and I encourage the authors to improve it. I think that the specifics of the spectroscopic modelling, apparently described in Jaconi et al., need to be included in this manuscript, particularly because the Jaconi et al. manuscript isn't yet published. But, even if the Jaconi et al paper were published, I think that at the very least, readers will need a clear summary of their methods and findings–not simply a report of their assessment statistics.

Answer: We agree that the reader needs more information on the spectroscopic modelling and as we are not sure when the review process for the paper of Jaconi et al. will be finished, we will include a more detailed description in the methods section of the revised version.

Second, the authors do not convincingly show that the spectroscopic models were sufficiently robust for predicting the 'unknowns', which I presume were the '...>2500 sites with mineral soil all over Germany' (mentioned only in the Introduction, line 106). Additional validation of the models with an independent test set will help, however, I would also encourage the authors to implement either a repeated cross validation, or to bootstrap the models to quantify their robustness and the uncertainty of their predictions (see for instance Viscarra Rossel, 2007). To this end, the authors might find it useful to read Viscarra Rossel & Hicks (2015). There, we proposed an approach for modelling the carbon fractions of a large continental scale dataset, reporting the robustness of the models, the (propagated) uncertainties of the predictions, and relating the spectroscopy to the chemistry of soil organic C.

Answer: As described in our answer to the comment of L. Cécillon, the models have been validated using an independent test set and the results will be included in the revised version of the manuscript. Both datasets, the calibration and the validation data set cover the area of interest (Germany). We will check the recommended papers for the options to further quantify the model uncertainty. However, with an independent validation dataset we already quantified the model uncertainty.

Quantifying uncertainty is particularly important when predicting 'unknown' samples. Without quantified uncertainty, the predictions will definitely be less valuable. This is particularly relevant for this study because the predictions are being used in subsequent analysis to potentially gain new understanding.

Answer: We agree that the quantification of uncertainty is crucial for gaining trust in the predicted values. Therefore, we propose to include a summary of the calibration and validation results in the supplement material of the revised version.

Finally, I would like to suggest some minor corrections:

- In lines 182–183, the Jaconi et al reference is cited as 'in prep' while in line 194 it is cited as 'submitted'

Answer: Thank you for noticing this mistake. We will change this in the revised version of the manuscript. However, we hope to get this paper to be published soon.

-The mention of the'...>2500 sites with mineral soil all over Germany.', in the Introduction, line 106, is inadequate. This should be described and made clear in the Methods section–possibly in section 2.4 after a (better) description of the spectroscopic modelling.

Answer: We agree that the methods need to be clear. However, there is a section on the soil inventory (2.1) and we will add more in the spectroscopic method section. In this case we do not agree with the comment, as it is good practice to give a very short overview in the introduction on how the research questions shall be answered. Of course the number of sites should also be stated in the methods section, which is the case.

- In lines 185–187: '... In addition, residual prediction deviation (RPD) was calculated, using the classification system devised by Viscarra Rossel et al. (2006)....' – I am quite sure that Viscarra Rossel et al. (2006) did not devise a classification for the RPD. Williams (1987) originally devised the RPD for assessing spectroscopic calibrations

of agricultural and food products. Later, Chang et al. (2001) suggested an arbitrary classification specifically for soil. It is very likely that Viscarra Rossel et al. (2006) simply used that classification, but I could not confirm one way or the other because the Viscarra Rossel et al. (2006) reference is not listed in the references.

Answer: Thank you for this clarification. We will revise this and change it to Chang et al. mentioning that the classification is arbitrary but can serve as indicator for the model quality.

- In terms of the RPD, Bellon-Maurel et al. (2010) suggested that the RPD should only be used if the data is normally distributed, otherwise, they propose the use of the RPIQ (Bellon-Maurel et al., 2010).

Answer: We will also include the RPIQ in the revised version.

- Following from that, in our spectroscopic modelling of soil carbon and fractions (Viscarra Rossel & Hicks, 2015), we found that their statistical distributions were often not normal and required logarithmic transformations. For this reason, it would be useful for the authors to report the distributions of the carbon and fractions data–but also because the PLSR algorithm assumes normally distributed data.

Answer: We agree with this and we log-transformed the data for model development. We will add information on this in a revised manuscript version.

Interactive comment on SOIL Discuss., https://doi.org/10.5194/soil-2017-30, 2017.