

## Comment on **essd-2022-291**

Jinshi Jian (Referee)

---

Referee comment on "Improving the Latin America and Caribbean Soil Information System (SISLAC) database enhances its usability and scalability" by Sergio Díaz-Guadarrama et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2022-291-RC1>, 2022

---

The manuscript "Improving Latin American Soil Information Database for Digital Soil Mapping enhances its usability and scalability" submitted to ESSD described a method to identify the main problems in the SISLAC profiles occur systematically in Latin American countries, and provided a work flow to identify the errors in SISLAC, and finally, the authors carefully checked the errors in the SISLAC database and provided a quality improved SISLAC. This work shows the potential of improved soil databases for the generation of spatial information such as SOC or any other property which have been surveyed in existing regional or national scale soil datasets, and it has the potential to improve the global scale soil datasets. I only have few minor suggestions for the authors to consider and to correct. Other than that, I believe this work contributed to improve the quality of an existing soil dataset and their works is important in data science community.

Some minor suggestions:

Line 160: how about the sites coincided with their respective country, but may have other issues?

Line 162: Figure 3c is an example of coordinates inverted, but why it was marked as correct in the figure (marked as  $\checkmark$ )?

Line 174: can you explain when and why gaps exist?

Line 314: "This work is a effort" should be "This work is an effort".

Line 314-324: this paragraph talked about improving SISLAC contribute to a better data in the region (national results such as Colombia, Ecuador, and Argentina), how about its contribution to the global soil dataset? Are SISLAC be included in the global soil datasets such as SoilGrid, SoilGrid2, HWSD? How and whether the approach used in this study can be applied to improve global soil datasets?

Line 322: "Y. Zhang (2020)" should be "Zhang (2020)", check this issue for the entire manuscript, please.

Discussion: I suggest that subtitles can be added to increase the readability of the discussion.

Captions of some tables and figures are too simple, and the necessary descriptions should be added to make the tables and figures self-explanatory.

Table 1: it has a period sign (.) at the end of the table caption, but table 2 does not has one, same issue for figures, please check all figure and table captions.

Table 2: PDDL, ODC-By, ODC-ODbL, CC-BY, CC-BY-NC, CC-BY-NC-ND; those are all acronyms, they should be explained.

Table 4: can you also give an example of gaps between layers exist?

Table 5: "Assign the value of the upper limit of the last layer plus 10", need to explain why "plus 10".

Table 6: for the first case (Organic layer), I see no difference between "Inconsistency" and "Correction Guideline". Should the top be "-5" in the correction guideline column? (i.e., organic layer should be -5 to 0).

Figure 3: in the brackets, panel a, b, and c were explained, why there is no description about panel d? Panel c was an example of coordinates inverted, why labeled as  $\sqrt{}$  ?

Figure 8: this figure looks not correct, should y axis "Residual" rather than "Predicted values"? And what are dashed lines and solid lines? They should be explained in the figure caption. Why the solid line is necessary in this figure?