Earth Syst. Sci. Data Discuss.,
https://doi.org/10.5194/essd-2020-149-SC1, 2020

Open Access Earth System Science Data Discussions

Interactive comment on "SoilKsatDB: global soil saturated hydraulic conductivity measurements for geoscience applications" *by* Surya Gupta et al.

Attila Nemes

attila.nemes@nibio.no

Received and published: 3 September 2020

After reading the paper I take the liberty of submitting a few uninvited recommendations – not a full review - to the authors while this paper is still in the review phase. I give a lot of credit to Reviewer #2's remarks and I strongly encourage the authors to clarify a substantial number of issues around the database in order to prevent avoidable criticism later.

I congratulate the authors on the initiative and effort – assembling any large and heterogeneous database of the like is a never-ending fight. Yet, I think the documentation of the data currently stops short of where it should be and leaves too many doubts about the actual contents and its meta-information. I try to add rather that repeat earlier

Printer-friendly version

Discussion paper

comments by the Reviewers.

In terms of the data and the database, my first focus is primarily but not solely on Table 2a. It is cited that the 'codes', which I interpret as the field names that are adopted from the USDA NCSS database. I can recognize some of that, yes. However, I need to warn that most of the larger data sources taken advantage here will not hold data that adhere to many of those codes and the definitions behind them in the USDA NCSS database. Just as examples, those fields that have 'clod' in their names will likely not be possible to match due to methodological differences (i.e. clod vs core measurements), and therefore this documentation will be misleading and infuses confusion for later users.

Ever since the first such international databases were published – including those with my involvement – the need and quest remains to be clear and specific about such details as methods, definitions, and the like. The USDA NCSS Soil Characterization Database sets some great example in that sense, but it cannot be unconditionally followed when the data in question are either mixed or do not adhere to those definitions/standards. I strongly suggest revising the documentation accordingly. This is better done now than later exploited by users and/or potentially hindering advancement in science.

Some additional specifics based on Table 2a, which does not cover the entire extent of the database (38 columns of information/data): - hzn_top/bottom appears to refer to horizon/layer designation, and not sample depths as suggested by the description and as also suggested by the examples in Table 2b - db_od: are all the data surely from oven dried samples? - Water retention data (w6, w10, w3, w15): Please clarify the methods and change the code/field names to the appropriate ones, once USDA NCSS is emulated. They have multiple data columns for several of those, differing in methodology. - Particle-size data: were all the data really given in the FAO/USDA format, and if not, then possible to interpolate with no specific challenges? Please confirm. - OC: this has been a source of grand confusion in more than one past database, and the language used throughout the paper is soft about it (at some point only calling it (OC

– organic content). Please be explicit about handling this variable – to what extent conversion was needed from the publications and how it was done. - Ksat: Was Ksat always published in the source? Did it have to be calculated from infiltration data? Please be explicit about the methods, I do not recall seeing it.

Some comments/questions with respect to the pedotransfer part of the paper: With respect to the PTF comparisons, I think the authors left a lot on the table and stripped themselves from greater potential impact. The temperate-tropical comparison is well known, and the field-lab aspect could have been explored much deeper with not too excessive work.

I concur with Reviewer #2 that the PTF part raises more questions than what is gained from it at this stage.

I invite the authors to include discussion on any locations/data for which field and lab Ksat was co-existent and whether those were handled/explored in some specific way. It is rare to have that capability.

I think excluding 15% of the data in exchange for OC to be part of the models could have been an affordable loss – but the authors will likely offer a big-picture response to that. Bad correlation with Ksat does not seem to be unique for this variable.

Could the authors include a third metric for a measure of bias? I can see greater spread in Figure 5 b and d than in a and c, but I cannot readily comprehend the claimed 'bias' from those two plots.

With respect to the offered discussion on lab vs. field results: I can accept the offered reasons as part of the big picture but lack any mention of e.g. measurement scale. Let me simply refer to the work by Ghanbarian et al. (2016) (10.1016/j.catena.2016.10.015) who explored the effect of sample dimensions on Ksat measurements – and that is only the laboratory part of this question. The presence of top-to-bottom connected (macro-)pores in a soil sample can also go both ways! Yes,

Printer-friendly version

Discussion paper

the taker of the sample may be tempted to avoid marcopores/cracks, but a short sample has greater chances for top-bottom connected clusters than a tall one. I just wanted to indicate that there is much more that could/should be added here. In terms of field measurements, methodology may matter a lot as well.

With respect to the offered discussion on temperate vs. tropical findings: Again, I can accept the offered points here, but there is likely more to the differences, and the authors could profit from expanding on this, in case PTFs remain part of this data paper. To mention one – a well-known one – the min-max range of particle-size metrics typically does not allow one to appreciate the differences in textural distribution between prevailing soils of those two climate regions. That very simply makes the tropical soils – and potentially their pore network types un- or underrepresented in any temperate PTF.

And finally two short comments on the text:

I suggest rewriting/reorganizing lines 1-15 of page 13 a bit. I found it very difficult to comprehend it because of the order of values and the many subsequent mentions of CCC and RMSE. Many values are very similar, and for CCC high value is good, for RMSE it is the opposite. Are any of the metrics significantly different between the MPR and RF methods?

I suggest including an explicit warning to the user about the scale of applicability, especially where the assigned quality metric is high (meaning location is uncertain). A difference of 10km looks small at the world scale but may not serve any smaller scale work too well. The true point may almost fall into a different country in some cases.

Interactive comment on Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2020-149, 2020.