

Interactive comment on “Evaluation of 18 satellite- and model-based soil moisture products using in situ measurements from 826 sensors” by Hylke E. Beck et al.

Anonymous Referee #1

Received and published: 7 June 2020

This paper describes the performance of various gridded soil moisture products with in situ surface soil moisture measurements. There is a lot thrown into this comparison and the methodology seems solid, but I am not sure about what we learned in the end.

1. Why is this particular subset of products selected? It is mixing spatial (horizontal & vertical) resolutions, operational and research products, etc., and makes a fair comparison questionable. Furthermore, it is not possible to stratify the results based on this random mix of features. Please provide more justification for the evaluation setup or refocus the paper.

For example, why was SMAPL3E included and not the coarser-scale L2/3 product,

C1

why SMOS v650 and not SMOS-IC, why GLEAM, etc. There is no reason given for the chosen products, even though the various products serve different purposes and have very different characteristics (e.g. SMOS retrievals offer both SM and VOD, DA products offer much more than only surface soil moisture, just to name a few).

Perhaps this paper should move its focus towards evaluating the new MeMo product and its underlying HBV modeling system, rather than shuffling that product into a general analysis that tries to vaguely address a list of too general questions for a non-representative or inconsistent subset of data products?

Random example: what is the relative performance of the single-sensor satellite products? If “all” available soil moisture products would be compared, or some meaningful features would be targeted, then we could learn something from this, but for the 4 discussed products, more than half of the answer was already given in earlier papers and the added value of the answer in this paper is minimal.

2. The 5-day filter is used to reduce noise, but has also been used to derive root-zone soil moisture in the past. Why are the results compared to surface soil moisture and not root-zone in situ measurements? Would that not be fairer?

3. In general, there is very little mentioning of the vertical representativity of the various products. It cannot possibly be that all products produce a consistent ~5 cm surface product. For example, how deep is the HBV soil moisture store? Is it comparable in volume to the volume observed by satellite data or other model-satellite surface soil moisture products? Due to their different wavelengths, the AMSR2, ASCAT and SMOS/SMAP products must be sensitive to different vertical surface layers. Is it fair to compare them all to the same ~5-cm surface in situ measurements?

4. The temporal resolution is also questionable: how is it possible to do a 3-hourly evaluation for all products (p.3, L.20)? Satellites only pass over every so many days.

5. Please provide more information on the quality screening of the satellite data. The

C2

text only mentions screening for frozen conditions, but each product comes with its own flags that need to be applied. For example, it is mentioned that AMSR2 and SMOS are more vulnerable to RFI: how did you screen these data for RFI? Did you screen for dense vegetation, topographic complexity, etc?

6. The consideration of both high and low frequency signals for the calculation of R is a good idea, but why is there no evaluation of the interannual variability, using a simple state-of-the-art anomaly R?

7. Not understood: "only HBV and the Catchment model underlying SMAPL4 have been calibrated". Is it fair to say that Catchment would be "calibrated" (for soil moisture, just like HBV?) in order to hardwire a single parameter (a constant)? Wouldn't all models then ever have been 'calibrated' to chose some hardwired parameters?

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., <https://doi.org/10.5194/hess-2020-184>, 2020.