ESSD reviewer's comments reply:

Reviewer 1

As mentioned in my earlier review, I think that this dataset is timely and very useful. I find the manuscript much improved and recommend it for publication after a few additional changes. In particular, I appreciate the new table 4. The dataset itself is also clearer and better aligned with what is written in the text. However, I have several minor comments (and some editorial suggestions) to improve the manuscript further (see annotated pdf). My main comments are:

The use of many t-tests to determine the significance of the difference in the mean values for each individual pair of texture classes is not ideal. To me an ANOVA with a post-hoc test would be a much more logical (an appropriate) test.

RE: We have now used the ANOVA with a post-hoc test as mentioned in P11L1-3.

The exact search terms that were used need to be stated explicitly

RE: Now we have stated them explicitly in the text P3L19-20.

The number of outliers that were removed needs to be stated explicitly

RE: We removed about 700 values (now stated on P6L13-14).

4) Make it clearer that the PTFs are based on measurements at different depths and highlight the dominance of the topsoil measurements by giving the % of all data points that are top-soil measurements.

RE: We give the percentage of the samples from topsoil (see P7L4-5 and P8L5-6).

5) Table 2b doesn't seem to show all the actual columns in the dataset (as given in Table 2a) – make it clearer that it is just a selection of the data columns or better show an example that has all columns and headers.

RE: We changed it and Table 2b shows now all the actual columns in the dataset.

6) The use of the word sample for data point is not ideal. The use of either 'measurement' or 'data point' would be much clearer than 'sample' as you also talk about laboratory samples and some of the data points are not based on laboratory samples but on field measurements. Similarly, I find the mixing of the word location and site inconsistent.

RE: We agree and we replaced "samples" by "measurements" "sites" by "locations".

7) Storing the actual datafiles as a compressed csv file is not the greatest option for data archiving. Perhaps these .gz files can not be opened in the future? Why not upload the .csv file directly?

RE: GZ is an open source file format (https://en.wikipedia.org/wiki/Gzip). It does not alter the csv file in any way, but it drastically speeds up data download from https://doi.org/10.5281/zenodo.3752721. Users in R can directly read the file by using x = read.csv (gzfile("sol_ksat.pnts_horizons_rm.csv.gz"). We hope the reviewer understands that the gzip is only used here to speed up data transfer.

Reviewer 2

Review on the revised submission of the manuscript “SoilKsatDB: global soil saturated hydraulic conductivity measurements for geoscience applications” by Surya Gupta et al.

The authors have revised and considerably improved their manuscript based on the received comments and suggestions. However, I still see a number of issues to be resolved to meet the high standards of ESSD publications.

P4L5f: I assume that PDFs refer to plots of the value distribution in the original publications and a respective cross check? I see the improvements of the section, however I find it difficult to make sense of this sentence: I fully agree that mistakes happen when digitising - and especially when digitising data. However, I do not see which kind of artefacts or errors the authors expected to identify when comparing PDFs? A bias or any change in the distribution is rather unlikely in my understanding of the matter. More probable are single number twists or misattribution of sample IDs. Exactly such errors cannot be found by visual inspection of PDFs.

Again, I see the required efforts and I value your work. However, I suggest to change the paragraph to:

In the case of legacy datasets (non-digital tabular format, non-peer-reviewed data), we invested a significant effort to digitize, clean and cross-check it.

RE: Thank you. We added the sentence as you suggested (see P5L1-2).

One general remark to the label “peer-reviewed” data: I would assume that the data is rather difficult to be judged in a peer-review process. Also in the current case, I cannot judge any of the reported data to be correct or not. I can only judge the transparency and reproducibility of the data compilation or processing. Hence I would actually suggest to drop this criterion at large but to include a respective label

in the database. I would think of something like: part of largely checked database, case study data from peer-reviewed study, data from public archive without reviewed publication,

RE: Thank you for the suggestion. Now, we have changed the terminology

P4L11f: I have an idea what the authors aim to express, but I would like to challenge the statement that uncertainty in spatial maps is reduced when “some values or range” is provided. The uncertainty about some property is always a matter of the respective application. Moreover, the debate about how many samples are representative for a pedon, catchment, soil type or country is similarly difficult to answer. Again, I value the intensive search of the authors, but I find the presentation difficult to digest. Maybe the authors could find means to present the data sources in a more plain manner?

RE: Thank you for the suggestion. We agreed and removed the sentence.

P3L32f: I suggest to drop this sentence here and include the conversion of the geo references in section 2.2

RE: We shifted the sentence to section 2.2.

P6L2f: I know several hydrological models which cannot use georeferenced points directly. Maybe the authors could merge the first two sentences towards something like: Georeferencing of Ksat measurements is important for using the data for local, regional or global hydrological and land surface models.

RE: We have now rephrased the sentence (see P5L11-12).

P6L19: If you assumed an application at the surface, why do you assign a depth of 0-20 cm? Does this imply that the database does not discern between surface and top-soil measurements? I appreciate the

listing of methods (Table 4) and I would be astonished to find most field methods to be applied in deeper positions if not explicitly mentioned.

RE: We rephrased the sentence to state that in case of missing depth information we assigned the measured values to the topsoil.

P2L25ff: I assume that it only holds for the infiltration data from the SWIG database. This could be formulated more clearly as many infiltration measurements actually fit the Gardner equation.

RE: The reviewer is right that (i) we are referring to SWIG database and (ii) many methods use the Gardner conductivity function. We rephrased the paragraph and link to comprehensive review papers.

P6L16ff: I see that overlapping data entries is an issue. However, this is more a question of data organisation from the various sources (2.1) than of quality. I suggest to move this to section 2.1

Standardisation would also include the texture reclassification (P3L28), which I suggest to move here.

RE: Thank you. We have rephrased and shifted the sentences as you suggested (see P3 L30-33 and P6L15-18).

P6L20ff: The “position accuracy” remains an issue. First of all I do not understand the last category if spatial reference is a criterion to enter the data base. What do the 142 samples refer to if the data is not available? I also cannot understand how this classification has been derived - neither based on which concept nor on which data.

RE: We have carefully read each of the publications describing the data source and then have assigned spatial location accuracy based on the description provided. This is of course somewhat subjective, but we consider that spatial location accuracy can be in fact often determined e.g. by opening the location description, and then navigating to sites in Google Earth. In some cases, the authors indicate that they

have used hand-held GPS or similar to geolocate points; hence we use a standard error for GPS for ±10 m (https://www.nstb.tc.faa.gov/reports/PAN96_0117.pdf#page=22). For 142 samples authors unfortunately did not provide any location hence we could not estimate the location errors but we assumed that they are >10 km. In any case, we were consistent in documenting all steps and location errors, as explained now in modified section 2.2.

P6 Sec 2.3: I still find “quality assignment” a very difficult term. Given the raised issues above it might be still worth some revision to really sort the methods which are a little scattered in the introduction and across the method sections. Since this is a data publication, the readers have to be able to understand how the data has been compiled and what it can provide and where limits exist. To me this means quite a little more precision in the presentation of the data. I assume that the authors can easily streamline the presentation of the first two sections by some sorting and removing of general statements.

RE: This is a good point and indeed, we should remove use term "quality assignment". We have now instead added a section on "Completeness assessment" (modified section 2.2). This means that we do not get into subjective assessment of quality of dataset, but simply assign a completeness level based on the availability of data and metadata.

P6L28ff.: I have asked this already in the first round of reviews to clarify the PTFs. I only see that you used RF to derive ksat from clay, sand and BD for two sets as a diagnosis tool for the data. As such, I find it legitimate and valuable. But I would suggest to remove much of the “overhead”. Your methods could simple explain the use-case of the data and why and how you have done the cross checks temperate vs. tropic // lab vs. field. This can be done in a quite direct and concise manner. Moreover, I suggest to include Sec. 2.5 in Sec. 2.4 and revise the level of precision.

RE: We merged the two sections and rephrased as suggested.

P6 Sec. 2.4: What is completely missing in the methods is the inspection of the derived database, which comes in Sec. 3.2 and Fig. 3 and 4. I guess that a large part of the figure caption (Fig. 3) is actually methods. Hence I suggest to sort this out and to be precise about the manuscript's structure.

RE: This is correct. We have now moved some of the method description text shown in Fig 3 back to section 2.4. In addition, we have shortened the text in captions of Fig. 3 and now clearly separate method description from results.

P9 Tab. 4: The hood infiltrometer is referenced to Schlüter et al. 2020. Since this paper is definitely not about the hood infiltrometer technique, I would really urge the authors to check their used references. The correct citation appears to be Schwärzel, K., and J. Punzel (2007), Hood Infiltrometer-A New Type of Tension Infiltrometer, Soil Science Society of America Journal, 71(5), 1438–1447, doi:10.2136/sssaj2006.0104.

RE: Thank you for noticing this. We have changed it now and checked the other references as well (see Table 4).

Another citation I spotted which I found not really to the point is at P2L14: You cite Schindler and Müller 2017 for a global soil physical dataset, which is correct in general. But in the context of Ksat I do not see how that fits if they report measured soil water retention curves and UNsaturated hydraulic conductivity which happen to be georeferenced…

RE: We have now removed this citation from the text.

For the field methods: I would be interested in the difference between a Guelf Infiltrometer and Guelph Permeameter. Likewise there are several disc infiltrometers. Long story short: I think the main issue to resolve here is: What is measured with a tension apparatus, what gives at least a constant head, what is

falling head but double ring, what is falling head and single ring, what is implicitly calculated from rainfall simulations. Maybe you can classify these accordingly?

In the lab, I consider also the sample size to be relevant (100ml vs 250ml). It remains cryptic to me what is behind the named methods eg. what is a hydraulic head in comparison to a constant/falling head? What is a cylinder method? In my view, it is not sufficient to just drop some citations here (which I did not follow up one by one after spotting the error above).

Since I find this to be a crucial information in your study, I strongly suggest to really work out the details.

RE: With 'infiltrometers' the flux through the soil surface is measured, with permeameters the flux below the surface is quantified (flow between two locations or percolation through an interface below the surface). The typo ('Guelf infiltrometer') was corrected. We agree that the classification in the lab-based methods in Table 4 were ambiguous and we modified them. For the field-based methods, we inherited the classification of Rahmati et al. 2018. For the convenience of the reader, we give for each Ksat measurement in the database the reference to the original paper describing how the values were measured. Please see sol_ksat.pnts_metadata.csv at https://doi.org/10.5281/zenodo.3752721.

P11L7ff.: bias and RMSE are common knowledge and do not require to be given as equations.

RE: We agreed, however, Dr. Atila Nemes asked in his online comment to provide the equations for accuracy parameters. We also think that it is important for the readers.

P12 Fig 2.: I still do not like the diagram since it reports incorrect proportions. If you insist to use it, then maybe extend the figure to the other classes like climate region shares and the other classes named in this section (in separate sub-plots). The caption is too excessive.

RE: This is a good point. We have now also added climate classes to Fig 2.

P12L6f: I do not get this. First of all, the methods could be listed similar to Table 4. Second, the table appears highly redundant and difficult to link to the samples to me. I cannot see a common ID (but the position). The respective studies appear to have used a respective set of techniques which could be given in one line and linked through the citation or a respective common key. Again, since this is a data publication, I do not see this as minor but mandatory to clarify.

RE: We agree and added a common ID to each measurement.

P13L3f: And yet another method. If you use a t-test I guess this should be motivated in the methods including a clarification why it has been applied. I do not see why this should be more insightful than boxplots or the given violin plots.

RE: Reviewer 1 asked for this test to show the significance test between soil texture classes.

Moreover, I would suggest to consider to rearrange some of your tests: One of your questions is if lab and field methods are comparable and meaningful. I find this important and you approach it by means of simple statistics plus the RF cross-application. The second question which I find is about differences between texture classes which again follow this pattern. So maybe you could frame this more holistically

in the methods section, which could make your manuscript much more easy to follow and comprehend? The same holds for the results section.

RE: Thank you. We have now modified the method section accordingly.

P13L15ff.: This repeats the methods and adds the aspect of an evaluation of the relative importance of the covariates in the RF models. I suggest to sort methods and results and to keep it quite concise here, explaining the figures in preparation for the discussion.

RE: We agreed and separated the method and result section concisely.

P14 Fig. 3a: Maybe a heatmap like Fig. 4ff. would be more insightful given the many overlapping points?

RE: We provided a heat map.

P14 Fig. 3: Please revise the caption. E.g. the t-test is not required to repeat here. The soil classes are ordered inversely. Interpretation is more a matter of the text…

RE: We revised the caption.

P15 Fig. 4: Maybe include that we see heatmaps here?

RE: We have now included in the caption of figure 4.

P15L1: Since I find many statements in the introduction, methods and results realting to the difficulties to derive such a data set valuable but often misplaced, I suggest to use the first discussion subsection for these. The matter about spatial attribution and precision, the many different methods, etc. could be pointed out here.

RE: We summarize these points now in section 4.3

P15ff. Sec. 4.1 and 4.2: I find it not really convincing that these “application examples” results are fully covered as discussion. I suggest to check again, what is results and what findings really suit the discussion about your dataset. It is not really covered in the data, if the data points from the tropical or temperate regions cover more or less sites with swelling clays. Similarly, sample sizes and pore connectivity have not been addressed so far. E.g. the tension infiltrometers should at least avoid such bias. Thus your general remarks do nor really cover the actual data you should discuss.

RE: We are confident that the application of the PTFs show the effect of soil structures and soil formation processes on Ksat. In the modified version, we motivate the application of the PTF in section 2.4

Finally, I would like to quote the conclusion of the Youngs (1991) review on infiltration measurements with respect to the notion of your manuscript:

”Classical soil physical theory that assumes the soil to be a uniform inert porous material leads to a physical understanding of the infiltration process and allows the three-dimensional flow that occurs during most infiltrometer measurements to be interpreted in terms of hydraulic soil properties or infiltration parameters appropriate for one-dimensional flow from large surface areas. However, complicating factors that make simple soil physical theory inapplicable, can vitiate the interpretation of measurements to give meaningful results for use in predicting the infiltration behaviour of an area. There is a need for the theory of soil-water movement in general, and infiltration theory in particular, to be extended to take into consideration these factors so that the infiltration process can be properly described when there is soil variability, when there is soil swelling and shrinking, and when there is soil aggregation, so that sound physics-based hydrological modelling can advance, Meanwhile, reproduceable values of hydraulic soil properties and infiltration parameters need to be obtained from measurements on different sized infiltrometers located over the site to make confident predictions.”

I find it quite astonishing that the awareness about the methods for infiltration and ksat has been much more sharp 30 years ago than it is in your manuscript…

All the best for your revisions. I have full confidence that you can turn your manuscript in a very worthwhile publication.

Youngs, E. G. (1991), Infiltration measurements—a review, Hydrological Processes, 5(3), 309–319, doi:10.1002/hyp.3360050311.