Earth Syst. Sci. Data Discuss., referee comment RC3
https://doi.org/10.5194/essd-2022-80-RC3, 2022

Comment on essd-2022-80

Anonymous Referee #3

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-RC3, 2022

This paper presents an improved seamless global daily soil moisture dataset from 2002 to 2022 (SGD-SM 2.0) based on the three satellite soil moisture sensors AMSR-E, AMSR2 and WindSat and Global daily precipitation products. A new convolutional neural network approach is used to fill the gaps and missing regions and ISMN data is used for the validation of SGD-SM 2.0.

The topic is of general interest due to the increasing drought and overexploitation of water resources in many regions of the world due to global climate change and fits well within the scope of the ESSD. The applied methods mostly appropriate and the manuscript is mostly well written but contains some incorrect wording and phrasing (see specific comments). My main concern is that the authors used only six stations for the validation of the global SGD-SM 2.0 data set, which is not inappropriate. The authors should make an effort to test whether SGD-SM 2.0 data accuracy is independent on the environmental conditions. The SGD-SM 2.0 data product would be well received by the science community working on Global Change issues and can be recommended for publication after all issues detailed below have been appropriately addressed.

General comments:

The authors only show the averaged evaluations indicators from all selected ISMN stations and only six in-situ soil moisture stations were actually used for the validation of global SGD-SM 2.0 data set. In my view, this is not enough to appropriately demonstrate the accuracy of a global SM data set. There much more data is available at ISMN. In addition, other in-situ soil moisture data products are freely available, e.g. Bogena et al. (2022). In this way, potential users could also see if the SGD-SM 2.0 data accuracy is independent on the environmental conditions, e.g. soil properties, vegetation coverages, climate zone.

Some soil moisture data shown in Fig. 8 show extremely SM high values of more than 80 Vol.%. Such high values are very unlikely, as soil porosity in most soil is typically between 40-50 Vol.%, indicating measurement errors in the in-situ data or soils with extremely high organic matter or clay content. Indicating a reference site description will help to understand this better. On the other hand, the SGD-SM 2.0 data the same high values, which is astonishing. In my view, these data outliers could be the result of SM overestimation by the CNN procedure due to the precipitation consideration. In addition, single outliers can be found in Figs. 9d and 10a. Again, this indicates the influence of precipitation. Maybe the data should be cleaned with an outlier detection method? Please add at least a discussion on these issues.

The in-situ soil moisture data from ISMN are treated anonymously in this work. However, the site owners that work hard to maintain the soil moisture stations should be better cited. This will help the site owners to ensure funds for the costly operation of the stations and data management. Therefore, the authors should add at table with basic information on the soil moisture data using, including the name of the site owners and/or monitoring networks instead of just presenting the station coordinates. See Bogena et al. (2022) for a great example. The necessary information is available in the metadata descriptions at ISMN.

Throughout the manuscript, you use the term "assimilation" in the context of including precipitation data in your CNN based data interpolation method. However, I think this is not appropriate as the term "data assimilation" is generally used optimally combine numerical models with observations.

Specific comments:

L17: Please cite the more recent ISMN publication of Dorigo et al. (2021)

L21-22: Incorrect phrasing.

L23: Change to "approximately 20% to 80%"

L30: "words"

L31: Change "destroys" to "degrades" or similar

L35-36: Citation is missing.

L43-44: Reads awful, please rewrite.

L56-57: Incorrect phrasing.

L70: Please mention the source of the in-situ data.

L78: The GES DISC website should be referenced

L85: Reads awful, please rewrite.

L97: Please cite the more recent ISMN publication of Dorigo et al. (2021)

L103: Change here and elsewhere to “long and short-term”

L130: Change to “soil moisture and precipitation products”

L132-133: Can you estimate the average time scales of the long and short-term memories and their variabilities? It would be interesting to know how different the time scales are.

L190: The term “epoch number” should be explained.

L290: Change to “the soil moisture time-series of”

Figure 11: Please show the precipitation in reverse order and as bar chart, which the standard way of presenting precipitation and much better to understand.

Literature

Bogena, H.R., M. Schrön, J. Jakobi, P. Ney, S. Zacharias, M. Andreasen, R. Baatz, … and H. Vereecken (2022): COSMOS-Europe: A European network of Cosmic-Ray Neutron Soil Moisture Sensors. Earth Syst. Sci. Data 14: 1125–1151. DOI: 10.5194/essd-14-1125-2022

Dorigo, W., I. Himmelbauer, D. Aberer, L. Schremmer, I. Petrakovic, L. Zappa, W. Preimesberger, A. Xaver, F. Annor, J. Ardö, D. Baldocchi, M. Bitelli, G. Blöschl, H. Bogena, … and R. Sabia (2021): The International Soil Moisture Network: serving Earth system science for over a decade. Hydrol. Earth Syst. Sci. 25: 5749–5804. DOI:10.5194/hess-25-5749-2021