General comments:

The article deals with the generation of spatial maps of soil texture fractions at the country scale (Colombia) using an ensemble machine learning approach. The best explanatory covariates are selected for each transformation at all standard depths through a recursive feature elimination. The modeling process is explained descriptively and global SoilGrids products are also included in the validation process.

The article is well written with the introduction and discussion sections connected. However, it is far from a basic verification. As stated in the article, a logarithmic function was applied before modeling, and predictions were made with an ensemble algorithm such as MACHISPLIN. Then, % Clay, Silt, and Sand were obtained with a transformation function. If I'm missing something so far, please let me know if you reply.

You know, in the beginning, you checked a total of 100, considering consistency, in each layer (2.2 Data harmonization and transformation, Line 130). Well, I invite you to think about it by doing this on the raster maps you produce and putting your head between your hands. For this, for example, add three fractions at 5 cm multiples with any GIS program and raster calculation tool. Are the results in all 100? What does it mean in terms of uncertainty to see a distribution of this?

If you put the result of this in the article, you can find the opportunity to compare with Amirian- Chakan et al., (2019) and perhaps deepen your discussion. https://doi.org/10.1016/j.still.2019.06.006

The article will be carefully examined from different parts by many valuable anonymous reviewers. Verification with large-scale enterprise and global products is appreciated. However, making these processes more understandable is necessary for journal readers. The special issues I have mentioned below are related to writing.

Specific comments:

Line 37-38; Any soil scientist can use this phrase. So how do we measure it in the field? *with our hands*. However, the fractions are estimated separately for the study. In line with this, a new study that can be referenced to this sentence has been published very recently. You can take care of this. Use such a general sentence if you are interested in predicting classes of soil texture. We can't know how many % Clay by hand, we can separate the "Clay" class (or Clay Loam).

Related work: https://doi.org/10.1016/j.catena.2022.106155

Line 43-44; Are you the first to state that soil texture is costly, and the sampling process is time-consuming in large-scale studies? Please cite the source where you read this or could be an example.

Line 51; weathering or formation? Maybe, covariates can be surrogate soil formation factors. "Weathering" defines the higher deeply effective process through the profile.

Line 58-59; This is because soil texture has a compositional data structure. Clay, silt, and sand can be mapped separately in "continuous" data type, as well as spatial maps of the probability of occurrence of classes can be produced. You can show sources working probability maps of soil particle fractions and Categorical soil texture classes in continuous data type.

Related works: (Of course, you are not limited to these.)

https://doi.org/10.2136/sssaj2017.04.0122

https://doi.org/10.1016/j.geoderma.2016.09.019

https://doi.org/10.1007/978-3-030-85577-2_55

Line 67-68; Have these words been said before? for example;

https://doi.org/10.1111/ejss.13071

Line 75-80; Include references to the existence of this information.

Line 144 Table??

Line 150; Section 2.3. How did you generate these environmental variables? Which programs did you use? SAGA or QGIS with open access? In addition to these, did you use a commercial product? (ESRI, ArcGIS). Or it could be on a different program around the world.

Line 194; Section 2.5. Why did you resample after downloading the Soilgrids data in the validation part? Didn't you use the data in each depth layer at the locations where your current approximately 4000 data falls?

Are your actual values the ones in your dataset, and the predictions from the Soil grids? Of course, the Ml predictions are estimated with the ensemble model. Can you elaborate here? Because including the downloaded layers of soil grids in the "resample" process changes those values to a certain extent.

Line 225; Table ???

Line 235 Table ??? There are many like this in the article.

You stated that you have subtracted the estimations from the observed values for the error. It is given as % on the maps (Figure 4-8). Clarify