Reply to Referee (ESSDD)

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

Anonymous Referee #1

Review of SoilKsatDB: global soil saturated hydraulic conductivity measurements for geoscience applications by Surya Gupta et al. The saturated hydraulic conductivity Ksat dataset that the authors compiled is extremely useful and highly needed. The paper describes the datatset clearly and is well written and easy to follow. The initial analyses done with the new dataset are interesting as well. Some of the figures in the paper can easily be used in lectures on soil hydrology. I checked the csv file of the database (from the website given at the end of the paper) and it contains more columns than described in the paper. This is a bit confusing. I have very few comments on the paper itself and highly recommend publication of the paper after some minor revisions.

RE: We thank the Reviewer for the positive assessment of our manuscript and for the numerous comments and suggestions. In the revised version, we have clarified the methodology and the database description we explain now all columns in the paper that are shown in the database. We have also modified the manuscript based on your feedback/edits on the manuscript. In addition, we have provided answers to your questions as listed below (in red color).

Dataset:

Q: I checked the csv file of the database and found the use of the ? to indicate missing data a bit annoying (even though it can be easily replaced by NaN or some other identifier). In column "hzn_desgn" both "no data" and "?" are used for no data. This is a bit confusing. Also, there are columns that seem to only have missing data and aren't defined in Table 2a: "usiteid", "labsampnum", "layer_sequence", "db_13b", "COLEws", "w15bfm", "adod wrd_ws13", "cec7_cly", "w15cly", "ph_kcl", "cec_sum", "cec_nh4". The column "site_obsdate" isn't defined and explained in Table 2a and it isn't clear what this is as it clearly isn't a date. Similarly, the columns "hzn_desgn", "w15bfm", wpg2" are not described in Table 2a, nor shown in Table 2b.

RE: We made several changes of the database. We have now removed all empty columns from the KSatDB. "?" and "No data" was replaced by "NA". All columns sown in the database are now explained in Table 2a in the paper.

Q: I would find it very useful if the database also contained a column with your classification of the climate and the calculated texture % based on the Nemes et al. method. This would mean less double work for other researchers who want to use the data (and possible errors).

RE: Thank you for this suggestion! We have now added an overlay of points and majority of www.OpenLandMap.org layers in the table "sol_hydro.pnts_horizons_rm.rds" (https://doi.org/10.5281/zenodo.3752721). This will be continuously maintained and extended with other layers. A complete list of layers and their codes is available at: https://gitlab.com/openlandmap/global-layers. Following the Reviewer's suggestion, we have now incorporated climate classification information in the database: file name "sol_ksat.pnts_cl_pedo.csv" (see version 0.3, https://doi.org/10.5281/zenodo.3752721). Regarding the calculation of texture % based on the Nemes et al. method, we calculated the texture % only for the UNSODA dataset (as this was not provided in first place) using the methodology by Nemes and co-authors. All other datasets already have texture % information. We have modified the methodology section to make this point clear (see P3L28-30).

Q: I would find it useful if the headers contained not only the name but also the units but this is just a personal preference that helps to avoid errors when reusing the data.

RE: We prefer to separate the units from column names. Instead we recommend the user to refer to the documentation/metadata which is now listed both on the dataset repository (https://gitlab.com/openlandmap/compiled-ess-point-data-sets/-/tree/master/themes/sol/SoilHydroDB) and in the paper (see Table 2a and 2b). We believe no users should have a problem locating the metadata and using the data correctly.

Paper:

Q: In the introduction, the authors argue that it is important to have accurate information on the location of the data points but this argument is not clearly supported by examples. The authors invested a lot of effort in obtaining this data for sites that were already included in other databases but for which the database didn't have the location information. I think that this is highly useful but the argument could be stronger. The PTF example for the use of the database doesn't use any detailed information on the location of the measurements. The paper would be stronger if examples were given or if there was (more) discussion of applications for which this spatial precision is indeed necessary.

RE: We give some specific examples in section 4.4.

Q: In addition to the compilation of existing (national-scale) databases, the authors also actively searched for data from underrepresented areas. This is very useful but it is, however, not fully clear how underrepresented areas were defined or how exactly they searched for these additional data points. Was there a certain cutoff in terms of publication date? Did they search for data from specific countries or was it based solely on soil type or climate? A bit more information on how they searched for these studies and thus which studies were included (and which were not included) would be useful.

RE: Based on the global map shown in Figure 1, we looked for countries and regions without values reported in the existing databases. We made then specific literature research on "Ksat" values for a specific country (or region like 'arid regions in Africa). In some cases, we also contacted colleagues that worked in these regions to ask for data support. We have better clarified this in the revised manuscript (P4L10-15).

Q: The paper contains several very useful figures that compare Ksat values for different soil types. It would be useful if it was indicated on these figures for which soil types the mean values are statistically significantly different.

RE: We thank the reviewer for this suggestion. We have prepared a table to show the significant differences between each soil texture class for table 5 as well as Figure 3b. (See Tables ST1 and ST2 in the Supplementary Files).

Q: On P3L26, it is mentioned that the sand silt clay fractions were estimated based on the method of Nemes et al. but from the text and Venn diagram in Figure 2, it appears that these data were available for

most of the papers/databases. Were they only estimated when they were not available already from the database? This is not so clear. How well did the Nemes method estimate the fractions when data were available?

RE: We apologize for the imprecision. We computed texture % using the method by Nemes et al. (2001) only for the UNSODA database, for which soil texture information was not directly available. We have now clarified this point in the manuscript (see P3L28-30).

Q: The authors develop a subjective accuracy score based on the location accuracy and the method. They state (P9L8) that they consider lab measurements more accurate than field measurements. Even though I understand what they mean, this was still a bit surprising to me as samples may be disturbed, suffer from compaction or smearing and are generally too small to contain a network of macropores. This is partly addressed in the discussion but some discussion (perhaps with a focus on accuracy vs precision?) and acknowledgements of the issues with soil samples in this part of the paper would be useful.

RE: We agree with the Reviewer. In the revised version, we removed the confidence degree based on the measurement method and only provided the positional accuracy based on the location. See subsection "Standardization and quality assignment".

Q: I know that there are different ways to use the word "sample" but here it is confusing to use the word for different things. I therefore suggest not to use the word sample for a datapoint, and to only use it to mean a soil sample (and thus for the laboratory measurements). In particular, "field measured soil samples" is a confusing use of the word sample. Also "temperate soil samples" seems to be used to indicate both field and lab (sample based) values from sites in a temperate climate. It would be better to reword these types of sentences to avoid any confusion.

RE: We have modified the text accordingly and used "field measured Ksat value" or "temperate-climate based Ksat values".

Q: The annotated pdf contains some additional suggestions (all minor) and highlights where the text can perhaps be improved a bit (these are just suggestions though, the paper is well written as it is).

RE: Thank you so much for the additional helpful suggestions. We have modified the text accordingly.