# Discussion on referee 2's comments: Miniaturised visible and near-infrared spectrometers for assessing soil health indicators in mine site rehabilitation by Shen et al.

We thank the referee for the comments. Below we provide a discussion (in blue text and preceded by **Authors:**). We denote 'manuscript' as 'MS', 'page' as 'P', 'line' as 'L', when referring locations in the manuscript.

**Comment 1**: The paper discusses about several statistical and machine learning algorithms for evaluation of the spectrometers and the model prediction accuracy. Many soil physical, chemical and biological properties are targeted. It would give a reference for further NIR application. The paper should be improved before publication. Some suggestions are listed below

**Authors:** We thank the referee for reviewing our manuscript.

**Comment 2**: Please give details of the experimental design for spectroscopy measurement. How can you observe the data in Figure 2.

**Authors:** The experimental design for the spectroscopic measurements are given in the Methods subsection 2.3 Soil spectroscopy and the spectroscopic measurements (P6, L141-148). However, we agree that the description of our experiments and analyses are complicated, making parts of the manuscript difficult to understand. Thus, in a revision, we propose to include a diagram to summarise the experimental design and experiments conducted. For instance, we could include a diagram like that shown in Figure 1.

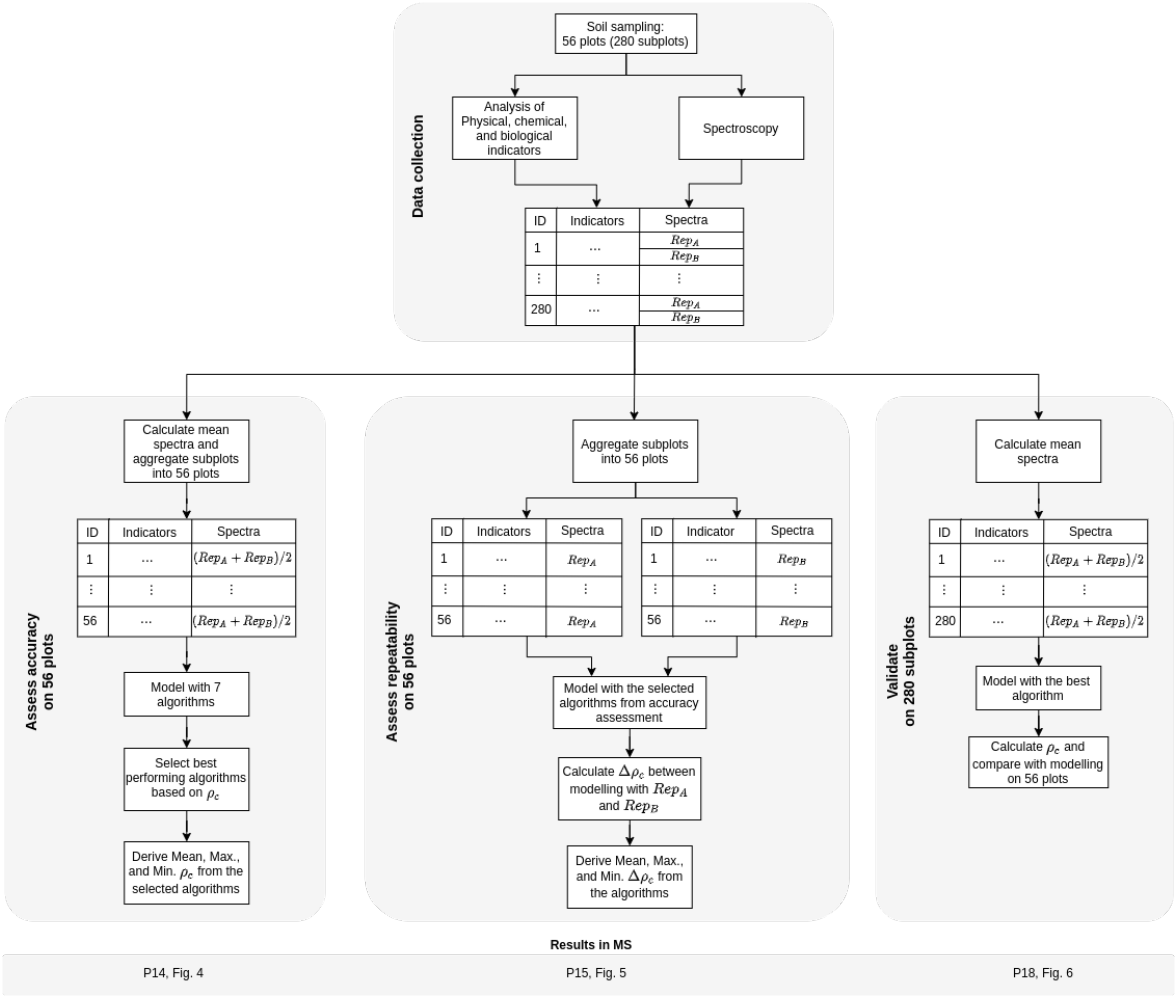


Figure 1: Study design

One such figure should also help to clarify how we produced our results, including Fig. 2 in the submitted manuscript. For MS Fig. 2, we calculated the mean of the two replicates ($Rep_A$ and $Rep_B$) for each spectrometer, which gives the plots in the first column of Fig. 2 (MS, P11). We also derived the difference between $Rep_A$ and $Rep_B$ using equation (1) (MS, P9) for each sample, which gives the second column in Fig.2. These plots show the repeatability of the spectroscopic measurements for each spectrometer. A smaller difference suggests better repeatability. The third column in Fig. 2 (MS, P11) shows results for the combined spectrometers.

**Comment 3**: I do not think that 56 samples for modeling is enough. Please prove and validate it, or I am not convinced of the results.

**Authors:** In total, we collected, analysed, and measured with the spectrometers 280 soil samples from subplots (described in MS Methods, P5–7). We performed the assessments and modelling in two ways. First, we aggregated the 280 subplots into 56 plots and we performed the modelling and validation on the aggregated data and using 10-fold cross validation (described in sub-section 2.4.1 Assessment of the spectroscopic modelling algorithms with data from plot). We aggregated the data for two reasons, (i) soil samples from a single plot were assumed to be somewhat similar and (ii) computational efficiency, since we assessed in terms of accuracy and repeatability seven spectrometers and combinations using seven algorithms and 29 soil properties, which produced approximately 4263 model evaluations with 10-fold cross validation).

Second, to ensure that our validation with the 56 data from plots was reasonable, we also performed the modelling and validation with the 280 data from the subplots but this time using 10-fold-plot-out cross validation (described in sub-section 2.4.4 Assessment of the spectroscopic modelling with data from subplots). The results from both of these approaches were similar and demonstrate the robustness of the results.

**Comment 4**: The definition of Lin's concordance correlation is not given.

**Authors:** We did provide a reference for the coefficient but in a revision we can of course also include a definition in L185 (MS P8) as follows:

'$\rho_c$ measures the deviation from a 45-degree line of perfect agreement between the observed and predicted values. It ranges from -1 to 1, with 1 denoting perfect agreement.'

**Comment 5**: Analysis of prediction errors (like RMSE) are needed in results and discussions part.

**Authors:** Please note that we did provide analysis of the errors using the root mean squared error (RMSE), mean error (ME), and standard deviation of the error (SDE) (Table 5 in the MS). The RMSE quantifies the inaccuracy, ME the bias, and SDE for

the imprecision such that $RMSE^2 = ME^2 + SDE^2$. And we do discuss these in the Results and Discussion sections. However, in a revision we can further emphasise the implications of these results.