

## Reviewers' comments (Second round)

### Reviewer #1

Some minor issues:

Line 11, "their accuracies were very similar to the PSF at each standard depth". I think it should be either "their accuracies were very similar at each standard depth" or "their predicted values were very similar to the measured PSF at each standard depth". You can't say accuracies are similar to PSF

*Response: You are correct. We improved the writing of this sentence.*

Line 126, use "predicted results" instead of "predictive results"

*Response: the sentence was corrected.*

Figure 2, I think the framework is a little confusing. The input maps, selection criteria and procedure, and final maps should be better distinguished from each other by different shapes, and the flow should be clearer. Here, it seems the parallelograms are used for maps, but it is also used for "Min MAE" and "Model selected". While the "Min MAE" should be the criterion to "select Model".

*Response: We agree. The flowchart was a little confusing. We made the figure again and used a different color for inputs and outputs. We restructured figure 2, and we specified the criteria for each process.*

### Reviewer #2

General comments:

I acknowledge that the authors have heavily revised the manuscript and have addressed most of my previous comments. The manuscript has improved significantly from the previous version. I only have a few minor comments – mainly clarification. Beside this, **the grammar and sentence structure are still partially incorrect and sometimes hinder the flow of reading. Therefore, I recommend a careful language edit before final publication.**

*Response:*

Specific comments:

Line 9f: I think there are already national texture maps based on digital soil mapping. You could highlight that this is the first national texture map based on digital soil mapping in Colombia.

*Response: We agree. The sentence was rewritten to clarify the statement.*

**Line 28f:** I disagree that clay content always has to increase with depth. Although this is quite common, it is not a requirement for mapping – not like that all three texture classes always have to sum up to 100%.

*Response: We agree. Clay content not necessarily increases with depth. We wanted to say that PSF could vary in proportion while depth increasing. We rewrote the sentence like this:*

*“Additionally, the soil texture study must include two principal statements: first, it is compositional data, which means that PSF sum to 100% (%clay+%sand+%silt), and this statement must be satisfied at each location (Amirian-Chakan et al., 2019); second, the proportion of PSF could vary between horizons depending on soil forming factors interactions (Orton et al., 2016; Poggio and Gimona, 2017).”*

**Line 43:** Does not every statistical model needs a response variable?

*Response: We can find two principal categories of statistical learning: supervised and unsupervised. In supervised learning, for each observation of the predictor measurement, there is an associated response measurement (response variable). While unsupervised learning describes the somewhat more challenging situation in which, for every observation, we observe a vector of measurements but no associated response; in this case, we can seek to understand the relationships between the variables or between the observations (James et al., 2013). Some tools for unsupervised statistical learning are cluster analysis and principal components analysis. Some applications of unsupervised statistical learning in digital soil mapping could be 10.1016/j.geoderma.2021.115012 and 10.1016/j.geomorph.2020.107305.*

*We rewrote the sentence to clarify the general statement:*

*“These predictions or probability estimates are derived from the use of supervised statistical learning (in the presence of training data for a response variable) or unsupervised statistical learning (in the absence of a response variable) (James et al., 2013).”*

**Line 73:** What does “RMSE  $\approx$  122,200” mean?

*Response: We agree; the sentence is not clear. 122.200 are the number of points or observations. We rewrote the sentence to clarify the statement.*

*“One representative case was developed by Hengl et al. (2021) for the continent of Africa at three depths (0, 20, and 50 cm) and 30 m spatial resolution. They produced predictions using two scale 3D ensemble machine learning (EML) framework and 122,200 training samples (approximately); their study utilized an improved predictive mapping framework: spatially-adjusted EML, that better accounts for spatial clustering of points. The spatial cross-validation methodology was a special point of their work, obtaining RMSE values of 9.6%, 13.7%, and 8.9% for clay, sand, and silt, respectively.”*

**Line 98:** Could you provide a short description of the soil texture classes? How are they defined in terms of size (e.g. silt: 0.002 to 0.05 or 0.002 to 0.063 mm).

*Response: We added a brief description for soil particle-size classification. The textural classification in the Soil survey studies in Colombia is made according to the United States Department of Agriculture (USDA). We rewrote the sentence as follows:*

*“A total of 4,203 georeferenced (EPSG:4326) soil profiles were collected from Sistema de Información de Suelos de Latinoamérica y el Caribe - SISLAC, a soil information system developed by FAO \citep{FAO:2020:Online}, that all contained information about soil particle-size fractions (PSF). These PSF are classified according to the United States Department of Agriculture (USDA) system; clay (particles smaller than 0.002 mm in diameter), silt (particle sizes from 0.002 to 0.05 mm in diameter), and sand (particle sizes from 0.05 to 2.0 mm in diameter).”*

**Line 162: Only a side-comment. The package mlr is not maintained anymore and the most recent version of this package is called mlr3 (<https://mlr3.mlr-org.com/>).**

*Response: Effectively, package mlr is considered retired from the mlr-org team. They suggest to use the new mlr3 framework from now on and for future projects. However, the creator of landmap package indicated that landmap provides methodology for automated mapping using Ensemble Machine Learning extends functionality of the mlr package (Hengl, T.:2021. landmap, url:<https://github.com/envirometrix/landmap>). Nevertheless, we adjusted the name of the new package.*

**Line 202: See my comment about (line 9).**

*Response: The sentence was rewritten as follows:*

*“This study represents the first effort to provide a national map of soil texture using a digital soil mapping framework in Colombia.”*

**Line 217: Figure 2: I find the flow chart really helpful. However, I was wondering why you show PSF error as an input into the models? Did you not use the transformed PSF values?**

*Response: In the spatial ensemble procedure, the inputs are the PSF predictions (MACHISPLIN, landmap, and SG), and the selection is based on the minimum MAE (previously calculated). We made this figure again.*

**Line 259: This should read Figure 5 to 9**

*Response: Corrected*

**Line 266: What do you mean by “extensive and continuous areas”?**

*Response: We agree with you: the term is not clear. We made a mistake with this expression because Amazon region is the largest in Colombia, but Orinoquia is not, and neither is continuous. We rewrote the sentences including the characteristics of MACHISPLIN predictions.*

**Line 273: What do you mean with the acronym EM? Should it read ME? But then the sentence does not make sense.**

*Response: You are right, the acronym was incomplete. It should be EML (Ensemble Machine Learning), as was defined in line 147 (methodology section): MACHISPLIN and landmap. The sentence was corrected as follows:*

*“The ME values were closer to zero, showing an improvement in the prediction; however, in this ensemble model, predictions of silt fraction had the highest bias, which is a different behavior of EML algorithms (MACHISPLIN and landmap), where sand fraction had the most biased predictions”*

**Line 373:** In line 273 you wrote: “RMSE values decreased for all PSF and standard depths, which means a raising in the precision of the map.” And here you write: “However, layers of 0-5, 5-15, and 15-30 cm obtained the best results”. Maybe I missed something, but to me this is contradictory.

*Response: In line 273 we were taking about the comparison between spatial ensemble maps and EML or SG products maps. While, in line 373 we are referring only for EML maps. To clarify this, we rewrote the first sentence in line 273 follows:*

*“The external validation for the spatial ensemble maps showed an improvement in their metrics vs. the maps using EML algorithm or SG product (Table 5)”*

**Line 392:** I really appreciate it that the authors provide the R code for this work.

*Response: We recognize that provide the R code is necessary and helpful.*