Response with revised manuscript: "Developing the Swiss mid-infrared soil spectral library for local estimation and monitoring" by Baumann et al.

https://soil.copernicus.org/preprints/soil-2020-105/

Letter of response

Dear Prof. van Wesemael,

We appreciate the comments that we have received from the reviewers. Thank for your kind editorial support. We have carefully revised our manuscript and have incorporated all the relevant suggestions into the revisions.

Please find in our response below the outline of the more substantial changes and edits that we have made in this round of review. Also note that small and simple changes that we could implement in straight-forward manner are highlighted in the track changed version that we have uploaded.

In our study, we have developed a realistic performance-driven transfer scenario to scale a national-level mid-infrared soil spectral library for monitoring soils at plot-level and over time. In that sense, our approach is novel and the revised manuscript is therefore more conclusive. We are sincerely looking forward that our work is soon accepted for publication in SOIL. Thereby, we are hoping that our effort is a valuable contribution to further develop local approaches with focus to use large and diverse soil spectral libraries for systematic soil monitoring.

We again thank the editor and the reviewers for the valuable time invested for feedback, which has helped to improve our manuscript substantially.

On behalf of all co-authors,
Philipp Baumann

Main edits

Improvements made throughout the manuscript

We have decided to remove the results of total organic matter entirely because is is a property calculated from organic C content using the fixed conversion factor, 1.72. We agree that the simple conversion of organic C to estimate organic matter is questionable and often inaccurate mainly because of variable functional composition of soil organic carbon compounds. Further, such linear transformation would not bring any complementary information in the model.

Both reviewers have indicated that the ratio of performance to interquartile distance (RPIQ) should be also reported besides RMSE to better account for data distributions when doing relative model comparisons. We agree with this comment and have accordingly edited the manuscript and added RPIQ in all relevant parts of the results, discussion, and conclusion.

We now consistently report all results in the International System of Units.

Title

We have taken the suggestion from reviewer 2 and change the title as follows: *"Developing the Swiss mid-infrared soil spectral library for local estimation and monitoring"*

Abstract

Page 1, lines 6–7: We have simplified the term "committees" with *"ensembles of rules (committees)"* and have specified *"nearest-neighbors"*.

Page 1, lines 14–16: The comparison between the general learning approach and the local transfer is now more succinct to make sure we do not compare algorithms but rather strategies: *"Compared to the general statistical learning approach, the local transfer approach — using two respective training samples on average — reduced the RMSE of total C per site fourfold."*

Introduction

Page 4, lines 18–20: We have changed these lines as follows to emphasize the main goal of local monitoring more clearly: "Specifically, we aimed for showing *local models' capacity to reproducing* time-series measurements (starting from 1985) of soil C at the Swiss agricultural long-term monitoring sites based on spectral analysesand two calibration samples per site"

Page 4, lines 25–26: We have followed the advice of reviewer two and recapitulate the objectives: *"In brief, our work addresses three objectives: (1) developing a national SSL, (2) building general prediction models using* CUBIST*, and (3) build site-specific (local) prediction models using* RS-LOCAL*. "*

Material and methods

Page 6, lines 10–19: We have completed the missing details on the chemical reference analyses done previously.

Page 6, 26–28: The sentence was changed to: "The reflectance spectra were acquired between 7500 cm^{-1} (1333.3 nm) and 600 cm^{-1} (16666.7 nm) at an effective resolution of 2 cm^{-1}, *and trimmed to the mid-IR range between 3996* cm^{-1} *and 600* cm^{-1} *before further processing (see below).*"

Page 7, line 3: We have changed it to: *"The resulting reflectance spectra (R; background referenced) were converted to apparent absorbance (A) by* $A = log_{10}(1/R)$*."*

Page 7, lines 18–22: We have added the required references: "We chose this algorithm since it has shown excellent performance for *modelling soil information* and developing SSLs with rather large soil variability and multicollinear spectral variables *(Bui et al.; Viscarra Rossel and Webster, 2012; Stevens et al., 2013; Miller et al.; Peng et al.; Viscarra Rossel et al., 2016; Dangal et al., 2019; Padarian et al., 2019)*, and because its interpretation is mechanistically more intuitive as it is a form of data partitioning (simple conditions and linear equations)."

Page 7, lines 25–33: We have drafted a short and concise paragraph that describes the general principles of CUBIST and introduces the its empirical parameters: "committees" and "neighbors".

Page 8, lines 7–8: We have resolved the confusion around the pseudo-random training and validation splits: *"The division into training and validation proportions of the data was done in consistent and repeatable manner (pseudo-random number generation)."*

Page 8, lines 25–26: We have clarified the variable importance method used: "We recursively eliminated subsets of variables with low CUBIST variable importance, *calculated as the average relative usage frequencies of a particular variable in split conditions and regressions*"

Page 10, lines 14–15, and page 11, line 1: We have addressed the missing model information: *"The* RS-LOCAL *procedure is based on partial least squares regression (PLSR) (Wold et al., 1983). For the* RS-LOCAL *tuning during the subset selection procedure and final calibrations, we tested 1 to 10 PLSR components."*

Section 2.6 Local soil information for plot-level monitoring: We have clarified the inconsistent terminology and variables for the different roles of the RS-LOCAL data.

Results

Page 12, lines 10–11; Table 1: We have added skewness to the summary of reference measurements. Table 1 has been updated with the correct SI units and values for measurements.

Figure 3. Table 3: We have changed the units and values for total C, OC, N, $CaCO_3$, clay, silt and sand. Further, we have added RPIQ to Table 3 and have renamed "bias" to *"ME"* in Figure 3 for consistency reasons. We also have mentioned the LOESS smoothing for indicating the trends in predictions in the caption of Figure 3.

Page 12, lines 26–29: We have fixed the wrong statement made previously: *"Figure 4 shows that the test RMSE of total C first slightly decreased and then steadily increased from all (209) to less spectral variables using* CUBIST *and RFE. The lowest error (*$\mathrm{RMSE}_{\mathrm{test}}$ *= 8.10* $\mathrm{g\,kg^{-1}}$ *total C) of spectroscopic estimation was achieved with the spectra with 105 variables. For the subsequent variable reduction steps, model performance steadily dropped until one wavenumber was left (*$\mathrm{RMSE}_{\mathrm{test}}$ *= 18.8* $\mathrm{g\,kg^{-1}}$ *total C)."*

Figure 4, page 15: We have fixed the units for total C in the panel showing results of the recursive feature elimination.

Page 18, lines 12–13: We have additionally compared the performance of the local transfer vs. the general rules approach on all validation data of the NABO data set: *" Across all validation data points of the NABO set, the* RS-LOCAL *transfer was 5.6 times more accurate for total C than the general rules in terms of RMSE and RPIQ (RMSE = 0.9*$\mathrm{g\,kg^{-1}}$ *C; RPIQ = 31.7)."*

Figure 6, page 19: The units and the plot axes were corrected to $\mathrm{g\,kg^{-1}}$ for total C.

Table 5, page 20: We have fixed the units and numbers.

Figure 7, page 21: We have corrected the units for total C and renamed the plot legend with the correct sample notations corresponding the information provided in the main text.

Discussion

Page 22, lines 17–30: We have written a new paragraph that address committees, neighbors, and the complexity in rules. We compare the latter results to the findings made by Viscarra Rossel et al., 2012.

Page 24: lines 32–24; page 25: lines 1–2: We have added more useful content: *"Although Gubler et al. (2019) reported only minor changes for the ensemble of permanent cropland or cropland-meadow monitoring sites (30), there were four sites with declining trends and nine sites with increasing trends in OC (*-11% *to* $+16\%$ *relative change per decade, respectively). Here, the trend of spectroscopic predictions could be investigated with respect to specific research questions on agronomic management-induced changes, further physicochemical soil characterization (e.g., OC fractions)."*

Page 26, lines 13–21: We have edited specific findings made by Helfenstein et al. (2021), that have further addressed the representation of organic soils in the current SSL by doing a regional transfer of this SSL for organic soils in two new ecoregions in Switzerland.

Appendix A

Figure A1, page 29:: We have corrected the unit for total C to $\mathrm{g\,kg^{-1}}$ and transformed the RMSE values accordingly.

References

Bui, E. N., Henderson, B. L., and Viergever, K.: Knowledge Discovery from Models of Soil Properties Developed through Data Mining, 191, 431–446, https://doi.org/10.1016/j.ecolmodel.2005.05.021.

Dangal, S. R. S., Sanderman, J., Wills, S., and Ramirez-Lopez, L.: Accurate and Precise Prediction of Soil Properties from a Large Mid-Infrared Spectral Library, Soil Systems, 3, 11, https://doi.org/10.3390/soilsystems3010011, https://www.mdpi.com/2571-8789/3/1/11, 2019.

Gubler, A., Wächter, D., Schwab, P., Müller, M., and Keller, A.: Twenty-five years of observations of soil organic carbon in Swiss croplands showing stability overall but with some divergent trends, Environmental Monitoring and Assessment, 191, 277, https://doi.org/10.1007/s10661-019-7435-y, 2019.

Miller, B. A., Koszinski, S., Wehrhan, M., and Sommer, M.: Comparison of Spatial Association Approaches for Landscape Mapping of Soil Organic Carbon Stocks, 1, 217–233, https://doi.org/10.5194/soil-1-217-2015.

Padarian, J., Minasny, B., and McBratney, A. B.: Using Deep Learning to Predict Soil Properties from Regional Spectral Data, Geoderma Regional, 16, e00 198, https://doi.org/10.1016/j.geodrs.2018.e00198, http://www.sciencedirect.com/science/article/pii/S2352009418302785, 2019.

Peng, Y., Xiong, X., Adhikari, K., Knadel, M., Grunwald, S., and Greve, M. H.: Modeling Soil Organic Carbon at Regional Scale by Combining Multi-Spectral Images with Laboratory Spectra, 10, e0142 295, https://doi.org/10.1371/journal.pone.0142295.

Stevens, A., Nocita, M., Tóth, G., Montanarella, L., and van Wesemael, B.: Prediction of Soil Organic Carbon at the European Scale by Visible and Near InfraRed Reflectance Spectroscopy, PLoS ONE, 8, e66 409, https://doi.org/10.1371/journal.pone.0066409, 2013.

Viscarra Rossel, R., Behrens, T., Ben-Dor, E., Brown, D., Demattê, J., Shepherd, K., Shi, Z., Stenberg, B., Stevens, A., Adamchuk, V., Aïchi, H., Barthès, B., Bartholomeus, H., Bayer, A., Bernoux, M., Böttcher, K., Brodský, L., Du, C., Chappell, A., Fouad, Y., Genot, V., Gomez, C., Grunwald, S., Gubler, A., Guerrero, C., Hedley, C., Knadel, M., Morrás, H., Nocita, M., Ramirez-Lopez, L., Roudier, P., Campos, E. R., Sanborn, P., Sellitto, V., Sudduth, K., Rawlins, B., Walter, C., Winowiecki, L., Hong, S., and Ji, W.: A Global Spectral Library to Characterize the World's Soil, Earth-Science Reviews, 155, 198–230, https://doi.org/10.1016/j.earscirev.2016.01.012, 2016.

Viscarra Rossel, R. A. and Webster, R.: Predicting Soil Properties from the Australian Soil Visible-near Infrared Spectroscopic Database, European Journal of Soil Science, 63, 848–860, https://doi.org/10.1111/j.1365-2389.2012.01495.x, 2012.

Wold, S., Martens, H., and Wold, H.: The Multivariate Calibration Problem in Chemistry Solved by the PLS Method, in: Matrix Pencils, edited by Kågström, B. and Ruhe, A., vol. 973, pp. 286–293, Springer Berlin Heidelberg, https://doi.org/10.1007/BFb0062108, http://link.springer.com/10.1007/BFb0062108, 1983.