

Earth Syst. Sci. Data Discuss., referee comment RC4
<https://doi.org/10.5194/essd-2022-80-RC4>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on essd-2022-80

Anonymous Referee #4

Referee comment on "SGD-SM 2.0: an improved seamless global daily soil moisture long-term dataset from 2002 to 2022" by Qiang Zhang et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2022-80-RC4>, 2022

This paper addresses the emergent need to increase soil moisture information access, quality, and quantity for multiple users and applications. The authors present an interesting study about reporting a new data version of a seamless global soil moisture product that increases both the quality and accuracy of the previous version of this product. The methods are sound and novel, particularly the development of a deep learning algorithm to fill daily gaps in soil moisture estimates. The authors compare the old and new versions of the datasets, and they provide a robust quantitative accuracy benchmark between versions.

The paper is generally well written. However, from the narrative, I feel that there are missing technical details. For example, the use and role of the three passive microwave sensors in modeled soil moisture values in the presence of the precipitation dataset is unclear. Also, can the authors elaborate on prediction variance or model-based uncertainties? I feel that uncertainty of estimates is commonly not presented in soil moisture gap-filling efforts despite being helpful for assessing the reliability of soil moisture predictions.

It is also my opinion that the accuracy limitations or advantages of the new product version are relative to the reader. For example, the authors poorly discuss their accuracy findings against previous research or gap filling efforts of satellite soil moisture estimates across scales.

The first version of the product has a relatively good number of citations, meaning that the community uses the product and that the methodological approach is being compared with similar research. The authors provide a thorough comparison between product versions, but they do not present the discussion of findings against previous research. I would appreciate more discussion about the potential implications of using the product's old or new version in multiple applications in terms of other available soil moisture estimates.

The paper leaves the value of this product relative to the reader as the comparison is made only between versions one and two, and it does not consider the large availability of other soil moisture estimates for multiple uses and applications. Many (hundreds if not thousands) studies currently report alternatives to downscale or fill gaps in satellite soil moisture data. I invite the authors to provide a more extensive literature review and discussion of previous research to support the value of their product.

I invite the authors to discuss the main implications of accuracy metrics to assess modeled soil moisture values. Can the authors describe the accuracy of the soil moisture sensors used? I invite the authors to use community accepted standards to report errors on soil moisture products, e.g., ubRMSE <https://www.sciencedirect.com/science/article/pii/S0034425720301760>, and discuss the accuracy of the soil moisture modeled values against other products or gap-filling efforts. A simple demonstration of the new knowledge that users can obtain from the new product would increase substantially the value of this excellent modeling framework applied to soil moisture satellite estimates.

Finally, but more importantly in my opinion (considering that this is a dataset journal), please consider publishing your code in order to fulfill the FAIR principles and contribute to the open-science culture transparently e.g., <https://bg.copernicus.org/preprints/bg-2021-323/bg-2021-323.pdf> .

Specific comments

L30 I recommend to avoid the word 'destroy' as aggregated soil moisture values are useful for multiple applications (e.g., to constrain long-term Earth system models). The cited references do not deal with gap-filling daily soil moisture values, please revise.

L35 sentence relative to the reader, can the authors be more specific and quantitative and include supporting references, e.g., which quantitative indexes?

L4065 Consider combining each weakness or limitation in v1 with their corresponding advantages in v2, instead of two separated lists.

L83 how they are employed?

L98 What is the criteria to select those sites?

L102 It seems to me that the authors solve a regression problem (where soil moisture is a response of precipitation and time) using deep learning, but they use the word assimilation, which is relatively fine for me given how the algorithm they use works. However I recommend to elaborate on the concept of data assimilation applied here for a

better and broader understanding of narrative flow.

L108 what soil moisture product? the authors use three products

L170 40x40 what?

L119 What was the criteria to select those 124 sites? Can the authors provide a map of points showing in colors the correlation between in-situ and their product for all the stations? I like the presented information but this is a global product and I think it will be useful to interpret the reliability of the product elsewhere. Also for bias indicators (MAE, RMSE), it would be nice to see a map of errors to identify areas with high or low quality of predictions. Please consider also the *ubRMSE* as it has been a widely discussed metric validating local to global soil moisture predictions. Discuss please the values of accuracy metrics in this and other products.

L225 can the authors highlight these points in figure 3b?

L285 the temporal resolution depends on the application.

Figures 9 and 10, consider using lines instead of points.

