

Comment on essd-2022-230

Anonymous Referee #2

Referee comment on "Harmonized Soil Database of Ecuador (HESD): data from 2009 to 2015" by Daphne Armas et al., Earth Syst. Sci. Data Discuss.,
<https://doi.org/10.5194/essd-2022-230-RC2>, 2022

The HESD is the result of the harmonization of two previous data sets from Ecuador. HESD is more accessible than the previous data sets from which it was generated. This type of data set is very useful and is much needed in regions like South America that are usually underrepresented in global databases. The HESD presented in this manuscript has the potential of being useful for future research and management efforts in Ecuador. Methods are described in a sufficient detail and scripts are supplied in the appendixes. This article is appropriate to support the publication of the HESD data set. Although, the authors should work on many aspects of the manuscript before publishing, listed below:

Specific comments:

Line 21: correct "how make it" for "how to make it".

Line 24: delete space in "13 542" and throughout in this kind of numbers.

Line 42: it should be "for digital soil mapping and for evaluating".

Line 59: it should be "and even less".

Line 64: delete "a" in "into a standard".

Line 71: I suggest rewriting this sentences as "Interoperability is defined as the collective effort of sharing information that can be used to produce and apply newly gained knowledge, and this is achieved by removing conceptual, technological, organizational, and cultural barriers".

Line 78: it should be "and to evaluate".

Line 101: I suggest rewriting this sentence as "In this way, we developed a new soil database with the purpose of constituting a national soil information system following international standards for archiving and sharing soil data."

Line 108: I suggest giving a translation of the projects' names and of the Ministry. This will give the reader a better idea of what the projects are about.

Figure 1: The insert with the American map would look better with an outline, same as in Figure 4. Also, province names are too small to read. I suggest erasing them anyways.

Figure 2: The figure description should say "harmonized database structure". Also, the figure is not easy to understand. In the bottom half of the figure, I suggest adding more informative names for column headers, and delete the example data. This figure should give more an idea of the structure and the meaning of variables in the database. Avoid using abbreviated names for each variable unless you explain the meaning in the figure description. More detailed information on the actual configuration of the dataframe file should be given in supplement material.

Line 132: delete "Wes" since it's his first name and correct this accordingly in the References section.

Line 138: delete "presented in this manuscript".

Line 139: delete "through"

Line 155: Did you mean "we used our own scripts" or something else in the lines of "we used scripts written for this purpose"? Either way, modify this sentence accordingly.

Line 162: I would change it for "Our only option to extract the information with this format was the free access program Smallpdf v 0.19.1".

Line 202: Change this either for "Soil dataset overview" or "Overview of soil dataset"

Line 204: Change this sentence for "With over 4.7 million records that include numeric (e.g., clay content, organic material, soil pH) and class soil properties (e.g., horizon designation, geology), HESD represents the most complete data compilation for mainland Ecuador".

Line 223: it should be "and abundance".

Line 227: it should be "relevant for the evaluation of soil health"

Figure 3: In the description of the figure, what do you mean by "average profile"? Do you mean "profile average"? Also, you should provide the units in the description and also in the figure itself. It is not clear what the percentages in the right-y-axis mean. Font size in general in this figure is too small to read.

Line 243: Try to make more connections between these lines in the paragraph. It seems very disconnected as it is now. I suggest: "Information in HESD could be used to evaluate how land use and management could affect soil properties (Beillouin et al., 2022). As an example, Table 3 shows a statistical analysis of different variables within two different ecosystems: cropland and forest. Although HESD presents the most complete 245 information at the national level, we recognize that there are still information gaps. One of the reasons behind this is that the two original projects from which the soil information was extracted were focused on agricultural areas, for this reason it is assumed that HESD does not fully represent all ecosystems across Ecuador. Further, we emphasize that there is bias in the data since croplands have 9675 points and forests, only 3694. With this in mind, the forest ecosystem presents higher average SOC (27.9 g. kg⁻¹)."

Line 250: How much SOC do croplands have in average?

Line 252: add a "," after HESD.

Line 260: it should be "at a genus".

Figure 4: in the legend it should be "Cordillera Costera Pacífico Ecuatorial"

Line 267: it should be "data points compiled".

Line 270: it should be "with the highest".

Line 271: it should be "24.7 %".

Line 275: it should be "27.8 g/kg".

Line 281: it should be "its representativeness"

Line 287: Did you mean "organic soil layer" instead of just "organic soil"?

Figure 5: Explain in the figure description what is the meaning of the index.

Line 314: I suggest changing "and so represent a more certain reality" with "to better represent the entire geographical range of Ecuador" or something along these lines.

Line 316: correct the citation format, it should be "(Armas et al., 2022)".

Line 323: can you expand on the global scale aspect?

Line 325: it should be "possibility to rescue".

Line 334: is suggest changing "to support" with "which could help to assess".

Line 336: it should be "Oyonarte worked".

Line 339: it should be "Jiménez helped".

Line 341: delete final ".

Line 345: "financiado por la" is in spanish, please translate this.

Line 476: Wes is his first name, this should be McKinney W.

Table 1: the units in Altitude should be masl. Also, correct "Local pending" for "Local slope".

Line 515: Use a proper citation style here.

Table 2: the description of the table should be "HESD Horizons coding conventions and soils property names, units of measurement and their description". Also, under Organic carbon, correct the variable description with "organic carbon. Measured". Under Exchangeable sodium, potassium, calcium, and magnesium correct "hold" for its past participle form "held".

Table 3: Correct the meaning of the initials in "CO = organic carbon" and in "PRES = Effective Depth".

Table 4: In the table description, delete ")" before (2013). Also, the names of the sectors of a same province are difficult to tell apart. It is not clear where one name ends and another one begins. Maybe add another horizontal line between the names or try to make them as distinct as possible. Also, when you give values for a region in its totality, please specify that on the same row by adding the word "Total", for example.