SOIL Discuss.,
https://doi.org/10.5194/soil-2018-32-AC2, 2019

Interactive comment on "A review on the global soil datasets for earth system modeling" *by* Yongjiu Dai et al.

Yongjiu Dai et al.

shgwei@mail.sysu.edu.cn

Received and published: 24 January 2019

1. General comments Comment: A review of soil datasets available for Earth system modeling is extremely useful, given the wide application of ESMs in important projects such as the coupled model intercomparison projects (CMIP) serving the IPCC reports, and in view of the challenges of observing soil properties covering the globe. However, the manuscript does in fact not fulfill what it promises in the title. It does not review datasets and compares them quantitatively (apart from selected maps in Fig. 1-2, but a systematic comparison is missing). Instead it discusses in length linkage and digital mapping methods, then how soil observational data in general can be incorporated in ESMs and what challenges arise. This is valuable *ancillary* information, and the manuscript summarizes a lot of important information on these topics. But the main

purpose of the paper is missed. A careful review of available datasets needs to be added, which is of course a major revision: there should be more than the 3 datasets in Tab. 3, unless justified that these 3 are special (for example it would be very illustrative to include all the currently used old datasets as well to know what a difference the new datasets might make). There should be a review also of other data than global maps, as needed e.g. for parameters. Most importantly, however, a quantitative comparison of at least key variables should be included, with useful statistical measure (maps, global mean and variability, latitudinal means, comparison against selected observational high-quality sites, . . .). Ideally, model sensitivity simulations would be run, but this latter point is not essential.

Reply: The purpose of the view is to offer insights to both soil data developer and ESM users. So we discussed contents may be interesting to both sides. We agree that a systematic quantitative comparison is a very important aspect this review should cover. So we will compare a selection of global soil data sets with a focus on the most recent developed ones. We will use one or two 'old' data sets which are used in ESMs in the comparison. It is not necessary to compare all the old data sets because they are based limited and outdated source data. All the old soil data are based on the FAO soil map and no more than 5,800 soil profiles. We are sure that they have poor quality. That is why we highly recommend the most recent developed data sets even without a quantitative comparison. We can see that the newly developed soil data in fig.1-2 have some major differences. It is valuable to compare them even though they may not be so comparable, because the datasets developed by the linkage method give the mean value of a soil class while the datasets developed by digital soil mapping simulated the soil as a continuum (though in nature soil may change abruptly). Nevertheless, we plan to use site observations of WOSISS to evaluate the soil data sets, though these observations are used or partly used in the development of global distribution of soil data. We will compare the key variables used in ESM with useful statistical measure as the review recommended. However, model sensitivity simulations will not be done in this reviews and need to be done in other studies. This review focuses on the global soil property maps in ESMs. We will not extend the content to other data including parameters which is a different topic but valuable. As we mentioned in the manuscript, variables such as soil temperature and To avoid misunderstanding, we will change the title to 'a review on the global soil property maps for earth system model' and modify the corresponding expression in the manuscript. However, we will use the term soil datasets for brevity.

Comment: A method for the review is missing, which leaves the reader in doubt whether he/she has been reading an opinion piece or a comprehensive review. Currently both the selection of mentioned datasets and the selection of ESMs is incomprehensive and not justified in its selection. For the models one could imagine to do a review of all TRENDY LSMs or of all CMIP5 (or even better CMIP6) ESMs and the datasets they are using. For the available datasets some objective criteria should be given as well, e.g. a list of criteria that datasets need to fulfill to be included in Tab. 3 (global, soil type and property x, y, z need to be included,. . .)

Reply: We will describe the selection of mentioned datasets and the selection of ESMs. As we mentioned in the above reply, datasets are chosen according to their data quality and developing time. In addition, the datasets should be freely available. Our currently list of ESMs covers the major LSMs but not all of them because a complete list will be too lengthy. We will check the list of TRENDY and CMIP5 to see if they used different soil datasets as we are only interested in the soil data rather than itself.

Comment: The organization of the sections does not appear logical: Datasets and their usage in ESMs (Section 2) is very good, presenting PTFs as Section 3 promises in l. 105 is also very useful but these PTFs are in fact never presented and compared, just discussed. Section 4 deals with data from the linkage method. why? Why not data from digital mapping as well? Section 5 deals with upscaling to the coarse ESM resolution. This is an important point, but there are many other challenges related to application of soil datasets in ESMs: One obstacle is that observations are not covering the soil depth as deeply as the ESMs and in other layer distributions. Another

that soil observations are derived from present-day, which has confounding effects of both environmental changes (climate, CO_2, nutrient deposition, . . .) and historical land use changes. Would this affect soil thermal and other properties needed as input to ESMs? How should one evaluate ESMs only for present-day then?

Reply: PTFs is not the major focus of this review while there is a very good review on PTF in ESMs which we cited as Looy et al., (2017). Section 4 do not discuss data from digital mapping because it does not have the aggregating problem like the data by the linkage method. Data by the linkage method are derived for each soil map unit and data by the linkage method are derived for each grid. We will add a sentence to clarify this at the beginning of this section. We agree that we should also discuss the lack of deep soil data and the changing of soil properties. For the lack of deep soil data, we do not have very good solution other than extrapolate the values based on the observations of shallower layers. There is not any global soil property map in time-series because we do not have available data. All the available soil observations in the last decades are used in the development of soil property maps without considering the changing environment. So these datasets should be considered as an average state. The effect of environmental change is the topic of quantitative modeling of soil forming processes (soil landscape and pedogenic models). Soil properties are changing but we are now taken it as static in ESMs. As some ESMs already simulate the soil carbon, this may be considered in PTF used to estimate soil hydraulic and thermal parameters. Other soil properties affecting soil hydraulic and thermal parameters include soil texture, bulk density, soil structure and so on, but the change is relatively slow. If we need to simulate the change of soil properties, a coupling of ESMs and soil landscape and pedogenic models will be needed. Otherwise, we need to predict the soil properties in the future using soil landscape and pedogenic models which are small scale models and has high uncertainty. This prediction may also be done by digital soil mapping taken the changing (especially for the future) climate and land use as covariates, which I think is the most feasible one to do.

Comment: How should ESMs deal with observational uncertainty (see comment below)?

Reply: See the reply to the specific comment.

Comment: I think what this paper needs to cover is (0) specifying what ESMs need, i.e. which spatial and temporal coverage, which variables (extending the list of parameters, initial state, evaluation/benchmarking in the introduction) (1) general methodology of deriving this soil information (mostly Sec. 2, PTFs would go in this section as well.) (2) comprehensive, quantitative comparison of available global soil datasets (largely missing) (3) discussion of existing challenges of data usage in ESMs, where one should come back to the list of usages in the introduction: evaluation data for example does not have to have global coverage. The upscaling would be one of several points here.

Reply: We will reorganize this manuscript as the reviewer recommended.

Comment: The paper is not very well written. First, the use of English language is incorrect or uncommon. Second, many expressions are not accurate. Just taking the first sentence as example: “Soil or pedosphere is a key component of Earth system, and plays an important role in the water, energy and carbon balances and biogeochemical processes.”First, it should read “The soil or pedosphere is a key component of the Earth system, . . .” (where “Earth” is correctly written in capitals, while it is not in the title: : :). Second, the carbon cycle is one example of biogeochemical processes, so it should read “ and *other* biogeochemical processes”. I am not correcting any of these language and accuracy errors in the following because they are too numerous. Reply: Thanks for pointing out these errors. We will revise this manuscript and take a language service.

More detailed comments: p. 1 Comment: * "Soil datasets function as model parameters": do the authors mean that model parameters can be derived from soil carbon maps? What parameters are they thinking of?

Reply: we corrected the expression to model inputs. This is the major usage of soil property maps in ESMs (table 1).

Comment: * "are preferred to those by the linkage method for ESMs": not understandable at this point in the manuscript - what is the “linkage method”?

Reply: we modified as “taxotransfer rule-based method’, which may be a more understandable terminology.

Comment: * "to provide secondary soil parameters to ESMs": what are secondary soil parameters?

Reply: we modified it to “derived soil properties”, which includes soil hydraulic, thermal and biogeochemical parameters. And we explained this when “secondary soil parameter” first appear in the manuscript.

Comment: Generally, the abstract does not read like a review of datasets, but like a commentary on challenges of integrating soil carbon datasets in ESMs. As a reader I would have expected an abstract here of types of data, see general comment above.

Reply: we will revise the abstract adding the related contents.

p. 2 Comment: * "However, soil dataset used in ESMs is not well updated nor well utilized yet.": This needs citation of which datasets are used and felt by the authors to be outdated.

Reply: To make this more objective, we added some citation from FAO and globalsoilmap (a community joint effort project). We have explained this in section 2.2: Except GSDE, HWSD and STATSGO (Miller and White, 1998) for USA in Table 1, these datasets were derived from the Soil Map of the World (note that large sections of GSDE and HWSD still used this map as a base map because there are no available regional or national maps) (FAO, 1971-1981) and limited soil profile data (no more than 5,800 profiles), which gained popularity because its simplicity and ease of use. But these are outdated and should no longer be used because much better soil information as introduced in Section 2.1 can be incorporated (Sanchez et al., 2009; FAO/IIASA/ISRIC/ISS-CAS/JRC, 2012).

Comment: * l. 45-48: Kearney & Maino are one specific study for Australia for soil moisture using one new soil dataset. Using this as reference for the entire "Earth system" and for "will improve" in the future is a stretch. Better look for a couple of references and spell them out explicitly.

Reply: This is only an example. We added more citations here: (eg. Livneh et al., 2015; Dy and Fung, 2016; Kearney and Maino, 2018). More examples are given with brief description in section 2.2.

Comment: * "could avoid the possibility of the non-linear singularity evolution of the modeling": this needs to be explained in one more sentence. Do the authors mean that models may have multiple equilibria?

Reply: Yes, it means models may have multiple equilibria. And we also added a sentence: The setting of initial nutrient stocks is a major factor leading to model-to-model variation in the simulation (Todd-Brown et al., 2014).

p. 3 Comment: * "for multiply layers rather than a global constant": This mixes up vertical resolution (-> layers) and horizontal resolution (-> global constant). Be more explicit in your description.

Reply: we modified it as: As a result, ESMs usually incorporate soil property maps (i.e., horizontal spatial distribution) for multiply layers rather than a global constant or a single layer.

Comment: * Is "linkage method" really the proper technical term here? It seems to me it is used in the literature rather for remapping than for linking soil observations to environmental variables. The paper would benefit from a clearer overview of technical terms and methods, if it is meant to serve as a review.

Reply: We used this term for brevity. But it may be misleading if readers are not

familiar with soil mapping. So we also used the other term taxotransfer rule-based method. We added this term when it first appears in the manuscript: The traditional way (i.e., the linkage method, also called taxotransfer rule-based method) is to link soil profiles and soil mapping units on soil type maps, sometimes with ancillary maps such as topography and land use (Batjes, 2003; FAO/IIASA/ISRIC/ISS-CAS/JRC, 2012).

Comment: * paragraph starting l. 93: Vector to raster conversion and remapping to a different resolution are certainly not the biggest or at least not the only obstacles to including soil datasets in models. That models need different variables than those directly observable or that observational datasets cover only a certain depth, which most often is different from the one ESMs cover, are examples of other important challenges. Overall, I feel the sections internally should be a bit better structured, with one topic being covered comprehensively by one paragraph.

Reply: This paragraph served as a brief introduction of the obstacles to including soil datasets in models and detailed description will be given in the later sections. That models need different variables than those directly observable is related to the PTF development. So we did not put here. We added the challenge of incorporating uncertainty in ESMs here. We modified the paragraph as follows: There are many challenges related to application of soil datasets in ESMs. (1) First, soil datasets are usually not appropriated scaled or formatted for the use of ESMs and some upscaling issues, which is the most frequently encountered, need to be addressed. The soil datasets produced by the linkage methods are polygon-based and need to be converted to fit the grid-based ESMs. This conversion can be done by either subgrid method or spatial aggregation. The up-to-date soil data are provided at a resolution of 1km or finer, while the LSMs are mostly ran at a coarser resolution. So upscaling of soil data is necessary before it can be used by ESMs. Proper upscaling methods need to be chosen carefully to minimize uncertainty in the modeling results introduced by them (Hoffmann and Christian Biernath, 2016; Kuhnert et al., 2017). (2) Second, all the current global soil datasets represent the average state of last decades, but soil properties change over time. Soil landscape and pedogenic models are developed to simulate soil forming processes and soil property changes, which can be incorporated into ESMs. The prediction of changing soil properties can be also done by digital soil mapping taken the changing climate and land use as covariates. (3) . (4) Last but not least, the depth of soil observations in soil survey are usually less than 2 m, but the ESMs usually covers much deeper. A pragmatic solution to this issue is to extrapolate the values based on the observations of shallower layers.

Comment: * "Two kinds of soil data are generated from soil surveys: soil polygon maps representing distribution of soil types and soil profiles with observations of soil properties. ESMs usually require the spatial distribution of soil properties, or soil property maps rather than soil classification information.": It is unclear how the information of the two sentences relates. Would this be correct: "Two kinds of soil data are generated from soil surveys: a classification of soil type (usually in the form of polygon maps) and soil profiles with observations of soil properties. ESMs usually require the spatial distribution of soil properties (soil property maps) rather than a classification of soil type." If so please always use the same term for the same information.

Reply: you are right. we modified as follows: Two kinds of soil data are generated from soil surveys: a map (usually in the form of polygon maps) representing main soil types in a landscape unit and soil profiles with observations of soil properties which are considered representative for the main component soils of the respective mapping units. ESMs usually require the spatial distribution of soil properties (i.e., soil property maps) rather than information about soil types. Two kinds of methods, i.e. the linkage method and the digital soil mapping method, are used to derive soil property maps.

p. 4 Comment: * "Soil maps show the geographical distribution of soil types,": I think this is too general, the term "soil map" is not a technical, well specified term. Rather speak explicitly of "soil type maps" to distinguish it from maps of soil properties.

Reply: In soil science, if it is not clarified, soil map refers to soil type map. To clarify

this, we modified to: Soil maps (the term soil map refers to soil type map in this paper) show the geographical distribution of soil types

Comment: * l. 153 ff linkage method: this is a useful description, but hard to read for non-experts. Please improve the clarity of the text. For example: * my understanding is that pedotransfer functions map well-observable to less-wellobservable properties, but here it sounds as if the PTFs are needed to link site-level (profile) observations of soil properties to soil type maps.

Reply: Sorry for the description leading to the misunderstanding. Pedotransfer here has nothing to do with Pedotransfer functions discussed in the late section. We added some notification here: The linkage method (called the taxotransfer rule-based method) is to link soil mapping units or soil polygons and soil profiles according to taxonomy-based pedotransfer (taxotransfer in short, note that pedotransfer here does mean pedotransfer functions which is a different thing) rules (Batjes, 2003).

Comment: * "The criteria used in the linkage could be one or many factors as following [. . .] and so on": this is very vague. Which type of criteria is this: soil physical and chemical properties?

Reply: this is related to the above comment. This is the criteria for linking soil map and soil profiles with all soil properties together.

Comment: * "Each soil type is represented by one or a group of soil profiles that meet the criteria, and usually the median or mean value of a soil property is assigned to the soil type.": Criteria and properties are mixed up here. Isn't it choosing one (or several) property as criteria, then mapping the rest?

Reply: this is related to the above comment. We hope the readers will be clear with the above clarification.

Comment: * l. 165-172: how do these references relate to the examples of "major soil maps" in the introduction?

Reply: these references include both soil type maps and soil property maps, while "major soil maps" in section 2.1 (not the introduction) refers to soil type map only.

Comment: * l. 188 ff: Again, please add clarity. The difference between linkage method and digital soil mapping is not just that the first has the same values across a polygon, but also in what information is used as criteria for mapping: the digital soil mapping uses environmental information, not just physical and chemical properties if I understood it correctly.

Reply: This is also related to the misunderstanding of the term pedotranfer.

p. 5 Comment: * “purity of soil map units is likely to be around 50 to 65%”: which statistical measure is meant by “purity”?

Reply: This is a term used in soil science, which means the percentage of the dominant soil type in a soil map unit. Details can be found in the following website: https://esdac.jrc.ec.europa.eu/ESDB_Archive/ESDBv2/esdb/sgdbe/metadata/purity_maps/p

p. 6 Comment: * Why is IGBP-DIS mentioned here the first time? It should have been mentioned under the linkage or the digital mapping methods (depending on what method is used) before.

Reply: IGBP-DIS is listed in Table 1. It is produced by the linkage method. We added IGBP-DIS under the linkage method.

Comment: * “soil organic carbon stocks at 1m depth”: is it meant “carbon stocks down to a depth of 1m”?

Reply: Yes, we corrected it.

Comment: Fig. 1: remove superfluous information that costs the reader time to read and hides the differences between the panels (the datasets): since the legend is the same for all sand (clay) panels it does not have to be repeated; same for “sand (clay) at 0-30cm (%)”, which is even stated in the caption. “Longitude” (typo!) and “latitude”

are also superfluous information.

Reply we will remove superfluous information.

Comment: Fig. 2: Same comment as for Fig. 1. “s” missing in soilgrids. Why is IGBP not included here as well? A more useful information for modelers would be the total carbon content down to a certain depth rather than units of g/kg.

Reply: we will remove superfluous information and add IGBP and total carbon content.

p. 6 cont’d Comment: * "several most popular ESMs”: give objective criteria for “popular"

Reply: Here we do not have objective criteria. The list will be extended according CMIP5 and TRENDY, but it is necessary to show all of them, only those with different soil datasets.

Comment: * l. 227-229: Again, it should be stated in how far the new datasets are superior over previous datasets.

Reply: This will need quantitatively assessment in the revision.

Comment: * l. 231: "This was started: : :” sounds a bit like advertisement and subjective, certainly other groups have been working on this to some extent for a long time as well. Reformulate more neutrally?

Reply: we modified it to: The Land-Atmosphere Interaction Research Group at Beijing Normal University (BNU, now at Sun Yat-sen University) has put much efforts on this topic.

Comment: * l. 245-253: What is the purpose of these references? Only if they prove model results have improved by the usage of the new soil map is it useful to cite them here.

Reply: These citations are showing the application of the new soil datasets in ESMs, which is stated in the first sentence of the paragraph: In recent years, efforts were taken to improve the soil data condition in ESMs.

Comment: Tab. 1: please fix typos (inconsistent punctuation and capitalization). Add version numbers to LSMs, as usage of soil information may change between versions. Not sure the references are always correct, e.g. LeQuere et al., ESDD 2018 a and b ("Global carbon budget 2017" and "2018", resp) state Reick or Mauritsen as JSBACH references, not Giorgetta.

Reply: we will check this table and make corrections.

p. 7 Comment: * l. 299 ff: if there are no uncertainty estimates, how can you judge soilgrids to be the most accurate one?

Reply: this is depended on the source data and method they used, and will need quantitative comparison in the revision.

Comment: * l. 305: not all models apply PTFs, some directly require these less observable variables as input, as you show in Tab. 1

Reply: It is true. But these variables are also derived by PTFS. To be precise, we modified it to: Earth system modellers have employed different pedotransfer functions (PTFs) to estimate soil hydraulic parameters (SHP), soil thermal parameters (STP), and biogeochemical parameters (Looy et al., 2017;Dai et al., 2013) or used these parameters as model inputs.

p. 9 Comment: * l. 359: The methods have been introduced before, so technical terms like "SMU"should have been introduced in these earlier chapters.

Reply: We introduced it in describing soil type maps: There are many soil mapping units (SMU) in a soil map and a SMU is composed of more than one component (i.e. soil type) in most cases.

Comment: * l. 365: A problem of using subgrid soil information is that ES modelers

do not know how to map them with land use information, which is also subgrid level. This may be the more fundamental obstacle than the computational issues that are mentioned.

Reply: Yes, this will increase the model complexity, too. A possible solution of mapping them with PFT or land cover is here: classify soil according soil properties and get a number of defined soil classes (SFT, n classes) like PFT (m classes); overlay the defined soil classes with PFT and get n by m combinations assuming PFT and SFT are independent.

p. 11 Comment: * "The temporal variation of global soil is quite challenging due to lack of data.": the aspect of temporal changes has not been addressed before and seems out of place in the summary.

Reply: We will add a section about this.

Comment: * "Soil image fusion is also needed to merge the local and global soil maps.": What is soil image fusion? Don't bring new methods in the summary section...

Reply: This is about outlook instead of summary. Soil image fusion is proposed by Hengl et al. (2017), which consider local and global soil maps as components of soil variation for ensemble predictions. We modified this to: Soil image fusion is also needed to merge the local and global soil maps, which consider them as components of soil variation for ensemble predictions (Hengl et al., 2017). A system for automated soil image fusion might take years before an operational system for global soil data fusion is fully functional.

Comment: * " Uncertainty estimation should be included in the soil datasets developed in the future.": of course uncertainty estimates build trust in an observational dataset. But how do the authors recommend should ESMs use such uncertainty estimates other than as criterion for which dataset to choose in the first place? Running multiple simulations combining upper and lower bounds in all possible combinations is too expensive...

Reply: We agree that running ESMs with all possible combinations is too expensive. An alternative to quantify effects of the uncertainty of soil properties on ESMs may be to use adaptive surrogate modeling based on statistical regression and machine learning which costs much lower computing time (Gong et al., 2015; Li et al. 2018). We will discuss this using a section.

Comment: * "The gap between soil data existence and data availability is huge": Reads awkward. Better "The gap between the amount of data that has been taken in surveys and the amount of data freely available is large."

Reply: we modified this to: The gap between the amount of data that has been taken in soil surveys and the amount of data freely available is large.

p. 12 Comment: * l. 482 "like many other data": Too general a statement, remove.

Reply: removed

Comment: l. 481 ": : : which has the most: : :": how do you know? Add reference or justify in other ways

Reply: We added a citation here (Hengl et al. 2017).

Comment: * Arbitrary last sentence. l. 465 already mentions the subgrid issue in ESMs. Is there no more general conclusion that can be given? Otherwise just delete the last paragraph and end with the more "outlook"-like previous paragraph.

Reply: the last paragraph is deleted.

Interactive comment on SOIL Discuss., https://doi.org/10.5194/soil-2018-32, 2018.