Open Access

Earth System

Science

Data

Discussions

ESSDD

Interactive
comment

Interactive comment on "EstSoil-EH v1.0: An eco-hydrological modelling parameters dataset derived from the Soil Mapof Estonia" *by* Alexander Kmoch et al.

Anonymous Referee #1

Received and published: 26 November 2019

The manuscript of Kmoch et al. describes a methodology for deriving high resolution 3D soil property data of Estonia, which is published with the manuscript. The data basis for the methodology is the National Soil Map of Estonia and the soil properties are derived with a special focus on the parameters necessary for running the SWAT model. These parameters include the saturated hydraulic conductivity, field capacity, wilting point and the USLE K erodibility factor. Such large-scale soil data is highly valuable for soil hydrological and water quality modeling on a scale relevant for decision makers (e.g., national scale). Organizing, homogenization and distribution of soil properties on such a large scale is very challenging and I acknowledge the work the authors did here. However, there are some points that prevent the manuscript from being published at

Printer-friendly version

Discussion paper

the current state.

General comments

1) Structure: The manuscript is very technical and includes too many details. The main step is the transformation of a text based soil classification (Soil Map of Estonia) to soil texture, which is then (together with SOC, bulk density and topographic information) used for deriving the soil hydraulic properties. All the details (especially the grammar definition parts) makes it difficult to follow these main steps. The explanation of all codes for transforming the letter codes to texture could e.g. be a part of the dataset itself as a documentation.

2) Texture: The step of transforming a text coded soil classification into a numerical texture value is a very crucial point of the methodology. All the focus on the "grammar definition" hides the main step of the transformation, which is done with Table 2. However, it is nowhere cited or mentioned how this table was derived. Is it based on the literature or on own data? This table is the main factor influencing your final results, hence it should be carefully described how you come up with this values. Furthermore, this rises the general point of the missing validation of your final soil texture product (that then influences the hydraulic properties). You only validate your grammar-generated codes and do an "expert check". However, the texture itself is not validated with measured data (as far as I understand). You mention on page 9 line 3-9 that you validated it against SoilGrids250m, but it is important to show this validation. An expert check alone is not enough, since other user of your data cannot assess the uncertainty. You need a reliable texture database for validating your results and hence Table 2 (and Table 3). For your SOC prediction you show such a validation and you correctly mention the relevance of validation of your other data on page 15 line 27-29.

3) Data quality: A data paper should be supportive for the dataset and help the users to evaluate the data and its quality (e.g. uncertainties). This is missing at the moment and instead of focusing on the grammar methodology you should rather present your final

Printer-friendly version

Discussion paper

derived hydraulic data with e.g. appropriate diagrams. This includes the uncertainties derived from the texture + SOC classification and also the uncertainty introduced by the pedotransfer functions you used (here ROSETTA). This is also relevant for me as a referee. At the moment for me it is really difficult to evaluate your data in a feasible time. You also mention the problems to derive USDA texture from the old soviet-era based texture system, which ignores the silt fraction and has a different definition for the gravel-sand boundary (page 7 line 29-31). This of course includes a lot of uncertainty, but I understand the benefits of transforming the texture to the often used USDA classification (e.g. usability of pedotransfer functions). I suggest to also include the soviet texture into your data. This can help to evaluate the error introduced by the two different systems and potentially allows to use the data with another "texture transfer function" (different from Table 2).

4) Dataset check: By checking randomly sampled polygons in the final GIS product (.shp) I recognized some problems with the soil layers. E.g. FID 96775 has two layers with SOC and bulk density values are shown in layer 1 and 2. However, texture values are indicated in layer 1 and 3, whereas layer 2 is empty. Similar problem was found in FID 178514 with only one layer but texture values in layer 1 and 2. Please check your data again.

5) SWAT focus: The manuscript focuses too strongly on SWAT. Although the dataset was created for using it with the SWAT model, this is not important in the data paper. Of course you can mention that the presented data is enough for many modeling purposes (e.g. SWAT), but at the moment the focus on SWAT makes the manuscript difficult to understand. E.g. on page 7 line 6-15, just mention that you have defined different layers.

6) Highlight the need for your dataset: You mention similar global or regional datasets (page 2 line 4-28). However you miss to highlight the need of your dataset. What is different from the others or "better" in your dataset? Why it needs a new dataset? For calculating the available water capacity you use the dataset of Tóth et al. (2017) which

is not mentioned in this section. Why is this dataset not usable for parameterizing models in Estonia?

In summary the manuscript should rather focus on the quality of the data than on the methodology of the grammar definition. That does not mean that the grammar definition should not be part of the data or manuscript, but it should be less prioritized. If the authors are able to provide quality and uncertainty measures of the data, I suggest major revisions. Otherwise, although I think such a large scale soil hydraulic dataset is very valuable and I acknowledge the amount of work, the manuscript should be rejected since the quality cannot be guaranteed.

Specific comments

Page 2 line 12-15, 27-28: Please explain the datasets at least a little if you mention them (e.g. what is SOTER or WISE?) Page 3 line 6-8: If you cannot proof it, please delete this sentence. Page 3 line 23-page 4 line 7: Out of context. Please give some introduction and change the structure. Page 9 line 3: What is the second source? SoilGrids250m is just one. Page 11 line 9-10: Please provide a reference for this calculation (SOC = SOM / 1.724). Where does the 1.724 come from? Page 12 line 24: Add reference for the permeability classes. Figure 1 in the lower blue box: "wilting point" not "witing point" Data file "texture_error_lookup.xls": In row 13 (index 11) the erroneous item is "=50/LS2". Is this correct? Because it is displayed as a "#DIV/0!" in Excel.

Data structure

I suggest to reorganize the structure of your data in the repository to make it more structured:

- the main derived map (.shp or other format)

- metadata (e.g. EstSoil-EH_v1.0_attribute_fields)

- folder with figures

- folder that contains all other information used to derive this map (e.g. SOC rf Model;original estionian soilmap, texture errors, rosetta outputs etc.)

- README

Printer-friendly version

Discussion paper