SOIL Discuss., author comment AC3
https://doi.org/10.5194/soil-2022-17-AC3, 2022

Reply on EC1

Danilo César de Mello et al.

Author comment on "Weathering intensities in tropical soils evaluated by machine learning, clusterization, and geophysical sensors" by Danilo César de Mello et al., SOIL Discuss., https://doi.org/10.5194/soil-2022-17-AC3, 2022

Nicolas P.A. Saby, 05 Sep 2022

I made an in depth reading of the paper and I have several major concerns about the work and the paper.

The abstract is quite vague and not very informative. Please explain the overall goal, the assumption, the approach implemented and give precise results and findings.

A: Rereading the abstract, we agreed with the reviewer and made the necessary adjustments in order to clarify the research objectives, main results and conclusions of the study.

The English does not seem to be ok.

A: We sent the for a general review of English (American English) to a specialized company, where a geoscience specialist also reviewed the entire manuscript (Proofreading service). A certificate attesting to the new revision of the manuscript was inserted in the "supplementary material" field.

The work is not enough clearly presented. A lot of questions arise when reading the actual version.

A: Thanks for the suggestion, we agree with the reviewer. This may have occurred given the number of covariates and the explanations of the relationship between them with weathering, in addition to the clustering and modeling technique used. However, we revised the entire manuscript and some paragraphs that we thought were confused in the explanations, and/or missing some details were rewritten in order to clarify the reader. Another problem may have been English, which was also revised.

The overall goal of this study is not enough clearly presented in the introduction. First (i), modelling with ML of weathering index (WI) is quite vague. But for what? Mapping, monitoring, statistical application? In (ii), it is said that importance of the covariates will be discussed but what is these covariates. This is not clear if these are the one for the digital soil mapping approach or the observed ones using sensor data. It is also a general comment about this paper where a confusion is made between these 2 types of information. The description of the proximal and remote sensing data should be better explained. The first ones are not spatially explicit.

A: Thanks for the review. Dear reviewer, we have re-evaluated the objectives and they are clearly met. Modeling the intensity of weathering means predicting and spatializing the data (maps shown in figures 5 and 6, for example, and table 2). The applications of obtaining weathering intensities, or in other words, more and less weathered soils, were briefly explained in the introduction (paragraph x). The importance of the variables used were those used in the modeling and which were duly presented and discussed in section 3.1 (3.1 Evaluation of model's performance, uncertainty and variables importance), including a figure (figure 4) demonstrating what these variables were. We believe that these variables should not be presented in the objectives of the work, but they should be explained and demonstrated in the methodology, as well as presented and discussed in the sessions indicated in the manuscript.

We have added the following paragraph after the objectives to clarify the relationship between digital mapping, remote and near sensing, and weathering:

"In this work we use the digital soil mapping approach (using geophysical sensors and satellite imagery) to estimate different intensities of weathering in an area that are comparable in terms of geology and soil types."

We have added a new paragraph in the introduction to clarify the concept of remote and near sensing:

"The proximal sensing is the use of field-based sensors to obtain signals from the soil when the sensor's detector is in contact with or close to (within 2 m) the soil (e.g., geophysical sensors). The sensors provide soil information because the signals correspond to physical measures, which can be related to soils and their properties. When the distance at which a particular sensor acquired a data is greater than 2m from the target, there is remote sensing (e.g., satellite images) (Rossel et al., 2011)."

The way geophysical data are used is rather difficult to understand and does not seem at all appropriate. It is quite difficult to understand why these data are mixed with the WI in the PCA step. This means that Geophysical data are not used as covariates but as a covariable. The PCA is not a method to evaluate (as it written in the abstract) but a multivariate analysis to explain the multivariate relations between variables. Moreover, the results of the PCA are not presented. It is not then possible to understand the signification of the principal components.

A: In this work we used the weathering index obtained with laboratory data, as explained in section 2.3. Weathering rates. We then combined this index with data obtained via geophysical sensors (in the field) and satellite (SYSI), which correspond to soil attributes generated by different intensities of weathering and, consequently, pedogenesis. Then we de-correlated the variables (Table 3). Subsequently, the PCA was performed using the kmeans method, where clustering indices corresponding to the different weathering

indices were generated. We used the cluster and PCA concomitantly to reduce the number of variables, which were later used in the cluster analysis, allowing the choice of uncorrelated groups, improving the performance of the analyses. The information used to create the clusters were: 6 parameters derived from geophysical sensors data (eU, eTh, K^{40}, magnetic susceptibility and, ECa), and the weathering index.

In addition, we adjusted the abstract following the reviewer issues about PCA.

We did not show PCA because it was used only to de-correlate sensor data.

Why did you use a clustering approach? We could expect a more quantitative digital soil mapping approach of the raw value of WI instead. Why are you creating clusters of WI and sensor data to map them after? Again, as the results of the PCA and the clustering are not presented, it is difficult to understand the signification of the cluster. It could happen that no correlation occurs between these data and the clusters represents only one of 2 variables. The results table 3 seem to validate this assumption as the discrimination between the different cluster are not very significant. However, an extensive discussion is provided about the signification of these clusters based on an interpretation of the *dsm* maps and by expertise. However, all the maps are wrong. Maybe a point map of the clusters would be better to be interpreted first. Finally, it is difficult to understand the adding value of the geophysical data.

A: We use a cluster approach in order to separate weathering intensity into classes.

In fact, we could expect a more quantitative digital soil mapping approach with raw WI values. However, it was not the focus of this work, since there are already other works that have already done this in the area of pedometrics.

In this work we used the weathering index obtained with laboratory data, as explained in section 2.3. Weathering rates. We then combined this index with data obtained via geophysical sensors (in the field) and satellite (SYSI), which correspond to soil attributes generated by different intensities of weathering and, consequently, pedogenesis. Then we de-correlated the variables (Table 3). Subsequently, the PCA was performed using the kmeans method, where clustering indices corresponding to the different weathering indices were generated. We used the cluster and PCA concomitantly to reduce the number of variables, which were later used in the cluster analysis, allowing the choice of uncorrelated groups, improving the performance of the analyses. The information used to create the clusters were: 6 parameters derived from geophysical sensors data (eU, eTh, K40, magnetic susceptibility and, ECa), and the weathering index.

Indeed, it could happen that no correlation occurs between these data and the clusters represents only one of 2 variables. However, as the reviewer observed, we performed a non-parametric assessment of sensor data and WI between groups (the Kruskal-Wallis test) to assess this issue and found that virtually all of them had statistical differences.

We believe that the reviewer is a little mistaken, since the values in table 3 indicate differences for most classes. Regarding the issue that the maps are wrong, we would like to know based on what arguments the reviewer asserts this issue. Wrong about what? As such, we are unable to respond.

The geophysical data contributed towards having significant relationships with weathering

and the soil attributes associated with them. as discussed in the work.

In the 2.1, two sets of data are presented (16 and 79 points) but only the 79 points are used for the study. Could you explain? finally, it is topsoil map of WI?

A: Yes, the reviewer is right. Sixteen soil profiles were described, as shown in figure 1. Seventy-nine points where samples were collected for physical-chemical laboratory analysis and, geophysical data. Data from the 16 soil profiles were used for soil classification, for further analysis between the intensity of weathering and the type of soil.

In fact, the weathering intensity was evaluated in the first 20cm of the soil, so yes, for the top soil, although the gamma gets values of readings of 30cm of depth.

The equation 1 is not explained enough. What is SIO2 and TIO2?

A: Thanks for the suggestion. We better explained the equation and the relationship between the elements and, add a new paragraph in material and methods section, as follow:

"The W1 index (Eq. 1) is based on the principle that during chemical weathering, there is an intense leaching of mobile elements such as basic cations and silica/ silicon oxide (SiO2) and residual concentration of less mobile and soluble oxides such as Fe2O3, Al2O3 and TiO2, mainly in tropical environments. In other words, these oxides gradually increase with increase in weathering intensity while all other elements are gradually reduced. This is the basis of calculating the weathering indices of soil which actually indicate the degree of weathering (Wani et al., 2016)."

In 2.7, two approaches are presented for the clustering, eg scott lethod and k-means. Which one did you use? Do you mean that a kmeans approach was implemented and the optimal number of clusters was selected using a scott approach?

A: The Scott test was used to determine the ideal number of clusters, while the k-means test was used to classify the samples in the group determined by scott.

The 2.8.2 should be titled "validation of the map".

A: We adjusted the sentence following the reviewer suggestion.

The explanation of the nested approach is quite difficult to follow. Do you mean that there is a tuning of the ML algorithm at each step of the LOOCV? This is not indicated in the fig 3. How this tuning is done? Bye cross validation?

A: The LOOCV (cross-validation) was used to optimize the hyperparameters of the evaluated algorithms, in each inner loop run. This phase is located inside the dotted square in figure 3. We are changing the flowchart to make this information clearer.

In addition, we tried to explain the nested-LOOCV method as follow:

"The nested-LOOCV method is a double loop process, where in the internal loop, the model is trained with a data set of size n-1, using the LOOCV for the optimization of the final model. On the other hand, the external loop corresponds to the test. In this loop, the remaining sample is predicted using the final model calculated in the inner loop. This prediction result is stored with the observed value of the remaining sample and later used to calculate the algorithm's performance (Jung et al., 2020; Neogi and Dauwels, 2019). The two loops are run n times (n = total number of samples, in our case 75). All samples are inserted into the outer loop, where the values predicted by the final model of each algorithm are calculated with the predicted and observed values of each sample. Then, the final result of the machine learning algorithm's performance will be obtained by predicted and observed values stored in the external loop. This is a robust method to evaluate the algorithm's performance and detects possible samples with problems in the collections or outliers. The training set generated in each loop went through the process of selecting covariates for importance and subsequent training".

283 should be titled validation not training.

A: We adjusted the sentence following the reviewer suggestion.

2.8.3 Why are you explaining that you are using a LOOCV now? This is very confusing as it is not explained in the fig 3. All the indicators explained in this section are based the result of the confusion matrix where the 4 kinds of results are computed, e.g., FP TP, FN and FN.

A: The LOOCV is used to optimize the hyperparameters of the evaluated algorithms. It is in this item that this point is made. In figure 3 this phase is the two-color square. The performance parameters used in this work (indicators) were all taken from the confusion matrix from which the FP TP, FN and PN indices are obtained. This point is inside the dotted square in figure 3.

It is not explained how the different maps produced during the LOOCV are combined at the end of the process. Do you compute the dominant value? Are using a probability approach?

A: Dear reviewer, we believe that this question is duly explained in item 2.8.3:

"...Then, the final map was created by combining the 71 prediction maps generated for each algorithm tested. In addition, the mode value for each pixel of the final map was calculated. The prediction error map was elaborated, considering the number of times that each algorithm chose the mode value in each map pixel normalized by the number of final maps (%)...."

"...The best final map chosen by the previous statistical tests was used to extract the geophysical sensor data, weathering indexes values at the sampling points...."

The "pls" is usually a regression approach and is not adapted to a classification

exercise. Could you detail which algorithm you used?

A: Thank you for this revision. In fact, this algorithm (pls) can be used as a classifier. As detailed in the text, the algorithms used, as well as their explanation, can be found in the caret manual at the link https://topepo.github.io/caret/available-models.html "6 Available Models". The pls is defined as "Partial Least Squares"- PLS. This algorithm can be used in classification and regression. The detailed explanation of the use of this algorithm is not relevant in this article, since it is an application of the algorithms (modeling) in soil science.

The size of the dataset is quite not adapted to the use of ML algorithm like random forest. You may consider more classical algorithm like the multinomial algorithm. As you use a LOOCV, we could not exclude an overfitting of the data. A k-fold cross validation would be more adapted. The hyper parameter of the different algorithms are not described so it is difficult to understand the quality of the models.

A: Firstly, we would like to have a larger number of samples than our current number and, we believe that all researchers would like to, but in our field conditions it was not possible as many other field conditions around the world. In addition, there is no minimum number of samples in the literature, indicated as a correct reference value for modeling soil and/or geophysical attributes. In addition, there are several researches published in good quality scientific journals, including Geoderma, which the authors also used a small number of samples in modeling processes (Fabijańczyk et al., 2017; Gebauer et al., 2020; Granger et al., 2017; Peukert et al., 2012; dos Santos Teixeira et al., 2021; Zhang and Zhu, 2019). In addition, LOOCV is more suitable for datasets with a small number of samples than k-fold.

In addition, we added the following sentence in material and methods:

The hyperparameters of each algorithm are described in the caret package manual in chapter 6. "Models described" available at https://topepo.github.io/caret/train-models-by-tag.html.

The sampling design is also not all adapted to a digital soil mapping approach. The locations were selected by expert judgement and are not very well spread over the area. There is extensive discussion about how to collect data in the recently published paper. This not really discussed in this paper. It is only acknowledged that the sample are not enough even but the spread of the data into the covariate space should be checked to discuss the validity of the map produced.

A: This sampling theory is valid and most used in geoesthetics. In ML, the use of conditioned Latin hypercube is being more used to define sampling sites. In addition, we would like to emphasize that during data collection, we advanced as far as possible due to the limitations of the sugarcane crop. However, we tried to carry out a distributed and representative sampling of the area, collecting data on all types of soils and lithology existing in the area, besides the topossequences.

I think this article needs to be thoroughly revised before publication. The use of the clustering step and the geophyical data should be better justified.

A: We agree with the reviewer and, we revied and explained in more detail and justification the two approaches (issues) highlighted by the reviewer.