SOIL Discuss., author comment AC4
https://doi.org/10.5194/soil-2022-17-AC4, 2022

Reply on RC1

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-AC4, 2022

Anonymous Referee #1,

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

A: Thanks for the acknowledgment, we checked all the issues about the study proposal pointed out by the reviewer.

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 techiniques to evaluate directly.

A: Weathering operates in a multitude of spatial (from a nanometer to a planetary scale) and temporal (from thousands to only a few years) scales and its action impacts several, if not all, Earth systems. Weathering releases solutes that nourish every terrestrial ecosystem, triggers the biogeochemical cycles of every chemical element and thus control both water (e.g., river, ocean, groundwater) and soil chemistry. From a larger scale perspective, weathering is one of the major planetary sinks for CO_2, and thus, play a pivotal role on climate change and life itself on Earth. As an extremely intricate and complex process, and in response to its interface nature, weathering has been assessed, quantified, measured, and studied from innumerous techniques and perspectives; all originated from different scientific fields and backgrounds. Many of such approaches and techniques are expensive and time-consuming and thus make it difficult to gather information about large areas in an accessible, reliable, and fast way. In the current scenario, where science must tackle many Earth science challenges in which weathering is a central process, the development of novel tools to study weathering is crucial to achieve the expected goals (Egli et al., 2001; Frings and Buss, 2019; Ruiz et al., 2020, 2022).

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.

A: We agreed with the reviewer and added a new paragraph:

"The combined use of geophysical sensors, satellite images and morphometry, by different machine learning algorithms proved to be a robust method and were able to model different weathering intensities."

The other parts of the objectives we believe are answered along the other parts of the conclusions.

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

-Machine learning is not a geotechnology.

A: We adjusted the sentences following the reviewer suggestions.

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

A: We adjusted the sentences following the reviewer suggestion.

Spectroscopy method is not a geotechnology

A: We adjusted the sentences following the reviewer suggestion.

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?

A: We agree with the reviewer and add the following paragraph:

"Weathering operates in a multitude of spatial (from a nanometer to a planetary scale) and temporal (from thousands to only a few years) scales and its action impacts several, if not all, Earth systems. Weathering releases solutes that nourish every terrestrial ecosystem, triggers the biogeochemical cycles of every chemical element and thus control both water (e.g., river, ocean, groundwater) and soil chemistry. From a larger scale perspective, weathering is one of the major planetary sinks for CO2, and thus, play a pivotal role on climate change and life itself on Earth. As an extremely intricate and complex process, and in response to its interface nature, weathering has been assessed, quantified, measured, and studied from innumerous techniques and perspectives; all originated from different scientific fields and backgrounds. Many of such approaches and techniques are expensive and time-consuming and thus make it difficult to gather information about large areas in an accessible, reliable, and fast way. In the current scenario, where science must tackle many Earth science challenges in which weathering is a central process, the development of novel tools to study weathering is crucial to achieve the expected goals (Egli et al., 2001; Frings and Buss, 2019; Ruiz et al., 2020, 2022)."

Dear reviewer, in relation to the time in which weathering acts and affects soil properties, there are numerous works in the literature that report how weathering can be extremely fast, especially in the tropical environment and under more susceptible parent materials.

Egli, M., Mirabella, A., & Fitze, P. (2001). Clay mineral transformations in soils affected by fluorine and depletion of organic matter within a time span of 24 years. Geoderma, 103(3-4), 307-334.

Ruiz, F., Andrade, G. R. P., Sartor, L. R., dos Santos, J. C. B., de Souza Júnior, V. S., & Ferreira, T. O. (2022). The rhizosphere of tropical grasses as driver of soil weathering in embryonic Technosols (SE-Brazil). Catena, 208, 105764.

Ruiz, F., Sartor, L. R., de Souza Júnior, V. S., dos Santos, J. C. B., & Ferreira, T. O. (2020). Fast pedogenesis of tropical Technosols developed from dolomitic limestone mine spoils (SE-Brazil). Geoderma, 374, 114439.

Table 2 – value of accuracy from kknn is wrong.

A: The text was adjusted following the reviewer suggestion.

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

A: We added a short sentence in text:

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.

Here we presented the hyperparameter:

Algorithm	Hyperparameters	Definition
RF	*"mtry"*	number of covariates that are chosen at random for each node in the tree. Being the only parameter optimized.
ntree	number of trees. BREIMAN, L, 2002, considers keeping the	

	value constant at 100.	
nodesize	the minimum number of data points on each terminal node. BREIMAN, L, 2002, considers keeping the value constant at 1 and 5 for classification and regression.	
KKNN	*kmax*	Number of neighbors within the search area. Optimized.
distance	Maximum search distance from the center of the test area. Optimized.	
kernel	Sample space transformation kernel. Optimized.	
Partial Least Squares (PLS)	*ncomp*	the number of main components to be used in modeling. Optimized.
avNNet (Model Averaged Neural Network)	*size*	Number of units (neurons) in the hidden layer. Optimized.
decay	parameter for weight drop using neural network optimization. optimized	
bag	Logical index of the use of bootstrap methodology in	

training optimization.
Optimized.

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?

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).

We are aware of the possibility of overfitting, so applying the nested-LOOCV methodology, which we believe is more suitable for a small number of samples.

The nested-LOOCV method is indicated as a set for small data sets, which other methods of evaluation of test samples would not be viable due to the low number of samples in a test samples (Ferreira et al., 2021), being more used in the field of medicine in human experiments or where the number of samples is limited, providing an unbiased estimate of the true error (Chen et al., 2017; Li et al., 2018; Xing et al., 2011; Xu et al., 2020). The Nested-LOOCV method is a double loop process, where in the first loop the model is trained with a data set of size n-1, and the test is done in the second loop with the missing sample and used to validate the test and the training performance.

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

A: Here, 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 soil types, toposequences and, lithology existing in the area.

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.

A: We understand the reviewer concern. However, we do have a few observations to note: 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.

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 such as: (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)

Finally, 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 soil types, toposequences and, lithology existing in the area.

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

A: The entire article was revised following the reviewer suggestion.

References

Chen, X., Zhang, H., Lee, S.-W. and Shen, D.: Hierarchical high-order functional connectivity networks and selective feature fusion for MCI classification, Neuroinformatics, 15(3), 271–284, 2017.

Egli, M., Mirabella, A. and Fitze, P.: Clay mineral transformations in soils affected by fluorine and depletion of organic matter within a time span of 24 years, Geoderma, 103(3–4), 307–334, 2001.

Fabijańczyk, P., Zawadzki, J. and Magiera, T.: Magnetometric assessment of soil contamination in problematic area using empirical Bayesian and indicator kriging: A case study in Upper Silesia, Poland, Geoderma, 308, 69–77, doi:https://doi.org/10.1016/j.geoderma.2017.08.029, 2017.

Ferreira, R. G., da Silva, D. D., Elesbon, A. A. A., Fernandes-Filho, E. I., Veloso, G. V., de Souza Fraga, M. and Ferreira, L. B.: Machine learning models for streamflow regionalization in a tropical watershed, J. Environ. Manage., 280, 111713, 2021.

Frings, P. J. and Buss, H. L.: The Central Role of Weathering in the Geosciences, Elements, 15(4), 229–234, doi:10.2138/gselements.15.4.229, 2019.

Gebauer, A., Ellinger, M., Brito Gomez, V. M. and Ließ, M.: Development of pedotransfer functions for water retention in tropical mountain soil landscapes: spotlight on parameter tuning in machine learning, Soil, 6(1), 215–229, 2020.

Granger, S. J., Harris, P., Peukert, S., Guo, R., Tamburini, F., Blackwell, M. S. A., Howden, N. J. K. and McGrath, S.: Phosphate stable oxygen isotope variability within a temperate agricultural soil, Geoderma, 285, 64–75, doi:https://doi.org/10.1016/j.geoderma.2016.09.020, 2017.

Jung, Y., Lee, J., Lee, M., Kang, N. and Lee, I.: Probabilistic analytical target cascading using kernel density estimation for accurate uncertainty propagation, Struct. Multidiscip.

Optim., 1–19, 2020.

Lee, J. and Park, K.: GAN-based imbalanced data intrusion detection system, Pers. Ubiquitous Comput., 25(1), 121–128, 2021.

Li, Y., Liu, J., Huang, J., Li, Z. and Liang, P.: Learning brain connectivity sub-networks by group-constrained sparse inverse covariance estimation for Alzheimer's disease classification, Front. Neuroinform., 12, 58, 2018.

de Mello, D. C., Safanelli, J. L., Poppiel, R. R., Veloso, G. V., Cabrero, D. R. O., Greschuk, L. T., Mello, F. A. de O., Francelino, M. R., Ker, J. C. and Leite, E. P.: Soil apparent electrical conductivity survey in different pedoenvironments by geophysical sensor EM38: a potential tool in pedology and pedometry studies, Geocarto Int., (just-accepted), 1–24, 2022.

Neogi, S. and Dauwels, J.: Factored Latent-Dynamic Conditional Random Fields for Single and Multi-label Sequence Modeling, arXiv Prepr. arXiv1911.03667, 2019.

Peukert, S., Bol, R., Roberts, W., Macleod, C. J. A., Murray, P. J., Dixon, E. R. and Brazier, R. E.: Understanding spatial variability of soil properties: a key step in establishing field□to farm□scale agro□ecosystem experiments, Rapid Commun. Mass Spectrom., 26(20), 2413–2421, 2012.

Ruiz, F., Sartor, L. R., de Souza Júnior, V. S., dos Santos, J. C. B. and Ferreira, T. O.: Fast pedogenesis of tropical Technosols developed from dolomitic limestone mine spoils (SE-Brazil), Geoderma, 374, 114439, 2020.

Ruiz, F., Andrade, G. R. P., Sartor, L. R., dos Santos, J. C. B., de Souza Júnior, V. S. and Ferreira, T. O.: The rhizosphere of tropical grasses as driver of soil weathering in embryonic Technosols (SE-Brazil), Catena, 208, 105764, 2022.

dos Santos Teixeira, A. F., Silva, S. H. G., Soares de Carvalho, T., Silva, A. O., Azarias Guimarães, A. and de Souza Moreira, F. M.: Soil physicochemical properties and terrain information predict soil enzymes activity in phytophysiognomies of the Quadrilátero Ferrífero region in Brazil, CATENA, 199, 105083, doi:https://doi.org/10.1016/j.catena.2020.105083, 2021.

Wani, S. A., Najar, G. R., Wani, J. A., Ramzan, M. and Hakeem, K. R.: Weathering and approaches to evaluation of weathering indices for soil profile studies–An overview, Soil Sci. Agric. Environ. Prospect., 183–198, 2016.

Wardhani, N. W. S., Rochayani, M. Y., Iriany, A., Sulistyono, A. D. and Lestantyo, P.: Cross-validation metrics for evaluating classification performance on imbalanced data, in 2019 international conference on computer, control, informatics and its applications (IC3INA), pp. 14–18, IEEE., 2019.

Xing, J., Wang, S. X., Jang, C., Zhu, Y. and Hao, J. M.: Nonlinear response of ozone to precursor emission changes in China: a modeling study using response surface methodology, Atmos. Chem. Phys., 11(10), 5027–5044, 2011.

Xu, X., Li, W., Tao, M., Xie, Z., Gao, X., Yue, L. and Wang, P.: Effective and Accurate Diagnosis of Subjective Cognitive Decline Based on Functional Connection and Graph Theory View, Front. Neurosci., 14, 2020.

Zhang, G. and Zhu, A.-X.: A representativeness heuristic for mitigating spatial bias in existing soil samples for digital soil mapping, Geoderma, 351, 130–143,

doi:https://doi.org/10.1016/j.geoderma.2019.05.024, 2019.