Earth Syst. Sci. Data Discuss.,
https://doi.org/10.5194/essd-2019-192-AC1, 2020

Earth System Science Data Discussions
Open Access

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

Alexander Kmoch et al.

alexander.kmoch@ut.ee

Received and published: 23 January 2020

Response to the reviewers' comments

on 'EstSoil-EH v1.0 An eco-hydrological modelling parameters dataset derived from the Soil Map of Estonia', posted to the Interactive discussion. RC1: 'Referee comment', Anonymous Referee #1, 26 Nov 2019 RC2: 'Referee Comment', Anonymous Referee #2, 21 Dec 2019 We thank the reviewers for their valuable comments. We will address the concerns raised by the reviewers point by point below. The response to the Referees are structured in the suggested sequence: (1) comments from Referees RC1/RC2, (2) author's response A: and author's changes in manuscript.

Printer-friendly version

Discussion paper

Anonymous Referee #1 Received and published: 26 November 2019

(1) RC1: 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 ati the current state.

(2) A: We appreciate the initial assessment and thank the reviewer for the value comments. We will address the concerns raised by the reviewer point by point below.

(1) RC1: 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) A: We agree, that we have included a lot of very technical information. We believe, that for the soil community in Estonia, and for countries which have similar national datasets like the original soil map of Estonia, these technical descriptions are very interesting and a topic for deeper discussion and evaluation as well. With that said, it is important to acknowledge and emphasize, that the source dataset – the original soil map of Estonia- as described in the article, is not based on modelled but on fully

observed data (e.g. texture, soil profile depth, rockiness, presence of organic layer etc). Systematic mapping of Estonian soils to produce soil map in scale 1:5 000 and 1:10 000 was started in 1954 (Reintam, L., Rooma, I., Kull, A. & Kõlli, R. 2005. Soil information and its application in Estonia. In: European Soil Bureau. Research report. 9, 121-132), with most intensive field studies in period 1965-1969. Generally field mapping was carried out in scale 1:10 000 but in hilly or undulating areas with higher soil diversity in scale 1:5000. In 1982-1988 older mapping data was updated and new areas were included with full-area soil quality (primarily fertility, rockiness, water regime, texture, erodability) assessment. In 1988-1990 soil field studies were performed in non-arable land and new mapping of ameliorated land. Forest soils were mapped in period 1976-1989. During large-scale field mapping of soils the texture was determined in situ based on organoleptic methods and for reference profiles laboratory analyses were performed. This enabled calibration between texture defined by organoleptic method by each researcher participating in field survey and texture determined in laboratory (Estonian Land Board, Explanation to the soil map, https://geoportaal.maaamet.ee/docs/muld/mullakaardi_seletuskiri.pdf?t=20091211092214). As a result of large-scale soil mapping, 119 soil varieties in Estonian national classification system have been distinguished and more than 500 combinations of textural status have been described. About 10,000 profiles (1 profile per 330ha) have been sampled and analysed for characteriation of mineral soils (Reintam, L., Rooma, I., Kull, A. & Kõlli, R. 2005. Soil information and its application in Estonia. In: European Soil Bureau. Research report. 9, 121-132; Reintam, Loit; Kull, Ain; Palang, Hannes; Rooma, Igna (2003). Large-Scale Soil Maps and a Supplementary Database for Land Use Planning in Estonia. Journal of Plant Nutrition and Soil Science-Zeitschrift Fur Pflanzenernahrung Und Bodenkunde, 166 (2), 225−231.). Thus, the texture codes and soil types assigned to the ca. 750000 mapped soil units (polygons) are based on many decades of in-situ land surveying practices and describe quite literally the physical state of the soil based on in-situ assessment. Now, in the article and new dataset, the quality of the extraction and derivation process of the initial texture values is obviously a very critical step for the whole dataset. In many cases, scientists in Estonia have used their own “scripts” to get some numerical data out that they need for their study areas, however, no standardised approach as presented in our paper is currently available. (3) We added this clarification to the article in section 2.1, However, we came to the decision that reducing the section 2.2.2 of preprocessing the texture codes, where the bulk of the already condensed technical description of the grammar is contained, would not improve overall readability in contrast to one of the original intentions of this dataset – to demonstrate the creation of a numerical dataset from an existing observed dataset.

(1) RC1: 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.

(2) A: We acknowledge the case presented here by the reviewer. As there are several arguments outlined by the reviewer we respond line by line:

Ll1-3: We partially agree with the reviewer that the transformation step from the extracted Estonian texture codes to the assignment of numerical values for the fine texture fractions was not well explained, however assigning the numerical values of the sand, silt and clay fractions per texture class is based on Estonian soil experts' knowledge. The emphasis on the grammar was intended to make clear that we had to put more work in to actually retrieve these codes in the first place. We added a description to make it more explicit.

Ll4-6: Lines 5-9 on page 8 are intended to describe that Estonian soil experts (co-authors Arno Kanal & Alar Astover, from the two main natural sciences universities in Estonia which are actively doing research and teaching in soil sciences) technically can assign these fractions to the historically recorded Estonian texture codes. And here again, the emphasis is on the fact, that the original soil map of Estonia is representing observed data. The recorded texture codes represent the in-situ assessment of the national soil mapping and surveying efforts. Explaining the Estonian texture code system would include a lot of technicalities and Estonian language terms, which was not desired. However, we acknowledge that this was not as clearly stated as it could have been. We improved the clarity on the creation of the table.

Ll7-15: "the texture itself is not validated with measured data (as far as I understand)" - We believe, the concern raised here again touches the understanding that the original soil map of Estonia is indeed observed data. But that is why we have to be more clear on how well we extracted the existing texture information from the original soil map. Therefore, we found it important to validate the textual extraction and not so much the decision for the assignment of fraction values for each code. Thus, we described it as our best efforts. In addition, there are detailed studies on reference soil profiles in Estonia, Latvia and Lithuania that relate original soil texture, so called Katchinsky texture system (Kachinsky NA. 1965. Fizika potchv [Soil physics], Vol. 1. Moscow: Moscow University Press[in Russian].) to USDA soil system (Calhoun, T.E., Ellermäe, O., Kõlli, R., Lemetti, I., Penu, P. & Smith, C.W., 1998. Benchmark Soils of Estonia Researched thru Baltic –American Collaboration. Problems of Estonian Soil Classification. Transactions of Estonian Agricultural University, 198, 76-114) and erosion modelling case studies where based on laboratory analyses transfer functions from Katchinsky to USDA texture classes were developed (Laas, A. & Kull, A. 2003, Application of GIS for soil erosion and nutrient loss modelling in a small river catchment. In: E. Beriatos, C.A. Brebbia, H. Coccossis, A.G. Kungolos (Eds.). Sustainable Planning and Development (525−534).. Southampton, Boston: Wessex Institute of Techonology Press.). Relationship between Katchinsky and Atterberg systems are provided by R. Kask (Kask, R. 2001. On the English Equivalents of the Estonian Terms for the Textural Classes of Estonian Soils. Journal of Agricultural Science, Vol. 14, 93-96. http://agrt.emu.ee/pdf/proceedings/toim_2001_14_kaskr.pdf) But we acknowledge the reviewers desire for numerical statistical hard data which could reveal how well and accurately the soil surveying has captured soil texture in these descriptive codes through the times. But we think, this would be beyond the scope of the current dataset because it would require inclusion of additional independent georeferenced datasets with USDA soil texture classes defined based on laboratory analysis of particle size distribution.

We added additional explanations and background to section 2.2.4.

(1) RC1: 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 derived hydraulic data with e.g. appropriate diagrams.

(2) A: We agree, the data quality is not consistently reported. The texture data and soil types are observed data of many years of national surveying activities. The original observations were classified into the Estonian texture code system based on Katchinsky (1965) soil particle size standards at the time of observation (not by us). We "just" translate" the texture codes back into numbers. Therefore, we don't see a possibility to explicitly display uncertainties related to that process, as we take these as observed data and thus we can make only short reference to another study what shows that achieving 5% accuracy in organoleptic determination of clay content for lower value classes while possible error increased in case of heavy texture classes (Kokk, R.

1987.Soil texture of Estonian soils, its determination and classification. Estonian soils in figures VI. In Estonian: Eesti muldade lõimis, selle määramine ja klassifitseerimine. Eesti mullastik arvudes VI). However, we will compare the sand, silt, clay and coarse fragments fractions with other soil datasets and report on standard deviations and R2. The uncertainties for the SOC predictions are reported. The uncertainties of BD are directly related to the SOC uncertainty. AWC is directly derived from Toth et al. 2017 EU-HydroSoilGrids and we will report the cumulative uncertainty based on our aggregation. K is predicted by ROSETTA based on our derived numerical texture values. We will report the uncertainty from these predictions.

(1) RC1: 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).

(2) A: There is no error to be assessed from the translation from one system into the USDA system. The numerical values for sand, silt and clay fractions were assigned to the Estonian texture system codes, then the values were also used to select the appropriate USDA classes based on the texture triangle. .). Relationship between Katchinsky and Atterberg systems are provided by R. Kask (Kask, R. 2001. On the English Equivalents of the Estonian Terms for the Textural Classes of Estonian Soils. Journal of Agricultural Science, Vol. 14, 93-96. http://agrt.emu.ee/pdf/proceedings/toim_2001_14_kaskr.pdf). It is not possible retro-

spectively to redefine minor differences in boundaries between different classes between texture systems, but we consider natural variation of texture within the soil mapping unit in scale 1:10 000 more significant than that of different texture systems. Texture transfer rules (Table 2 in manuscript) to get from Estonian texture classes to USDA particle size distributions were composed by authors according to Estonian guideline “Field Soil Survey – Muldade väliuurimine” (Astover et al. 2013) where matches of Estonian/Soviet and USDA/FAO classes for field survey is provided. In our opinion it is appropriate approach for data conversion in used mapping scale. We agree that it might increase uncertainty for point data but should be not major problem in case polygon data.

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.

(2) A: Thank you for pointing these errors out. We believe they were introduced when making assumptions about layers that have no depth reported, but are above or below a layer that has a depth reported. We are uploading an updated dataset.

(1) RC1: 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.

(2) A: We acknowledge the reviewer’s suggestion, we generalised to the need for ecohydrological modelling, and reduced the focus on SWAT throughout the manuscript.

(1) RC1: 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?

(2) A: Other datasets are available for use in Estonia-based modelling contexts, that is correct. However, vector-based EU soil datasets are very coarse and excessively generalise large parts of the diverse Estonian landscape. High-detail datasets such as Soilgrids are themselves based predicted on a grid 1km/250m, and not based on observed data. The presented dataset is of very high spatial detail based on the original Estonian national soil map, which was created from directly surveying all of Estonia. Thus, our presented dataset much more spatially related to the landform/landuse observed there. Furthermore, the textures and SOC/BD values are directly derived from reliable observed data samples from Estonia, with a reproducible workflow, whereas this is not true for many other reported soil dataset that covers the area of Estonia. Furthermore, the method created to translate original hard copy soil map (with traditional textual codes) to digitally readable GIS-based map can be used by several other countries (e.g. Latvia, Lithuania, Ukraine etc) and this enables spatially more explicit modelling of ecosystems.

We added the explanation as highlights to the manuscript in the introduction.

(1) RC1: 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.

(2) A: We thank the reviewer for the valuable comments. However, we want to turn the interest of the reader also to one of our original motivations, which was to provide methodological approach of getting from legacy qualitative soil map to quantitative functional maps.

(1) RC1: Specific comments:

(1) RC1: 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?) (2) A: added to the manuscript

(1) RC1: Page 3 line 6-8: If you cannot proof it, please delete this sentence.

(2) A: deleted.

(1) RC1: Page 3 line 23-page 4 line 7: Out of context. Please give some introduction and change the structure.

(2) A: added to the manuscript

(1) RC1: Page 9 line 3: What is the second source? SoilGrids250m is just one.

(2) A: The two sources were the manually decoded dataset from the paragraph before and the second was SoilGrids250m. We rephrased.

(1) RC1: Page 11 line 9-10: Please provide a reference for this calculation (SOC = SOM / 1.724). Where does the 1.724 come from?

(2) A: Coversion factor 1.72 is widely used universal value (Soil Analysis: An Interpretation Manual. Eds. K. I. Peverill, L. A. Sparrow, D. J. Reuter. CSIRO Publishing, 2001; Soil Carbon Dynamics: An Integrated Methodology. Eds. W. L. Kutsch, M. Bahn, A. Heinemeyer. Cambrige. 2012), however we acknowledge that the real value varies slightly between soils.

(1) RC1: Page 12 line 24: Add reference for the permeability classes. Figure 1 in the lower blue box: "wilting point" not "witing point"

(2) A: we updated the figure and added the reference

(1) RC1: Data file "texture_error_lookup.xls": In row 13 (index 11) the (1) RC1: erroneous item is "=50/LS2“. Is this correct? Because it is displayed as a “#DIV/0!” in Excel.

(2) A: it is not erroneous. The cell code is “=50/ls”, which is an invalid texture code. Excel interpretes the “=”

(1) RC1: 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

Anonymous Referee #2 Received and published: 21 December 2019 Summary

(1) RC2: In the manuscript by Kmoch et al. a new countrywide soil dataset for Estonia at 1:10000 scale is presented. Those soil properties are provided which are the most frequently required soil input variables for eco-hydrological modelling, focusing on providing soil data for the SWAT model. The data originates from the Soil Map of Estonia vector dataset (1:10000), which includes information on soil types according to Estonian soil classification, soil quality, number and depth of soil layers, information on course fragments and Estonian texture classes. Numerical soil properties are derived or through using characteristic values of certain soil groups or computing them from available information, or if data is not available for calculation, data of external dataset is used.

General comments

The scale of the presented soil dataset is outstanding. Detailed information about coarse fragments is unique. Descriptive or categorical type information originating from soil survey is very valuable even if uncertainty is generated when those are converted into quantitative data. The manuscript presents method to derive input information from soil survey data for those models, which require quantitative information about soil properties. This kind of data transformation has several difficulties which authors had to face. Significant amount of work has been put into the construction of the presented dataset, which has to be acknowledged. The work deserves to be published after major revision. Please find hereinafter suggestions for consideration.

(2) A: We appreciate the initial assessment and thank the reviewer for the value comments. We will address the concerns raised by the reviewer point by point below.

(1) RC2: Terminology used in international literature should be adapted in the manuscript. It is not clear what authors mean by "complex text codes" in the abstract.

(2) A: The Estonian texture information field in the original soilmap's attribute table is comprising not only of one actual texture class, but joined with classifiers for the rock content, peat soils and distinct compositional layers and their depth. Visual examples of the meaning "complex text code" of the soil map are shown by Reintam, L., Rooma, I., Kull, A. & Kõlli, R. 2005. Soil information and its application in Estonia. In: European Soil Bureau. Research report. 9, 121-132 and Reintam, Loit; Kull, Ain; Palang, Hannes; Rooma, Igna (2003). Large-Scale Soil Maps and a Supplementary Database for Land Use Planning in Estonia. Journal of Plant Nutrition and Soil Science-Zeitschrift Fur Pflanzenernahrung Und Bodenkunde, 166 (2), 225−231.) and as an example of texture code of Skeletic Leptosol is shown "ls_110-20/pk;r_4ls_1" which indicates presence of sandy silt loam, layer depth, solid limestone bedrock, rock content 50-7%. We have included more information in the manuscript regarding the texture attribute field in the abstract, and sections 2.2.2 and 2.2.4.

(1) RC2: Please provide more precise information about the meaning of "soil profiles (e.g., layers, depths)" "layer information", which is mentioned in the abstract and introduction. Under materials and methods section authors mention that potential fertility

was mapped, in the abstract and introduction soil quality is mentioned. It has to be clarified which soil property with which method was mapped, and reference or detailed description on how it was derived is needed.

(2) A: The field surveys for each polygon were made by sampling the soil to a default depth of one metre and the describing the visual and organoleptically defined texture of the soil from the surface to the depth of one metre. This is considered the observed soil profile. The assessed information was then noted on the field data sheet.

We added more information to the section 2.1.

(1) RC2: A table including metadata would be very informative in the manuscript, in which variable name, file name, description of variable, units of measure, reference, etc. could be included, e.g. meta file of SoilGrids. The "EstSoil-EH_v1.0_attribute_fields.txt" file could be a starting point for that.

(2) A: We thank the reviewer for the suggestion and will add an overview table for the derived variables.

(1) RC2: The authors could put into context the novelty of providing data at 1:10000 scale – which scale is outstanding. Information on other national soil datasets – which are considered detailed or high-resolution e.g. https://dl.sciencesocieties.org/publications/sssaj/pdfs/82/1/186, etc. – could be referenced, and the progress presented by EstSoil-EH v1.0 could be highlighted.

(2) A: We added the reference to the introduction in order to highlight how the EstSoil-EH v1.0 dataset preparation relates.

(1) RC2: Regarding the mapped soil properties, the following specific comments could be considered for the manuscript: 1. Soil type: Is it not clear why new soil types were added to the original dataset, how soil type was extended, e.g. was original soil type recoded based on soil profile information included in the dataset? How were Estonian soil types translated into WRB reference groups? Is there a reference document for it?

Based on which soil classification system did you add new soil types and how? Please write down how many soil types were included initially and how many soil types were added. It is not clear how you got 7067 soil types in the attribute table if 120 soil types exist in Estonia. Maybe you meant something different.

(2) A: The Estonian national soil data set describes a base set of soil types, unfortunately in different literature 120-130. The difference is mainly caused because some researchers add to soil types also non-soil surface types (e.g. soilless bedrock; in total 7 types) and/or distinguish some subtypes of main soil types (e.g. Krf – recultivated open pit mining soil). It describes the most generalized level of Estonian soil classification for mapping that can be extended from these main 120-130 classes, incl. level of erosion, slope position, similar to FAO secondary identifiers, but instead of keeping them separate they just extended the main class, as the inherent "grammar" is well known in Estonia. This increased the overall number of explicit soil classes to more than 7000. We "just" extracted the main soil classes again.

(1) RC2: P4 L28: why "Overall soil type group" is differentiated from "Soil type" which is in L20?

(2) A: Original soil type is main category in Estonian soil classification and soil type group includes several soil types with main similar features. We clarified with more information in sections 2.1 and 2.2.1.

(1) RC2: 2. Texture classes: Clarification is needed on how USDA soil textural classes and then sand, silt and clay content were derived. Based on present manuscript Estonian soil textural classes were available from the official 1:10000 scale National Soil Map of Estonia. Estonian soil texture class names were translated using USDA terminology. Based on the Estonian texture class names average sand, silt and clay content were added to each soil layers. Please consider to add USDA texture class names based on the average sand, silt and clay content which characterize the Estonian texture classes. Please provide reference for the definition of the Estonian texture

classes.

(2) A: The process followed a different order: The Estonian texture classes (Katchinsky, 1965 system), based on their known composition (based on how they were encoded at the time of survey) have numerical values for sand, silt, clay fractions assigned. Then, based on these numbers the USDA texture names were assigned. The single important reference is table 2 that designates the sand silt and clay fractions for each known Estonian texture types

(1) RC2: 3. Coarse fragments content: It is not clear how - “skeleton indicator number” was derived from the shape and size of the stones and - “inferred rock content (% of volume)” was derived from “skeleton indicator number”.

(2) A: We acknowledge that this was not well explained. Similar to the Estonian texture classes there exist Estonian stoniness classes, that describe a certain type of coarse fragments within the soil profile. An additional number in connection with this rock type identifier indicates the amount/volume of these rocks in 1kg of soil. We used this to characterise the coarse fragments. We clarified this in section 2.2.4

(1) RC2: 4. Soil organic carbon content: It has to be described why measured SOC data was averaged by soil units in the training dataset for deriving SOC prediction. Was not it possible to use soil profile data to derive the prediction? Predictors used in the random forest method could be listed under materials and methods section. Performance of SOC prediction could be included in a table. Variable importance could be shown in a figure. (2) A: As several of the Estonian soil units (which contain the predictor soil profile information) would have contained a significant amount of sampling points for SOC (basically a few experimental trenches on agricultural fields) all the predictor variables (including the soil type, texture and topographical variables) aggregated for these units would be identical, this would have created a very strong bias in the training of the model. Therefore, SOC data points that lie within the same soil unit were averaged.

(1) RC2: 5. Bulk density: It is mentioned that BD is calculated based on texture and SOM, but texture is not included in Equation 4. It has to be considered that moist bulk density is required for SWAT.

(2) A: Depending on the formula indeed we are not using texture at all. Our real SOC/SOM field measurements and thus the modelled values are for dry bulk density. Removing the focus on SWAT as suggested by Reviewer1 would mitigate the conflict to provide moist bulk density. We state that this is dry BD, based on the dried SOM measurements.

(1) RC2: 6. Potential fertility: It is listed under materials and methods, but not included under the results. Reference or description for the computation would be needed.

(2) A: It was a historical data field in the original soil map. We did not use this field as the reliability and its original calculation is questionable. We remove notion of this field.

(1) RC2: 7. Organic horizon thickness: Similarly to potential fertility, it is mentioned under materials and methods, but not discussed in results section. Do you mean thickness of A horizon or thickness of soil horizon with accumulation of humified organic matter? Please add reference.

(2) A: See above, that is correct. For now we are not considering this field. Possibly we should? It could improve some data regarding organic matter/SOC? However, we are not decoding it. Same for the rockiness ('kivisus')

(1) RC2: 8. Please clearly state for which soil properties the performance could not be analysed because of lacking measured data.

(2) A: Yes, we have to be more precise/consistent about this. We added the reporting of data quality to manuscript.

(1) RC2: Some parts of the manuscript could be simplified by decreasing mainly technical descriptions, e.g.: P6 L3-P7 L3.

(2) A: We partially agree with RC2. However, we deem these technical details are already greatly reduced.

(1) RC2: More descriptive plots and tables could be provided for the readers about the derived dataset.

(2) A: Reviewer 1 also suggested plots regarding uncertainties and data quality. We will add these to the manuscript.

(1) RC2: Language revision of the manuscript would improve its readability. (2) A: The manuscript had already been edited by English native professional scientific editor.

Interactive comment on Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2019-192, 2019.

Printer-friendly version

Discussion paper