

## Dear revisor (RC1):

We greatly appreciate the time you have devoted to our manuscript. Below we respond to the suggestions you indicated, which served to improve our manuscript. There are two important changes in the document; the first is that in a joint effort with FAO's Latin America and the Caribbean Soil Partnership, a review was made of the databases available in the region and we were able to consolidate a larger database, which has grown from 41,000 records to almost 67,000. This is reflected in the new manuscript along with the new DOI of the dataset, which are made available to the soil science scientific community under the FAIR (Findable, Accessible, Interoperable and Reusable) principles. The second is that at the suggestion of another reviewer, the digital soil mapping part has been excluded in order to focus on the description of the SISLAC database and the methodology used for its analysis. Once again, we thank you for your time and comments that have enriched this work.

## General comments from RC1

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:

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

Response: We recognize that there could be other issues with the sample location. However, we are working with a database of a continental region; at this level, the only spatial check that we can perform is the one of belonging. There is no way to check that they are not correctly located because, in the original database, we just have the "country" as a variable; we do not have another specific variable.

### 2. Line 162: Figure 3c is an example of coordinates inverted, but why it was marked as correct in the figure (marked as v)?

Response: (Line 162) the symbol in the image has been changed, the circle is marked with "!" indicating alert. In the paragraph, it is clarified that they were checked by inverting coordinates, and when these profiles coincided, they were validated.

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

Response: this part was explained in Line 173 Gaps can occur for reasons such as: the data was not taken at the site, loss of data in the office, or error or omission in transcription.

**4. Line 314: “This work is a effort” should be “This work is an effort”.**

Response: This part was removed.

**5. 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?**

Response: The new version is free to use without restrictions (we provided a DOI for downloading the dataset), and one of the motivations for this work is that it can be used by as many users as possible, including those who can integrate it into global initiatives such as WoSIS or GloSIS. These additions are not within the scope of this work at the moment.

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

Response: There are two authors with the same surname and the same year (Y. Zhang (2020) and Z. Zhang (2020)), so it is specified with the initial of the first name. Additionally, the list of references was reviewed.

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

Response. Subtitles have been added.

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

Response: The necessary descriptions were complemented.

**9. 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.**

Response: The titles of tables and figures have been standardized: *Table (or Fig.) N: Description*.

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

Response: Acronyms were explained on the table.

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

Response: For this table, as there is no error, the gaps are not included in the demonstration.

**12. Table 5: “Assign the value of the upper limit of the last layer plus 10”, need to explain why “plus 10”.**

Response: This rule was defined by expert judgement. Complementated in the table (Now is Table 7)

**13. 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).**

Response: **(Table 8, line 190)** We apologize for a mistake when pasting the image. The image has been changed to the correct one.

**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  $v$  ?**

Response: **(Line 160)** The descriptions are specified in the paragraph above the table and the symbol was changed in panel 3.

**14. 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?**

This comment corresponds to the excluded part and is therefore omitted.