

Earth Syst. Sci. Data Discuss., referee comment RC4  
<https://doi.org/10.5194/essd-2021-437-RC4>, 2022  
© Author(s) 2022. This work is distributed under  
the Creative Commons Attribution 4.0 License.

## Comment on **essd-2021-437**

Anonymous Referee #4

---

Referee comment on "Colombian soil texture: building a spatial ensemble model" by Viviana Marcela Varón-Ramírez et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2021-437-RC4>, 2022

---

This study used digital soil mapping framework to generate the first texture maps at five depths of Colombia. They used additive log-ratio (ALR) transformation on sand, silt, clay content to develop models. Two ensemble machine learning methods (MACHISPLIN and landmap) and predicted maps from SoilGrids were compared and a spatial ensemble function was created to select the best model for each pixel in the final maps.

I have some specific questions:

**Introduction:** The introduction is too long, and it should be significantly shortened. Some general sentences can be removed, for example, lines 97–101, Models or algorithms for digital soil mapping are evaluated ...

The **language** should be improved. Some sentences are incorrect. For example, line 109, "Understanding which are the prediction algorithms and approaches yield lower error levels at the pixel level..." delete "are the".

**Methodology:** This section is clear and easy to follow, but more details should be provided.

Line 123, how did you sample the profiles? Based on identified horizons in the field?

Line 124, what method was used to measure the clay, sand, and silt in the laboratory?

Line 131, five standard depths (5, 15, 30, 60, and 100 cm). I think it is better to change it to 0–5, 5–15, 15–30, 30–60, 60–100 cm throughout the paper.

Line 146, why was clay used as the denominator? Is there any difference if you use sand or silt?

Line 150, 83 environmental covariates, please provide information on how you obtained these covariates and description of these covariates. For example, there are 6 lithology and 10 soil orders, what are these? What are the 5 oblique geographic coordinates? How did you adjust the covariates to 1 square kilometer? Which resample method did you use?

Line 154, Table ??. None of the Tables are clearly mentioned in the text.

Line 194, add the original resolution of SG and which method did you for the resample? Do you know roughly how many profiles the SG used in Colombian area?

Line 195, “external validation”. I think it’s more common to say “independent validation”.

Line 200, is AVE the same as R<sup>2</sup> (coefficient of determination)?

Line 206, so the error maps in Figs. 4, 5, 6, 7, 8 are from interpolation of residuals using ordinary kriging. In a widely used method – regression kriging, the final map is obtained by adding up the regression map and residual kriging map. Have the authors considered doing the same thing? In addition, the error map is not regarded as uncertainty map. Have you considered calculating model uncertainty using other methods, e.g., bootstrapping?

## **Results:**

Line 250, it should be “decreased with increasing depth”. Similarly, line 257, “the RMSE increased with increasing depth”.

Table 4 and 5, why do you use “adjustment parameters”?

## **Discussion:**

Line 288, "While many studies focus on mapping soil properties such as pH and organic matter, less studies focus on comparing and testing approaches for maximizing accuracy." I think there are many papers focusing on method development and improving the accuracy.

Line 298, "As depth increases, the soil thins, and the proportion of clay and silt rises.". Unclear sentence. What does "soil thins" mean?

Line 380, Fig. 1 is first cited in the discussion. It should be mentioned before Fig. 2.