Earth Syst. Sci. Data Discuss.,
https://doi.org/10.5194/essd-2020-59-RC2, 2020

Open Access Earth System Science Data Discussions

Interactive comment on "A new dataset of satellite observation-based global surface soil moisture covering 2003–2018" *by* Yongzhe Chen et al.

Anonymous Referee #2

Received and published: 29 June 2020

The authors tried to generate long-term surface soil moisture at a global scale, via data fusion of 11 microwave remote sensing-based soil moisture products since 2003 through neural network approach, and SMAP soil moisture products were used as the training target. The idea is very interesting and should be encouraged to explore further how much extent the machine learning can help in Earth Observation for delivering physically-consistent (or physic-aware) products. However, the way the current manuscript is written, organized is still far from clarity, structured for this reviewer to comprehend their contributions. I would suggest rejection and encourage the author to continue along this line of effort. In the following, I listed some major concerns: 1. The author claimed that "This new dataset, once validated against the International Soil Moisture Network (ISMN) records, is supposed to be superior to the existing products

Printer-friendly version

Discussion paper

(ASCAT-SWI, GLDAS Noah, ERA5-Land, CCI/ECV and GLEAM), and is applicable to studying both the spatial and temporal patterns. " This assumption is too strong. On the other hand, it seems the author referred to the validation of the NN-based 10-d soil moisture products versus the 10-d averaged ISMN in-situ observations (as seen Figure 5, Figure 8, Figure S3, S6, and S9). Is it true? In any case, it should be specified under what conditions the generated product is performing better than other products. "supposed to be superior" is really not a scientific statement.

2. There were some strange 'terminologies' the author used for discussion, for example: a. 'penetrability of microwave' (which is seldom found in the literature. A more widely used term is 'microwave penetration depth'); b. "Soil moisture retrieval from passive microwave sensors is based on the correlation between soil dielectric conductivity, that is influenced by soil moisture . . . .". Following the theoretical development of soil moisture retrievals from remote sensing, the relationship between soil moisture and dielectric constant is the fundamental (not soil dielectric conductivity).

3."However, this data is regional, with a large temporal gap, and cannot be seen as observational-based only since precipitation data is incorporated." This is a very strange argument. We all know there is a strong link between precipitation and soil moisture variation. Physically speaking, one used the antecedent precipitation index to understand how precipitation events drive the variation of soil moisture. This is like one of 'quality impact factors'. If the above argument is true, we can argue that the author's approach in this manuscript is also not 'observation-based', as they used LAI, land cover, LST, and many other factors.

4. " are these factors used as direct spatial predictors of soil moisture or just because they are related to the errors of satellite soil moisture retrievals (i.e., the quality impact factors of soil moisture)? We insist on the latter, proposing two main reasons for the incorporation of environmental factors." This is very confusing and not necessarily correct, and not well grounded. We know the soil moisture retrieval from remote sensing is using a radiative transfer model to account for scattering and emissions from both

soil and vegetation, which is conflicting with the author's statements.

5. 'Water Body' was used as one of the predictors (it should be predictor, rather than quality impact factors). This is very strange. As we know, water body map in either SMOS or SMAP soil moisture products were used to mark out those locations to avoid soil moisture retrievals over these water bodies (otherwise, it would be physically no sense, in terms of soil moisture). This is wrong and not physically sound to include water bodies as one of predictor for predicting surface soil moisture.

6. For 'topographic complexity' 'soil texture', the author used from different sources, one from ASCAT ancillary data and the other use SMAP ancillary data. This reviewer is wondering why such a choice? Why not making it consistent (i.e., get ancillary data from one single product, instead of two?)

7. '3σ denoise'. what is the effect of such a filter on identifying extreme years? for example, during 2003, 2010, 2018, 2019 there are extreme heat events in Europe and the soil moisture is so dry which can be beyond the 3 standard deviations.

8.NN design. SMAP is only available after 2015, so I am not sure what is the meaning of simulation period 2012D19$\sim$2013D36, but also 2014-2018. I guess this is constrained by the available data (PROBA-V and GLASS LAIs)? But in any case, it does not represent any physical meaning to predict 2015data with 2012-2013 data. At least, the NN design is not clear on why it is designed as such.

There are some other specific comments can be found in the attached PDF.

Please also note the supplement to this comment:
https://essd.copernicus.org/preprints/essd-2020-59/essd-2020-59-RC2-supplement.pdf

Interactive comment on Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2020-59, 2020.