Resubmission of revised Manuscript:

Original title: Dataset variability and carbonate concentration influence the performance of local visible-near infrared spectral models (egusphere-2023-1087)

Dear Editor, dear Reviewers,

Many thanks for your constructive and insightful comments on our manuscript. We revised the manuscript according to your comments as outlined in our review answer (12.09.2023). We are happy to resubmit an improved manuscript. As you suggested we removed research question 3 about the influence of dataset variability and expanded in exchange the analysis how field characteristics influence the performance of local models. Therefore, major changes in the manuscript text were necessary and we describe them in the following list. Additionally, a track-changed version of the manuscript is also provided.

We hope we addressed all comments to your satisfaction.

On behalf of all co-authors,

Simon Oberholzer

List of relevant changes

Title

Since we removed the variability part of the analysis, we changed the title to: "Best performances of visible-near infrared models in soils with little carbonate - a field study in Switzerland"

Abstract

The abstract was rewritten to put a stronger focus on absolute prediction performance (RMSE) and include the additional aspects of field characteristics that influence model performance (correlation between target variables).

Introduction

Descriptions of dataset variability in literature was removed as well as research question three and the former research question four became now number three.

Methods

Section titles (2.1 – 2.8) are still the same as in the first version of the manuscript. Since we included the variability in soil texture in our analysis, we describe the measurement in section 2.2. Section 2.7 "Assessment of site characteristics influencing model performance" was rewritten according to the expanded analysis.

30 **Results**

The old section 3.5 "Influence of dataset variability on model performance" was replaced with former section 3.6 "Site characteristics influencing model performance". The new section 3.5 contains three new figures (Fig. 6, 7 and 8) that were not part of the former manuscript. Additionally, Figure 4 was rearranged to group the VIP analysis per dataset and not per soil property like before. We discovered a small calculation error for estimating the model metrics in cross-validation. Therefore,

35 the model performance metrics in Table 2 slightly changed but did not influence the main results. The error was part of a function in the *simplerspec* R-package and has already been corrected by its creator Philipp Baumann.

Discussion

The sections 4.4 and 4.5 of the old manuscript were removed. The former section 4.6 became the new section 4.4 and was

40 expanded to three subsections: 4.4.1 Mean carbonate content, 4.4.2 Correlation between target variables, 4.4.5 Variability of clay content. The other discussion parts (4.1- 4.3) contain only minor changes.

Conclusion

We included the results of the expanded analysis also in the conclusion.

45

References

All the DOIs were carefully tested.

Supplementary material

50 Dataset descriptions of soil texture were included in Table S1. Table S2 is new and provides an example of pre-processing optimization on a specific dataset. Table S3 is also new compares to two methods that were used to measure soil texture. The Euclidean distances in Fig. S1 were recalculated with pre-processed spectra. Figures S3, S4, S5 and S6 are new and provide supplementary information about the expanded analysis how field characteristics influenced model performance.

55 **Data and code availability**

We decided to make the data and important R codes available upon publication on a zenodo repository.