

## **Comment on soil-2022-17**

Anonymous Referee #1

---

Referee 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-RC1>, 2022

---

The article has an interesting general idea, but it raises some questions about the study proposal.

The first general question in the article that I didn't see answered is: What would be the real importance of evaluating soil weathering by different techniques? Soil fertility? This can be measured directly. 'environmental issues'? there are also techniques to evaluate directly.

Listing the objectives of this study, not all of them were answered in the course of the article, nor in the conclusions section, leaving readers without a final answer to certain points raised.

Some errors in concepts and terms were noticed, such as:

-Machine learning is not a geotechnologie.

-'proximal remote sensing'? satellite image is not proximal and geophysical is not remote.

-Spectroscopy method is not a geotechnology

I miss a paragraph explaining how the fact of weathering affects everything you said. Bearing in mind that this process takes years and years to affect soil properties, would that make any difference now? And if you say this can be seen in the difference in soil types, then isn't it easy to study the difference between soils?

Table 2 – value of accuracy from kkn is wrong.

When comparing the algorithms, I didn't see you talking about which parameters were used in each of them.

In terms of modeling and mapping, as you mentioned, the number of samples is very low, not being a great number of samples to work in an area of almost 200 ha. In addition to using it to calibrate a model with RF. Don't you think this would affect the construction of the model? don't you think the data wouldn't be overfitted?

Speaking now about the mapping exercise, in my opinion the distribution of samples is not adequate to carry out the mapping exercise.

My main concern is: Even knowing that the amount of samples was not enough and as the arrangement of samples is not suitable for the DSM approach, you still decided to carry out the article.

I suggest improving the quality of the data and methodologies used and strongly revising the article