Response to the Comments of Referee #4

Dear Referee #4:

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-2020-353 (SGD-SM: Generating Seamless Global Daily AMSR2 Soil Moisture Long-term Productions (2013–2019)). 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:

The authors present a method and dataset to fill daily AMSR2 soil moisture product gaps with a CNN for the years 2013-2019. The abstract jumps quickly into the topic, but is somehow ambiguous by not being clear to which current soil moisture products ("...the acquired daily soil moisture productions") the introduction relates? It would be good to explicitly state that it relates to the AMSR2 products (not productions) ...this lack of language clarity (e.g. "reliable cooperativity..." etc.) traces through the whole manuscript and needs to be strictly revised before considering acceptance. Otherwise, it is really hard for the reader to understand, and thus, to estimate the usefulness of this dataset. I think there is a lot of potential in this paper and dataset (even if "only" 6 years length). However, many technical aspects make it hard to grasp the content and estimate the quality in the first place. A few other comments as follows.

Response: We are particularly grateful to the referee's careful reading and detailed suggestions!

"the acquired daily soil moisture productions" has been corrected as "the acquired daily AMSR2 soil moisture products" in the abstract part.

For the language clarity, we have revised the whole manuscript sentence by sentence in the updated version. According to the comments, we have tried our best to improve the previous manuscript. An item-by-item response to each constructive comment follows.

Major comments:

Q4.1: ***Still the abstract creates more questions than answers. The evaluation measures are difficult to interpret. Why stating 2 evaluation measures, with one for original data? Also the choice of units (m^3/m^3) is not immediately clear as the dataset only produces percent values for soil moisture?***

Response: Thanks for these comments. For in-situ validation (2 evaluation measures), we compare the reconstructed with original AMSR2 daily soil moisture products as 'A (B)'. 'B' refers to the evaluation index of orginal products. 'A' stands for the evaluation index of reconstructed products after gap-filling. Compared with 'A' and 'B', the difference is that the reconstrcting values of gap regions need also to be evaluated in 'A', while 'B' needn't. Overall, the accuracy of reconstructed AMSR2 daily soil moisture products is generally accorded with the original products. The differences of these indexes R: 0.683 (0.687), RMSE: 0.099 (0.095), and MAE: 0.081 (0.078) are minor between the reconstructed and original soil moisture products. In other words, this validation ensures the reliability and availability of the proposed seamless global daily AMSR2 soil moisture products.

For the units of AMSR2 soil moisture products, we have corrected the "m^3/m^3" value as the percent value in the whole manuscript.

Q4.2: ***The introduction is not clearly introducing the AMSR2 original dataset. I think it would be of great value if at least basic technical cornerstone of the original dataset is described.***

Response: Thanks for this meaningful suggestion. We have described the basic technical cornerstone of the original AMSR2 soil moisture products in sect 2.1 as follow:

"In this work, we focus on dealing with AMSR2 soil moisture products. This sensor was onboard on the Global Change Observation Mission 1-Water (GCOM-W1) satellite, launched in May 2012 (Kim et al., 2015). The released datasets include three passive microwave band frequencies: 6.9 GHz (C1 band), 7.3 GHz (C2 band, new frequency compared with AMSR-E), and 10.7 GHz (X band). It can observe the global land two times within a day (Wu et al., 2016): ascending (day-time) and descending (night-time) orbits. The primary spatial resolution of this datasets denotes 0.25° global grids. And the AMSR2 soil moisture retrieval algorithms include

Land Parameter Retrieval Model (LPRM) and Japan Aerospace Exploration Agency (JAXA) (Du et al., 2017; Kim et al., 2018). Besides, the uncertainty of soil moisture for each band were also given in AMSR2 products.

In our study, we choose LPRM AMSR2 descending level 3 (L3) global daily 0.25° soil moisture products as research data. This dataset could be obtained at https://hydro1.gesdisc.eosdis.nasa.gov/. For instance, the original AMSR2 0.25° soil moisture data obtained in April 2, 2019 is displayed in Fig. 1. Due to the satellite orbit coverage and the limitations of soil moisture retrieving algorithms in tundra areas (Muzalevskiy et al., 2020), the acquired AMSR2 daily soil moisture products are always incomplete in global land (about 30%∼80% invalid ratio, excluding Antarctica and most of Greenland), as shown in Fig. 1. The daily global land coverage ratio of AMSR2 soil moisture data in 2019 is listed in Fig. 2. Distinctly, the global land coverage ratio is low in wintertime, and high in summertime. The mean global coverage ratio of 2019 is just about 56.5% in AMSR2 soil moisture daily products. Apparently, these incomplete soil moisture data cannot be directly applied for subsequent spatial and time-series analysis, as mentioned in previous Sect 1."

Q4.3: ***Stating that the codes are also published is misleading so far. "The related Python codes of this dataset are also available at https://github.com/qzhang95/SGD-SM." (authors) only holds an example code of extracting data. I think it would be really helping the transparency of the data quality if the physical network implementation (TensorFlow, Keras, SKLearn, Pytorch?) would be also open in the spirit of open data and open source and reproducibility. Understandably, a trained neural network model is not 100% reproducible, but the model could be also archived on Zenodo? It only makes sense, because it would be very feasible to update the dataset on a yearly basis with the developed model. In the end the idea of ESSD is "living data".***

Response: Thanks for this issue. We have revised the sentence "The related Python codes of

this dataset are also available at https://github.com/qzhang95/SGD-SM." as "An example Python code of extracting this dataset are also available at https://github.com/qzhang95/SGD-SM." The training and testing procedure of the proposed model are implemented by Pytorch platform. The implemented code of this work will be released on Zenodo after possible acceptance, to flexibly update the dataset on a yearly basis with the developed model. We couldn't agree more with the referee's opinion that the idea of ESSD is "living data". This can also facilitate the development and utilization of soil moisture products.

Q4.4: ***The year 2013 folder only contains 362 files, not 365. May 2013 only seems to have 28 files? Please check your upload on Zenodo.***

Response: Many thanks for your careful checking of the released dataset! The reason is that the NASA's official LPRM AMSR2 L3 soil moisture products indeed only have 28 daily files in May 2013 (missing data files in date May 11, 12, and 13). Referee can also verify this issue at NASA's GES DISC website at: https://hydro1.gesdisc.eosdis.nasa.gov/data/WAOB/LPRM_AMSR2_D_SOILM3.001/2013/05/. We have supplemented this explanation in the updated products (https://doi.org/10.5281/zenodo.4417458).

Q4.5: ***(Zhang et al., 2020) This citation is not in the references list, there are only Zhang 2020a and 2020b. Please add, presumably it is your data citation: DOI: 10.5281/zenodo.3960425***

Response: Thanks for this comment. We have added this citation into the reference of this work: Zhang, Q., Yuan, Q., Li, J., Wang, Y., Sun, F., Zhang, L. (2021). SGD-SM: Generating Seamless Global Daily AMSR2 Soil Moisture Long-term Products (2013-2019) (Version 1.0) [Data set]. Zenodo. DOI: 10.5281/zenodo.4417458.

Q4.6: ***"More details of this work are released at https://qzhang95.github.io/Projects/Global-Daily-Seamless-AMSR2/..." your current paper should reflect the most important and up to date source of information until it is published. It would be ok though to refer to it as a "technical supplement" maybe? However, those URLs are not reliable. Thus, if you have a technical supplement with more details, it could be added to your Zenodo archive (which will be reliable and has a DOI).***

Response: Thanks for this comment. We have deleted this URL in the revised manuscript, to avoid misunderstanding this work. Some technical explanations with more details have been supplemented in our Zenodo archive (DOI: 10.5281/zenodo.4417458).

Q4.7: ***Section 3.1 starting on page 7 is the methodological main part of your neural network implementation. While I'd like to acknowledge that technical level of description, I have two contradicting issues with it:***

1) p7 ll137-139: "This network includes 11 layers (3D partial CNN unit and ReLU (Rectified Linear Unit)) in Fig. 4. The size of 3D filters is all set as 3×3×3. Number of feature maps before ten layers is fixed as 90, and the channel of feature map in the final layer is exported as 1".

That is very technical, yet it is not clear why those dimensions where chosen. The discussion section does not discuss the CNN and the design choices at all and what effect they have. For example, how can this capture the comparatively big gap areas of the original AMSR2 dataset?

2) On the other hand, much of this could also go into a technical supplement and you could provide a much higher-level overview for the reader in the paper. The paper is the data description, and many readers and future users of the dataset will not have the technical understanding of judging or even reading through the technical low-level design of the CNN – nevertheless this still also needs to be documented.

Response: Thanks for these meaningful issues. We have added a technical supplement of the network implementation. Detailed effects of 3D filters, layers, and feature maps in the proposed model are depicted below:

1) 3D CNN filters: 3D CNN filters are employed to simultaneously capture both spatial and temporal soil moisture information in time-series products. For large gap areas, partial CNN is developed to exclude the invalid AMSR2 soil moisture information.

2) Layers: More layers in deep neural network can extract more intrinsic feature information for soil moisture products gap-filling.

3) Feature maps: Feature maps get the description of the original soil moisture products from multiple angles, through different 3D CNN filters.

For clearly understanding the parameters chosen in the proposed network such as 3D CNN filters, layer numbers, and feature maps, we have supplemented the sensitivity analysis of these parameters in discussion part. As listed in Table 4, Table 5, and Table 6, discussions for the 3D CNN filters, layer numbers, and feature maps are investigated in simulated missing regions validation, respectively. Accordingly, the optimal indexes are chosen as the setting value.

Table 4. Discussion for the 3D CNN filters in simulated missing regions validation

Parameter	Evaluation index		
	R	RMSE	MAE
3×3×3	**0.968**	**0.068**	**0.047**
5×5×5	0.957	0.076	0.048
7×7×7	0.949	0.081	0.050

Table 5. Discussion for the layer numbers in simulated missing regions validation

Parameter	Evaluation index		
	R	RMSE	MAE
10	0.962	0.072	0.048
11	**0.968**	**0.068**	**0.047**
12	0.966	0.070	0.049

Table 6. Discussion for the feature maps in simulated missing regions validation

Parameter	Evaluation index		
	R	RMSE	MAE
60	0.963	0.071	0.048
90	**0.968**	**0.068**	**0.047**
120	0.967	0.069	0.047

Q4.8: ***Also, why not an LSTM type network?***

Response: Thanks for this interesting query. LSTM type network indeed plays an important role for time-series data. In fact, the spatial and temporal information are both significant on spatial consistency between the valid and invalid soil moisture regions, and temporal continuity in time-series curve. Therefore, we develop the spatio-temporal convolutional network in this study, to simultaneously utilize the spatial and temporal soil moisture information. Through this 3-D strategy, we can both exploit the spatial consistency and temporal continuity for soil moisture products gap-filling. In our future work, we will combine the LSTM network with 3-D partial convolutional network, to futher ullize the spatio-temporal soil moisture information.

Q4.9: ***Calling it spatio-temporal 3D might be misleading, as it is areal 2D and then a temporal dimension. Spatio-temporal indicates that already, the added 3D might lead to think of spatial 3D plus time.***

Response: Thanks for this comment. To avoid misleading, we have corrected "spatio-temporal 3D" as "spatio-temporal" in the whole manuscript.

Q4.10: ***What was the reason to choose the 0.25 dec degrees as spacing for the data files?***

Response: Thanks for this query. The spatial resolution of the original global daily AMSR2 soil moisture products is 0.25 dec degrees. To avoid introducing additional error and uncertainty, we didn't carry out the downscaling operation of the generated SGD-SM products. We have supplemented this explanation in the revised manuscript.

Q4.11: ***Last but not least, I'd like to advocate for a bit more metadata in the netcdf files, because netcdf provides great means for metadata. For example, you could adhere a bit more to the NetCDF-CF conventions, or at least add e.g. attributes such as title, reference and a time stamp in the dataset, not relying on the filename for example. You could also join at least the yearly slices into a "cube" that follows conventions of the Earth Sciences community (e.g. longitude instead of lon as variable name). Also the Zenodo deposit could have more fields filled out for improved discovery, more keywords (e.g. "soil moisture"?) and terms from controlled vocabularies, such as GEMET (https://www.eea.europa.eu/help/glossary/gemet-environmental-thesaurus) or similar.***

Response: We are very grateful for referee's detailed guidance on our released dataset! We have regenerated the SGD-SM products and updated them on Zenodo flatform (DOI: 10.5281/zenodo.4417458). The title, reference, and time stamp have been added into the metadata in daily NetCDF files, as shown in the following table. In addition, we have also replaced the abbreviation variables "lon" and "lat" with the full names "longitude" and "latitude" in all the NetCDF files.

More keywords like soil moisture, AMSR2, seamless, global, daily, and SGD-SM have been supplemented in Zenodo flatform (https://doi.org/10.5281/zenodo.4417458), for better improving the utilization of our products in Figure A.

```
netcdf file: D:/SGD-SM/2019/LPRM_AMSR2_20190101.nc
{
  float Latitude(Latitude=720);
       :units = "degree_north";
  float Longitude(Longitude=1440);
       :units = "degree_east";

  global attributes:
       :reference = "SGD-SM: Generating Seamless Global Daily AMSR2
          Soil Moisture Long-term Products (2013-2019)";
       :url = "https://doi.org/10.5281/zenodo.3960425";
       :time_stamp = "2021-01-04 20:32:22";
       :author = "Processed by Qiang Zhang, Wuhan University";
       :date = "20190101";
       :source = "netCDF4 python module tutorial";
}
```

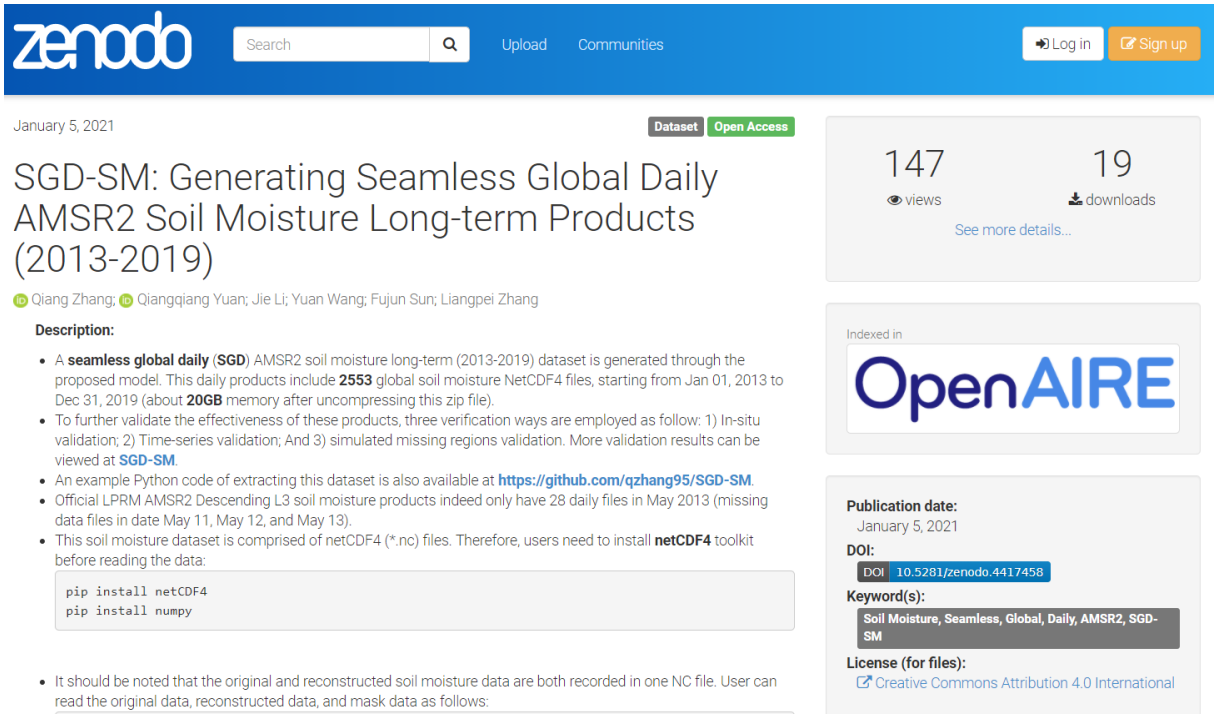

Figure. A. Updated keywords in Zenodo flatform (DOI: 10.5281/zenodo.4417458).