Comment on gmd-2021-153
Anonymous Referee #2

Referee comment on "A new methodological framework for geophysical sensor combinations associated with machine learning algorithms to understand soil attributes" by Danilo César de Mello et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2021-153-RC2, 2021

The manuscript aims to provide a novel methodological framework for combining terrain attributes and data from geophysical sensors with machine learning algorithms in order to understand the pedosphere system and model soil attributes. An analysis of the importance of pedoenvironmental variables in predictive modelling is also presented.

Although the study has scientific significance to it, it is undermined by the writing issues present in the manuscript. The manuscript contains numerous grammatical and spelling errors (Eg - misspelt hyperparameters in 244 etc.). Furthermore, some ideas are not well fleshed out which hinders the understanding. For example, the discussion about the null model could be further elaborated while discussing the results.

Additionally, I agree with the comments that the LOOCV is not a rigorous method of validating the true performance of the modelling since it lacks a true test set to evaluate the model against. It is recommended that the authors use a separate test set for validation purposes.

Some of the specific issues are highlighted below:

- Grammatical errors and spelling mistakes need to be fixed.
- The methodological flowchart - In Pearson’s test 95% threshold branching, at least one branch must have ‘yes’.
- The mathematical notations in the paper need to be more consistent. For Eg, in eq 4, change RMSE_NULL to NULL_RMSE.
- Many of the symbols used have not been clearly explained earlier. For example, it is not specified what the values $O_m$ and $O_i$ mean in Eq 4 (I believe these must be $Q_m$ and $Qobs_i$, respectively). Similarly, in eq 5, it is not clear what $Q_{train}$ means.

Overall, the presentation of the manuscript needs to be improved significantly in further revisions, along with a diligent model validation approach.