

Comment on soil-2021-138

Anonymous Referee #1

Referee comment on "Miniaturised visible and near-infrared spectrometers for assessing soil health indicators in mine site rehabilitation" by Zefang Shen et al., SOIL Discuss., <https://doi.org/10.5194/soil-2021-138-RC1>, 2022

The paper number soil-2021-138 of Zefang Shen and co-authors regards the repeatability of the spectrometers and the accuracy of the spectroscopic models built with seven statistical and machine learning algorithms. The authors found that miniaturised spectrometers and combinations predicted 24 of the 29 soil properties with moderate or greater accuracy. The repeatability of the miniaturised NIR spectrometers was similar to that of the full-range, portable spectrometer. The manuscript is potentially a contribution of interest for remote sensing application and it is within its specific scopes but in my opinion the manuscript don't fit in the scope of this journal. In my opinion the work lacks of a clear application of soil, same part of the ms. are very difficult to understand and don't permit to comprehend the originality of the paper.

As is written do not explain the main results and should be connected with the aim of the research. Instead, the aim (L85-90, pg. 4)) focus only in several approach of multivariate procedure, seems just a mere exercise of multivariate statistics applied to remote sensing equipment's. .

The novelty is not explained as well as the gap in our knowledge that the manuscript with its objectives should fill. For example if the novelty are studies on 1) hyperspectral quantitative analysis (L155 to provide comparison data for the 29 soil chemical, physical and biological soil properties to be assessed using spectroscopic methods) 2) compare Spectral range, resolution, price, weight, and dimensions of miniaturised and portable spectrometers used in this study. (?), 3) to assess the spectroscopic modelling with different statistical and machine learning algorithms, as well as 4) the accuracy of the spectrometers estimates and their repeatability, 5) assessment of the spectroscopic modelling algorithms (L160) with data from plots or (finally!) a model analysis of physicochemical indicators of polluted soil (?). The gap that results from a state-of-the-art topic should be clarify form the beginning .

In my opinion the authors have to clarify in the state of art: 1) why they us this soil health indicators in mine site rehabilitation 2) if the procedure adopted is able to predict the level of contamination or soil health , 3) what are the limits of their predictive model, 4) if the model can be used for other place, because the authors do not compare their results with similar study.

The second problem of this work is in the preprocessing method that is very confused. The Materials and Methods section (L160-185) does not provide sufficient detail to follow the progress of the manuscript. Regarding methods, there is a use of PLSR, RF, SVM, GBXBoost, CUBIST, SVM, GPRL, GPRP an incredible set of algorithms without explaining

the criteria or limits or even if they are designed for these tools. So the reader is assuming or just have to imagine if the spectra were preprocessed. So the raw data were filtered with Savitzky-Golay (SGR) may be with multiplicative scatter correction MSC, or standard normal variate SNV, if they are treated with linear baseline correction LBC, peak normalization N, mean center MC. All this pre-processing are without explanation, whereas all the rest of the methods are only to be found in the results and conclusion discussion section. It is not explained why these methods were used and not others, how they are related to each other, nor is there sufficient detail to understand what was done and how these methods achieved the objectives of the manuscript. I am sorry for the Authors but no revision can at this point improve this work. Many other comments would be possible both for the sections 'Materials and methods' and 'Results and discussion', but it is useless because the comments made are more than sufficient to recommend the rejection of the manuscript