Comment on essd-2022-21
Anonymous Referee #2

Referee comment on "An in situ observation dataset of soil hydraulic properties and soil moisture in a high and cold mountainous area on the northeastern Qinghai-Tibet Plateau" by Jie Tian et al., Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2022-21-RC2, 2022

Overall, the paper is well written, and the data are valuable. However, I still have some main concerns:

- The authors give a good description of the SHP and SM datasets. However, they fail to clarify the accuracies of the ground-based datasets. Errors exist in either ground measurements or products. Direct comparisons between ground measurements and products can not help us understand the quality of the ground measurements. Although it is still challenging to quantify errors within the in-situ measurement, methods like triple collocation do exist that can give uncertainties of the in-situ measurements. I would strongly recommend the authors to try to explain the accuracy of the ground measurements, at least to cite some previous validation work to prove that the validation results of SHP and SM products in this work are consistent with them. This will make the quality of the ground measurements to be convincing.

- I believe that the long-time series point SM measurements are valuable and meaningful. However, I do not think they are suitable for validating coarse SM products. Since the in-situ measurements are all obtained from single stations, spatial heterogeneity impacts on the validation results can not be ignored. They should be considered, especially when evaluating SM products with a spatial resolution of tens kilometers using in-situ point measurements. Actually, dense in-situ SM observation networks are an effective way to minimize the impacts of spatial heterogeneity. Several dense in-situ SM observation networks within the Qinghai Tibet Plateau, such as Heihe network constructed during HiWATER, Naqu and Pali of the CTP-SMTMN networks and Maqu and Ngari of the Tibet-Obs networks, have provided long time-series SM measurements which can be well used for SM evaluation. Therefore, I would suggest the authors use point SM measurements in different applications on a small scale to clarify their quality.
Minor comments:

- L40, it is arbitrary to say “highly uncertain”.
- L105, in our study area.
- L130, make sure that it is “at the long-term SM monitoring stations” or “at the random sampling site”?
- Table 1, suggest to list spatial resolutions for HWSD, SoilGrid, ShangYG and DaiYJ.
- Table 2, suggest to list spatial resolutions for GLDAS, ERA5 and SMAP SM products.
- Line 230, double-check the numbers here. I can not well relate some of the numbers here to those listed in Table 3. For example, why n ranges within 0.09 and 0.12? why cv of clay ranges within (0.18, 0.28)? etc.
- L250, is it -0.66?
- L255, wrong space place in “… significant, except…”
- Figure 5, keep the soil property name consistent with those in the text. E.g., bulk to bulk density or BD, s in logks should be a subscript.
- L280, Theta_r and theta_s should be written formally.
- Figure 7, bulk, theta_s, Theta_r, and alpha should be written formally.
- L330, Figure 11 here should Figure 10? The “higher” in “…and this is higher than the temporal…” should be “lower”?
- Figure 9, the text 0.75 in the legend is wrong? Should it be 0.075? In addition, clarify in the text how to obtain figure 9? Same as figure 6 using Kriging method in ArcGIS to interpolate?
- L350, two meanings for PBIAS here. One is positive bias, the other is percent bias.
- Figure 11, use BD instead of bulk.
- L410, how can you conclude that “our SM dataset provides new accurate in-situ SM measurements covering …”?