

## Reply to Reviewer 2

12.09.2023

### General answer

We thank you for your comments and recommendations. We numbered your comments and provided a specific answer for each one about how we will implement it in the manuscript.

We hope we addressed all comments to your satisfaction.

On behalf of all co-authors,

Simon Oberholzer

### Comment 1:

“The paper is well-written and addresses a topical issue i.e., the scale of the data sets for optimizing spectral models. However, the authors rely on quality performance parameters that are not independent: RMSE,  $R^2$  and RPD. Minasny et al. already demonstrated some years ago that there is an overlap between these parameters.”

We thank you for pointing out this important concern. We completely agree and will remove this entire part from the manuscript (chapter 3.5, 4.4 and 4.5; see review 1). We will expand the analysis about how site characteristics (carbonate content, soil texture) influence the performance of local models and possibly cause interferences of spectral features (see also answer to review 1).

### Comment 2:

In recent literature the RPD is mostly used for global comparison between studies. Bellon Maurel proposed to use the PIQ which is better suited for non-normally distributed data sets.”

We decided to use RPD for two reasons:

1. Our data do not show major deviations from normal distribution which provides robust results with PLSR and also justifies the use of RPD.
2. Most readers are more familiar with RPD than RPIQ.

### Comment 3:

Section 2.7. Regressions between RPD and CV are spurious, as the standard deviation appears in both terms. These relations should not be used (e.g., Fig. 5 and 7). The same applies to the relation between PRMSE and CV. Such relations always give a positive trend.

We will remove this analysis from the manuscript (see comment 1 and review 1).

### Comment 4:

Finally, the authors conclude that local models perform better than general models, but do not really give a recommendation on a sampling strategy for fields/regions with no prior knowledge of the soil properties.

We thank you for pointing this out. Local models with vis–NIR show overall good performance, but a high carbonate content decreases the model performance. The question of how to sample a

completely unknown area is a crucial one in soil science and difficult to answer. For local vis–NIR projects, we have the following recommendations, that are probably not described clearly enough in the manuscript:

- 1) Check for the carbonate content in the soil and then one can decide whether vis–NIR calibration modeling is feasible (if you decide for the latter your sample size will increase substantially).
- 2) Think about the model accuracy (absolute prediction error) that can be achieved as described in our data. Is that prediction error tolerable or is it too high?

If we know what exactly causes the loss of precision in carbonate rich soils (what we will do in the next step) we might even be able to draw clearer conclusions.

**Comment 5:**

I am not a native speaker, but I would avoid using 'high' and 'low', e.g., small and large errors. Please check with a native speaker.

Thank you very much. We will check that and adapt the language accordingly.

**Comment 6:**

Generally, 'concentration' refers to solutions, while 'content' is a more general term that can also refer to a solid in another solid. You use both terms, but I would advise to systematically use 'content'.

We will use content instead of concentration throughout the manuscript. Thank you very much.