SOIL Discuss.,
https://doi.org/10.5194/soil-2017-30-SC3, 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.

Dr Viscarra Rossel

raphael.viscarra-rossel@csiro.au

Received and published: 16 February 2018

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 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.

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.

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.

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'

- 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.

- 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.

- 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).

- 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.

I hope that this helps.

Raphael VISCARRA ROSSEL

References cited:

Bellon-Maurel, V., Fernandez-Ahumada, E., Palagos, P., Roger, J-M., McBratney, A.B., 2010. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. TrAC Trends in Analytical Chemistry 29, 1073-1081.

Chang, C.-W., Laird, D.A., Mausbach, M.J., Hurburgh C.R., 2001. Near-infrared reflectance spectroscopy - Principal components regression analyses of soil properties. Soil Science Society of America Journal 65, 480-490.

Viscarra Rossel, RA. 2007. Robust modelling of soil diffuse reflectance spectra by

bagging-partial least squares regression. Journal of Near Infrared Spectroscopy 15: 39–47.

Viscarra Rossel RA, Hicks WS. 2015. Soil organic carbon and its fractions estimated by visible–near infrared transfer functions. European Journal of Soil Science, 66(3): 438–450.

Williams, P.C. 1987. Variables affecting near-infrared reflectance spectroscopic analysis. In: Near- Infrared Technology in the Agricultural and Food Industries (eds P. Williams & K. Norris), pp. 143– 167. American Association of Cereal Chemists Inc., Saint Paul, MN.

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