Comments to “Best performances of visible-near infrared models in soils with little carbonate - a field study in Switzerland”

2023-12-29

Contents

1 General comments **1**

1.1 The analysis how model performance relates to soil properties is not replicable and unreliable 2

1.2 The study should emphasize results for the RMSE in the analysis of how model performance relates to soil properties 3

1.3 The Interpretation of bad predictive accuracy for the models for field A and F should be clarified 4

2 Specific comments **5**

References **7**

1 General comments

You have improved the manuscript considerably and addressed nearly all comments I gave to the first version of the manuscript. I would like to thank you for the detailed comments and explanations. I have some comments on the expanded analysis how model predictive performance relates to field and soil properties. Of these, the first section (The analysis of how model performance relates to soil properties is not replicable and unreliable) mentions points I think should definitely be addressed.

My comments in subsections “The study should emphasize results for the RMSE in the analysis of how model performance relates to soil properties” and “The Interpretation of bad predictive accuracy for the models for field A and F should be clarified” are — I hope useful — suggestions to clarify or expand the presentation of results and their discussion. Finally, in section “Specific comments” I list specific comments about the current version.

My main concern with the new analysis how model performance relates to soil properties is that it is not replicable and not reliable because it is not defined what a “clear visual influence on model performance” (l. 198 to 199) is, how it can be reliably identified, and because it is unclear how the uncertainty in the RMSE (respectively the R^2 or RPD) was considered in this analysis.

1.1 The analysis how model performance relates to soil properties is not replicable and unreliable

I think that the analysis how model performance relates to "field size, soil texture, carbonate content and with the correlation coefficients between SOC and total N in the dataset." (l. 198 to 199) is at the moment not replicable and unreliable, first because the reader is not told how to define a "clear visual influence on model performance" (l. 201 to 202), and second because it is unclear how the uncertainty in the RMSE (respectively the R^2 or RPD) was considered in this analysis. In particular:

1. l. 199 to 200: "With six local datasets as independent variables it is hardly possible to apply statistical tests and therefore, we relied on visible inspection."

   - This justification does not make much sense to me. First, it is possible to apply statistical tests even with few observations, even though the tests will probably show in at least some cases that more samples would be needed to analyze whether the relation is or is not compatible with some null hypothesis in a sensible way. This is exactly what good tests should do: tell us, when we are not allowed to draw conclusions because the uncertainties are too large.
     Second, a simple visual inspection is not convincing here: Why should we be licensed to give up estimating effect sizes and accept a visual inspection if sample sizes are small? Exactly in this case, formal procedures to estimate uncertainties are important to show the reader what may and what may not be concluded (and first of all how to arrive at a conclusion). Uncertainties may be large with only 6 fields, I agree, but we have to make a start somewhere.
     Please note that I do not imply that you should compute tests here, especially not for a null hypothesis of no effect, because I do not think this would make sense — no one would suggest that the variables you have analyzed are not at all related to model performance. I only suggest to remove this justification in favor of a visual inspection and to use a replicable procedure to analyze the data. Below, I make concrete suggestions for how to improve this analysis.

2. Regarding comment 14 of reviewer 1 (in your answers to reviewer 1, Oberholzer (2023)), you commented: "We did not consider the uncertainty in RMSE for this analysis, but we will do include it in our revised version (as already suggested above; see Concern 1)."

   - Currently, I can only see error bars for the RMSE in Fig. 6 and S3, however it is unclear how these uncertainties were considered in the analysis other than plotting error bars. This is also liked to the non-replicability of the visual analysis.

Instead, I suggest to estimate Pearson correlation coefficients or simple linear regression models of the RMSE (see also the next subsection "The study should emphasize results for the RMSE in the analysis of how model performance relates to soil properties") versus field size, soil texture, carbonate content and with the correlation coefficients between SOC and total N in the dataset. This analysis should consider the uncertainties in the RMSE (e.g., via Monte Carlo analysis) and report the average correlations or slope with uncertainties.

You may use tests or other decision criteria to decide which relations to interpret in the main text, but you should clearly define them in the text and show all results at least in the supporting information (as is currently also the case). This would give the reader more information about the uncertainties of this assessment, and about the effect of uncertainties in the estimated RMSE, and you would provide models one could test in the future (for example whether other studies find similar or higher correlations with other spectral measurement methods, etc.).

If you only provide a visual evaluation, you do not give opportunities for these comparisons and you do not show how big the influence of uncertainties in the RMSE are on the estimated relations and this makes the analysis unreliable.

1.2 The study should emphasize results for the RMSE in the analysis of how model performance relates to soil properties

Currently, when analyzing how model performance relates to "field size, soil texture, carbonate content and with the correlation coefficients between SOC and total N in the dataset." (l. 198 to 199), the manuscript de-emphasizes results for the RMSE at the expense of R^2 and RPD (for example, Fig. 5 only contains results for the relation of R^2 and RPD and results for the RMSE are shown fully only in the supporting information, Fig. S3). I suggest to emphasize results for the RMSE and not for R^2 and RPD because I think that the RMSE is the performance metric most interesting both for researchers and practitioners and because I do not agree with the justification you give for the de-emphaiss of results for the RMSE.

The justification you give for the de-emphasis of the RMSE in this analysis is that (l. 295 to 296) "… 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)." As stated in the previous subsection, this analysis is not replicable and not reliable.
Looking at Fig. S3, I see that field A seems to have, in many cases, a distinct pattern in the RMSE versus soil properties than the models for the other fields (including the model for field F). For example, the model for SOC for field A has the largest RMSE (and largest uncertainty in RMSE) in comparison to the other models, despite having a carbonate content similar to the other fields, except field F. I suggest that the behavior of the model for field A in Fig. S3 is particularly interesting and warrants a better interpretation than currently given (see the next subsection "The Interpretation of bad predictive accuracy for the models for field A and F should be clarified").

Thus, I think that the relation of the RMSE to soil properties is more interesting for researchers and practicioners and I think that absolute model performance offers more insights which have not been discussed sufficiently yet. For these reasons, I suggest to emphasize results for the RMSE. Please consider this as suggestion to clarify and improve the analysis because I think that currently, the analysis is less interesting (because it focuses on R^2 and RPD) and ambiguous (because I think the explanation why the models for field A often are less accurate should be a different one than why the models for field F often are less accurate, see the next subsection).

1.3 The Interpretation of bad predictive accuracy for the models for field A and F should be clarified

Your current analysis of the influence of carbonate concentrations on the predictive accuracy for the models for field A and F is in my opinion ambiguous. The manucsript states

- l. 293 to 295: "Fields A and F that showed lower model performance in terms of RPD had higher carbonate content, lower correlation coefficient between SOC and total N and higher variability in soil texture (compare also with density plots in Fig. 1)."
- l. 364: "We found an influence of carbonate content with lowest performance of local spectral models on fields A and F."

This can give the reader the impression that samples from fields A and F had a similar carbonate content, even though this is not the case. From Fig. 1 and Fig. S3, one can see that field A has an average carbonate content more similar to fields with good predictive performance (e.g. field E), but few samples with a high carbonate content. In contrast, field F has on average a high carbonate content and no (or very few) samples with lower carbonate content. All other fields do not seem to have samples with carbonate contents similarly large as in field F or A. Thus, a high average carbonate content does not seem to be necessary to cause bad predictive performance. It rather seems plausible that already few carbonate rich samples decrease model predictive accuracy.

Perhaps similar to you, I assume that the high variability in sample carbonate content (not the high average carbonate content per se) has caused the bad performance of prediction models for field A. Based on Fig. 1, one can see that only field A has such a large gradient in inorganic C content and despite average carbonate content similar to other fields with low carbonate content (for all samples), you have found a large RMSE and large uncertainty in the RMSE for field A (supporting Fig. S3). If you mean this, I suggest that you make clear that field A and field F differ in their properties and that it makes a difference whether the average carbonate concentration is high (field F) or whether the variability in carbonate content is high, even if there are only few samples with large carbonate content (field A): In the first case, at least total C content and pH value appear to be predictable quite accurately (with relatively low RMSE) (see Fig. S3), apparently because carbonate absorbance bands here can be used to explain total C content (which is mainly carbonate C), but carbonate bands cannot explain other soil properties related to SOM. In the latter case, only pH value appears to be predictable quite accurately (with relatively low RMSE) because SOC and carbonate both contribute to total C, but carbonate bands and bands related to SOC interfere.

Here, it would be easy to test the assumption that already few samples with large carbonate content suffice to decrease predictive performance: Simply drop the samples with high carbonate content from field A and compute the prediction models (with same preprocessing settings) with the reduced dataset. If the resulting models have a much better predictive performance and you can exclude measurement artifacts, bad performance of model for field A is at least related to the high variability in carbonate content.

Similarly, you could include some samples with high carbonate content from field A or F to the datasets from the other fields and test whether this increases the RMSE of the corresponding prediction models. Finally, you could add few samples with low carbonate content to the dataset from field F to test this from the other side, i.e. whether high variability in carbonate content with few samples with low carbonate content also increases the RMSE of models (in comparison to the model with samples from field F only).

2 Specific comments

1. l. 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.

2. l. 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 an 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) …".

3. l. 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.

4. l. 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.

5. l. 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".

6. l. 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"?

7. l. 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 unlcear whether "the higher lab measurement error with higher carbonate content can explain the lower model performance on soils with high carbonate content for SOC …".

8. l. 389: "… makes it more difficult to attribute absorption features …". I suggest to replace "attribute" by "relate".

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

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

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.

12. Fig. S5, field D: Variable POXC appears to have some missing values which results in `NA`s 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.

References

Oberholzer, Simon. 2023. "Reply on RC1." Peer Review. https://doi.org/10.5194/egusphere-2023-1087-AC1.