SOIL Discuss.,
https://doi.org/10.5194/soil-2018-32-RC2, 2018




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

**Anonymous Referee #2**

Received and published: 27 December 2018

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 inter-comparison 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



Printer-friendly version

Discussion paper

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.

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, . . .).

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 âĂŤ but these PTFs are in fact never presented and compared, just discussed. Section 4 deals with data from the linkage method âĂŤ 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, CO2, 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 âĂŤ only for present-day then?

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

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.

More detailed comments:

p. 1

* "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?

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

* "to provide secondary soil parameters to ESMs": what are secondary soil parame-

ters?

* 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.

p. 2

* "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.

* 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.

* "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?

p. 3

* "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.

* 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.

* 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.

* "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.

p. 4

* "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.

* 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-well-observable properties, but here it sounds as if the PTFs are needed to link site-level (profile) observations of soil properties to soil type maps.

* "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?

* "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?

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

* 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 âĂŤ if I understood it correctly.

p. 5

* "purity of soil map units is likely to be around 50 to 65%": which statistical measure is meant by "purity"?

p. 6

* 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.

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

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.

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.

p. 6 cont'd

* "several most popular ESMs": give objective criteria for "popular"

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

* 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?

* 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.

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.

p. 7

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

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

p. 9

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

* 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.

p. 11

* "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.

* "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...

* " 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...

* "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." p. 12

* l. 482 "like many other data": Too general a statement, remove. l. 481 "... which has the most...": how do you know? Add reference or justify in other ways

* 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.

---

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