Reply to Second Review from 29.12.2023 (minor revisions)

16.01.2024

1 General answer

We thank you again for your careful evaluation of the manuscript and very constructive comments. You raised three general comments and 12 additional specific comments. For more clarity between our answers and your comments, we have inserted your original statement in blue color and gave a point-by-point reply.

We hope we addressed all comments to your satisfaction and thank you for your suggestions that helped us to further develop the manuscript.

On behalf of all co-authors,

Simon Oberholzer

General comment 1: The analysis how model performance relates to soil properties is not replicable and unreliable

We thank you for this comment. You raised the important point that we should apply some statistical analysis to make the analysis between model performance and soil properties more reliable and replicable. So far, we relied on a visual inspection because (as you also mentioned) the sample size (six fields) is for most statistical tests too small to formulate or even reject a null hypothesis. However, besides statistical tests you mentioned the option to report correlation coefficients between soil properties and model performance. As you suggested we did linear correlations with a Monte Carlo simulation (1000 iterations) to account for the uncertainty of the RMSE and reported in Figure 6 the found Pearson's correlation coefficient with standard deviation.

General comment 2: The study should emphasize results for the RMSE in the analysis of how model performance relates to soil properties

We agree that RMSE is for real practitioners probably the most interesting parameter. Therefore, we show now RMSE in dependence of soil characteristics for all five properties and not only for SOC in the main text. Moreover, we extend the discussion on the RMSE. Still, we are convinced that we still need R2 and RPD as one of our main discussion points, since we rely on them to compare the performances on the different fields, relatively to their soil property ranges (e.g. a pH model for a field with small pH values and/or narrow pH ranges will also lead to low RMSE, whereas the RPD relates the standard deviation of the pH values to the RMSE).

General comment 3: The Interpretation of bad predictive accuracy for the models for field A and F should be clarified

Thank you very much for this feedback. Indeed, field A is quite different from field F and so far, these differences have not been clearly stated in the manuscript. You suggested that for field F the lower model performance might have been provoked by the high carbonate content in all samples whereas for field A only a few samples with high carbonate content might have decreased model performance. As you suggested we calculated a new model for field A where we removed the samples with high carbonate content. We tried two thresholds (> 15 g kg^{-1} and > 10 g kg^{-1} inorganic C) but the new models showed very similar performance to the original model, and we decided therefore not report these reduced models in the discussion but instead stress the differences between field A and F more clearly. Similarly, we underlined that there are many more field and site characteristics, that can affect model performance, than only the ones we analyzed. We guess that the lower performance of field A might be caused by an additional factor, we do not have information about.

2 Specific comments

Comment 1:

72 to 73: "Do field and soil characteristics (e. g. field size, soil texture, carbonate content, correlations of soil properties) of the target site influence the performance of spectral models?"

• I would suggest to replace this by "How do field and soil characteristics (e. g. field size, soil texture, carbonate content, correlations of soil properties) of the target site relate to the performance of spectral models?" or split it into two parts, for example: "How does carbonate content influence the performance of spectral models?" (first question) and "How do other field and soil characteristics (e. g. field size, soil texture, correlations of soil properties) of the target site relate to the performance of spectral models?" (second question). The current wording implies that your analysis could show that there is a causal relation between specific properties of the target samples (or site), but since the study currently is observational, this is not the case. In some cases, one clearly can argue that specific soil properties have a causal influence on the predictive performance of the models, e.g. carbonate content. But your study does not show this, but uses information from previous studies on peaks caused by molecular structures in carbonates and results from your own analysis to elaborate how exactly carbonate content controls model performance (a causal factor which is already known/suggested in previous studies) and corelation between spectral variables and the target variables. For other properties, I am sceptical if there is any theory for how they should be causal (or how one should infer and define such causal effects based on observational data and available prior knowledge). For example, field size may be related to variability in soil properties and thus spectral variability and variability in the relation between target variables and spectral variables for local fields. However, it is not hard to imagine that besides spatial variability, factors such as sampling design are important, too, and your analysis does not consider this in detail. For these reasons, I would either remove causal wording or split the question into two parts, one where your study can elaborate causal relations (because we already know that they exist and roughly how they work), and one where exploratory analyses are interesting, but your current analysis does not provide causal information.

Thank you very much for this suggestion that makes the structure of the manuscript more concise. We decided to choose your first suggestion.

Comment 2:

223: “However, the RMSE of the local models for pH of fields A (0.08 ± 0.02) …”. This is the first occurrence of a mean ± error in the text and I suggest to state here, what this means, e.g. “However, the RMSE of the local models for pH of fields A (0.08 ± 0.02) (mean ± standard deviation) …”.

Thank your for this clarification. We changed the text as you suggested.

Comment 3:

295 to 296: “However, in absolute prediction performance (RMSE) we only found a clear effect for SOC (Fig. 6) and not for the other properties (Fig. S3 in the supplementary material).”

• Since you used a non-replicable visual identification procedure, I cannot follow this conclusion. I suspect that some additional relations are “clear” when using a replicable identification procedure and even when considering uncertainty in the RMSE (e.g., with high probability, the Pearson correlation, even considering the uncertainty in RMSE, is larger than 0). I suspect this may be the case for example for the pH value.

Now we calculated the correlation coefficients and can refer to them. We also address now the strong correlation between RMSE and carbonate content for pH which is basically caused by the very low variability in pH on fields with high carbonate content and therefore low RMSE.

Comment 4:

297 to 298: “We did not observe an influence of field size absolute contents of sand, silt and clay or variability of carbonate content on model performance (see Fig. S4 in the supplementary material).”

• Again, it needs to be clarified how exactly (and reproducibly) the absence of an effect was identified.

For RPD and R^2 we did not calculate correlation coefficients because they are not linear and with the little samples size, we do not want to assume a specific relationship. We changed the sentence and talk now that the three characteristics we chose, showed graphically most impact.

Comment 5:

279 to 281: “It can clearly be seen that on field B and to a lower extent on field F, the same wavelengths were important in all soil properties related to soil organic matter (SOC, 280 total C, total N and POXC) …”. I would suggest to remove “clearly”.

We removed “clearly” as you suggested.

Comment 6:

366 to 368: “Looking at the correlation between spectral variables and inorganic C respectively SOC (Fig. 7) we can confirm this finding but have to add that on the local scale the absorption bands for carbonate and SOC varied substantially between different datasets.”

• Absorption bands (peaks) are caused at specific energy levels of the near infrared radiation because molecular bonds interact with the infrared radiation at specific energy levels. The position of specific absorption bands for carbonate and SOC is thus fixed. Do you mean here that the relative intensity of “the absorption bands for carbonate and SOC varied substantially between different datasets”?

Exactly. We enlarged the expression “absorption bands” to “relative intensity of absorption bands” as you suggested.

Comment 7:

374 to 376: “The higher lab measurement error with higher carbonate content can explain the lower model performance on soils with high carbonate content for SOC but not for the other four soil properties where model performance (in terms of RPD) still tended to be lower than on fields with little carbonate content (Fig. 5).”

• Actually, you have not shown that “the higher lab measurement error with higher carbonate content can explain the lower model performance on soils with high carbonate content for SOC …” (my emphasis). One way to analyze this would be to simulate for datasets with low carbonate content, but similar range in SOC content, SOC content values from the measured SOC contents and the lab measurement error if the samples had a high carbonate content. Please note that I do not say you should conduct such an analysis, I just describe how one could analyze whether the lab measurement error for SOC content could have caused bad model performance for SOC content to provide constructive criticism, and state that without such an analysis, it is unclear whether “the higher lab measurement error with higher carbonate content can explain the lower model performance on soils with high carbonate content for SOC …”.

We changed the term “can explain” to “might be a possible explanation”, to make clear that now statistical analysis was conducted.

Comment 8:

389: “… makes it more difficult to attribute absorption features …”. I suggest to replace “attribute” by “relate”.

Thank you. Your suggestion is more concise, so we replaced “attribute” with “relate”.

Comment 9:

Tab. 2: Why do the R2 values have no standard deviation for SOC and pH for field A?

Thank you for carefully reading the manuscript. We corrected that mistake and added the standard deviation.

Comment 10:

Fig. 3: What do error bars for model RMSE represent?

We added the meaning of the error bars in the figure caption (“The error bars for RMSE of spectral models represent standard deviations across the repeats in the cross-validation”).

Comment 11:

Fig. 5: What do the filled circles and error bars represent in the figure? This does not seem to be explained in the legend or in the figure caption.

We added the explanation in the caption (“The error bars represent standard deviations across the repeats in the cross-validation”). As well we added this explanation in the caption of the figures in the supplementary material.

Comment 12:

Fig. S5, field D: Variable POXC appears to have some missing values which results in NAs for all Pearson correlation coefficients involving this variable. Based on the plot, I assume that these are only very NA observations and I assume the values are missing at random. Thus, selection effects in omitting these samples are unlikely, and I recommend to recompute these correlations without the NA values.

Thank you very much. Field D contains one NA for POXC that disturbed the calculation of correlation coefficients. We excluded this sample from the analysis for POXC and renewed the Figure S4.