Response to the Comments of Referee #3

Dear Referee #3:

We are particularly grateful for your careful reading, and for giving us many constructive comments of this work!

According to the comments and suggestions, we have tried our best to improve the previous manuscript ESSD-2022-80 (SGD-SM 2.0: An Improved Seamless Global Daily Soil Moisture Long-term Dataset From 2002 to 2022). An item-by-item response follows.

Once again, we are particularly grateful for your careful reading and constructive comments. Thanks very much for your time.

Best regards,

Qiang Zhang

General comments:

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.

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.

Response: We are particularly grateful to the reviewer for his/her detailed suggestions! According to the comments, we have tried our best to improve the previous manuscript. For the in-situ validation issue, more specific explanations could be checked in Q3.1. For the environmental condition issue, the discussion on whether SGD-SM 2.0 data accuracy is independent could also be checked in Q3.1. An item-by-item response to each constructive comment follows.

Major comments:

Q3.1: ***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.

Response: Thanks for this comment. In this work, we select 124 stations from ISMN from 2002 to 2022 and match them with corresponding soil moisture product in SGD-SM 2.0. Actually, we chose six in-situ soil moisture stations as examples for scatter visualization. In other words, all the selected 124 in-situ sites are employed to validate the accuracy of SGD-SM 2.0. We match the hourly in-site values with the descending products. In consideration of validation reliability, we choose the two neighboring in-site values correspond with the observation time of soil moisture products. Then we average them as the ground-truth data.

Through all the 124 selected in-situ sites, Table 1 compares the original products with SGD-SM 2.0. The average evaluation indicators (R, RMSE, and MAE) of original soil moisture and SGD-SM 2.0 products are 0.679 (0.672), 0.094 (0.096), and 0.075 (0.078), respectively. Generally, the precision of SGD-SM 2.0 products performs similar with incipient products. The diversities of those indicators are little between the original and reconstructed SGD-SM 2.0 products in Table 1. To a certain extent, in-situ validation testifies the reconstructed accuracy and validity of the SGD-SM 2.0 products.

Table 1. Comparisons between the original and SGD-SM 2.0 products through 124 selected in-situ sites.

Soil moisture products	Average evaluation indicators			
	R	RMSE	ubRMSE	MAE
Original	0.679	0.094	0.058	0.075
SGD-SM 2.0	0.672	0.096	0.061	0.078

In terms of the independent on the environmental conditions (e.g. soil properties, vegetation coverages, climate zone), these 124 selected in-situ sites are widely distributed all over the world

(Europe, North America, South America, Asian, Africa and Australia). The soil properties, vegetation coverages and climate zones are diverse from each other. Through this in-situ validation way, we can test whether SGD-SM 2.0 data accuracy is independent on the environmental conditions. These descriptions have been supplemented in the revised manuscript.

Q3.2: ***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.***

Response: Thanks for these issues. Actually, the SM values in this work are the volume ratio (unit: $m^3 \cdot m^{-3}$, from 0% to 100%), rather than the mass ratio ($kg \cdot m^{-3}$, usually 0% to 50%). This phenomenon is normal because of the unit via volume ratio, not measurement errors or SM overestimation by the CNN. For the outliers in Figs. 9d and 10a, this indeed indicates the influence of precipitation for the proposed LSTM-CNN model. We also consider the outlier detection method, while filter strategy will also disturb the maximal/minimum value. Overall, these outliers are few with small impact for SGD-SM 2.0. Therefore, we don't clean the data with an outlier detection method. The future work on SGD-SM 3.0 will develop a new framework to restrain the outlier problem. These descriptions have been supplemented in the revised manuscript.

Q3.3: ***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.***

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

Response: Thanks for this significant suggestion. We have added a table with basic information on the in-situ soil moisture sites like Bogena et al. (2022). As listed in Table 2, it includes the name of the station, country, longitude/latitude, main land use, lattice water, and soil organic carbon. Due to the page limiting, we give the six COSMOS in-situ sites in Fig. 8 as follow:

Table 2. Basic information on the six COSMOS in-situ soil moisture sites in Fig. 8.

Station	Lon/Lat	Elevation (m)	main land use	lattice water	soil organic carbon
COSMOS-016	42.537, -72.171	316	Crop	4.50%	1.59%
COSMOS-055	0.2825, 36.866	1824	Bush	6.10%	1.11%
COSMOS-082	48.141, 15.171	73	Grass	2.10%	1.93%
COSMOS-096	-14.159, 131.388	169	Silty Sand	2.30%	1.24%
COSMOS-101	-21.617, -47.632	563	Grass	1.70%	1.87%
COSMOS-123	31.369, 91.899	1201	Forest	4.48%	2.36%

[Literature: Bogena, H. R., Schrön, M., Jakobi, J. et al.: COSMOS-Europe: a European network of cosmic-ray neutron soil moisture sensors, Earth Syst. Sci. Data, 14, 1125–1151, https://doi.org/10.5194/essd-14-1125-2022, 2022.]

Q3.4: ***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.***

Response: Thanks for this issue. We also agree that "data assimilation" is generally used optimally combine numerical models with observations. In this work, SGD-SM 2.0 introduces the global daily precipitation products into the reconstructing framework. Through the auxiliary precipitation data, SGD-SM 2.0 could lead in the daily extreme weather information for gap-filling. Therefore, we have replaced "assimilation" as "fusion" in the whole manuscript, to better embody the meaning of multi-source products fusion (precipitation and soil moisture).

Specific comments:

Q3.5: ***L17: Please cite the more recent ISMN publication of Dorigo et al. (2021).***

Literature: 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

Response: Thanks for this comment. We have cited this publication in the revised manuscript as follow:

[Citation: Dorigo, W., Himmelbauer, I., Aberer, D. et al.: The International Soil Moisture Network: serving Earth system science for over a decade, Hydrol. Earth Syst. Sci., 25, 5749–5804, https://doi.org/10.5194/hess-25-5749-2021, 2021.]

Q3.6: ***L21-22: Incorrect phrasing.***

Response: Thanks for this comment. We have revised this sentence as follow:

"As shown in Fig. 1(a) and (b), these soil moisture products exist plenty of gap regions."

Q3.7: ***L23: Change to "approximately 20% to 80%".***

Response: Thanks for this suggestion. We have rewritten this sentence as follow:

"Actually, the land coverage rate is only approximately 20% to 80% in daily AMSR-E/2 and WindSat quantitative products."

Q3.8: ***L30: "words".***

Response: Many thanks for the reviewer's careful reading and checking! We have revised "word" as "words" in this sentence.

Q3.9: ***L31: Change "destroys" to "degrades" or similar.***

Response: Thanks for pointing out this issue. We have changed "destroys" to "degrades" in this sentence.

Q3.10: ***L35-36: Citation is missing.***

Response: Thanks for this comment. We have supplemented the related citation in the revised manuscript as follow:

"Relevant quantitative indexes (R, RMSE and MAE) and results demonstrate that SGD-SM 1.0 products can be extended for global, daily and full-coverage soil moisture measurements (Zhang et al., 2021)."

[Citation: Zhang, Q., Yuan, Q., Li, J., Wang, Y., Sun, F., and Zhang, L.: Generating seamless global daily AMSR2 soil moisture (SGD-SM) long-term products for the years 2013–2019, Earth Syst. Sci. Data, 13, 1385–1401, https://doi.org/10.5194/essd-13-1385-2021, 2021.]

Q3.11: ***L43-44: Reads awful, please rewrite.***

Response: Thanks for this suggestion. We have rewritten this sentence as follow:

"SGD-SM 1.0 ignores the daily extreme weather condition. If one day occurs a sudden precipitation, SGD-SM 1.0 usually performs poor under this scenario."

Q3.12: ***L56-57: Incorrect phrasing.***

Response: Thanks for this comment. We have rewritten this incorrect phrasing in this sentence as follow:

"Through fusing auxiliary precipitation data, SGD-SM 2.0 could lead in the daily extreme weather information for gap-filling."

Q3.13: ***L70: Please mention the source of the in-situ data.***

Response: Thanks for this suggestion. We have given the detailed source of the in-situ data as follow:

"The in-situ soil moisture sites are employed to validate the reconstructing precision of SGD-SM 2.0. These in-situ data are downloaded from International Soil Moisture Network (ISMN)."

Q3.14: ***L78: The GES DISC website should be referenced.***

Response: Thanks for this comment. We have referenced The GES DISC website in this sentence: "These datasets are all recorded at GES DISC website (NASA GES DISC, 2022)."

[Reference: NASA GES DISC: https://disc.gsfc.nasa.gov/, last access: 06 June 2022.]

Q3.15: ***L85: Reads awful, please rewrite.***

Response: Thanks for this suggestion. We have rewritten this sentence as follow:

"Precipitation usually has a high correlation with soil moisture in the corresponding regions."

Q3.16: ***L97: Please cite the more recent ISMN publication of Dorigo et al. (2021).***

Response: Thanks for this comment. We have cited this publication in the revised manuscript as follow:

[Citation: Dorigo, W., Himmelbauer, I., Aberer, D. et al.: The International Soil Moisture Network: serving Earth system science for over a decade, Hydrol. Earth Syst. Sci., 25, 5749–5804, https://doi.org/10.5194/hess-25-5749-2021, 2021.]

Q3.17: ***L103: Change here and elsewhere to "long and short-term".***

Response: Thanks for this issue. We have revised this statement as "long and short-term memory" in this sentence and elsewhere of the updated version.

Q3.18: ***L130: Change to "soil moisture and precipitation products".***

Response: Thanks for pointing out this error. We have recorrected this sentence as "soil moisture and precipitation products" in the revised manuscript.

Q3.19: ***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.***

Response: Thanks for this interesting query. The proposed model uses long and short-term memory network to extract time-series information for generating SGD-SM 2.0. Actually, this network cannot estimate the average time scales of the long and short-term memories and their variabilities. The memory mechanism introduces the short-term memory to ensure the adjacent correction for the next node. The long-term memory is used to ensure the sequentiality of time-series nodes.

Q3.20: ***L190: The term "epoch number" should be explained.***

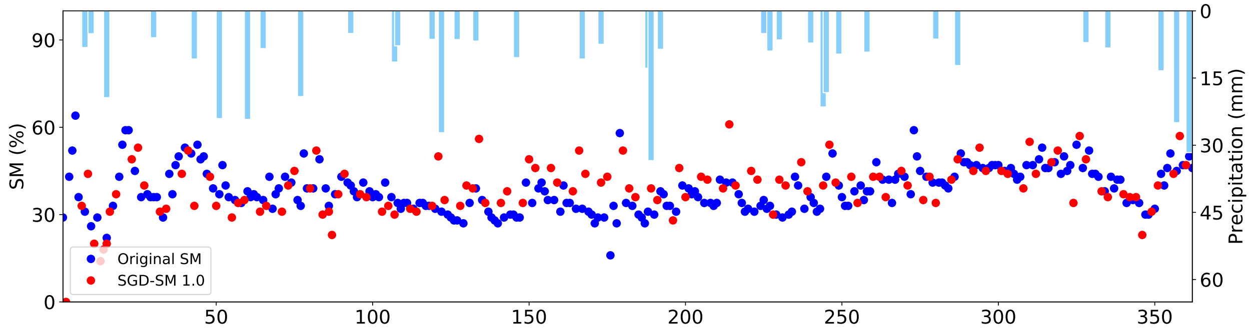
Response: Thanks for this issue. We have explained the definition of the term "epoch number" in the updated version as follow: "One epoch represents that all the samples in the training set have been utilized for the neural network optimization at one time."

Q3.21: ***L290: Change to "the soil moisture time-series of".***

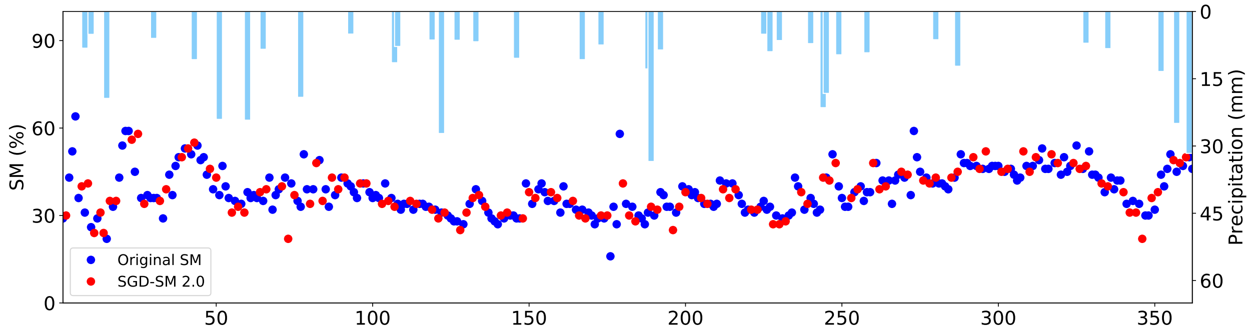
Response: Thanks for this comment. We have revised this sentence as "In-situ validation and time-series validation testify the soil moisture time-series of SGD-SM 2.0 products (R: 0.672, RMSE: 0.096, MAE: 0.078)" in Section 6.

Q3.22: ***Figure 11: Please show the precipitation in reverse order and as bar chart, which is the standard way of presenting precipitation and much better to understand.***

Response: Thanks for this meaningful suggestion. We have shown the precipitation in reverse order and as bar chart in Fig. 11. Current figure is much better to understand the significance of precipitation information.



(a) Time-series daily original soil moisture, SGD-SM 1.0, and precipitation results in 2013



(b) Time-series daily original soil moisture, SGD-SM 2.0, and precipitation results in 2013

Fig. 11. Time-series daily original soil moisture, SGD-SM 1.0/2.0, and precipitation results at location (48.875°N, 140.375°E) in 2013.