

SOIL Discuss., author comment AC2
<https://doi.org/10.5194/soil-2022-17-AC2>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Reply on CC4

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-AC2>, 2022

Rafael Siqueira, 05 Sep 2022

I've read the pre-print version of this manuscript with great interest, due to the novelty associated to the application of modern technologies to investigate soil weathering and pedogenetic development, a very welcome initiative in Soil Science. The authors successfully combine two powerful approaches or tools constantly applied in the Pedometrics field: computational statistical techniques (machine learning and multivariate analysis) and geophysics proximal sensing data, aiming to understand in a quantitative manner how the soils develop in a tropical landscape. Nonetheless, I would like to write down some comments as reflections and suggestions:

A: Thank you for taking the time to read our work and recognize the importance of our research.

I am not a native speaker, but a new look at the English writing for corrections would be important to improve the quality of the text. Furthermore, some parts of the text are a bit clumsy and a review would be welcome.

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. In addition, we reviewed the entire text to locate and modify the parts that were unclear.

Abstract

Line 17 - I am not quite sure if you could call machine learning as a "geotechnology", for the reason it was not originally developed for geosciences.

A: We agree with the reviewer and modify the sentence as suggested.

Line 19 – Satellite imaging is not a proximal sensing technique, but a remote sensing technique.

A: We agree with the reviewer and modify the sentence as suggested.

Line 21 – You must explain in the Abstract why you used the PCA and clusters analysis. At the same time, I recommend to cite what information have you used to create the clusters, and moreover, what they really mean, in the Abstract.

A: 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.

We modified a little a sentence in abstract and material in methods (*section 2.7 Principal component analysis and clusterization*) better explain this step.

Abstract: *"Afterwards, the principal component analysis and the ideal number of clusters was determined, in order to reduce the number of variables, which were later used in the cluster analysis. The data 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."*

Material and methods: *"... (PCA) was applied to the 6 parameters derived from geophysical sensors data (eU, eTh, K^{40} , κ and, ECa), and the weathering indexes."*

Line 22 – Change "we determine and used the ideal number of clusters" for "we used the ideal number of clusters". You've already cited you determined the number of clusters before.

A: We agree with the reviewer and modify the sentence as recommended.

Line 27 – "The Nitisol over East diabase presented greater weathering intensity than".

A: We do not understand this query. In the text the sentence is complete and correct. Please verify:

... the Nitisol presented greater weathering intensity than the Nitisol over West diabase areas."

The last sentence of the Abstract is clumsy.

A: We agree with the reviewer and modify the sentence as recommended.

Introduction

The Introduction is very well written, making a brief literature review about weathering and the geophysical sensors.

A: Thank you for read.

Line 94 – Would not be better if you used just “model weathering intensity using combined data from geophysical sensors” than “model weathering index using combined data from geophysical sensors”?

A: We agree with the reviewer and modify the sentence as recommended.

Line 97 – I suggest to remove the objective 4, since you did not explore it in your results.

A: We agree with the reviewer and modify the sentence as recommended.

Line 99 – I would remove the citation about satellite here.

A: At this point we judged to keep the "satellite" in the sentence due to the use of SYSI, generated from satellite images in the survey.

Material and methods

Line 109 - I guess the name of the municipality cited in the text is with just one “f”.

A: We agree with the reviewer and modify the sentence as recommended.

Line 135 - You could standardize the Embrapa citation. Sometimes you cite Embrapa, 2011 and sometimes Embrapa, 2017.

A: We agree with the reviewer and modify the sentence as recommended.

Line 135 – In the laboratory analysis topic, you make a complete description about physical and chemical analysis of soils according to Embrapa. However, you do not use any of them, unless the oxides contents. I suggest to remove all analyses which you did not apply in this paper.

A: At this point, we decided to keep all the analyzes that were carried out, because data such as granulometry, fertility and others, allow us to correlate the intensity of weathering with the environment where the samples were collected, and relate them to geology and soil type, which was one of the approaches of the article.

Line 146 - I think it'd be productive you write one or two more lines explaining better how the oxides contents, the only analysis you really used, are obtained. Moreover, you cite Fe₂O₃ content but you also do not use it, only the SiO₂ and TiO₂ contents.

A: On this specific point, we disagree with the reviewer, as these analyzes are already traditionally well known in soil science and are already duly referenced. We apply the principle of objectivity in the writing of the work while ensuring the reproducibility of the methodology.

Line 150 - If you consider that index (WI) used in the text as sufficient for your goals, I suggest you to better characterize this index in particular and its meaning about weathering. Moreover, I see in many parts of the text you using "weathering indexes" when in this paper you use just one index.

A: The term was better explained following the reviewer's recommendations. The manuscript was revised and the passages where we cited "weathering indexes" were corrected to "weathering index". The follow sentence was added in the final of this section:

"...This index is based on the principle that during chemical weathering, there is an intense leaching of mobile elements such as basic cations and silica (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)."

Line 197 - You are very correct in saying that the use of the IGC data allowed the creation of a more detailed topographic data than orbital sources, such as SRTM, could do. But just a hint for future works: with the scale of IGC database you claimed (1:10.000) you'd be able to obtain an MDE with a way finer resolution than you obtained, which would be more appropriate to the goals that you wanted to achieve.

A: We thank the reviewer for this valuable tip. We are already adopting this suggestion in other work that is being developed.

Table 1 – very nice table explaining each morphometric variable. It will be a good reference for other authors.

A: We thank you for the recognition.

Line 217 - It will be interesting you cite in an only place the "7 parameters

derived from geophysical sensors data” which you used. According to your descriptions in the “Geophysical data collection” and the Table 3, you used 5 parameters from the geophysical sensors, not 7.

A: We have amended the sentence as per the reviewer's suggestion. In fact, there were 6 parameters, since we consider each gamma channel as a parameter. We add the following snippet to the sentence:

“... (PCA) was applied to the 6 parameters derived from geophysical sensors data (eU, eTh, K40, κ , ECa) and the weathering indexes...”

Line 228 - “argiluviation or ferralitization indexes”. Again, why two indexes if you supposedly used just one? Furthermore, it is the SiO₂/TiO₂ index that you described previously one of them? If so, you should have to indicate that before.

A: In this sentence we have made a mistake. Therefore, the terms argilluviation and ferralitization were excluded from the sentence.

Line 229 - “These groups characterize themselves by present similar values within the groups, but different values between one group to another”. You should refine the description about the statistical clustering.

A: Sorry, but on this particular point, we disagree with the reviewer. We believe that this sentence is well explained. This explanation was the same given in other previously published works. However, we added the word "statistics" to the sentence, as follows:

“These groups characterize themselves by present statistical similar values within the groups, but different values between one group to another.”

Line 275 - Very good insight about the combined use of the PCA and k-means analysis as well as the use of the Nested LOOCV for handling few samples.

A: We thank the reviewer for recognizing the work.

Line 233 - “This result was used to extract the values of covariates (morphometric and geological data)”. Not only these covariates, but also the Synthetic Soil image (SYSI).

A: The sentence was adjusted following the reviewer recommendation.

Line 236 – “base database”?

A: The sentence was adjusted following the reviewer suggestion.

Line 250 – “increases the performance of machine processing algorithms”. Not necessarily.

A: Thanks for the review, however on this point we disagree with the reviewer. We cite and find works where other researchers have found that there really are increases in the performance of machine processing algorithms. Furthermore, we verified in this and other works by our group that this phenomenon actually occurs. If the reviewer sends us references with contrary scientific convictions, we will review this point, otherwise, we decide to keep it.

Line 260 – the remotion by correlation is not only associated to “reduce computational time” but also to minimize the problem of multicollinearity.

A: We add this in the sentence following the reviewer recommendation.

Line 272 – I think it is “repeatecv”.

A: We corrected the sentence following the reviewer suggestion.

Line 350 – You cite here that you used the Kruskal Wallis test to choose the best model. But in your results, you only showed the Kruskal Wallis being used to differentiate the weathering clusters. You need to synchronize your material and methods with your results in this regard.

A: Thanks for the review, however we decided to keep the results more relevant and easier for readers to explain.

Line 359 – You don’t need to cite RF twice in the sentence.

A: We adjusted the sentence following the reviewer suggestion.

Line 360 – Before “The RF algorithm presented equivalent performances than other algorithms”, insert “In other studies,”.

A: We adjusted the sentence following the reviewer suggestion.

Line 381 – “performance precision”? Redundancy.

A: We adjusted the sentence following the reviewer suggestion.

Line 381 – Paragraph shows some clumsy sentences.

A: The paragraph was revised by proof-reading service and was improved.

Figure 5 - I've seen that the authors explain for the first time what the clusters really mean (weathering intensity) in a clear way only on the Figure 5. I suggest the authors to anticipate this important explanation. At the same time, I suggest the authors to label the clusters with the information of weathering, at last, the numbers 1,2,3 that the authors arbitrarily chose are just categorical, not expressing by themselves the weathering degree.

A: We improve the figure as the reviewer suggested and, also, we added a short paragraph explaining for what the clusters were created and what they mean, in material and methods section, as follow:

"It is important to realize that, the clusters were generated in order to demonstrate the different degrees of weathering on the different lithologies and soil types."

Line 451 – According to the rest of your text, the cluster 3 has higher weathering, followed by cluster 1 and then the cluster with lesser weathering, cluster 2. In this sentence, you have been confounded the order of the clusters.

A: We adjusted the sentence following the reviewer suggestion.

Line 458 – "West Rodic Nitisol" – I think the authors confounded the terms. Here, the correct would be "East Rodic Nitisol".

A: We corrected the sentence following the reviewer's recommendations.

Line 465 – "For the weathering index".

A: We corrected the sentence following the reviewer's recommendations.

Table 3 – interesting table. Could explain in a paragraph which of the parameters explain better the weathering, according with your data? For me, the most important were the W1, plus magnetic susceptibility and ECa.

A: A new paragraph has been added, following the reviewer's recommendations:

The greatest statistical differences and, consequently, the weathering intensity were better evidenced for magnetic susceptibility, ECa and radiometric data from gamma-ray spectrometry, respectively, evidencing the differences between the different weathering index.

Line 494 – East diabase

A: We corrected the sentence following the reviewer's recommendations.

Line 495 – West diabase and West Rhodic Nitisol

A: We corrected the sentence following the reviewer's recommendations.

Line 514 – free drainage

A: We corrected the sentence following the reviewer's recommendations.

Line 523 – but there is higher K^{40} in the supposedly more weathered Lixisols (cluster 1)

A: Dear reviewer, we believe that there was a mistake in the interpretation. All K^{40} values for area were low (less than 1), including for Lixisols areas.

Line 526 – decreasing or increasing ECa? You must explain that.

A: Dear reviewer, we added the follow sentence in this paragraph:

"...however, it is harder to explain how environmental variable contribute more and how they affect the ECa values during weathering intensity, because they all pedoenvironmental variable act concomitantly to produce ECa values (parent material, pedogenetic and weathering processes, mineralogy, texture, water dynamics on landscape, magnetic susceptibility, texture, CEC and base saturation) (Mello et al., 2022)."

Line 539 – you do not model with the nested LOOCV, but you evaluated it.

A: Dear reviewer, thanks for the review. We understand the reviewer's concern, however at this point we are referring that nested-LOOCV was suitable for modeling as a whole process, for small sample sets. The modeling in this work was carried out in three stages: training, validation and test. The nested-LOOCV *inner loop* is done in training and validation.

"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 71). 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".

My suggestions aim to improve the quality of this excellent paper. I congratulate the authors for their work and hope to read more on this very interesting topic.

A: Thank you for the excellent reviews that contributed to our work and for the recognition.