

Comment on soil-2022-17

Guilherme Oliveira

Community 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-CC3>, 2022

This article represents an excellent application of Machine learning techniques in geophysical-soil science data. The manuscript brings an excellent approach to the geophysical relationships with processes that occur in the soil (weathering) and variables of easy acquisition.

The methodological flowchart is very well defined, However, I had doubts in some specific parts.

First of all, the English writing needs to be improved. I suggest submitting to a specific proofreading company.

Why did you use the F1-score test instead of the Kappa metric or accuracy metric?

Is the data balanced? The accuracy is a good metric when the data is balanced. This is very important because all of these algorithms have a bias with unbalanced data.

Moreover, the algorithms used (excepted for RF) required standardized data.

Why didn't you employ the covariates in the unsupervised clustering? I dont understand why you compare the number of PCA dimensions with the number of clusters.

With proper revisions, I believe it to be an important contribution to the journal.