Dear Prof. Dr. Hendricks-Franssen,

We would like to sincerely thank you for handling the manuscript and the reviewers and Korbin Nelson for their valuable comments. We have provided detailed responses to the reviewers one week before the closure of the open discussion, so the reviewers have already been given an opportunity to respond. Additionally, we have submitted a revised version of the manuscript, including a tracked-changes version. The main changes are as follows:

1. Explained the calculation of the correlations using instantaneous satellite soil moisture retrievals; in response to all reviewers.
2. Added Kruskal-Wallis test probability values to Table 1; in response to reviewer 3.
3. Added a justification for our product selection; in response to reviewers 1 and 3 and Korbin Nelson.
4. Added information about the sensor types included in the ISMN archive; in response to reviewer 3 and Korbin Nelson.
5. Recalibrated HBV for ERA5 and IMERG and modified the discussion; in response to reviewers 2 and 3.
6. Discussed the layer depths of the models in the main text; in response to reviewer 1 and Korbin Nelson.
7. Added a statement about the generalizability of the performance of the calibrated models; in response to reviewers 2 and 3.
8. Added more precise latency values to Table 1; in response to reviewer 2.
9. Improved all maps (Figs. 1, 4, 5, and S2); in response to reviewer 3.
10. Provided information about the product quality flags; in response to reviewer 1.
11. Added a statement regarding the difficulty of satisfying the assumptions underlying triple collocation techniques; in response to reviewer 2.
12. Added details about the ERA5 temperature bias correction and the low- and high-frequency correlation coefficients; in response to reviewer 2.
13. Added details about the *in situ* data for the calibration; in response to reviewer 2.

Sincerely,

Hylke Beck (on behalf of all co-authors)

Review by Korbin Nelson

The manuscript of Beck et al. evaluated the temporal dynamics of 18 state-of-the-art (quasi-)global near-surface soil moisture products. I this study very interesting and up-to-date. Overall, the paper is well organized and well written, and provides new insights about the advantages and disadvantages of different soil moisture products and on the merit of various technological and methodological innovations.

We thank Dr. Nelson for reviewing our manuscript and providing thoughtful comments.

However, the introduction is not well written and more discussion and comparison to recent studies should be provided. In my opinion, the paper deserves publication once the following points are addressed with some more details.

We appreciate the comment; we have re-read the introduction with this in mind and made some improvements.

line 4-12: Provide the reason why you would like to address these questions. I like your way to express your purposes of your study. However, it's not appropriate to pose so many questions here without giving any reason.

We agree and have added that these questions are *"frequently faced by researchers and end-users alike."* References and further background on each question is provided in the subsections discussing addressing the questions (Sections 3.1 to 3.9).

Section 2: Why these datasets are chosen out for comparison? What are main differ-ences among the products within each group (i.e., satellites, open-loop models, and models with DA)?

Good question. We have added the following to justify our product selection: "*We evaluated six products per category, which was sufficient to compare the performance among and within product categories and address the questions posed in the introduction. We only considered widely used products with (quasi-)global coverage and we attempted to keep the selection of products in each category as diverse as possible. For example, we considered products based on several major satellite missions used for global soil moisture mapping (AMSR2, ASCAT, SMAP, and SMOS), models of various type and complexity (with and without calibration), different sources of* *precipitation data (satellites, reanalyses, gauges, and combinations thereof), and various data merging and assimilation techniques (with different inputs).*"

The authors missed some recent publications on soil moisture evaluation. For example:

Chen, Y., & Yuan, H. (2020). Evaluation of nine sub-daily soil moisture model products over China using high-resolution in situ observations. Journal of Hydrology, 125054.https://doi.org/10.1016/j.jhydrol.2020.125054

Tavakol, A., Rahmani, V., Quiring, S. M., & Kumar, S. V. (2019). Evaluation analysis of NASA SMAP L3 and L4 and SPoRT-LIS soil moisture data in the United States. Remote Sensing of Environment, 229, 234-246. https://doi.org/10.1016/j.rse.2019.05.006

Add a review on these publications in introduction and more discussions with these papers in Section 4 will add much value to this manuscript.

Thanks for pointing us to these very interesting studies. We have added them to the introduction and to other relevant sections of the paper. Even though our paper has already well over 200 references, the body of literature on soil moisture estimation is so vast that it's easy to miss studies.

line 30: What are the sensor types? Are there all FDR sensors?

We have added the following text: *"The measurements were performed using various types of sensors, including time-domain reflectometry sensors, frequency-domain reflectometry sensors, capacitance sensors, and cosmic-ray neutron sensors, among others.*"

Add a map showing the observation length and the frequency of in-situ observation.

Please see Supplementary material Fig. S1 for a figure showing the observation length and the frequency of *in situ* observation.

Table 1: Add one column to describe the vertical layers for the soil moisture products. Since soil moisture data of model products or satellites are not representative at 5 cm, have you done some vertical interpolation?

The depths of the soil layers of the models are provided in the “Details” column. The penetration depth of microwave signals can differ significantly depending on the observation frequency and the land surface conditions, and therefore cannot be listed in the table. To improve the vertical representation of the satellite products, we used the SWI filter (see Section 2.1). We have added the following text to the revised manuscript to discuss the vertical support of the models: *“The vertical support is physically consistent with in situ soil moisture measurements at 5-cm depth for most models. The average depth of the soil layer (i.e., half the depth of the lower boundary) is 2.5 cm for SMAPL4, 3.5 cm for ERA5 and ERA5-Land, 5 cm for GLEAM, 8.5 cm for HBV-ERA5, 6.6 cm for HBV-IMERG, 7.3 cm for HBV-MSWEP, and 15 cm for VIC-PGF (Table 1; Supplement Table S1). The soil layers of HBV may seem too deep, especially since they represent conceptual “buckets” that can be fully filled with water, in contrast to the soil layers of the other models which additionally consist of mineral and organic matter. However, the soil layer depths of HBV were calibrated (see Section 2.3) and are thus empirically consistent with in situ measurements at 5-cm depth.*”

Review #1

This paper describes the performance of various gridded soil moisture products within 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.

We thank the reviewer for their thorough assessment and helpful comments.

Briefly summarized, we evaluated the largest and most diverse selection of soil moisture products to date, to the best of our knowledge. This allowed us to gain several novel insights into the relative advantages and disadvantages of a broad range of methodologies and data sources used to estimate soil moisture, as well as into techniques to evaluate the estimates. Of course, one outcome of our comparison is quantitative information on the relative performance of different products, which will be helpful for researchers deciding which product(s) to use in their analysis. However, there are several other findings:

- a smoothing filter helps to avoid disadvantaging noisy satellite products in product evaluations;
- the new ERA5 reanalysis precipitation data provide performance close to gauge-based precipitation estimates;
- satellite products perform worse in cold climates than in warmer climates, and so do model products;
- a simple, calibrated model can outperform substantially more complex, data-intensive models;
- precipitation data quality is the main factor determining the benefit of data assimilation;
- satellite data assimilation provides greater performance improvements for models with a poor soil moisture simulation efficiency;
- model calibration can be more beneficial than satellite data assimilation into an uncalibrated model;
- satellite products tend to exhibit larger regional performance differences than models;

We believe that these findings, which are succinctly summarized in the conclusions, are of value to the general readership of HESS.

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.

This is a good question. We have added the following to the revised manuscript to justify our product selection: "*We evaluated six products per category, which was sufficient to compare the performance among and within product categories and address the questions posed in the introduction. We only considered widely used products with (quasi-)global coverage, and we attempted to keep the selection of products in each category as diverse as possible. For example, we considered products based on several major satellite missions used for global soil moisture mapping (AMSR2, ASCAT, SMAP, and SMOS), models of various type and complexity (with and without calibration), different sources of precipitation data (satellites, reanalyses, gauges, and combinations thereof), and various data merging and assimilation techniques (with different inputs).*"

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

Please see our preceding response. There are too many candidate soil moisture products available for us to include all of them in our analysis. Some admittedly subjective selection is therefore necessary.

We appreciate that different products have different design objectives and characteristics and offer different auxiliary data, but do not see how that invalidates our evaluation.

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 anon-representative or inconsistent subset of data products?

Thanks for the comment. A paper just about MeMo and HBV would be of interest to a much smaller part of the community than the present evaluation. The MeMo product was included primarily to assess the effectiveness of a different merging approach compared to ESA-CCI. The HBV model products were added to (i) examine how well a simple calibrated model performs, (ii) assess the impact of different precipitation forcing datasets on the overall performance, and (iii) quantify the benefits of satellite data assimilation for different precipitation forcing datasets.

We believe the nine questions posed in the paper are pertinent to numerous researchers and end users of soil moisture data and not “too general.” We are not entirely sure why the reviewer refers to our way of addressing the questions “vague” as we provide clear, concise, and well-referenced objectives and findings.

As an aside, HBV is not underlying the MeMo product, they are completely independent 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.

This example pertains to just one of the nine questions addressed in the paper. Earlier papers did not apply smoothing filters, stratified the results in different ways, most did not explicitly assess high- and low-frequency fluctuations, and most did not compare the performance of these single-sensor products to other types of products. As such, we believe our analysis does provide significant added value. Furthermore, even if part of our answer to this question was already given in earlier papers, we do not see it as a bad thing to replicate the findings of previous papers.

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?

We actually applied the 5-day filter to make the comparison with the *in situ* measurements at 5-cm depth more fair. The 5-day filter serves two purposes, to reduce noise and to deepen the vertical support of the superficial satellite observations (not to the root zone but to approximately 5 cm). Without the 5-day filter, the satellite products would perform, on average, significantly worse than the two other major product categories, and some particularly noisy satellite products would be severely disadvantaged (e.g., SMOS; see Fig. 2). If we had used *in situ* measurements of the root zone as reference we probably would have used a filter with a longer temporal window.

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?

Thanks for the comment. We believe our study represents a fair comparison. The 5-day filter deepens the vertical support to make the superficial satellite observations more representative of *in situ* measurements at 5-cm depth (please see our previous response). Previous soil moisture product evaluations tended to compare soil moisture retrievals directly to *in situ* measurements at 5-cm depth, and therefore may have underestimated the 'true' skill of products. We considered optimizing the time lag constant *T* for each product but decided against this, because we wanted to make statements about the accuracy of the original data, not a post-processed product.

We have added the following text to the revised manuscript regarding the vertical support of the models: *"The vertical support is physically consistent with in situ soil moisture measurements at 5-cm depth for most models. The average depth of the soil layer (i.e., half the depth of the lower boundary) is 2.5 cm for SMAPL4, 3.5 cm for ERA5 and ERA5-Land, 5 cm for GLEAM, 8.5 cm for HBV-ERA5, 6.6 cm for HBV-IMERG, 7.3 cm for HBV-MSWEP, and 15 cm for VIC-PGF (Table 1; Supplement Table S1). The soil layers of HBV may seem too deep, especially since they represent conceptual "buckets" that can be fully filled with water, in contrast to the soil layers of the other models which additionally consist of mineral and organic matter. However, the soil layer depths of HBV were calibrated (see Section 2.3) and are thus empirically consistent with in situ measurements at 5-cm depth."*

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.

Thank you for the comment. We have added the following text into the revised manuscript: *"For the satellite products without SWI filter, we matched the instantaneous soil moisture retrievals with coincident 3-hourly in situ measurements to compute the R values."*

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

We appreciate the suggestion. We will provide more information about the quality flags used for the satellite products in the revised manuscript. Thanks for the suggestion.

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?

The "state-of-the-art anomaly R" measures both the (seasonal-scale) interannual variability and short-term deviations from the long-term mean seasonal cycle. The skill of short-term variations in the soil moisture products is assessed in our high-frequency filter.

We did not separately evaluate the (seasonal-scale) interannual variability due to the short temporal span of some of the products (less than 5 years), which precludes us from calculating reliable correlation coefficients.

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?

Both HBV and Catchment have been explicitly calibrated against independent *in situ* soil moisture measurements by optimizing a certain performance metric. The same may be true but has not been similarly documented for the other models included in the evaluation. The calibration procedure of HBV is described in Section 2.3, while the calibration procedure of Catchment is described in Reichle et al. (2019b).

Review #2

This is my first review of the manuscript “Evaluation of 18 satellite- and model-based soil moisture products using in situ measurements from 826 sensors”.The study is very interesting and fits well with the scope of the HESS journal. It is well written and structured with relevant research questions answered in details in the results section. The literature cited is updated and figures and tables well formatted.

We thank Dr. Massari for his thorough assessment of our manuscript.

Despite this I have different MAJOR comments the authors should seriously consider:

1. 826 sensors is quite a large number for soil moisture stations and gives the impression that this evaluation is very general. However, by looking at the locations where these sensors are located the reader realizes that the majority are located over US and Europe, that is, over very data rich regions (i.e., where models tend to perform better). I think in title is much more important to highlight where the analysis is carried out rather than the number of sensors used. This give also a clearer picture of the results obtained in the study.

Thanks for the suggestion. We considered replacing *“from 826 sensors”* with *“from the US, Europe, and Australia”* in the title. However, since this would make the title less concise, we did not make this change. We do, however, clearly highlight in the paper that our results may not generalize to the entire global land surface and have devoted an entire subsection (3.9) to this issue.

We do not fully agree with the generalization that models perform better over data-rich regions, as this depends on the precipitation forcing used to drive the models. Our evaluation includes six models with non-gauge-based precipitation forcings (ERA5, ERA5-Land, HBV-ERA5 with and without data assimilation, and HBV-IMERG with and without data assimilation), and the performance of these models is largely representative of data-poor regions.

2. Following point 1 the results can be a bit biased towards models (also considering the type of evaluation the authors chose, see my comment 3e) and product that require use calibration (e.g., HBV runs). The product evaluation is in practice carried out exactly where in situ observations are more dense and where are more dense more calibration

stations are present. This is partly highlighted by the authors but only at the end of the document while I would add more discussion about this issue.

Thanks for the comment. We have changed several existing sentences and added the following sentence to Section 3.9: *"The calibrated models (HBV and the Catchment model underlying SMAPL4) may, however, perform slightly worse in regions with climatic and physiographic conditions dissimilar to the in situ sensors used for calibration (but probably still better than the uncalibrated models).*"

3.The overall methodology needs to be strongly improved and detailed as many aspects are not clear and/or not well discussed and justified:

a.The evaluation is carried by considering the temporal dynamic which is fine for the considerations done in the paper and from previous literature (see Koster et al. 2009), however, it is not clear how the evaluation at 3 hour resolution is done for satellite data with a revisit time larger than 1 day (e.g. SMAP, SMOS) and for model forced with rainfall with daily resolution. This must be clarified.

We agree and have added the following text to explain this more clearly in the revised manuscript: *"For the satellite products without SWI filter, we matched the instantaneous soil moisture retrievals with coincident 3-hourly in situ measurements to compute the R values.*"

b.The Triple Collocation (TC) is a foundation of one integration technique (i.e., the one of MeMo) and is a well known technique for the readers this manuscript point to. I am surprised that its theoretical foundation has not introduced in a more rigorous way and the assumptions made not tested. For instance, line 24 page 4 reads "with ASCATSWI and HBV-MSWEP, which are independent from each other and from the passive products". First the requirements are not the independence among the products but independence of their errors as well as their mutual linearity. These assumptions might not hold even for the products chosen (Gruber et al. 2016) as here, in addition, also SWI is systematically applied to at least two products of the triplet. This can falsify the results obtained via TC. I think some additional discussion and testing of the validity of the assumption is needed. The authors can consider the application of the Quadruple collocation technique (Gruber et al. 2016) for testing this assumption which many authors of this manuscript are familiar with.

We thank the reviewer for his comment. We do not entirely agree with the statement that *"the requirements are not the independence among the products but independence* *of their errors,"* because if the products are fully independent, it follows that the errors will be fully independent as well. Unless of course the reference is imperfect (which is the case if *in situ* data are used as reference), in which case the errors reflect both the product and the reference.

We recognize that the error independence assumption and other assumptions may not be fully satisfied in our study and we have therefore added the following statement to the revised paper: "*Triple collocation-based merging techniques rely on several assumptions (linearity, stationarity, error orthogonality, and zero cross-correlation; Gruber et al., 2016) which are generally difficult to fully satisfy in practice, affecting the optimality of the merging procedure.*"

We carefully examined the Quadruple Collocation (QC) methodology presented in Gruber et al. (2016). They note that QC still requires *"zero error cross covariance between some specific data set combinations"* (Section 2.4), which means that expert judgement is still needed to determine which products have correlated errors and which don't prior to estimating the correlations between two products. Pan et al. (2015) also highlighted the need for expert pre-judgement. QC is therefore only useful to estimate the correlation after already having "assumed" that particular products are more likely to be correlated than others. In light of this, we believe QC offers limited independent insight into the TC assumptions. The developer of QC makes a similar statement in Gruber et al. (2017): *"Recently, Gruber et al. (2016) proposed an extension to TCA where the inclusion of more than three data sets in the analysis allows for — at least partly — resolving nonzero error cross-correlation structures, yet a demonstration of the robustness of the method on a global scale is still pending. Therefore, one may for practical reasons neglect error cross correlations between different active or passive data sets at the cost of non optimal SNR improvements, or make a conservative educated guess for error cross-correlation levels for data sets where they are expected.*"

The application of the SWI filter was necessary to temporally match the different satellite products, which would not have been possible using instantaneous retrievals at non-overlapping irregular times (Gruber et al., 2020). We agree, however, that the SWI filter does not need to be applied to both satellite products in the triplets, and therefore in the revised manuscript we use unfiltered ASCAT data.

c.Maybe this is just a technical matter and nothing major but talking about the climatology of SMAP sounds quite weird with only four four/ five years of observations.

From an evaluation point of view I think it still fine, however, from a longer perspective this climatology is likely to not consider the real climate variability.

We agree and have replaced *"climatologies"* with *"averages."*

d.Many terms and procedures are just mentioned without specifying important details. This makes the study hardly reproducible. Examples: line 17 pag. 5 "Temperature estimates were taken from ERA5, downscaled to 0.1 and bias-corrected on a monthly basis through an additive approach". How the downscaling and the bias correction has been done exactly? "Additionally, we calculated Pearson correlation coefficients for the low- and high-frequency fluctuations of the 3-hourly time series...". Please tell what correlations of high and low fluctuations would provide in addition to classical correlation?

Thanks for bringing this up; we read the manuscript again to make sure no details are missing. We added the following to explain the ERA5 correction: *"To improve the representation of mountainous regions and ameliorate potential biases, the ERA5 air temperature data were matched on a monthly climatological basis using an additive (as opposed to multiplicative) approach to the comprehensive station-based WorldClim climatology (V2; 1-km resolution; Fick and Hijmans, 2017).*"

The following sentence was added to better highlight the added value of the Pearson correlation coefficients for the low- and high-frequency fluctuations: *"Additionally, to quantify the performance of the products at different time scales, we calculated Pearson correlation coefficients for the low-frequency fluctuations (i.e., the slow variability at monthly and longer time scales;* R_{lo}*) and the high-frequency fluctuations (i.e., the fast variability at 3-hourly to monthly time scales;* R_{hi}*).*"

e.This is an important aspect: "We did not average sites with multiple sensors to avoid potentially introducing discontinuities in the time series." Line 31 pag. 6. This means that if the satellite footprint of a specific product includes multiple in situ stations multiple correlations values are considered? If so, this makes the process of evaluation very random and not really under control as different products are characterized by a different spatial sampling and might include a different number of stations. Moreover, this exacerbates the problem of biased results towards model or products working well over US as many correlation values would originates from stations located in United States with an additional penalization of other locations which have already less stations. For a fair evaluation each pixel must count one correlation value. In this respect the product collocation is a crucial aspect that has not properly discussed and described in the manuscript. For example in Su et al. (2015) and Massari et al. (2017) the co-location of the satellite data and model data was determined by nearest-neighbour association and a screening step for removing ground sensors non-representative at the coarse scale was implemented. In their study, if multiple valid stations co-located in a satellite pixel were present, the station with the highest mean correlation was retained (see section 2.6 of Su et al. 2015 for further details).

We thank the reviewer for this thoughtful comment. This issue is commonly referred to as the collocation issue (Gruber et al., 2020) and unfortunately there are no satisfactory solutions, particularly when the products have such a wide range of grid-cell and footprint sizes. After much deliberation we decided not to change the current approach for the following reasons:

1. A coarser spatial sampling should, in our opinion, be penalized (as is currently the case), since it reflects a technical limitation in the ability of the product to represent heterogeneous areas.
2. We believe that grid-cells or footprints with multiple *in situ* sensors should be assigned more weight (as is currently the case), because the presence of multiple sensors reduces the sampling uncertainty and thus leads to a more reliable performance estimate.
3. The removal of *in situ* sensors that are not representative of the coarse scale is not straightforward in our evaluation due to the substantial variety in model grid-cell and satellite footprint sizes. We are not in favor of resampling all products to a common grid as this would penalize products with a higher spatial resolution.
4. The removal of 'unrepresentative' *in situ* sensors is further confounded by the fact that the location of satellite footprints varies over time (i.e., the footprint of today's satellite overpass is not exactly the same as the footprint of the next overpass). Su et al. (2015) and Massari et al. (2017) did not have this issue as their products were all gridded.
5. Retaining only the *in situ* sensors with the best performance may paint an overly rosy picture of the products.

We would like to note that our approach has also been used by numerous other researchers (e.g., Albergel et al., 2012; Karthikayan et al., 2017; Al-Yaari et al., 2019), which thus implicitly agreed with our view. Nevertheless, we agree about the importance of highlighting that several dense measurement networks exert a strong influence on the overall results and we therefore expanded the first sentence of Section 3.9 as follows: *"The large majority (98 %) of the in situ soil moisture measurements used as reference in the current study were from dense monitoring networks in the USA and Europe (Fig. 1) and therefore our results will be most applicable to these regions."*

f."We calibrated the 7 relevant parameters of HBV using in situ soil moisture measurements between 2010 and 2019 from 177 independent sensors from the International Soil Moisture Network (ISMN) archive that were not used for performance assessment (Section 2.5; Supplement Fig. S2)." Line 20 pag. 5. How the selection of these stations was carried out? Why 177? Why such a spatial distribution? Does a different choice provide similar results? I think all these aspects need to be clarified.

We have added the following to the revised manuscript: *"These sensors did not have enough measurements during the evaluation period (March 31, 2015, to September 16, 2019) and thus were available for an independent calibration exercise."* A different selection of *in situ* sensors would have provided similar results due to the low degrees of freedom (just 7 parameters were calibrated using 177 sensors). Note that HBV has been recalibrated for ERA5 and IMERG in the revised paper.

g."T was set to 5 days for all products, as the performance did not change markedly using different values, as also reported in previous studies". The application of the exponential filter with a constant parameter T=5 days might be not appropriate for all the satellite products as the different products have a different vertical support. Since the calibration was carried out for the model why T was not calibrated also for the satellite products?

We strongly considered optimizing the time lag constant *T* for each product in the revised manuscript but in the end decided against this for two main reasons. First, we did not want to deviate too much from the original data because we want to make statements about the accuracy of the original data, not a post-processed product. Secondly, we did not want to give the satellite products an unfair advantage compared to the uncalibrated models, which would likely also benefit from the application of the SWI filter (though likely not as much).

The calibration of HBV was carried out because the model cannot be run without calibration, as it is a conceptual model with parameters that do not represent physical properties of the land surface. Note that we added the following regarding the generalization of the performance of the calibrated models to Section 3.9: *"The calibrated models (HBV and the Catchment model underlying SMAPL4) may, however,* *perform slightly worse in regions with climatic and physiographic conditions dissimilar to the in situ sensors used for calibration (but likely still better than the uncalibrated models).*"

4."As forcing, we used the MSWEP precipitation dataset because of its favourable performance in numerous evaluations .... The calibrated parameter set was used for all HBV runs, including those forced with ERA5 or IMERG precipitation." I think proceeding in this way is not fair for the cross-validation. As HBV is basically a conceptual model, its parameters tend to correct also for errors contained in the data used to force it. Indeed, it has been largely demonstrated in the scientific literature (e.g., Zeng et al., 2018) that the impact of imperfect precipitation estimates on model efficiency can be reduced to some extent through the adjustment of model parameters. In other words, If you calibrate the parameters for MSWEP rainfall, then, when you force HBV with others precipitation inputs the results might be sub-optimal. Thus for a fair evaluation different sets of parameters should be used each one referring to the specific rainfall product used to force the hydrological model.

Our initial reason for not recalibrating HBV for ERA5 and IMERG was that we did not expect the resulting parameters to realistically represent the transformation of precipitation to soil moisture, because ERA5 and IMERG do not incorporate any gauge data and exhibit systematic errors (in mean, occurrence, and magnitude; Beck et al., 2019a). Conversely, the calibration of MSWEP has likely resulted in parameters that relatively realistically represent the transformation of precipitation to soil moisture, since MSWEP incorporates vast amounts of daily gauge data and exhibits almost no systematic errors in the study area (Beck et al., 2019a).

However, since we agree that the recalibration of HBV for ERA5 and IMERG might potentially lead to a small performance improvement, we followed the reviewer's suggestion and carried out the recalibration. The following text was added: *"To avoid giving one of the precipitation datasets an unfair advantage, we recalibrated the model for each of the three precipitation datasets (ERA5, IMERG, and MSWEP).*" The negligible performance improvement after calibration for ERA5 and IMERG (0.00 and 0.01, respectively) probably reflects the low degrees of freedom (just 7 model parameters were calibrated using data from 177 sensors) and thus limited ability of the parameters to correct for systematic errors.

5.MeMo integration. The study is based on similar conceptual framework presented in Kim et al. 2018 here (maximization of correlation) with the difference that in Kim et al. correlations are calculated with a benchmark while here are obtained from TC. Beside the satisfaction of the underlying assumptions related to TC which I have discussed on point 3b, Eq. 3-5 of the study of Kim et al. demonstrates that for the maximization of R when merging two products (but this holds for multiple products also), cross-correlation terms must be taken into account (it is also demonstrated in Gruber et al. 2017 already cited in the manuscript) thus the framework described in MeMo integration is not theoretically optimal. However, if the products are independent the framework collapses into a simple weighing average as cross-correlation are zero. I assume the authors consider null cross correlations within SMAP, SMOS and AMSR2 which I think is statistically not demonstrated. So i strongly suggest to provide some additional details and justifications about the integration framework used. This can explain why MeMo " MeMo performed only marginally better in terms of R than the best-performing single-sensor product SMAPL3ESWI" (Line 15 pag. 12).

We do indeed, implicitly, assume null cross-correlations among AMSR2, SMAPL3E, and SMOS. This is an assumption to all TC applications that may not be fully met, similar to the assumption of perfectly Gaussian distributions. The null cross-correlations assumption cannot be formally tested as the truth is not known. One could evaluate the correlation in deviations versus *in situ* data but of course they do not represent the truth either and they are not available everywhere, so this does not solve the issue.

6."The satellite products provided the least reliable soil moisture estimates and exhibited the largest regional performance differences on average, whereas the models with satellite data assimilation provided the most reliable soil moisture estimates and exhibited the smallest regional performance differences on average.". I think the authors should highlight again here that this result is expected given the high density gauge observations used in the study area. Highlighting this is very important as for instance ground validation conducted in data-rich areas does not adequately reflect the added values of satellite observations (Dong et al. 2019).

Thanks for the comment. Even when excluding the three models with data assimilation using gauge-corrected precipitation forcings (GLEAM, SMAPL4, HBV-MSWEP+SMAPL3E), the remaining three models with data assimilation (ERA5, HBV-ERA5+SMAPL3E, and HBV-IMERG+SMAPL3E) still provide more reliable soil moisture estimates and smaller regional performance differences on average. This conclusion is thus not simply attributable to the inclusion of gauge observations in some of the precipitation forcings.

Minor comments:

Line 24 pag. 3. Every satellite product contains proper quality flags for removing these low quality data while doing this with an external dataset might not guarantee optimal results. Please at least discuss this.

We will expand our discussion of this.

Line 12 pag. 4. “Three-hourly soil moisture time series of AMSR2SWI, SMAPL3ESWI, SMOSSWI”. No clear how these time series are created or extracted from products having revisit times larger than 1 day. This is unknown in the paper.

The last paragraph of Section 2.1 explains that the SWI filter was applied on a 3-hourly basis and that *“the SWI at time t was only calculated if ≥1 retrievals were available in the interval (t−T; t] and ≥3 retrievals were available in the interval [t−3T; t−T].”* Application of the SWI filter is thus certainly possible for products with revisit times longer than 1 day.

Line 20 pag. 6. So the triplet is the same as above except for the presence of SMAPL3E in place of SMAPL3ESWI?

This is correct.

Figure 1 caption: Stations in Europe are not really visibile (e.g., Denmark). Can you make a bit darker?

Thank you for the comment. We have increased the size of the stations and completely revised the figures.

Figure 2 caption: Please explain better panels b, c and d.

We have expanded the caption with a few additional details.

Line 6-9 pag. 11. I think this is the main reason.

The vertical representativeness could well be the main reason, however, we believe the noise reduction is also an important reason, given the often substantial seemingly random variability between consecutive instantaneous retrievals.

Line 24 pag. 12, “First, ESA-CCISWI incorporates ASCAT, which performed less well in the present evaluation, whereas”. This cannot be a reason if the integration is "optimal" as the different parent products are weighed according to their relative performance. So the second one is more likely the reason. Please rephrase or justify with more solid arguments.

The reviewer is right in theory; as discussed earlier in our response, given the difficulty of satisfying all triple collocation assumptions, our merging approach is unlikely to be fully “optimal,” and we did not claim it was. For this reason, the inclusion of a product of lower quality results in a performance degradation. As mentioned before, we have added the following statement to the preceding paragraph to highlight this: “*Triple collocation-based merging techniques rely on several assumptions (linearity, stationarity, error orthogonality, and zero cross-correlation; Gruber et al., 2016) which are generally difficult to fully satisfy in practice, affecting the optimality of the merging procedure.*”

Line 3 pag. 13. “and satellite-based GPCP V1.3 Daily Analysis (Huffman et al., 2001)” How a daily rainfall can provide 3-hourly estimates?

Good question. This is explained in Section 2.1: *“Since the evaluation was performed at a 3-hourly resolution, we downscaled the two products with a daily temporal resolution (VIC-PGF and GLEAM) to a 3-hourly resolution using nearest neighbor resampling.*” We realize that this is not ideal, but there was no other solution.

Line 20 pag. 14. Please explain what is the meaning of efficiency here.

Thanks for the comment. By efficiency we refer to how realistically the model represents the transformation of precipitation into soil moisture. We have rephrased *“the model efficiency”* to *“the soil moisture simulation efficiency.*”

Line 11 pag. 16. Check this sentence, it appears out of place.

Deleted, thanks.

Table 3: Latency of the products. Change to a more precise value or remove. Several does not provide enough information. I think ERA5 is now available with a delay of three days.

We have provided more precise latency values. The latency of ERA5 appears to be 6 days at this moment.

Table 3: Spatial and temporal resolution. With such a diverse range of products I suggest to replace “temporal resolution and spatial resolution” with spatial and temporal sampling.

Done.

References

Gruber, A., Su, C. H., Crow, W. T., Zwieback, S., Dorigo, W. A., & Wagner, W. (2016). Estimating error cross-correlations in soil moisture data sets using extended collocation analysis. Journal of Geophysical Research: Atmospheres, 121(3), 1208-1219.

Zeng, Q., Chen, H., Xu, C. Y., Jie, M. X., Chen, J., Guo, S. L., & Liu, J. (2018). The effect of rain gauge density and distribution on runoff simulation using a lumped hydrological modelling approach. Journal of hydrology, 563, 106-122.

Su, C. H., Narsey, S. Y., Gruber, A., Xaver, A., Chung, D., Ryu, D., & Wagner, W. (2015). Evaluation of post-retrieval de-noising of active and passive microwave satellite soil moisture. Remote Sensing of Environment, 163, 127-139.

Massari, C., Su, C. H., Brocca, L., Sang, Y. F., Ciabatta, L., Ryu, D., & Wagner, W. (2017). Near real time de-noising of satellite-based soil moisture retrievals: An intercomparison among three different techniques. Remote Sensing of Environment, 198, 17-29.

Dong, J., Crow, W., Reichle, R., Liu, Q., Lei, F., & Cosh, M. H. (2019). A Global Assessment of Added Value in the SMAP Level 4 Soil Moisture Product Relative to Its Baseline Land Surface Model. Geophysical Research Letters, 46(12), 6604-6613.

Review #3

This is a very interesting and promising paper certainly useful to document the biblio-graphical effort on soil moisture evaluation. I feel however that authors have skimmed over some essential explanations and was sometimes wondering if I had the latest version of the manuscript from HESSD (?). The bullet points format of the manuscript does not help and a lot of discussion is missing prior it can be considered for publication. I recommend major revisions, please see below an attempt to help.

We thank the reviewer for their thorough assessment and helpful comments.

Although very important, this kind of evaluation is by design almost never in favour of the satellite based products. It has been highlighted several time in the literature in the past decade that in data rich areas where models are highly constrained by high quality observations, their soil moisture is of better quality that the one retrieved from spatial remote sensing. As the in situ measurements sensors you are using are largely located in those data rich areas, this should be emphasize in the manuscript.

We agree that this issue affects many previous studies. We designed our study to give the satellite products a fair opportunity in two ways:

1. We included six models with non-gauge-based precipitation forcings (ERA5, ERA5-Land, HBV-ERA5 with and without data assimilation, and HBV-IMERG with and without data assimilation). The performance of these models is largely representative of data-poor areas.
2. We evaluated versions of the satellite products processed with SWI filter which generally performed substantially better (Section 3.2). Previous soil moisture product evaluations tended to compare instantaneous soil moisture retrievals directly to the *in situ* measurements, and may therefore have underestimated the 'true' skill of satellite products.

However, despite this, the satellite products still generally performed worse.

We agree with the reviewer that it is important to highlight that models with gauge-based precipitation forcings may not perform as well in data-poor areas, which we have done multiple times in the paper:

- *"It should be kept in mind, however, that these studies, including the present one, used in situ soil moisture measurements from regions with dense rain gauge networks, and hence likely overestimate model performance (Dong et al., 2019)."*
- *"In sparsely gauged areas the four models using precipitation forcings that incorporate daily gauge observations (GLEAM, HBV-MSWEP, HBV-MSWEP+SMAPL3E, and SMAPL4; Table 1) will inevitably exhibit lower performance (but not necessarily lower than the other models)."*

Page 1, Lines 9-8 : a) It gives the false impressions that data assimilation brings an improvement going from 0.69 to 0.72 while models with data assimilation do not all have open-loop counterparts (the opposite being true as well). I know it is the abstract but perhaps you should already give scores that can highlight the added value of data assimilation by considering the mean R values of their open-loop counterpart (HBV+ERA5, HBV+IMERG, HBV+MSWEP). b) I am personally not a big fan of such statement in an abstract and I am not sure it is well supported by your results particu-larly regarding the large distribution of your scores (boxplots of figures 2 & 3) and the lack of discussions on score difference significance

We appreciate the comment. Since the abstract is already quite long we won't be able to present median R improvement scores for each of the products with and without data assimilation. As suggested, we have deleted the statement referred to by the reviewer.

We have added probability (p) values (calculated using the Kruskal-Wallis test) in the manuscript to Table 1 where we compare the performance of ascending and descending overpasses of the single-sensor products. However, we follow the soil moisture product validation recommendations set out by Gruber et al. (2020) and avoid making any statement or interpretation about statistical significance or non-significance, because *"a label of statistical significance does not mean or imply that an association or effect is highly probable, real, true, or important. Nor does a label of statistical nonsignificance lead to the association or effect being improbable, absent, false, or unimportant."*

We did not present p-values for all 254 product combinations for all three performance metrics (R, R_{hi}, and R_{lo}) and did not explicitly report p-values when comparing the medians scores of different products, as this would significantly hamper the readability of the paper. Additionally, we carried out some experiments using the Kruskal-Wallis test on synthetic R distributions with properties similar to the actual R distributions, and found that even small differences in median R of just 0.02 tend to be statistically significant at the *p*=0.05 level, whereas greater differences of 0.03 tend to be statistically significant at the *p*=0.001 level. Thus, the reviewer can safely assume that differences in median *R* of ≥0.02 will be statistically significant at (at least) the *p*=0.05 level.

Regarding the medians of the major product categories (discussed in section 3.8), these are all significantly different at at least the $p=10^{-11}$ level.

Page 1, Line 14 (also Line 16 and true for many part of the manuscript): Are those differences significant? why didn't you provide confidence intervals? Also according to figure 2 it is ESA-CCI_SWI that has a median R value of 0.67 while ESA-CCI has a median R value of 0.56, please clarify. The notion of with/without SWI does not appear in the abstract (?).

The SWI subscript was indeed missing in the abstract and was added to ESA-CCI and the other satellite products in the revised manuscript. Additionally, we added a line about the SWI results. Thanks for the comment.

The difference in median *R* between $\text{ESA-CCI}_{\text{SWI}}$ and MeMo is quite large and indeed highly statistically significant ($p=10^{-7}$). As explained in the preceding response, we prefer to refrain from making statements about statistical significance or non-significance.

Page 2, Line 14 "Additionally, many had a regional (sub-continental) focus [...]" I would not say yours is different (?) Particularly looking at figure 1, please clarify. Also you could add a lot of recent references that had looked at very similar dataset to like to yours. You are only slightly discussing towards the end of your manuscript, please revised

There are numerous soil moisture product evaluations that focused on a single country or a small area (i.e., a sub-continental region), whereas we tried to use all available *in situ* data globally to draw the most generalizable conclusions possible. Our *in situ* data covers the entire conterminous US and thus can be considered at least "continental." We recognize, of course, that the coverage of the *in situ* sensors is far from fully global, and to we have devoted an entire subsection to discussing the generalizability of our results. Our study has already well over 200 references and we cite numerous recent studies that also use ISMN data as reference. We are not sure which recent references are missing.

Page 2, Lines 25-26 “Furthermore, several new or recently reprocessed products have not been thoroughly evaluated yet, such as ERA5 (Hersbach et al., 2020), ERA5-Land (C3S, 2019), and ESA-CCI V04.4 (Dorigo et al., 2017).” For ERA5, Li et al have used 842 qualified sites covering 25 networks (rather recent paper I must admit): https://rmets.onlinelibrary.wiley.com/doi/10.1002/joc.6549 For ESA-CCI, Did you check the product website and documentation? https://www.esa-soilmoisture-cci.org/validation

Thanks for pointing us to the paper and website which are both very interesting. However, our paper was already finalized before Li et al. (2020) appeared online. We were aware of the online ESA-CCI evaluation, but we did not include it in the paper primarily because it has not been peer-reviewed.

Page 3, Line 1 “[. . .] from 826 sensors located primarily in the USA and Europe [. . .]” Thus as for previous studies you have mentioned the extent to which your findings can be generalized is unclear (?), please revise as this sentence could be misleading.

Agreed. We have replaced *“and thus the extent to which their findings can be generalized is unclear”* with *“potentially leading to conclusions with limited generalizability.”*

Page 3, Line 5 Question on SWI appears only here, seems a bit out of the blue (?)please introduce SWI earlier not to confuse readers.

We agree, and have added the following to introduce the SWI: *“There is also still uncertainty around [...] the impact of smoothing filters such as the Soil Wetness index (SWI; Wagner et al., 1999; Albergel et al., 2008) on the performance ranking of products.”*

Page 3, Line 14, section 2.1 I am wondering here if I have the correct version of the manuscript as several dataset are not presented? It is a general comment that you have to justify why you have used those 18 dataset and not others, otherwise it looks like cherry-picking. While some are state-of-the arts, others are self-made, please revised the choice and presentation of the dataset.

All products are introduced in Table 1, which we refer to in the first sentence of Section 2.1. We have added the following to justify our selection of products: *“We evaluated six products per category, which was sufficient to compare the performance among and within product categories and address the questions posed in the introduction. We only considered widely used products with (quasi-)global coverage and we attempted to* *keep the selection of products in each category as diverse as possible. For example, we considered products based on several major satellite missions used for global soil moisture mapping (AMSR2, ASCAT, SMAP, and SMOS), models of various type and complexity (with and without calibration), different sources of precipitation data (satellites, reanalyses, gauges, and combinations thereof), and various data merging and assimilation techniques (with different inputs).*"

Page 3, Line 24 I assume you have used soil temperature of the first layer of soil between 1-7cm, is so please say it. Alternatively you could have discarded in situ measurements of soil moisture when associated measurements of soil temperature (if available) was < 4 dC

We agree and have added 0–7 cm in reference to the ERA5 soil temperature estimates.

Page 4, Lines 16-17 Add references I appropriate

We would be happy to add relevant references but we are not aware of any. This is a relatively simple part of our methodology that we believe can be understood and replicated without references.

Page 5, Line 10 "The model was run twice for 2010–2019 [. . .]" Please clarify if this was done for each forcing dataset (I assume so)

Yes, the initialization was performed for each precipitation dataset. In the revised manuscript HBV is recalibrated for each precipitation dataset. We have added the following text: *"To avoid giving one of the precipitation datasets an unfair advantage, we recalibrated the model for each of the three precipitation datasets (ERA5, IMERG, and MSWEP).*"

Page 5, Line 20 "We calibrated the 7 relevant parameters of HBV [. . .]" This will have to be discuss further already if it impacts your results wrt to the land surface model based product?

We have added the following to the revised manuscript: *"The calibrated models (HBV and the Catchment model underlying SMAPL4) may, however, perform slightly worse in regions with climatic and physiographic conditions dissimilar to the in situ sensors used for calibration (but likely still better than the uncalibrated models).*" Section 3.7 discusses the benefits and limitations of model calibration in detail, including implications with respect to the land surface model-based products, as suggested by the reviewer.

Page 6, section 2.5 Are they all using the same measurement methodology?

Thanks for the comment. We added the following text: *"The measurements were performed using various types of sensors, including time-domain reflectometry sensors, frequency-domain reflectometry sensors, capacitance sensors, and cosmic-ray neutron sensors, among others."*

Page 7, figure 1 In such study this kind of global maps tend to show areas with no data more than areas with data. It is not obvious than 2 two zooms over North America and Europe add anything, perhaps you could have one figure with 3 panels, North America,Europe and Australia (?)

Agreed; we have revised this figure (as well as the other figures) as proposed by the reviewer, but with four panels instead of three (Alaska, Europe, conterminous US, and Southeastern Australia). Thanks for the suggestion.

Also I suspect here that most of the stations in the "cold" class over North America are from the SNOTEL network located in mountainous area where the retrieval of soil moisture from space is rather complex. This should be emphasise in the text at it is biasing your results.

We agree; thanks for the comment. The retrieval may indeed be more complex in cold regions, which we mention in the paper: *"the confounding influence of dense vegetation cover (de Rosnay et al., 2006; Gruhier et al., 2008; Dorigo et al., 2010), highly organic soils (Zhang et al., 2019b), and standing water (Ye et al., 2015; Du et al., 2018) on soil moisture retrievals."* The influence of mountainous terrain on the retrievals is also mentioned in the paper: *"Most satellite products performed worse in terms of R in areas of steep terrain (Fig. 2d), consistent with previous evaluations (Paulik et al.,2014; Karthikeyan et al., 2017a; Ma et al., 2019), and attributed to the confounding effects of relief on the upwelling microwave brightness temperature observed by the radiometer (Mialon et al., 2008; Pulvirenti et al., 2011; Guo et al., 2011)."*

An additional explanation for the lower performance in cold regions (missing from our original submission) may be that the sensors are less representative of the coarse scale of the products. We therefore added the following: *"it could also be that the in situ measurements are [...] less representative of satellite footprints or model grid-cells."*

Page 9, figure 2 I may have missed a point but I did not understand how did you obtain 3-hourly data for e.g. ASCAT, SMOS, ESA-CCI, SMAP...please revise.

This was indeed not clearly explained. We have added the following text: *"For the satellite products without SWI filter, we matched the instantaneous soil moisture retrievals with coincident 3-hourly in situ measurements to compute the R values."*

It would have been easier to have them close to one another (SWI and not SWI) on your figures but has you have several questions to answer it was probably not easy to pick up the correct order of products for those figures.

We agree that having the SWI and non-SWI products close to each other in the figure would be useful for answering the SWI-related question but less useful for the other questions addressed in the study.

Page 11, section 3.2 My personal opinion is that this is a low pass filter smoothing the time-series, nothing more

We agree with this observation; the SWI filter is in essence a low-pass filter smoothing the time series.

Page 11, section 3.3 Are you R values significant? I may have missed something here but from your figures 2 and 3 (boxplots distribution) it is difficult for me to give a clear answer to this question (while you are doing it in the abstract)

The large majority of R values are highly statistically significant, since an R value of just 0.14 tends to be needed to obtain a statistically significant correlation (at the 0.05 level) for a sample size of 200 (the minimum sample size before an R value is calculated in this study; see Figure 1). Our R values are, however, generally much higher (Fig. 2) and our sample sizes much greater, and therefore our R values will be much more statistically significant.

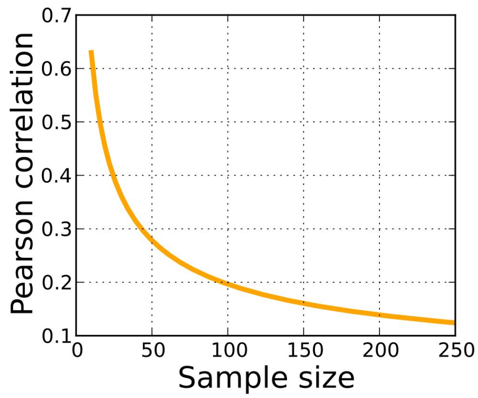

Figure 1. Plot showing the minimum value of Pearson's correlation coefficient (*R*) that would be significant at the 0.05 level for a given sample size. Source: https://commons.wikimedia.org/wiki/File:Correlation_significance.svg.

As explained in the beginning of this response letter, following recommendations of Gruber et al. (2020) we refrain from making statements about statistical significance or non-significance.

Page 12, section 3.4 Line 21 “[. . .] the central Rocky Mountains [. . .]” This are usually area where it is difficult to retrieve soil moisture form space. Memo perhaps does better than ESA-CCI but is it good? are we talking about R values going from 0.2 to 0.3 orfrom 0.6 to 0.8? From figure 4 it is difficult to see anything (at least to me). Again, are the differences significant? Lines 22-23 Confidence interval would help Line 29 Please clarify “[. . .] from the best sensor each day[. . .]”

Please see Fig. 2d of our paper for median *R* values for mountainous versus flat areas (denoted by the letters S and F, respectively) for the different products. The median *R* is 0.61 for $\text{ESA-CCI}_{\text{SWI}}$ versus 0.73 for MeMo, which is a substantial difference (statistically significant at the $p=10^{-6}$ level).

Page 12, Lines 31-32 Is it surprising to find the 3 calibrated HBV models leading this ranking? Again I would not claim such a best to worst ranking without discussing the significance of scores.

This was somewhat surprising given the simplicity of HBV and the fact that HBV has been designed for runoff estimation in cold regions. Conversely, numerous studies have demonstrated the flexibility and effectiveness of HBV, and the model has been calibrated against *in situ* soil moisture measurements. See our discussion in the second paragraph of the subsection in question: *“This demonstrates that soil moisture estimates from complex, data-intensive models (H-TESSEL underlying ERA5 and ERA5-Land, GLEAM, and the Catchment model underlying SMAPL4) are not* *necessarily more accurate than those from relatively simple, calibrated models (HBV).*" Note that we have devoted an entire subsection to discussing the benefits and limitations of calibration (Section 3.7).

Page 13, figure 4 (also true for figure 5) Not sure this figure is very helpful as hardly visible (?) Perhaps you could use scatterplots, e.g. x-axis R for ESA-CCI vs in situ,y-axis R for MeMo vs in situ and then use color-codes for any classification you like.

We appreciate the comment but a scatterplot would not tell us *where* the products perform better or worse and thus would be much less informative. We could indeed use color codes to denote the locations, but we feel a map is more clear. That said, we have completely redesigned the figure and hope it is more useful now.

Page 13, Line 1 ERA5 is a coupled land atmosphere system where ASCAT has been assimilated. Could you comment on the impact it may (or may not) have when using it to force HTESSEL land surface model in ERA5-Land? is it fully independent from ASCAT?

The assimilation of ASCAT soil moisture is unlikely to have influenced the precipitation generated by ERA5, given (i) the small influence of the assimilation on the soil moisture simulations (Muñoz Sabater et al., 2019) and (ii) the vast amounts of other observations (ground and satellite) also assimilated (Hersbach et al., 2020).

P.14, Line 1 Data intensive models could also be calibrated don't you think? I personally thing it is wrong to oppose land surface model and calibrated hydrological models. Their objectives are different.

We may be misunderstanding the comment, but we do not consider it unfair to include both land surface models and calibrated hydrological models in the same evaluation. For users simply looking for the most accurate product — probably the most common type of end-user — the data source or modeling approach is not important. We fully agree that the design objectives can be different and that data-intensive models can be calibrated as well. The calibration of computationally demanding models is however more challenging, as mentioned at the end of Section 3.7.

P.14, Line 8 There is more to say from such figure as figure 5 (?) e.g. discuss the geographical patterns

Thank you for the suggestion. We have added the following: "*For HBV-IMERG, the greatest improvements were found over the central Rocky Mountains (Fig. 5), where IMERG performs relatively poorly (Beck et al., 2019a).*"

P.14, Lines 21-23 Please discuss if it is likely to be because of the inputs quality (AS-CAT/SMOS) or a methodological matter.

We explain in the sentence thereafter that it is probably a methodological issue: *"They attributed this to the adverse impact of simultaneously assimilated screen-level temperature and relative humidity observations on the soil moisture estimates.*"

P.14, Lines 26 There is also a study showing that the assimilation of ESA CCI inGLEAM leads to a decrease of quality (Brecht et al., 2018 GMD?)

We suspect the reviewer might be referring to Martens et al. (2016). However, this study shows small (not negligible) improvements in the soil moisture simulations after DA.

Page 14, Lines 32-33 Which was expected right?

This was indeed in accordance with our expectations, but this has not been explicitly discussed in previous studies (to our knowledge).

P.15, section 3.7 Perhaps this could be moved few sections above?

We appreciate the comment. However, since we compare the benefits of model calibration and data assimilation in this section, we have to discuss the data assimilation results first. It is therefore not possible to move this subsection.

P.16, Lines 16-19 In agreement with many previous studies (e.g. Albergel et al., 2010,HESS, Dorigo et a;., 2017, RSE...)

We agree and list eight previous studies that agree with our results: *"Our performance ranking of the major product categories is consistent with previous studies for the conterminous USA (Liu et al., 2011; Kumar et al., 2014; Fang et al., 2016; Dong et al., 2020), Europe (Naz et al., 2019), and the globe (Albergel et al., 2012; Tian et al., 2019; Dong et al., 2019).*"

P.17, section 3.9 Perhaps worth referencing / discussing Reichle et al,. 2019 ? Verification of the SMAP Level-4 Soil Moisture Analysis Using Rainfall Observations in Australia, https://ieeexplore.ieee.org/document/8898398

Thanks for the suggestion. We are not sure which statement of Section 3.9 is supported by the results of Reichle et al. (2019). Note that the author of that study, Rolf Reichle, is also co-author of the present study.

Evaluation of 18 satellite- and model-based soil moisture products using *in situ* measurements from 826 sensors

Hylke E. Beck[1], Ming Pan[1], Diego G. Miralles[2], Rolf H. Reichle[3], Wouter A. Dorigo[4], Sebastian Hahn[4], Justin Sheffield[5], Lanka Karthikeyan[6], Gianpaolo Balsamo[7], Robert M. Parinussa[8], Albert I.J.M. van Dijk[9], Jinyang Du[10], John S. Kimball[10], Noemi Vergopolan[1], and Eric F. Wood[1]

[1]Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ, USA
[2]Hydro-Climate Extremes Lab (H-CEL), Ghent University, Ghent, Belgium
[3]Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, MD, USA
[4]Department of Geodesy and Geoinformation (GEO), Vienna University of Technology, Vienna, Austria
[5]School of Geography and Environmental Science, University of Southampton, Southampton, United Kingdom
[6]Centre of Studies in Resources Engineering, Indian Institute of Technology, Bombay, Powai, Mumbai 400 076, India
[7]European Centre for Medium-Range Weather Forecasts (ECMWF), Reading, UK
[8]School of Geographic Sciences, Nanjing University of Information Science & Technology, Nanjing, Jiangsu, People's Republic of China
[9]Fenner School of Environment and Society, Australian National University, Canberra, Australian Capital Territory, Australia
[10]Numerical Terradynamic Simulation Group, University of Montana, Missoula, MT 59801, USA

Correspondence: Hylke E. Beck(hylke.beck@gmail.com)

Abstract. Information about the spatiotemporal variability of soil moisture is critical for many purposes, including monitoring of hydrologic extremes, irrigation scheduling, and prediction of agricultural yields. We evaluated the temporal dynamics of 18 state-of-the-art (quasi-)global near-surface soil moisture products, including six based on satellite retrievals, six based on models without satellite data assimilation (referred to hereafter as "open-loop" models), and six based on models that assimilate satellite soil moisture or brightness temperature data. Seven of the products are introduced for the first time in this study: one multi-sensor merged satellite product called MeMo and six estimates from the HBV model with three precipitation inputs (ERA5, IMERG, and MSWEP) and with and without assimilation of SMAPL3E satellite retrievals, respectively. As reference, we used *in situ* soil moisture measurements between 2015 and 2019 at 5-cm depth from 826 sensors, located primarily in the USA and Europe. The 3-hourly Pearson correlation (R) was chosen as the primary performance metric. ~~The median $R \pm$ interquartile range across all sites and products in each category was 0.66 ± 0.30 for the satellite products, 0.69 ± 0.25 for the open-loop models, and 0.72 ± 0.22 for the models with satellite data assimilation~~Application of the Soil Wetness Index (SWI) smoothing filter resulted in improved performance for all satellite products. The best-to-worst performance ranking of the four single-sensor satellite products was SMAPL3E$_{\mathrm{SWI}}$, SMOS$_{\mathrm{SWI}}$, ~~SMOS,~~ AMSR2$_{\mathrm{SWI}}$, and ASCAT$_{\mathrm{SWI}}$, with the L-band-based SMAPL3E$_{\mathrm{SWI}}$ (median R of 0.72) outperforming the others at 50 % of the sites. Among the two multi-sensor satellite products (MeMo and ESA-CCI$_{\mathrm{SWI}}$), MeMo performed better on average (median R of 0.72 versus 0.67), mainly due to the inclusion of SMAPL3E$_{\mathrm{SWI}}$. The best-to-worst performance ranking of the six open-loop models was HBV-MSWEP, HBV-ERA5, ERA5-Land, HBV-IMERG, VIC-PGF, and GLDAS-Noah. This ranking largely reflects the quality of the precipitation forcing. HBV-MSWEP (median R of 0.78) performed best not just among the open-loop models but among all products. The calibration of HBV improved

the median R by $+0.12$ on average compared to random parameters, highlighting the importance of model calibration. The best-to-worst performance ranking of the six models with satellite data assimilation was HBV-MSWEP+SMAPL3E, HBV-ERA5+SMAPL3E, GLEAM, SMAPL4, HBV-IMERG+SMAPL3E, and ERA5. The assimilation of SMAPL3E retrievals into HBV-IMERG improved the median R by $+0.06$, suggesting that data assimilation yields significant benefits at the global scale.

1 Introduction

Accurate and timely information about soil moisture is valuable for many purposes, including drought monitoring, water resources management, irrigation scheduling, prediction of vegetation dynamics and agricultural yields, forecasting floods and heatwaves, and understanding climate change impacts (Wagner et al., 2007; Vereecken et al., 2008; Ochsner et al., 2013; Dorigo and de Jeu, 2016; Brocca et al., 2017; Miralles et al., 2019; Tian et al., 2019; Karthikeyan et al., 2020; Chawla et al., 2020). Over recent decades, numerous soil moisture products suitable for these purposes have been developed, each with strengths and weaknesses (see Table 1 for a non-exhaustive overview). The products differ in terms of design objective, spatiotemporal resolution and coverage, data sources, algorithm, and latency. They can be broadly classified into three major categories: (i) products directly derived from active- or passive-microwave satellite observations (Zhang and Zhou, 2016; Karthikeyan et al., 2017b), (ii) hydrological or land surface models without satellite data assimilation (referred to hereafter as "open-loop" models; ~~Cammalleri et al., 2015; Bierkens, 2015; Kauffeldt et al., 2016~~Cammalleri et al., 2015; Bierkens, 2015; Kauffeldt et al., 2016; Chen and Y), and (iii) hydrological or land surface models that assimilate soil moisture retrievals or brightness temperature observations from microwave satellites ~~(Moradkhani, 2008; Liu et al., 2012; Lahoz and De Lannoy, 2014; Reichle et al., 2017)~~(Moradkhani, 2008; Pan et al., .

~~A plethora of studies addressed the important task of evaluating~~ Numerous studies have evaluated these soil moisture products using *in situ* soil moisture measurements (e.g., ~~Dorigo et al., 2011; Bindlish et al., 2018~~Jackson et al., 2010; Bindlish et al., 2018), other independent soil moisture products (e.g., Chen et al., 2018; Dong et al., 2019), remotely-sensed vegetation greenness data (e.g., Tian et al., 2019), or precipitation data (e.g., Crow et al., 2010; Karthikeyan and Kumar, 2016). Pronounced differences in spatiotemporal dynamics and accuracy were found among the products, even among those derived from the same data source. However, most studies evaluated only one specific product or a small subset (≤ 3) of the available products (e.g., ~~Martens et al., 2017; Liu et al., 2019; Zhang et al., 2019b~~Martens et al., 2017; Liu et al., 2019; Zhang et al., 2019b; Tavakol et al., 2019). Additionally, many had a regional (sub-continental) focus (e.g., Albergel et al., 2009; Gruhier et al., 2010; Griesfeller et al., 2016), ~~and thus the extent to which their findings can be generalized is unclear~~potentially leading to conclusions with limited generalizability. Furthermore, several new or recently reprocessed products have not been thoroughly evaluated yet, such as ERA5 (Hersbach et al., 2020), ERA5-Land (C3S, 2019), and ESA-CCI V04.4 (Dorigo et al., 2017). There is also still uncertainty around, for example, the effectiveness of multi-sensor merging techniques (Petropoulos et al., 2015), the impact of model complexity on the accuracy of soil moisture simulations (Fatichi et al., 2016), ~~and~~ the degree to which model deficiencies and precipitation data quality affect the added value of data assimilation (Xia et al., 2019), and the impact of smoothing filters such as the Soil Wetness index (SWI; Wagner et al., 1999; Albergel et al., 2008) on the performance ranking of products.

Our main objective was to undertake a comprehensive evaluation of 18 state-of-the-art ~~(sub-)daily (~~quasi-)global near-surface soil moisture products in terms of their temporal dynamics (Section 2.1). Our secondary objective was to introduce seven new soil moisture products (one multi-sensor merged satellite product called MeMo introduced in Section 2.2 and six HBV model-based products introduced in Sections 2.3 and 2.4). As reference for the evaluation, we used *in situ* soil moisture measurements between 2015 and 2019 from 826 sensors located primarily in the USA and Europe (Section 2.5). We aim to shed light on the ~~strengths and weaknesses~~ advantages and disadvantages of different soil moisture products and on the merit of ~~different~~ various technological and methodological innovations by addressing nine ~~pertinent questions~~ key questions frequently faced by researchers and end-users alike:

1. How do the ascending and descending retrievals perform (Section 3.1)?
2. What is the impact of the ~~Soil Wetness Index (SWI)~~ SWI smoothing filter (Section 3.2)?
3. What is the relative performance of the single-sensor satellite products (Section 3.3)?
4. How do the multi-sensor merged satellite products perform (Section 3.4)?
5. What is the relative performance of the open-loop models (Section 3.5)?
6. How do the models with satellite data assimilation perform (Section 3.6)?
7. What is the impact of model calibration (section 3.7)?
8. How do the major product categories compare (Section 3.8)?
9. To what extent are our results generalizable to other regions (Section 3.9)?

2 Data and methods

2.1 Soil moisture products

We evaluated in total 18 near-surface soil moisture products, including six based on satellite observations, six based on open-loop models, and six based on models that assimilate satellite data (Table 1). We evaluated six products per category, which was sufficient to compare the performance among and within product categories and address the questions posed in the introduction. We only considered widely used products with (quasi-)global coverage and we attempted to keep the selection of products in each category as diverse as possible. For example, we considered products based on several major satellite missions used for global soil moisture mapping (AMSR2, ASCAT, SMAP, and SMOS), models of various type and complexity (with and without calibration), different sources of precipitation data (satellites, reanalyses, gauges, and combinations thereof), and various data merging and assimilation techniques (with different inputs).

The units differed among the products; some are provided in volumetric water content (typically expressed in m^3 m^{-3}; e.g., ERA5) and others in degree of saturation (typically expressed in %; e.g., ASCAT). We did not harmonize the units among the products, because the Pearson correlation coefficient — the performance metric used in the current study (Section 2.6) — is insensitive to the units. Since the evaluation was performed at a 3-hourly resolution, we downscaled the two products with a daily temporal resolution (VIC-PGF and GLEAM) to a 3-hourly resolution using nearest neighbor resampling. In contrast to the model products, the satellite products (with the exception of ASCAT) often do not provide retrievals when the soil is frozen or snow-covered (Supplement Fig. S1). To keep the evaluation consistent ~~, we discarded~~ (Gruber et al., 2020), we used ERA5 (Hersbach et al., 2020) to discard the estimates of all 18 products when the near-surface soil temperature of layer 1 (0–7 cm) was $< 4°$C and/or the snow depth was > 1 mm~~(both determined using ERA5; Hersbach et al., 2020)~~.

~~For all satellite products with the exception of MeMo~~

To deepen the vertical support of the superficial satellite observations and suppress noise, we also evaluated 3-hourly versions of the satellite products processed using the ~~Soil Wetness Index (SWI)~~ SWI exponential smoothing filter (Wagner et al., 1999; Albergel et al., 2008)~~, which reduces noise and improves the consistency with *in situ* measurements~~. MeMo was not processed as it was derived from SWI-filtered products. The SWI filter is defined according to:

$$\mathrm{SWI}(t) = \frac{\sum_i \mathrm{SM}_{\mathrm{sat}}(t_i)e^{-\frac{t-t_i}{T}}}{\sum_i e^{-\frac{t-t_i}{T}}}, \tag{1}$$

where $\mathrm{SM}_{\mathrm{sat}}$ (units depend on the product) is the soil moisture retrieval at time t_i, T (days) represents the time lag constant, and t represents the 3-hourly time step. T was set to 5 days for all products, as the performance did not change markedly using different values, as also reported in previous studies (Albergel et al., 2008; Beck et al., 2009; Ford et al., 2014; Pablos et al., 2018). Following Pellarin et al. (2006), the SWI at time t was only calculated if ≥ 1 retrievals were available in the interval $(t-T,\, t]$ and ≥ 3 retrievals were available in the interval $[t-3T,\, t-T]$.

The vertical support is physically consistent with *in situ* soil moisture measurements at 5-cm depth for most models. The average depth of the soil layer (i.e., half the depth of the lower boundary) is 2.5 cm for SMAPL4, 3.5 cm for ERA5 and ERA5-Land, 5 cm for GLEAM, 8.5 cm for HBV-ERA5, 6.6 cm for HBV-IMERG, 7.3 cm for HBV-MSWEP, and 15 cm for VIC-PGF (Table 1; Supplement Table S1). The soil layers of HBV may seem too deep, especially since they represent conceptual "buckets" that can be fully filled with water, in contrast to the soil layers of the other models which additionally consist of mineral and organic matter. However, the soil layer depths of HBV were calibrated (see Section 2.3) and are thus empirically consistent with *in situ* measurements at 5-cm depth.

2.2 Merged soil Moisture (MeMo) product

Merged soil Moisture (MeMo) is a new 3-hourly soil moisture product derived by merging the soil moisture anomalies of three single-sensor passive-microwave satellite products with SWI filter ($\mathrm{AMSR2}_{\mathrm{SWI}}$, $\mathrm{SMAPL3E}_{\mathrm{SWI}}$, and $\mathrm{SMOS}_{\mathrm{SWI}}$; Table 1). MeMo was produced for 2015–2019 (the period with data for all three products) as follows:

1. Three-hourly soil moisture time series of $\mathrm{AMSR2}_{\mathrm{SWI}}$, $\mathrm{SMAPL3E}_{\mathrm{SWI}}$, $\mathrm{SMOS}_{\mathrm{SWI}}$, the active-microwave satellite product $\mathrm{ASCAT}_{\mathrm{SWI}}$, and the open-loop model HBV-MSWEP were normalized by subtracting the long-term means and dividing by the long-term standard deviations of the respective products (calculated for the period of overlap).

2. Three-hourly anomalies were calculated for the five products by subtracting their respective seasonal ~~climatologies~~averages. The seasonal climatology was calculated by taking the multi-year mean for each day of the year, after which we applied a 30-day central moving mean to eliminate noise. The moving mean was only calculated if > 21 days with values were present in the 30-day window. Due to the large number of missing values in winter (Supplement Fig. S1), we were not able to compute the seasonality and, in turn, the anomalies in winter for some satellite products.

3. Time-invariant merging weights for AMSR2$_{\text{SWI}}$, SMAPL3E$_{\text{SWI}}$, and SMOS$_{\text{SWI}}$ were calculated using extended triple collocation (McColl et al., 2014), a technique to estimate Pearson correlation coefficients (R) for independent products with respect to an unknown truth. The R values for the respective products were determined using the triplet consisting of the product in question in combination with ASCAT ~~SWI~~ and HBV-MSWEP, which are independent from each other and from the passive products. The R values were only calculated if > 200 coincident anomalies were available. The weights were calculated by squaring the R values.

4. For each 3-hourly time step, we calculated the weighted mean of the available anomalies of AMSR2$_{\text{SWI}}$, SMAPL3E$_{\text{SWI}}$, and SMOS$_{\text{SWI}}$. If only one anomaly was available, this value was used and no averaging was performed. The climatology of SMAPL3E — the best-performing product in our evaluation — was added to the result, to yield the MeMo soil moisture estimates.

2.3 HBV hydrological model

Six new 3-hourly soil moisture products were produced using the Hydrologiska Byr Vattenbalansavdelning (HBV) conceptual hydrological model (Bergström, 1976, 1992) forced with three different precipitation datasets and with and without assimilation of SMAPL3E soil moisture estimates, respectively (Table 1). HBV was selected because of its low complexity, high agility, computational efficiency, and succesful application used in numerous studies spanning a wide range of climate and physiographic conditions (e.g., Steele-Dunne et al., 2008; Driessen et al., 2010; Beck et al., 2013; Vetter et al., 2015; Jódar et al., 2018). The model has one soil moisture store, two groundwater stores, and 12 free parameters. Among the 12 free parameters, 7 are relevant for simulating soil moisture as they pertain to the snow or soil routines, while 5 are irrelevant for this study as they pertain to runoff generation or deep percolation. The soil moisture store has two inputs (precipitation and snowmelt) and two outputs (evaporation and recharge). The model was run twice for 2010–2019; the first time to initialize the soil moisture store, and the second time to obtain the final outputs.

HBV requires time series of precipitation, potential evaporation, and air temperature as input. For precipitation, we used three different datasets: (i) the reanalysis ERA5 (hourly 0.28° resolution; Hersbach et al., 2020); (ii) the satellite-based IMERG dataset (Late Run V06; 30-minutes 0.1° resolution; Huffman et al., 2014, 2018); and (iii) the gauge-, satellite-, and reanalysis-based MSWEP dataset (V2.4; 3-hourly 0.1° resolution; ~~??). We calculated 3-hourly accumulations for~~ Beck et al., 2017b, 2019b). For the ERA5 and IMERG datasets, we calculated 3-hourly precipitation accumulations. Daily potential evaporation was estimated using the Hargreaves (1994) equation from daily minimum and maximum air temperature. ~~Temperature~~ The daily potential evaporation data were downscaled to 3-hourly using nearest neighbour resampling. Air temperature estimates were taken from

ERA5~~, downscaled to 0.1° and bias-corrected~~ . To improve the representation of mountainous regions and ameliorate potential biases, the ERA5 air temperature data were matched on a monthly ~~basis through an additive approach using~~ climatological basis using an additive (as opposed to multiplicative) approach to the comprehensive station-based WorldClim climatology (V2; 1-km resolution; Fick and Hijmans, 2017). ~~The daily potential evaporation data were downscaled to 3-hourly using nearest neighbour resampling.~~

We calibrated the 7 relevant parameters of HBV using *in situ* soil moisture measurements ~~between 2010 and 2019~~ from 177 independent sensors from the International Soil Moisture Network (ISMN) archive ~~that were not used for performance assessment~~ (Section 2.5; Supplement Fig. S2). These sensors did not have enough measurements during the evaluation period (March 31, 2015, to September 16, 2019) and thus were available for an independent calibration exercise. The parameter space was explored by generating $N = 500$ candidate parameter sets using Latin hypercube sampling (McKay et al., 1979), which splits the parameter space up into N equal intervals and generates parameter sets by sampling each interval once in a random manner. The model was subsequently run for all candidate parameter sets, after which we selected the parameter set with the best overall performance across the 177 sites (Supplement Table S1). As objective function, we used the median Pearson correlation coefficient (R) calculated between 3-hourly *in situ* and simulated soil moisture time series. ~~As forcing, we used the MSWEP precipitation dataset because of its favourable performance in numerous evaluations (e.g., Alijanian et al., 2017; Sahlu et al., 2017; Bai and Liu, 2018; Casson et al., 2018; Beck et al., 2017c, 2019a; Zhang et al., 2019a; Satgé et a). The calibrated parameter set was used for all HBV runs, including those forced with ERA5 or IMERG precipitation.~~ To avoid giving one of the precipitation datasets an unfair advantage, we recalibrated the model for each of the three precipitation datasets (ERA5, IMERG, and MSWEP).

2.4 Soil moisture data assimilation

Instantaneous soil moisture retrievals (without SWI filter) from SMAPL3E (Table 1) were assimilated into the HBV model forced with the three above-mentioned precipitation datasets (ERA5, IMERG, and MSWEP). Previous regional studies that successfully used HBV to assess the value of data assimilation include Parajka et al. (2006), Montero et al. (2016), and Lü et al. (2016). We used the simple Newtonian nudging technique of Houser et al. (1998) that drives the soil moisture state of the model towards the satellite observations. Nudging techniques are computationally efficient and easy to implement, and have therefore been used in several studies (e.g., Brocca et al., 2010b; Dharssi et al., 2011; Capecchi and Brocca, 2014; Laiolo et al., 2016; Cenci et al., 2016; Martens et al., 2016). For each grid-cell, the soil moisture state of the model was updated when a satellite observation was available according to:

$$\mathrm{SM}^{+}_{\mathrm{mod}}(t) = \mathrm{SM}^{-}_{\mathrm{mod}}(t) + kG\left[\mathrm{SM}^{\mathrm{sc}}_{\mathrm{sat}}(t) - \mathrm{SM}^{-}_{\mathrm{mod}}(t)\right], \qquad (2)$$

where $\mathrm{SM}^{+}_{\mathrm{mod}}$ and $\mathrm{SM}^{-}_{\mathrm{mod}}$ (mm) are the updated and *a priori* soil moisture states of the model, respectively, $\mathrm{SM}^{\mathrm{sc}}_{\mathrm{sat}}$ (mm) are the rescaled satellite observations, and t is the 3-hourly time step. The satellite observations were rescaled to the open-loop model space using cumulative distribution function (CDF) matching (Reichle and Koster, 2004).

The nudging factor k ($-$) was set to 0.1 as this gave satisfactory results. The gain parameter G ($-$) determines the magnitude of the updates and ranges from 0 to 1. G is generally calculated based on relative quality of the satellite retrievals and the open-loop model. Most previous studies used a spatially and temporally uniform G (e.g., Brocca et al., 2010b; Dharssi et al., 2011; Capecchi and Brocca, 2014; Laiolo et al., 2016; Cenci et al., 2016). Conversely, Martens et al. (2016) used the triple collocation technique (Scipal et al., 2008) to obtain spatially variable G values. Here we calculated G in a similar fashion according to:

$$G = \frac{R_{\text{sat}}^2}{R_{\text{sat}}^2 + R_{\text{mod}}^2}, \tag{3}$$

where R_{sat} and R_{mod} ($-$) are Pearson correlation coefficients with respect to an unknown truth for SMAPL3E and HBV, respectively, calculated using extended triple collocation (Section 2.2). R_{sat} was determined using 3-hourly anomalies of the triplet SMAPL3E, $\text{ASCAT}_{\text{SWI}}$, and HBV-MSWEP (Table 1) which are based on passive microwaves, active microwaves, and an open-loop model, respectively. R_{mod} was determined using 3-hourly anomalies of the triplet HBV (forced with either ERA5, IMERG, or MSWEP), $\text{ASCAT}_{\text{SWI}}$, and $\text{SMAPL3E}_{\text{SWI}}$. The anomalies were calculated by subtracting the seasonal ~~climatologies~~ averages of the respective products. The seasonal ~~climatologies~~ averages were determined as described in Section 2.2. The R_{sat} and R_{mod} values were only calculated if > 200 coincident anomalies were available. The resulting G values vary in space but are constant in time.

2.5 *In situ* soil moisture measurements

As reference for the evaluation, we used harmonized and quality-controlled *in situ* volumetric soil moisture measurements ($\text{m}^3\ \text{m}^{-3}$) from the ISMN archive (Dorigo et al., 2011, 2013; Appendix Table A1). The measurements were performed using various types of sensors, including time-domain reflectometry sensors, frequency-domain reflectometry sensors, capacitance sensors, and cosmic-ray neutron sensors, among others. Similar to numerous previous evaluations (e.g., Albergel et al., 2009; Champagne et al., 2010; Albergel et al., 2012; Wu et al., 2016), we selected measurements from sensors at a depth of 5 cm ($\pm$2 cm). Since the evaluation was performed at a 3-hourly resolution, the measurements in the ISMN archive, which have a hourly resolution, were resampled to a 3-hourly resolution. We only used sensors with a 3-hourly record length $>$ 1 year (not necessarily consecutive) during the evaluation period from March 31, 2015, to September 16, 2019. We did not average the measurements of sites with multiple sensors to avoid potentially introducing discontinuities in the time series. In total 826 sensors, located in the USA (692), Europe (117), and Australia (17), were available for evaluation (Fig. 1). The median record length was 3.0 years.

2.6 Evaluation approach

We evaluated the 18 near-surface soil moisture products (Table 1) for the 4.5-year long period from March 31, 2015 (the date on which SMAP data became available), to September 16, 2019 (the date on which we started processing the products). As performance metric, we used the Pearson correlation coefficient (R) calculated between 3-hourly soil moisture time series from the ~~sensor and the product~~*in situ* sensors and the products, similar to numerous previous studies (e.g.,

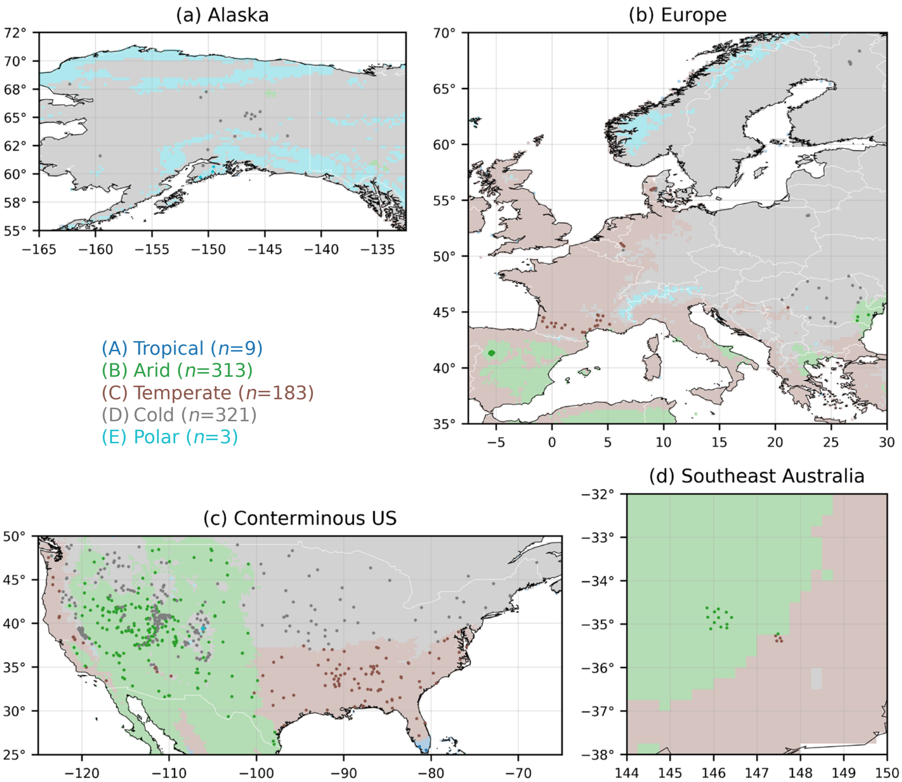

Figure 1. Major Köppen-Geiger climate class (Beck et al., 2018) of the 826 sensors used as reference. n denotes the number of sensors in each class.

Karthikeyan et al., 2017a; Al-Yaari et al., 2017; Kim et al., 2018). R measures how well the *in situ* and product time series correspond in terms of temporal variability, and thus evaluates the most important aspect of soil moisture time series for the majority of applications (Entekhabi et al., 2010; Gruber et al., 2020). It is insensitive to systematic differences in mean and variance, which can be substantial due to: (i) the use of different soil property maps as input to the retrieval algorithms and hydrological models ~~(Koster et al., 2009)~~(Teuling et al., 2009; Koster et al., 2009); and (ii) the inherent scale discrepancy between *in situ* point measurements and satellite footprints or model grid-cells ~~(Miralles et al., 2010; Crow et al., 2012)~~ (Miralles et al., 2010; Crow et al., 2012; Gruber et al., 2020).

Additionally, to quantify the performance of the products at different time scales, we calculated Pearson correlation coefficients for the ~~low- and~~ low-frequency fluctuations (i.e., the slow variability at monthly and longer time scales; R_{lo}) and the high-frequency fluctuations ~~of the~~ (i.e., the fast variability at 3-hourly ~~time series (~~$\cancel{R_{\mathrm{lo}}}$ ~~and~~ to monthly time scales; R_{hi}~~, respectively~~). The low-frequency fluctuations were isolated using a 30-day central moving mean, similar to previous studies (e.g., Albergel et al., 2009; Al-Yaari et al., 2014; Su et al., 2016). The moving mean was calculated only if > 21 days with ~~values~~

estimates were present in the 30-day window. The high-frequency fluctuations were isolated by subtracting the low-frequency fluctuations from the original 3-hourly time series. ~~We~~

To ensure a fair evaluation, we discarded the estimates of all products when the near-surface soil temperature was $< 4°C$ and/or the snow depth was > 1 mm (both determined using ERA5; Hersbach et al., 2020). For ~~each sensor and product~~the satellite products without SWI filter, we matched the instantaneous soil moisture retrievals with coincident 3-hourly *in situ* measurements to compute the R values. Since the evaluation was performed at a 3-hourly resolution, we downscaled the two products with a daily temporal resolution (VIC-PGF and GLEAM; Table 1) to a 3-hourly resolution using nearest neighbor resampling. To ensure reliable R values, we only calculated R, R_{hi}, or R_{lo} values if > 200 coincident soil moisture estimates from the sensor and the product were ~~present. Since the spatiotemporal coverage differed among the products (Table 1), the~~ available. The final number of R, R_{hi}, and R_{lo} values thus varied depending on the product.

To derive insights into the reasons for the differences in performance, median R values were calculated separately for different Köppen-Geiger climate classes, leaf area index (LAI) values, and topographic slopes. To determine the Köppen-Geiger climate classes, we used the 1-km Köppen-Geiger climate classification map of Beck et al. (2018; Fig. 1), which represents the period 1980–2016. To determine LAI, we used the 1-km Copernicus LAI dataset derived from SPOT-VGT and PROBA-V data (V2; Baret et al., 2016; mean over 1999–2019). To determine the topographic slope, we used the 90-m MERIT DEM (Yamazaki et al., 2017). ~~The~~ To reduce the scale mismatch between point locations and satellite sensor footprints or model grid-cells, we upscaled the Köppen-Geiger, LAI, and topographic slope maps ~~were upscaled~~ to 0.25° using majority, average, and average resampling, respectively~~, to make them more representative of satellite sensor footprints and model grid-cells.~~.

3 Results and discussion

3.1 How do the ascending and descending retrievals perform?

Microwave soil moisture retrievals from ascending and descending overpasses may exhibit performance differences due to diurnal variations in land surface conditions (Lei et al., 2015) and radio-frequency interference (RFI; Aksoy and Johnson, 2013). Table 2 presents R values for the instantaneous ascending and descending retrievals of the four single-sensor products (AMSR2, ASCAT, SMAPL3E, and SMOS; Table 1). Descending (local night) retrievals were more reliable for the passive microwave-based AMSR2, in agreement with several previous studies (Lei et al., 2015; Griesfeller et al., 2016; Bindlish et al., 2018), and consistent with the notion that soil-vegetation temperature differences during day-time interfere with passive microwave soil moisture retrieval (Parinussa et al., 2011). Descending (local morning) retrievals were more reliable for the active microwave-based ASCAT (Table 2), in agreement with Lei et al. (2015). The ascending and descending retrievals performed similarly for the passive microwave-based SMAPL3E and SMOS (Table 2). For the remainder of this analysis, we will use only descending retrievals of AMSR2. We did not discard the ascending retrievals of ASCAT as they helped to improve the performance of $\text{ASCAT}_{\text{SWI}}$.

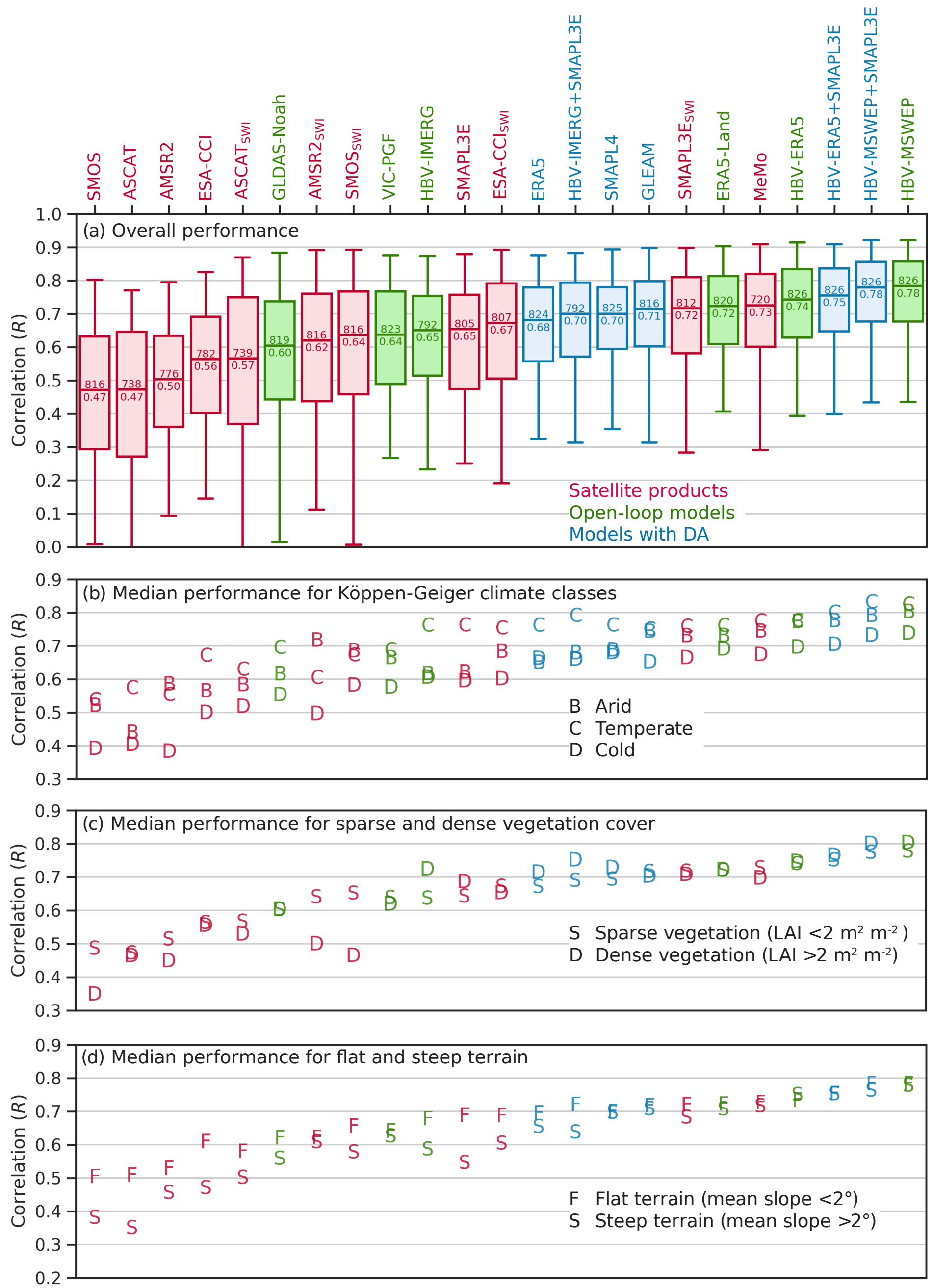

Figure 2. (a) Performance of the soil moisture products in terms of 3-hourly Pearson correlation (R). The products were sorted in ascending order of median R. Outliers are not shown. The number above the median line in each box represents the number of sites with R values and the number below the median line represents the median R value. Also shown are median R values for different (b) major Köppen-Geiger climate classes, (c) mean leaf area index (LAI) values, and (d) mean topographic slopes.

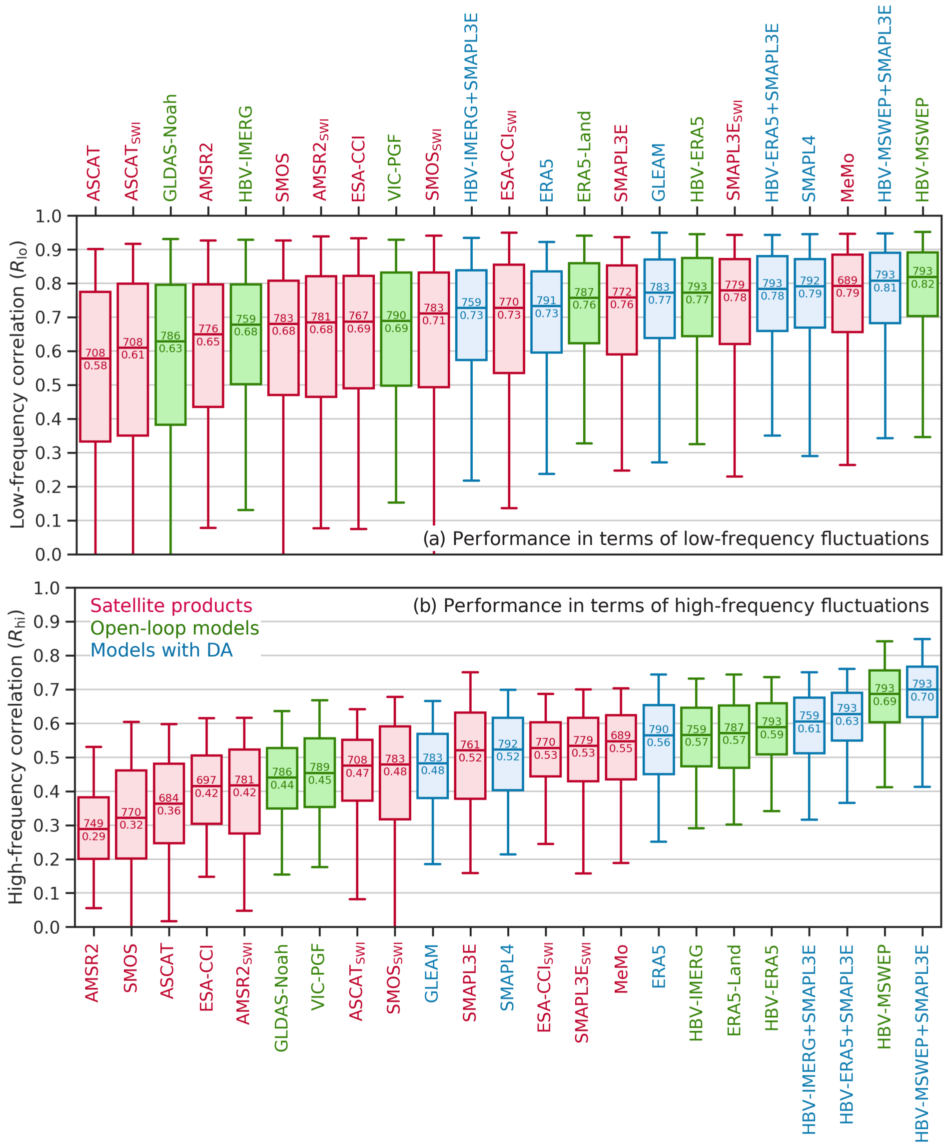

Figure 3. Performance of the soil moisture products in terms of 3-hourly Pearson correlation for (a) low-frequency fluctuations (R_{lo}) and (b) high-frequency fluctuations (R_{hi}). The products were sorted in ascending order of the median. The number above the median line in each box represents the number of sites with R_{lo} or R_{hi} values and the number below the median line represents the median R_{lo} or R_{hi} value. Outliers are not shown.

3.2 What is the impact of the Soil Wetness Index (SWI) smoothing filter?

The application of the SWI filter resulted in higher median R, R_{hi}, and R_{lo} values for all satellite products (Figs. 2a and 3; Table 1). The median R improvement was $+0.12$ for AMSR2, $+0.10$ for ASCAT, $+0.07$ for SMAPL3E, $+0.17$ for SMOS, and $+0.11$ for ESA-CCI (Fig. 2a). The improvements are probably mainly because the SWI filter reduces the impact of random errors and potential differences between ascending and descending overpasses (Su et al., 2015; Bogoslovskiy et al., 2015). Additionally, since the SWI filter simulates the slower variability of soil moisture at deeper layers (Wagner et al., 1999; Albergel et al., 2008; Brocca et al., 2010a), it improves the consistency between the *in situ* measurements at 5-cm depth and the microwave signals, which often have a penetration depth of just 1–2 cm depending on the observation frequency and the land surface conditions (Long and Ulaby, 2015; Shellito et al., 2016a; Rondinelli et al., 2015; Lv et al., 2018). Our results suggests that previous near-surface soil moisture product assessments (e.g., Zhang et al., 2017; Karthikeyan et al., 2017a; Cui et al., 2018; Al-Yaari et al., 2019; Ma et al., 2019), which generally did not use smoothing filters, may have underestimated the true skill of the products.

3.3 What is the relative performance of the single-sensor satellite products?

Among the four single-sensor products with SWI filter ($\mathrm{AMSR2_{SWI}}$, $\mathrm{ASCAT_{SWI}}$, $\mathrm{SMAPL3E_{SWI}}$, and $\mathrm{SMOS_{SWI}}$; Table 1), $\mathrm{SMAPL3E_{SWI}}$ performed best in terms of median R, R_{lo}, and R_{hi} by a wide margin (Figs. 2a and 3), in agreement with previous studies using triple collocation (Chen et al., 2018) and *in situ* measurements from the USA (Karthikeyan et al., 2017a; Zhang et al., 2017; Cui et al., 2018; Al-Yaari et al., 2019), the Tibetan Plateau (Chen et al., 2017), the Iberian Peninsula (Cui et al., 2018), and across the globe (Al-Yaari et al., 2017; Kim et al., 2018; Ma et al., 2019). The good performance of $\mathrm{SMAPL3E_{SWI}}$ is likely attributable to the deeper ground penetration of L-band signals (Lv et al., 2018), the sensor's higher radiometric accuracy (Entekhabi et al., 2010), and the application of an RFI mitigation algorithm (Piepmeier et al., 2014). $\mathrm{SMOS_{SWI}}$ is also an L-band product, while the $\mathrm{AMSR2_{SWI}}$ product used here was derived from X-band observations, which have a shallower penetration depth (Long and Ulaby, 2015). Both $\mathrm{AMSR2_{SWI}}$ and $\mathrm{SMOS_{SWI}}$ are more vulnerable to RFI, which may have reduced their overall performance (Njoku et al., 2005; Oliva et al., 2012). The active microwave-based $\mathrm{ASCAT_{SWI}}$ performed significantly better in terms of high-frequency than low-frequency fluctuations (Fig. 3), likely due to the presence of seasonal vegetation-related biases (Wagner et al., 2013). $\mathrm{ASCAT_{SWI}}$ showed a relatively small spread in R_{hi} values (Fig. 3b), although it showed the largest spread in R and R_{lo} values not just among the single-sensor products but among all products (Figs. 2a and 3a).

All single-sensor satellite products achieved lower R values in cold climates (Figs. 1 and 2b), in agreement with other global evaluations using ISMN data (Kim et al., 2018; Al-Yaari et al., 2019; Zhang et al., 2019b; Ma et al., 2019), and previously attributed to the confounding influence of ~~vegetation dynamics~~ dense vegetation cover (de Rosnay et al., 2006; Gruhier et al., 2008; Dorigo et , highly organic soils (Zhang et al., 2019b), and standing water (Ye et al., 2015; Du et al., 2018) on soil moisture retrievals~~(de Rosnay et al.,~~ . However, since the models also tend to exhibit lower R values in cold regions (Fig. 2b), it could also be that the *in situ* measurements are of lower quality ~~and/or~~ or less representative of satellite footprints or model grid-cells, or that our procedure to screen for frozen or snow-covered soils is imperfect. AMSR2 and particularly $\mathrm{AMSR2_{SWI}}$ performed noticeably better in terms of R in arid climates (Figs. 1 and 2b), as reported in previous studies (Wu et al., 2016; Cho et al., 2017), and likely due to the availability of coincident Ka-band brightness temperature observations which are used as input to the LPRM retrieval algorithm (Parinussa et al., 2011). AMSR2 and SMOS (with and without SWI filter) showed markedly lower R values for sites with mean leaf area index > 2 m^2 m^{-2} (Fig. 2c), confirming that their retrievals are affected by dense vegetation cover (Al-Yaari et al., 2014; Wu et al., 2016; Cui et al., 2018). Most satellite products performed worse in terms of R in areas of steep terrain (Fig. 2d), consistent with previous evaluations (Paulik et al., 2014; Karthikeyan et al., 2017a; Ma et al., 2019), and attributed to the confounding effects of relief on the upwelling microwave brightness temperature observed by the radiometer ~~(Mialon et al., 2008; Pulvirenti et al., 2011)~~(Mialon et al., 2008; Pulvirenti et al., 2011; Guo et al., 2011).

3.4 How do the multi-sensor merged satellite products perform?

The multi-sensor merged product MeMo (based on $\mathrm{AMSR2_{SWI}}$, $\mathrm{SMAPL3E_{SWI}}$, and $\mathrm{SMOS_{SWI}}$) performed better than the four single-sensor products for all three metrics (R, R_{lo}, and R_{hi}; Figs. 2a and 3; Table 1). These results highlight the value of multi-sensor merging techniques, in line with prior studies that merged satellite retrievals (Gruber et al., 2017; Kim et al., 2018), model outputs (Guo et al., 2007; Liu and Xie, 2013; Cammalleri et al., 2015), and satellite retrievals with model outputs (Yilmaz et al., 2012; Anderson et al., 2012; Tobin et al., 2019; Vergopolan et al., 2020). However, MeMo performed only marginally better in terms of median R than the best-performing single-sensor product $\mathrm{SMAPL3E_{SWI}}$ (which was incorporated in MeMo; Fig. 2a). The most likely reason for this is ~~probably that since all products incorporated in MeMo are based on passive-microwave remote sensing, their errors may to a certain degree be cross-correlated and hence may not fully cancel each other out (Yilmaz and Crow, 2014)~~that triple collocation-based merging techniques rely on several assumptions (linearity, stationarity, error orthogonality, and zero cross-correlation) which are generally difficult to fully satisfy in practice, affecting the optimality of the merging procedure (Yilmaz and Crow, 2014; Gruber et al., 2016).

Additionally, MeMo performed better than the multi-sensor merged product $\mathrm{ESA\text{-}CCI_{SWI}}$ (based on AMSR2, ASCAT, and SMOS) for all three metrics (Figs. 2a and 3). MeMo performed better in terms of R at 68 % of the sites, and performed particularly well across the central Rocky Mountains, although $\mathrm{ESA\text{-}CCI_{SWI}}$ performed better in eastern Europe (Fig. 4). The two products performed similarly in terms of high-frequency fluctuations (median R_{hi} of 0.55 for MeMo versus 0.53 for $\mathrm{ESA\text{-}CCI_{SWI}}$; Fig. 3b). The better overall performance of MeMo compared to $\mathrm{ESA\text{-}CCI_{SWI}}$ (Figs. 2a, 3, and 4) is probably due to two factors. First, $\mathrm{ESA\text{-}CCI_{SWI}}$ incorporates ASCAT, which performed less well in the present evaluation, whereas MeMo incorporates $\mathrm{SMAPL3E_{SWI}}$, which performed best among the single-sensor products (Figs. 2a and 3). The median R of MeMo dropped by 0.04 after ~~we excluded~~ excluding $\mathrm{SMAPL3E_{SWI}}$ (data not shown), which supports this explanation. The next version of ESA-CCI (V5) is anticipated to incorporate SMAP soil moisture estimates, and is therefore expected to perform better (Gruber et al., 2019). Secondly, MeMo merges soil moisture estimates from multiple sensors each day, whereas $\mathrm{ESA\text{-}CCI_{SWI}}$ uses only the soil moisture estimate from the 'best' sensor each day, resulting in a loss of information.

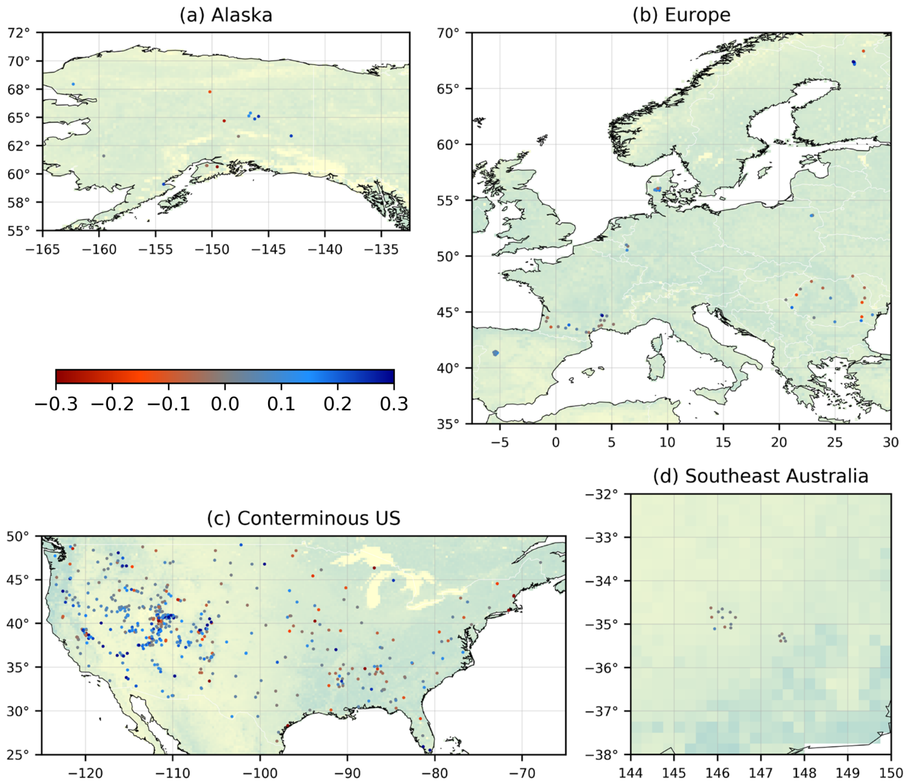

Figure 4. Three-hourly Pearson correlations (R) obtained by MeMo minus those obtained by ESA-CCI. Blue indicates that MeMo performs better, whereas red indicates that ESA-CCI performs better. A map of long-term mean LAI (Baret et al., 2016) is plotted in the background.

3.5 What is the relative performance of the open-loop models?

The ranking of the six open-loop models in terms of median R (from best to worst) was (i) HBV-MSWEP, (ii) HBV-ERA5, (iii) ERA5-Land, (iv) HBV-IMERG, (v) VIC-PGF, and (vi) GLDAS-Noah (Fig. 2a; Table 1). The models were forced with precipitation from, respectively: (i) the gauge-, satellite-, and reanalysis-based MSWEP V2.4 ~~(??)~~(Beck et al., 2017b, 2019b) , (ii) and (iii) the ERA5 reanalysis (Hersbach et al., 2020), (iv) the satellite-based IMERGHHE V06 (Huffman et al., 2014, 2018), (v) the gauge- and reanalysis-based PGF (Sheffield et al., 2006), and (vi) the gauge- and satellite-based GPCP V1.3 Daily Analysis (Huffman et al., 2001). This order matches the overall performance ranking of precipitation datasets in a comprehensive evaluation over the conterminous USA carried out by Beck et al. (2019a). Furthermore, the performance of HBV-ERA5 did not depend on the terrain slope, while HBV-IMERG performed worse in steep terrain (Fig. 2d), which is also consistent with the evaluation of Beck et al. (2019a). HBV-IMERG performed worse for low-frequency than for high-frequency fluctuations (Fig. 3), which likely reflects the presence of seasonal biases in IMERG (Beck et al., 2017c; Wang and Yong, 2020). Overall, these results confirm that precipitation is by far the most important determinant of soil moisture simulation performance ~~(Gottschalck et al., 2005; Liu et al., 2011; Beck et al., 2017c; Dong et al., 2019)~~

(Gottschalck et al., 2005; Liu et al., 2011; Beck et al., 2017c; Dong et al., 2019; Chen and Yuan, 2020). The superior performance of MSWEP is primarily attributable to the ~~daily gauge corrections (?)~~inclusion of daily gauge observations (Beck et al., 2019b) .

Among the three soil moisture products derived from ERA5 precipitation (ERA5, ERA5-Land, and HBV-ERA5), and among the three products forced with daily gauge-corrected precipitation (GLEAM, HBV-MSWEP+SMAPL3E, and SMAPL4; Table 1), the ones based on HBV performed better overall in terms of all three metrics (R, R_{lo}, and R_{hi}; Figs. 2a and 3). This demonstrates that soil moisture estimates from complex, data-intensive models (H-TESSEL underlying ERA5 and ERA5-Land, GLEAM, and the Catchment model underlying SMAPL4) are not necessarily more accurate than those from relatively simple, calibrated models (HBV). This is in line with several previous multi-model evaluations focusing on soil moisture (e.g., Guswa et al., 2002; Cammalleri et al., 2015; Orth et al., 2015), the surface energy balance (e.g., Best et al., 2015), evaporation (e.g., McCabe et al., 2016), runoff (e.g., Beck et al., 2017a), and river discharge (e.g., Gharari et al., 2020).

3.6 How do the models with satellite data assimilation perform?

The performance ranking of the models with satellite data assimilation in terms of median R (from best to worst) was HBV-MSWEP+SMAPL3E, HBV-ERA5+SMAPL3E, GLEAM, SMAPL4, HBV-IMERG+SMAPL3E, and ERA5 (Fig. 2a; Table 1). The assimilation of SMAPL3E retrievals resulted in a substantial improvement in median R of $+0.06$ for HBV-IMERG, a minor improvement of $+0.01$ for HBV-ERA5, and no change for HBV-MSWEP (Fig. 2a). Improvements in R were obtained for 90 %, 65 %, and 56 % of the sites for HBV-IMERG, HBV-ERA5, and HBV-MSWEP, respectively. For HBV-IMERG, the greatest improvements were found over the central Rocky Mountains (Fig. 5)~~. These~~ , where IMERG performs relatively poorly (Beck et al., 2019a). Overall, these results suggest that data assimilation provides greater benefits when the precipitation forcing is more uncertain~~(Beck et al., 2019a)~~. Since rain gauge observations are not available over the large majority of the globe (Kidd et al., 2017), we expect data assimilation to provide significant added value at the global scale, as also concluded by Bolten et al. (2010), Dong et al. (2019), and Tian et al. (2019). The lack of improvement for HBV-ERA5+SMAPL3E and HBV-MSWEP+SMAPL3E suggests that the gain parameter G (Eq. 3), which quantifies the relative quality of the satellite and model soil moisture estimates, can be refined further.

The ERA5 reanalysis, which assimilates ASCAT soil moisture (Hersbach et al., 2020), obtained a lower overall performance (median $R = 0.68$) than the open-loop models ERA5-Land (median $R = 0.72$) and HBV-ERA5 (median $R = 0.74$), which were both forced with ERA5 precipitation (Fig. 2a). This suggests that assimilating satellite soil moisture estimates (ERA5) was less beneficial than either increasing the model resolution (ERA5-Land) or improving the ~~model~~ soil moisture simulation efficiency (HBV). In line with these results, Muñoz Sabater et al. (2019) found that the joint assimilation of ASCAT soil moisture retrievals and SMOS brightness temperatures into an experimental version of the Integrated Forecast System (IFS) model underlying ERA5 did not improve the soil moisture simulations. They attributed this to the adverse impact of simultaneously assimilated screen-level temperature and relative humidity observations on the soil moisture estimates.

In line with our results for HBV-MSWEP+SMAPL3E, Kumar et al. (2014) did not obtain improved soil moisture estimates after the assimilation of ESA-CCI and AMSR-E retrievals into Noah forced with highly accurate NLDAS2 meteorological

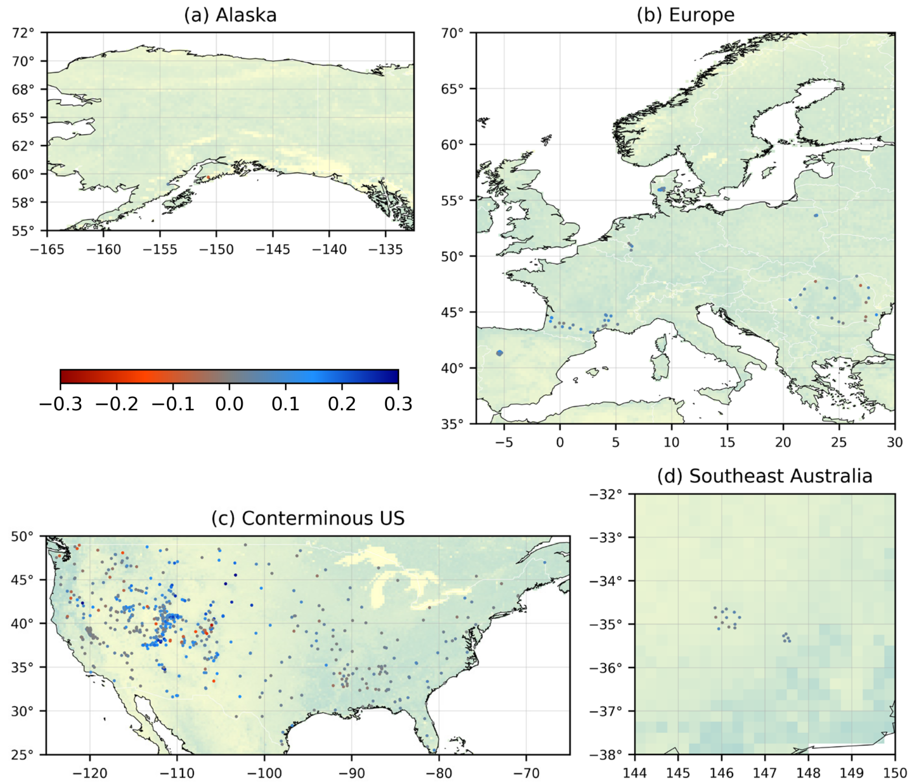

Figure 5. Three-hourly Pearson correlations (R) obtained by HBV-IMERG+SMAPL3E minus those obtained by HBV-IMERG. Blue indicates improved performance after data assimilation, whereas red indicates degraded performance after data assimilation. The sites in Alaska and Finland are not shown ~~because~~ as IMERG does not cover high latitudes. A map of long-term mean LAI (Baret et al., 2016) is plotted in the background.

data for the conterminous USA. Conversely, several other studies obtained substantial performance improvements after data assimilation despite the use of high-quality precipitation forcings (Liu et al., 2011; Koster et al., 2018; Tian et al., 2019). We suspect that this discrepancy might reflect the lower performance of their open-loop models compared to ours. Using different (but overlapping) *in situ* ~~measurements~~datasets, Koster et al. (2018) and Tian et al. (2019) obtained mean daily open-loop R values of 0.64 and 0.59, respectively, while we obtained a mean daily open-loop R of 0.75 (slightly lower than the 3-hourly median value shown in Fig. 2a). Overall, it appears that the benefits of data assimilation are greater for models that exhibit structural or parameterization deficiencies.

3.7 What is the impact of model calibration?

Among the models evaluated in this study, only HBV and the Catchment model underlying SMAPL4 have been calibrated, although only a single parameter out of more than 100 was calibrated for the Catchment model (Reichle et al., 2019b).

HBV-ERA5, HBV-IMERG, and HBV-MSWEP with calibrated parameters obtained ~~a~~ median R ~~of~~ values of 0.74, 0.65, and 0.78, respectively (Fig. 2a), whereas ~~HBV-MSWEP~~ the same three models with randomly generated ~~parameters obtained a~~ (uncalibrated) parameters obtained mean median R ~~of 0.66 (standard deviation 0.07~~values of 0.59, 0.53, and 0.62, respectively (standard deviations 0.17, 0.16, and 0.16, respectively; data not shown). The calibration thus resulted in ~~a mean increase~~ mean increases in median R of ~~-~~$+0.15$, $+0.12$, ~~which represents a substantial improvement~~ and $+0.16$, respectively, for the three models, which represent substantial improvements in performance. These results are in line with previous studies calibrating different models using soil moisture from *in situ* sensors (e.g., Koren et al., 2008; Shellito et al., 2016b; Thorstensen et al., 2016; Reichle et al., 2019b) or remote sensing (e.g., Zhang et al., 2011; Wanders et al., 2014; López López et al., 2016; Koster et al., 2018).

~~The mean improvement~~

The mean improvements in median R obtained for HBV-ERA5, HBV-IMERG, and HBV-MSWEP after calibration ($+0.15$, $+0.12$~~) was double the improvement obtained for HBV-IMERG~~, and $+0.16$, respectively) were significantly greater than the improvements obtained for the same three models after satellite data assimilation ($+0.01$, $+0.06$, and -0.00, respectively; Fig. 2a; Section 3.6), which suggests that model calibration ~~is more beneficial~~ results in more benefit overall than data assimilation. Additionally, model calibration ~~is likely to benefit~~ benefits regions with both sparse and dense rain gauge networks, whereas data assimilation mainly benefits regions with sparse rain gauge networks (Section 3.6). Conversely, only data assimilation is capable of ameliorating potential deficiencies in the meteorological forcing data (e.g., undetected precipitation).

Our calibration approach was relatively simple and yielded only a single spatially uniform parameter set (Section 2.3). Previous studies focusing on runoff have demonstrated the value of more sophisticated calibration approaches yielding ensembles of parameters that vary according to climate and landscape characteristics ~~(Beck et al., 2016, accepted)~~(Samaniego et al., 2010; Beck et al., 2016, acc . Whether these approaches have value for soil moisture estimation as well warrants further investigation. It should be noted, however, that many current models have rigid structures, insufficient free parameters, and/or a high computational cost ~~and are therefore~~ which makes them less amenable to calibration (Mendoza et al., 2015). Moreover, the validity of calibrated parameters may ~~become less valid~~ be compromised when the model is subjected to climate conditions it has never experienced before (Knutti, 2008). Care should also be taken that calibration of one aspect of the model does not degrade another aspect and that we get “the right answers for the right reasons” (Kirchner, 2006).

3.8 How do the major product categories compare?

The median $R \pm$ interquartile range across all sites and products in each category was 0.53 ± 0.32 for the satellite soil moisture products without SWI filter, 0.66 ± 0.30 for the satellite soil moisture products with SWI filter including MeMo, 0.69 ± 0.25 for the open-loop models, and 0.72 ± 0.22 for the models with satellite data assimilation (Fig. 2a; Table 1). The satellite products thus provided the least reliable soil moisture estimates and exhibited the largest regional performance differences on average, whereas the models with satellite data assimilation provided the most reliable soil moisture estimates and exhibited the smallest regional performance differences on average. Our performance ranking of the major product categories is consistent with previous studies for the conterminous USA (Liu et al., 2011; Kumar et al., 2014; Fang et al., 2016; Dong et al., 2020), Europe

(Naz et al., 2019), and the globe (Albergel et al., 2012; Tian et al., 2019; Dong et al., 2019). It should be kept in mind, however, that these studies, including the present one, used *in situ* soil moisture measurements from regions with dense rain gauge networks, and hence likely overestimate model performance (Dong et al., 2019).

The large spread in performance across the satellite products reflects the large number of factors that affect soil moisture retrieval, including, among others, vegetation cover, surface roughness, soil ~~texture~~composition, diurnal variations in land surface conditions, and RFI ~~, among others~~ (Zhang and Zhou, 2016; Karthikeyan et al., 2017b). The spread in performance across the open-loop models is lower as it depends primarily on the precipitation data quality, which, in turn, depends mostly on a combination of gauge network density and prevailing precipitation type (convective versus stratiform; Gottschalck et al., 2005; Liu et al., 2011; Beck et al., 2017c; Dong et al., 2019). The smaller spread in performance across the models with satellite data assimilation is due to the fact that individual errors in satellite retrievals and model estimates are cancelled out, to a certain degree, when they are combined, confirming the effectiveness of the data assimilation procedures (Moradkhani, 2008; Liu et al., 2012; Reichle et al., 2017).

3.9 To what extent are our results generalizable to other regions?

The large majority (98 %) of the *in situ* soil moisture measurements used as reference in the current study were from dense monitoring networks in the USA and Europe (Fig. 1) and therefore our results will be most applicable to these regions. We speculate that our results for the models (with and without data assimilation; Figs. 2, 3, and 5) apply to other regions with dense rain gauge networks and broadly similar climates (e.g., parts of China and Australia, and other parts of Europe; Kidd et al., 2017). The calibrated models (HBV and the Catchment model underlying SMAPL4) may, however, perform slightly worse in regions with climatic and physiographic conditions dissimilar to the *in situ* sensors used for calibration (but likely still better than the uncalibrated models). In sparsely gauged areas the four model products based on precipitation forcings that incorporate daily gauge observations (GLEAM, HBV-MSWEP, HBV-MSWEP+SMAPL3E, and SMAPL4; Table 1) will inevitably exhibit ~~reduced performance~~ lower performance (but not necessarily lower than the other model products). In convection-dominated regions models driven by precipitation from satellite datasets such as IMERG may well outperform those driven by precipitation from reanalyses such as ERA5 ~~(Massari et al., 2017; Beck et al., 2017c; ?)~~(Massari et al., 2017; Beck et al., 2017c, 2019b). Conversely, in mountainous and snow-dominated regions models driven by precipitation from reanalyses are likely to outperform those driven by precipitation from satellites ~~(Ebert et al., 2007; ?; Beck et al., 2019a)~~(Ebert et al., 2007; Beck et al., 2019b, a).

Our results for the satellite soil moisture products may be less generalizable, given the large spread in performance across different regions and products revealed in the current study (Figs. 2 and 3) and in previous quasi-global studies using triple collocation (Al-Yaari et al., 2014; Chen et al., 2018; Miyaoka et al., 2017). ~~Some predictors of retrieval performance were identified, with the most accurate estimates found for low-relief terrain with sparse vegetation (cf. Tian et al., 2019). Furthermore, outside densely-gauged~~ Outside developed regions we expect the lower prevalence of RFI to lead to more reliable retrievals for those satellite products susceptible to it (Njoku et al., 2005; Oliva et al., 2012; Aksoy and Johnson, 2013; Ticconi et al., 2017). At low latitudes the lower satellite revisit frequency will inevitably increase the sampling uncertainty and reduce the overall value of satellite products relative to models. In tropical forest regions passive products often do not provide soil moisture retrievals, and when they do, the retrievals are typically less reliable than those from active products due to the dense vegetation cover (Al-Yaari et al., 2014; Chen et al., 2018; Miyaoka et al., 2017; Kim et al., 2018). Shedding more light on the strengths and weaknesses of soil moisture products in regions without dense measurement networks — for example using independent soil moisture products (Chen et al., 2018; Dong et al., 2019) or by expanding measurement networks (Kang et al., 2016; Singh et al., 2019) — should be a key priority for future research (Ochsner et al., 2013; Myeni et al., 2019).

4 Conclusions

To ~~elucidate the strengths and weaknesses~~ shed light on the advantages and disadvantages of different soil moisture products and on the merit of ~~different~~ various technological and methodological innovations, we evaluated 18 state-of-the-art (sub-)daily (quasi-)global near-surface soil moisture products using *in situ* measurements from 826 sensors located primarily in the USA and Europe. Our main findings related to the nine questions posed in the introduction can be summarized as follows:

1. Local night retrievals from descending overpasses were more reliable overall for AMSR2, whereas local morning retrievals from descending overpasses were more reliable overall for ASCAT. The ascending and descending retrievals of SMAPL3E and SMOS performed similarly.

2. Application of the SWI smoothing filter resulted in improved performance for all satellite products. Previous near-surface soil moisture product assessments generally did not apply smoothing filters and therefore may have underestimated the true skill of the products.

3. $\mathrm{SMAPL3E_{SWI}}$ performed best overall among the four single-sensor satellite products with SWI filter. $\mathrm{ASCAT_{SWI}}$ performed markedly better in terms of high-frequency than low-frequency fluctuations. All satellite products tended to perform worse in cold climates.

4. The multi-sensor merged satellite product MeMo performed best among the satellite products, highlighting the value of multi-sensor merging techniques. MeMo also outperformed the multi-sensor merged satellite product $\mathrm{ESA\text{-}CCI_{SWI}}$, likely due to the inclusion of $\mathrm{SMAPL3E_{SWI}}$.

5. The performance of the open-loop models depended primarily on the precipitation data quality. The superior performance of HBV-MSWEP is due to the calibration of HBV and the daily gauge corrections of MSWEP. Soil moisture simulation performance did not improve with model complexity.

6. In the absence of model structural or parameterization deficiencies, satellite data assimilation yields substantial performance improvements mainly when the precipitation forcing is of relatively low quality. This suggests that data assimilation provides significant benefits at the global scale.

7. The calibration of HBV against *in situ* soil moisture measurements resulted in substantial performance improvements. The improvement due to model calibration tends to exceed the improvement due to satellite data assimilation and is not limited to regions of low quality precipitation.

8. The satellite products provided the least reliable soil moisture estimates and exhibited the largest regional performance differences on average, whereas the models with satellite data assimilation provided the most reliable soil moisture estimates and exhibited the smallest regional performance differences on average.

9. We speculate that our results for the models (with and without data assimilation) apply to other regions with dense rain gauge networks and broadly similar climates. Our results for the satellite products may be less generalizable due to the large number of factors that affect retrievals.

Appendix: *In situ* soil moisture measurement networks

Table A1 lists the measurement networks part of the ISMN archive from which we have used *in situ* soil moisture data.

Author contributions. H.E.B. conceived, designed, and performed the analysis and took the lead in writing the paper. E.F.W. was responsible for funding acquisition. All co-authors provided critical feedback and contributed to the writing.

Competing interests. The authors declare no competing interests.

Acknowledgements. We are grateful to the numerous contributors to the ISMN archive. The soil moisture product developers are thanked for producing and making available their products. We thank the editor, Christian Massari, Korbin Nelson, and two anonymous reviewers for their constructive comments which helped us to improve the manuscript. Hylke E. Beck and Ming Pan were supported through IPA support from the U.S. Army Corps of Engineers' International Center for Integrated Water Resources Management (ICIWaRM), under the auspices of UNESCO. Robert M. Parinussa was funded by National Natural Science Foundation of China grant 41850410492. Noemi Vergopolan was in part supported by NASA Soil Moisture Cal/Val Activities (NNX14AH92G). Noemi Vergopolan and Ming Pan were in part supported by NOAA's "Modernizing Observation Operator and Error Assessment for Assimilating In-situ and Remotely Sensed Snow/Soil Moisture Measurements into NWM" project (NA19OAR4590199). Rolf H. Reichle ~~was~~ and John S. Kimball were supported by the NASA SMAP Science Team. Diego G. Miralles acknowledges support from the European Research Council (ERC) DRY-2-DRY project (715254) and the Belgian Science Policy Office (BELSPO) STEREO III ALBERI project (SR/00/373). Wouter A. Dorigo acknowledges support from the ESA CCI Programme and ESA's IDEAS+ and QA4EO projects.

References

Aksoy, M. and Johnson, J. T.: A study of SMOS RFI over North America, IEEE Geoscience and Remote Sensing Letters, 10, 515–519, 2013.

Al-Yaari, A., Wigneron, J.-P., Ducharne, A., Kerr, Y., de Rosnay, P., de Jeu, R., Govind, A., Al Bitar, A., Albergel, C., noz Sabater, J. M., Richaume, P., and Mialon, A.: Global-scale evaluation of two satellite-based passive microwave soil moisture datasets (SMOS and AMSR-E) with respect to Land Data Assimilation System estimates, Remote Sensing of Environment, 149, 181–195, https://doi.org/10.1016/j.rse.2014.04.006, 2014.

Al-Yaari, A., Wigneron, J.-P., Kerr, Y., Rodriguez-Fernandez, N., O'Neill, P. E., Jackson, T. J., De Lannoy, G. J. M., Al Bitar, A., Mialon, A., Richaume, P., Walker, J. P., Mahmoodi, A., and Yueh, S.: Evaluating soil moisture retrievals from ESA's SMOS and NASA's SMAP brightness temperature datasets, Remote Sensing of Environment, 193, 257–273, https://doi.org/10.1016/j.rse.2017.03.010, 2017.

Al-Yaari, A., Wigneron, J.-P., Dorigo, W., Colliander, A., Pellarin, T., Hahn, S., Mialon, A., Richaume, P., Fernandez-Moran, R., Fan, L., Kerr, Y., and De Lannoy, G.: Assessment and inter-comparison of recently developed/reprocessed microwave satellite soil moisture products using ISMN ground-based measurements, Remote Sensing of Environment, 224, 289–303, https://doi.org/10.1016/j.rse.2019.02.008, 2019.

Albergel, C., Rüdiger, C., Pellarin, T., Calvet, J.-C., Fritz, N., Froissard, F., Suquia, D., Petitpa, A., Piguet, B., and Martin, E.: From near-surface to root-zone soil moisture using an exponential filter: an assessment of the method based on in-situ observations and model simulations, Hydrology and Earth System Sciences, 12, 1323–1337, 2008.

Albergel, C., Rüdiger, C., Carrer, D., Calvet, J.-C., Fritz, N., Naeimi, V., Bartalis, Z., and Hasenauer, S.: An evaluation of ASCAT surface soil moisture products with in-situ observations in Southwestern France, Hydrology and Earth System Sciences, 13, 115–124, 2009.

Albergel, C., de Rosnay, P., Gruhier, C., noz Sabater, J. M., Hasenauer, S., Isaksen, L., Kerr, Y., and Wagner, W.: Evaluation of remotely sensed and modelled soil moisture products using global ground-based in situ observations, Remote Sensing of Environment, 118, 215–226, https://doi.org/10.1016/j.rse.2011.11.017, 2012.

Alijanian, M., Rakhshandehroo, G. R., Mishra, A. K., and Dehghani, M.: Evaluation of satellite rainfall climatology using CMORPH, PERSIANN-CDR, PERSIANN, TRMM, MSWEP over Iran, International Journal of Climatology, 37, 4896–4914, 2017.

Anderson, W. B., Zaitchik, B. F., Hain, C. R., Anderson, M. C., Yilmaz, M. T., Mecikalski, J., and Schultz, L.: Towards an integrated soil moisture drought monitor for East Africa, Hydrology and Earth System Sciences, 16, 2893–2913, 2012.

Bai, P. and Liu, X.: Evaluation of five satellite-based precipitation products in two gauge-scarce basins on the Tibetan Plateau, Remote Sensing, 10, 2018.

Baret, F., Weiss, M., Verger, A., and Smets, B.: ATBD for LAI, FAPAR and FCOVER from PROBA-V products at 300 m resolution (GEOV3), INRA — Institut National de la Recherche Agronomique, Paris, France, 2016.

Beck, H. E., de Jeu, R. A. M., Schellekens, J., Van Dijk, A. I. J. M., and Bruijnzeel, L. A.: Improving curve number based storm runoff estimates using soil moisture proxies, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2, 250–259, 2009.

Beck, H. E., Bruijnzeel, L. A., van Dijk, A. I. J. M., McVicar, T. R., Scatena, F. N., and Schellekens, J.: The impact of forest regeneration on streamflow in 12 meso-scale humid tropical catchments, Hydrology and Earth System Sciences, 17, 2613–2635, 2013.

Beck, H. E., van Dijk, A. I. J. M., de Roo, A., Miralles, D. G., McVicar, T. R., Schellekens, J., and Bruijnzeel, L. A.: Global-scale regionalization of hydrologic model parameters, Water Resources Research, 52, 3599–3622, 2016.

Beck, H. E., van Dijk, A. I. J. M., de Roo, A., Dutra, E., Fink, G., Orth, R., and Schellekens, J.: Global evaluation of runoff from 10 state-of-the-art hydrological models, Hydrology and Earth System Sciences, 21, 2881–2903, 2017a.

Beck, H. E., van Dijk, A. I. J. M., Levizzani, V., Schellekens, J., Miralles, D. G., Martens, B., and de Roo, A.: MSWEP: 3-hourly 0.25° global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data, Hydrology and Earth System Sciences, 21, 589–615, 2017b.

Beck, H. E., Vergopolan, N., Pan, M., Levizzani, V., van Dijk, A. I. J. M., Weedon, G. P., Brocca, L., Pappenberger, F., Huffman, G. J., and Wood, E. F.: Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling, Hydrology and Earth System Sciences, 21, 6201–6217, 2017c.

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., and Wood, E. F.: Present and future Köppen-Geiger climate classification maps at 1-km resolution, Scientific Data, 5, https://doi.org/10.1038/sdata.2018.214, 2018.

Beck, H. E., Pan, M., Roy, T., Weedon, G. P., Pappenberger, F., van Dijk, A. I. J. M., Huffman, G. J., Adler, R. F., and Wood, E. F.: Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data for the CONUS, Hydrology and Earth System Sciences, 23, 207–224, 2019a.

Beck, H. E., Wood, E. F., Pan, M., Fisher, C. K., Miralles, D. M., van Dijk, A. I. J. M., McVicar, T. R., and Adler, R. F.: MSWEP V2 global 3-hourly 0.1° precipitation: methodology and quantitative assessment, Bulletin of the American Meteorological Society, 100, 473–500, 2019b.

Beck, H. E., Pan, M., Lin, P., Seibert, J., van Dijk, A. I. J. M., and Wood, E. F.: Global fully-distributed parameter regionalization based on observed streamflow from 4229 headwater catchments, Journal of Geophysical Research: Atmospheres, accepted.

Bell, J. E., Palecki, M. A., Baker, C. B., Collins, W. G., Lawrimore, J. H., Leeper, R. D., Hall, M. E., Kochendorfer, J., Meyers, T. P., Wilson, T., and Diamond, H. J.: U.S. climate reference network soil moisture and temperature observations, Journal of Hydrometeorology, 14, 977–988, 2013.

Bergström, S.: Development and application of a conceptual runoff model for Scandinavian catchments, PhD thesis, SMHI Reports RHO 7, Swedish Meteorological and Hydrological Institute (SMHI), Norköping, Sweden, 1976.

Bergström, S.: The HBV model—its structure and applications, SMHI Reports RH 4, Swedish Meteorological and Hydrological Institute (SMHI), Norrköping, Sweden, 1992.

Best, M. J., Abramowitz, G., Johnson, H. R., Pitman, A. J., Balsamo, G., Boone, A., Cuntz, M., Decharme, B., Dirmeyer, P. A., Dong, J., Ek, M., Guo, Z., Haverd, V., van den Hurk, B. J. J., Nearing, G. S., Pak, B., Peters-Lidard, C., Santanello, J. A., Stevens, L., and Vuichard, N.: The plumbing of land surface models: benchmarking model performance, Journal of Hydrometeorology, 16, 1425–1442, 2015.

Bierkens, M. F. P.: Global hydrology 2015: state, trends, and directions, Water Resources Research, 51, 4923–4947, https://doi.org/10.1002/2015WR017173, 2015.

Bindlish, R., Cosh, M. H., Jackson, T. J., Koike, T., Fujii, H., Chan, S. K., Asanuma, J., Berg, A., Bosch, D. D., Caldwell, T., Collins, C. H., McNairn, H., Martinez-Fernandez, J., Prueger, J., Rowlandson, T., Seyfried, M., Starks, P., Thibeault, M., Van Der Velde, R., Walker, J. P., and Coopersmith, E. J.: GCOM-W AMSR2 soil moisture product validation using core validation sites, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 11, 209–219, 2018.

Bogoslovskiy, N. N., Erin, S. I., Borodina, I. A., and Kizhner, L. I.: Filtration and assimilation of soil moisture satellite data, in: 21st International Symposium Atmospheric and Ocean Optics: Atmospheric Physics, edited by Romanovskii, O. A., vol. 9680, pp. 1411–1415, International Society for Optics and Photonics, SPIE, https://doi.org/10.1117/12.2205957, 2015.

Bolten, J. D., Crow, W. T., Zhan, X., Jackson, T. J., and Reynolds, C. A.: Evaluating the utility of remotely sensed soil moisture retrievals for operational agricultural drought monitoring, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 3, 57–66, 2010.

Brocca, L., Melone, F., Moramarco, T., Wagner, W., and Hasenauer, S.: ASCAT soil wetness index validation through in situ and modeled soil moisture data in central Italy, Remote Sensing of Environment, 114, 2745–2755, 2010a.

Brocca, L., Melone, F., Moramarco, T., Wagner, W., Naeimi, V., Bartalis, Z., and Hasenauer, S.: Improving runoff prediction through the assimilation of the ASCAT soil moisture product, Hydrology and Earth System Sciences, 14, 1881–1893, https://doi.org/10.5194/hess-14-1881-2010, 2010b.

Brocca, L., Crow, W. T., Ciabatta, L., Massari, C., de Rosnay, P., Enenkel, M., Hahn, S., Amarnath, G., Camici, S., Tarpanelli, A., and Wagner, W.: A review of the applications of ASCAT soil moisture products, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 10, 2285–2306, 2017.

C3S: ERA5-Land reanalysis, https://cds.climate.copernicus.eu, 2019.

Calvet, J., Fritz, N., Froissard, F., Suquia, D., Petitpa, A., and Piguet, B.: In situ soil moisture observations for the CAL/VAL of SMOS: the SMOSMANIA network, in: 2007 IEEE International Geoscience and Remote Sensing Symposium, pp. 1196–1199, 2007.

Cammalleri, C., Micale, F., and Vogt, J.: On the value of combining different modelled soil moisture products for European drought monitoring, Journal of Hydrology, 525, 547–558, https://doi.org/10.1016/j.jhydrol.2015.04.021, 2015.

Capecchi, V. and Brocca, L.: A simple assimilation method to ingest satellite soil moisture into a limited-area NWP model, Meteorologische Zeitschrift, 23, 105–121, 2014.

Casson, D. R., Werner, M., Weerts, A., and Solomatine, D.: Global re-analysis datasets to improve hydrological assessment and snow water equivalent estimation in a sub-Arctic watershed, Hydrology and Earth System Sciences, 22, 4685–4697, 2018.

Cenci, L., Laiolo, P., Gabellani, S., Campo, L., Silvestro, F., Delogu, F., Boni, G., and Rudari, R.: Assimilation of H-SAF soil moisture products for flash flood early warning systems. case study: Mediterranean catchments, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 9, 5634–5646, 2016.

Champagne, C., Berg, A., Belanger, J., McNairn, H., and De Jeu, R.: Evaluation of soil moisture derived from passive microwave remote sensing over agricultural sites in Canada using ground-based soil moisture monitoring networks, International Journal of Remote Sensing, 31, 3669–3690, 2010.

Chan, S. K., Bindlish, R., O'Neill, P., Jackson, T., Njoku, E., Dunbar, S., Chaubell, J., Piepmeier, J., Yueh, S., Entekhabi, D., Colliander, A., Chen, F., Cosh, M. H., Caldwell, T., Walker, J., Berg, A., McNairn, H., Thibeault, M., Martinez-Fernández, J., Uldall, F., Seyfried, M., Bosch, D., Starks, P., Holifield Collins, C., Prueger, J., van der Velde, R., Asanuma, J., Palecki, M., Small, E. E., Zreda, M., Calvet, J., Crow, W. T., and Kerr, Y.: Development and assessment of the SMAP enhanced passive soil moisture product, Remote Sensing of Environment, 204, 931–941, https://doi.org/10.1016/j.rse.2017.08.025, 2018.

Chawla, I., Karthikeyan, L., and Mishra, A. K.: A review of remote sensing applications for water security: quantity, quality, and extremes, Journal of Hydrology, p. 124826, https://doi.org/10.1016/j.jhydrol.2020.124826, 2020.

Chen, F., Crow, W. T., Bindlish, R., Colliander, A., Burgin, M. S., Asanuma, J., and Aida, K.: Global-scale evaluation of SMAP, SMOS and ASCAT soil moisture products using triple collocation, Remote Sensing of Environment, 214, 1–13, https://doi.org/10.1016/j.rse.2018.05.008, 2018.

Chen, M., Shi, W., Xie, P., Silva, V. B. S., Kousky, V. E., Higgins, R. W., and Janowiak, J. E.: Assessing objective techniques for gauge-based analyses of global daily precipitation, Journal of Geophysical Research, 113, D04 110, https://doi.org/10.1029/2007JD009132, 2008.

Chen, Y. and Yuan, H.: Evaluation of nine sub-daily soil moisture model products over China using high-resolution in situ observations, Journal of Hydrology, 588, 125 054, https://doi.org/10.1016/j.jhydrol.2020.125054, 2020.

Chen, Y., Yang, K., Qin, J., Cui, Q., Lu, H., La, Z., Han, M., and Tang, W.: Evaluation of SMAP, SMOS, and AMSR2 soil moisture retrievals against observations from two networks on the Tibetan Plateau, Journal of Geophysical Research: Atmospheres, 122, 5780–5792, 2017.

Cho, E., Su, C.-H., Ryu, D., Kim, H., and Choi, M.: Does AMSR2 produce better soil moisture retrievals than AMSR-E over Australia?, Remote Sensing of Environment, 188, 95–105, https://doi.org/10.1016/j.rse.2016.10.050, 2017.

Crow, W. T., Miralles, D. G., and Cosh, M. H.: A quasi-global evaluation system for satellite-based surface soil moisture retrievals, IEEE Transactions on Geoscience and Remote Sensing, 48, 2516–2527, 2010.

Crow, W. T., Berg, A. A., Cosh, M. H., Loew, A., Mohanty, B. P., Panciera, R., de Rosnay, P., Ryu, D., and Walker, J. P.: Upscaling sparse ground-based soil moisture observations for the validation of coarse-resolution satellite soil moisture products, Reviews of Geophysics, 50, https://doi.org/10.1029/2011RG000372, 2012.

Cui, C., Xu, J., Zeng, J., Chen, K.-S., Bai, X., Lu, H., Chen, Q., and Zhao, T.: Soil moisture mapping from satellites: an intercomparison of SMAP, SMOS, FY3B, AMSR2, and ESA CCI over two dense network regions at different spatial scales, Remote Sensing, 10, 2018.

de Rosnay, P., Calvet, J.-C., Kerr, Y., Wigneron, J.-P., Lemaitre, F., Escorihuela, M. J., noz Sabater, J. M., Saleh, K., Barrié, J., Bouhours, G., Coret, L., Cherel, G., Dedieu, G., Durbe, R., Fritz, N., Froissard, F., Hoedjes, J., Kruszewski, A., Lavenu, F., Suquia, D., and Waldteufel, P.: SMOSREX: a long term field campaign experiment for soil moisture and land surface processes remote sensing, Remote Sensing of Environment, 102, 377–389, 2006.

Dharssi, I., Bovis, K. J., Macpherson, B., and Jones, C. P.: Operational assimilation of ASCAT surface soil wetness at the Met Office, Hydrology and Earth System Sciences, 15, 2729–2746, 2011.

Dong, J., Crow, W., Reichle, R., Liu, Q., Lei, F., and Cosh, M. H.: A global assessment of added value in the SMAP Level 4 soil moisture product relative to its baseline land surface model, Geophysical Research Letters, 46, 6604–6613, 2019.

Dong, J., Crow, W. T., Tobin, K. J., Cosh, M. H., Bosch, D. D., Starks, P. J., Seyfried, M., and Collins, C. H.: Comparison of microwave remote sensing and land surface modeling for surface soil moisture climatology estimation, Remote Sensing of Environment, 242, https://doi.org/10.1016/j.rse.2020.111756, 2020.

Dorigo, W. and de Jeu, R.: Satellite soil moisture for advancing our understanding of earth system processes and climate change, International Journal of Applied Earth Observation and Geoinformation, 48, 1–4, https://doi.org/10.1016/j.jag.2016.02.007, 2016.

Dorigo, W., Wagner, W., Albergel, C., Albrecht, F., Balsamo, G., Brocca, L., Chung, D., Ert, M., Forkel, M., Gruber, A., Haas, E., D.Hamer, P., Hirschi, M., Ikonen, J., de Jeu, R., Kidd, R., Lahoz, W., Liu, Y. Y., Miralles, D., Mistelbauer, T., Nicolai-Shaw, N., Parinussa, R., Pratola, C., Reimerak, C., van der Schalie, R., Seneviratne, S. I., Smolander, T., and Lecomte, P.: ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions, Remote Sensing of Environment, 203, 185–215, https://doi.org/10.1016/j.rse.2017.07.001, 2017.

Dorigo, W. A., Scipal, K., Parinussa, R. M., Liu, Y. Y., Wagner, W., de Jeu, R. A. M., and Naeimi, V.: Error characterisation of global active and passive microwave soil moisture datasets, Hydrology and Earth System Sciences, 14, 2605–2616, 2010.

Dorigo, W. A., Wagner, W., Hohensinn, R., Hahn, S., Paulik, C., Xaver, A., Gruber, A., Drusch, M., Mecklenburg, S., van Oevelen, P., Robock, A., and Jackson, T.: The International Soil Moisture Network: a data hosting facility for global in situ soil moisture measurements, Hydrology and Earth System Sciences, 15, 1675–1698, 2011.

Dorigo, W. A., Xaver, A., Vreugdenhil, M., Gruber, A., Hegyiová, A., Sanchis-Dufau, A. D., Zamojski, D., Cordes, C., Wagner, W., and Drusch, M.: Global automated quality control of in situ soil moisture data from the International Soil Moisture Network, Vadose Zone Journal, 12, https://doi.org/10.2136/vzj2012.0097, 2013.

Driessen, T. L. A., Hurkmans, R. T. W. L., Terink, W., Hazenberg, P., Torfs, P. J. J. F., and Uijlenhoet, R.: The hydrological response of the Ourthe catchment to climate change as modelled by the HBV model, Hydrology and Earth System Sciences, 14, 651–665, 2010.

Du, J., Kimball, J. S., Galantowicz, J., Kim, S.-B., Chan, S. K., Reichle, R., Jones, L. A., and Watts, J. D.: Assessing global surface water inundation dynamics using combined satellite information from SMAP, AMSR2 and Landsat, Remote Sensing of Environment, 213, 1–17, https://doi.org/10.1016/j.rse.2018.04.054, 2018.

Ebert, E. E., Janowiak, J. E., and Kidd, C.: Comparison of near-real-time precipitation estimates from satellite observations and numerical models, Bulletin of the American Meteorological Society, 88, 47–64, 2007.

Entekhabi, D., Njoku, E. G., O'Neill, P. E., Kellogg, K. H., Crow, W. T., Edelstein, W. N., Entin, J. K., Goodman, S. D., Jackson, T. J., Johnson, J., Kimball, J., Piepmeier, J. R., Koster, R. D., Martin, N., McDonald, K. C., Moghaddam, M., Moran, S., Reichle, R., Shi, J. C., Spencer, M. W., Thurman, S. W., Tsang, L., and Van Zyl, J.: The Soil Moisture Active Passive (SMAP) mission, Proceedings of the IEEE, 98, 704–716, 2010.

Entekhabi, D., Reichle, R. H., Koster, R., and Crow, W. T.: Performance metrics for soil moisture retrievals and application requirements, Journal of Hydrometeorology, 11, 832–840, 2010.

Fang, L., Hain, C. R., Zhan, X., and Anderson, M. C.: An inter-comparison of soil moisture data products from satellite remote sensing and a land surface model, International Journal of Applied Earth Observation and Geoinformation, 48, 37–50, https://doi.org/10.1016/j.jag.2015.10.006, 2016.

Fatichi, S., Vivoni, E. R., Ogden, F. L., Ivanov, V. Y., Mirus, B., Gochis, D., Downer, C. W., Camporese, M., Davison, J. H., Ebel, B., Jones, N., Kim, J., Mascaro, G., Niswonger, R., Restrepo, P., Rigon, R., Shen, C., Sulis, M., and Tarboton, D.: An overview of current applications, challenges, and future trends in distributed process-based models in hydrology, Journal of Hydrology, 537, 45–60, https://doi.org/10.1016/j.jhydrol.2016.03.026, 2016.

Fick, S. E. and Hijmans, R. J.: WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas, International Journal of Climatology, 37, 4302–4315, 2017.

Ford, T. W., Harris, E., and Quiring, S. M.: Estimating root zone soil moisture using near-surface observations from SMOS, Hydrology and Earth System Sciences, 18, 139–154, 2014.

Gharari, S., Clark, M. P., Mizukami, N., Knoben, W. J. M., Wong, J. S., and Pietroniro, A.: Flexible vector-based spatial configurations in land models, Hydrology and Earth System Sciences Discussions, 2020, 1–40, https://doi.org/10.5194/hess-2020-111, 2020.

Gottschalck, J., Meng, J., Rodell, M., and Houser, P.: Analysis of multiple precipitation products and preliminary assessment of their impact on Global Land Data Assimilation System land surface states, Journal of Hydrometeorology, 6, 573–598, 2005.

Griesfeller, A., Lahoz, W., Jeu, R., Dorigo, W., Haugen, L., Svendby, T., and Wagner, W.: Evaluation of satellite soil moisture products over Norway using ground-based observations, International Journal of Applied Earth Observation and Geoinformation, 45, 155–164, 2016.

Gruber, A., Su, C.-H., Zwieback, S., Crow, W., Dorigo, W., and Wagner, W.: Recent advances in (soil moisture) triple collocation analysis, International Journal of Applied Earth Observation and Geoinformation, 45, 200–211, 2016.

Gruber, A., Dorigo, W. A., Crow, W., and Wagner, W.: Triple collocation-based merging of satellite soil moisture retrievals, IEEE Transactions on Geoscience and Remote Sensing, 55, 6780–6792, 2017.

Gruber, A., Scanlon, T., van der Schalie, R., Wagner, W., and Dorigo, W.: Evolution of the ESA CCI soil moisture climate data records and their underlying merging methodology, Earth System Science Data, 11, 717–739, 2019.

Gruber, A., De Lannoy, G., Al-Yaari, C. A. A., Brocca, L., Calvet, J.-C., Colliander, A., Cosh, M., Crow, W., Dorigo, W., Draper, C., Hirschi, M., Kerr, Y., Konings, A., Lahoz, W., McColl, K., Montzka, C., noz Sabater, J. M., Peng, J., Reichle, R., Richaume, P., Rudiger, C., Scanlon, T., van der Schalie, R., Wigneron, J.-P., and Wagner, W.: Validation practices for satellite soil moisture retrievals: What are (the) errors?, Remote Sensing of Environment, 244, 111 806, https://doi.org/10.1016/j.rse.2020.111806, 2020.

Gruhier, C., de Rosnay, P., Kerr, Y., Mougin, E., Ceschia, E., Calvet, J.-C., and Richaume, P.: Evaluation of AMSR-E soil moisture product based on ground measurements over temperate and semi-arid regions, Geophysical Research Letters, 35, https://doi.org/10.1029/2008GL033330, 2008.

Gruhier, C., de Rosnay, P., Hasenauer, S., Holmes, T., de Jeu, R., Kerr, Y., Mougin, E., Njoku, E., Timouk, F., Wagner, W., and Zribi, M.: Soil moisture active and passive microwave products: intercomparison and evaluation over a Sahelian site, Hydrology and Earth System Sciences, 14, 141–156, 2010.

Guo, Y., Shi, J., Du, J., and Fu, X.: Evaluation of terrain effect on microwave radiometer measurement and its correction, International Journal of Remote Sensing, 32, 8899–8913, 2011.

Guo, Z., Dirmeyer, P. A., Gao, X., and Zhao, M.: Improving the quality of simulated soil moisture with a multi-model ensemble approach, Quarterly Journal of the Royal Meteorological Society, 133, 731–747, 2007.

Guswa, A. J., Celia, M. A., and Rodriguez-Iturbe, I.: Models of soil moisture dynamics in ecohydrology: a comparative study, Water Resources Research, 38, https://doi.org/10.1029/2001WR000826, 2002.

H SAF: Metop ASCAT surface soil moisture climate data record v5 12.5 km sampling (H115), http://dx.doi.org/10.15770/EUM_SAF_H_0006, https://doi.org/10.15770/EUM_SAF_H_0006, EUMETSAT SAF on Support to Operational Hydrology and Water Management, 2019a.

H SAF: ASCAT surface soil moisture climate data record v5 extension 12.5 km sampling — Metop (H116), https://navigator.eumetsat.int/product/EO:EUM:DAT:METOP:H116, EUMETSAT SAF on Support to Operational Hydrology and Water Management, 2019b.

Hargreaves, G. H.: Defining and using reference evapotranspiration, Journal of Irrigation and Drainage Engineering, 120, 1132–1139, 1994.

He, X., Pan, M., Wei, Z., Wood, E. F., and Sheffield, F.: A global drought and flood catalogue from 1950 to 2016, Bulletin of the American Meteorological Society, 101, E508–E535, 2020.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horanyi, A., noz Sabater, J. M., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G. D., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Holm, E., Janiskova, M., Keeley, S., Laloyaux, P., Lopez, P., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, https://doi.org/10.1002/qj.3803, 2020.

Houser, P. R., Shuttleworth, W. J., Famiglietti, J. S., Gupta, H. V., Syed, K. H., and Goodrich, D. C.: Integration of soil moisture remote sensing and hydrologic modeling using data assimilation, Water Resources Research, 34, 3405–3420, 1998.

Huffman, G. J., Adler, R. F., Morrissey, M. M., Bolvin, D. T., Curtis, S., Joyce, R., McGavock, B., and Susskind, J.: Global precipitation at one-degree daily resolution from multi-satellite observations, Journal of Hydrometeorology, 2, 36–50, 2001.

Huffman, G. J., Bolvin, D. T., Braithwaite, D., Hsu, K., Joyce, R., Kidd, C., Nelkin, E. J., and Xie, P.: NASA Global Precipitation Measurement (GPM) Integrated Multi-satellitE Retrievals for GPM (IMERG), Algorithm Theoretical Basis Document (ATBD), NASA/GSFC, Greenbelt, MD 20771, USA, 2014.

Huffman, G. J., Bolvin, D. T., and Nelkin, E. J.: Integrated Multi-satellitE Retrievals for GPM (IMERG) Technical Documentation, Tech. rep., NASA/GSFC, Greenbelt, MD 20771, USA, 2018.

Jackson, T. J., Cosh, M. H., Bindlish, R., Starks, P. J., Bosch, D. D., Seyfried, M., Goodrich, D. C., Moran, M. S., and Du, J.: Validation of Advanced Microwave Scanning Radiometer soil moisture products, IEEE Transactions on Geoscience and Remote Sensing, 48, 4256–4272, 2010.

Jin, R., Li, X., Yan, B., Li, X., Luo, W., Ma, M., Guo, J., Kang, J., Zhu, Z., and Zhao, S.: A nested ecohydrological wireless sensor network for capturing the surface heterogeneity in the midstream areas of the Heihe river basin, China, IEEE Geoscience and Remote Sensing Letters, 11, 2015–2019, 2014.

Jódar, J., Carpintero, E., Martos-Rosillo, S., Ruiz-Constán, A., Marín-Lechado, C., Cabrera-Arrabal, J. A., Navarrete-Mazariegos, E., González-Ramón, A., Lambán, L. J., Herrera, C., and González-Dugo, M. P.: Combination of lumped hydrological and remote-sensing models to evaluate water resources in a semi-arid high altitude ungauged watershed of Sierra Nevada (Southern Spain), Science of The Total Environment, 625, 285–300, https://doi.org/https://doi.org/10.1016/j.scitotenv.2017.12.300, 2018.

Kang, C. S., Kanniah, K. D., Kerr, Y. H., and Cracknell, A. P.: Analysis of in-situ soil moisture data and validation of SMOS soil moisture products at selected agricultural sites over a tropical region, International Journal of Remote Sensing, 37, 3636–3654, 2016.

Kang, J., Li, X., Jin, R., Ge, Y., Wang, J., and Wang, J.: Hybrid optimal design of the eco-hydrological wireless sensor network in the middle reach of the Heihe river basin, China, Sensors, 14, 19 095, 2014.

Karthikeyan, L. and Kumar, D. N.: A novel approach to validate satellite soil moisture retrievals using precipitation data, Journal of Geophysical Research: Atmospheres, 121, 11 516–11 535, 2016.

Karthikeyan, L., Pan, M., Wanders, N., Kumar, D. N., and Wood, E. F.: Four decades of microwave satellite soil moisture observations: Part 2. Product validation and inter-satellite comparisons, Advances in Water Resources, 109, 236–252, https://doi.org/10.1016/j.advwatres.2017.09.010, 2017a.

Karthikeyan, L., Pan, M., Wanders, N., Kumar, D. N., and Wood, E. F.: Four decades of microwave satellite soil moisture observations: Part 1. A review of retrieval algorithms, Advances in Water Resources, 109, 106–120, https://doi.org/10.1016/j.advwatres.2017.09.006, 2017b.

Karthikeyan, L., Chawla, I., and Mishra, A. K.: A review of remote sensing applications in agriculture for food security: crop growth and yield, irrigation, and crop losses, Journal of Hydrology, p. 124905, https://doi.org/10.1016/j.jhydrol.2020.124905, 2020.

Kauffeldt, A., Wetterhall, F., Pappenberger, F., Salamon, P., and Thielen, J.: Technical review of large-scale hydrological models for implementation in operational flood forecasting schemes on continental level, Environmental Modelling & Software, 75, 68–76, https://doi.org/10.1016/j.envsoft.2015.09.009, 2016.

Kerr, Y. H., Waldteufel, P., Richaume, P., Wigneron, J. P., Ferrazzoli, P., Mahmoodi, A., Al Bitar, A., Cabot, F., Gruhier, C., Juglea, S. E., Leroux, D., Mialon, A., and Delwart, S.: The SMOS soil moisture retrieval algorithm, IEEE Transactions on Geoscience and Remote Sensing, 50, 1384–1403, 2012.

Kidd, C., Becker, A., Huffman, G. J., Muller, C. L., Joe, P., Skofronick-Jackson, G., and Kirschbaum, D. B.: So, how much of the Earth's surface is covered by rain gauges?, Bulletin of the American Meteorological Society, 98, 69–78, 2017.

Kim, H., Parinussa, R., Konings, A. G., Wagner, W., Cosh, M. H., Lakshmi, V., Zohaib, M., and Choi, M.: Global-scale assessment and combination of SMAP with ASCAT (active) and AMSR2 (passive) soil moisture products, Remote Sensing of Environment, 204, 260–275, https://doi.org/10.1016/j.rse.2017.10.026, 2018.

Kirchner, J. W.: Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology, Water Resources Research, 42, W03S04, https://doi.org/10.1029/2005WR004362, 2006.

Knutti, R.: Should we believe model predictions of future climate change?, Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 366, 4647–4664, 2008.

Koren, V., Moreda, F., and Smith, M.: Use of soil moisture observations to improve parameter consistency in watershed calibration, Physics and Chemistry of the Earth, Parts A/B/C, 33, 1068–1080, 2008.

Koster, R. D., Guo, Z., Yang, R., Dirmeyer, P. A., Mitchell, K., and Puma, M. J.: On the nature of soil moisture in land surface models, Journal of Climate, 22, 4322–4335, 2009.

Koster, R. D., Liu, Q., Mahanama, S. P. P., and Reichle, R. H.: Improved hydrological simulation using SMAP data: relative impacts of model calibration and data assimilation, Journal of Hydrometeorology, 19, 727–741, 2018.

Kruskal, W. H. and Wallis, W. A.: Use of Ranks in One-Criterion Variance Analysis, Journal of the American Statistical Association, 47, 583–621, 1952.

Kumar, S. V., Peters-Lidard, C. D., Mocko, D., Reichle, R., Liu, Y., Arsenault, K. R., Xia, Y., Ek, M., Riggs, G., Livneh, B., and Cosh, M.: Assimilation of remotely sensed soil moisture and snow depth retrievals for drought estimation, Journal of Hydrometeorology, 15, 2446–2469, 2014.

Lahoz, W. A. and De Lannoy, G. J. M.: Closing the gaps in our knowledge of the hydrological cycle over land: conceptual problems, Surveys in Geophysics, 35, 623–660, 2014.

Laiolo, P., Gabellani, S., Campo, L., Silvestro, F., Delogu, F., Rudari, R., Pulvirenti, L., Boni, G., Fascetti, F., Pierdicca, N., Crapolicchio, R., Hasenauer, S., and Puca, S.: Impact of different satellite soil moisture products on the predictions of a continuous distributed hydrological model, International Journal of Applied Earth Observation and Geoinformation, 48, 131–145, https://doi.org/10.1016/j.jag.2015.06.002, 2016.

Lei, F., Crow, W. T., Shen, H., Parinussa, R. M., and Holmes, T. R. H.: The impact of local acquisition time on the accuracy of microwave surface soil moisture retrievals over the contiguous United States, Remote Sensing, 7, 13 448–13 465, 2015.

Liu, J.-G. and Xie, Z.-H.: Improving simulation of soil moisture in China using a multiple meteorological forcing ensemble approach, Hydrology and Earth System Sciences, 17, 3355–3369, 2013.

Liu, Q., Reichle, R. H., Bindlish, R., Cosh, M. H., Crow, W. T., de Jeu, R. A. M., De Lannoy, G. J. M., Huffman, G. J., and Jackson, T. J.: The contributions of precipitation and soil moisture observations to the skill of soil moisture estimates in a land data assimilation system, Journal of Hydrometeorology, 12, 750–765, 2011.

Liu, Y., Weerts, A. H., Clark, M., Hendricks Franssen, H.-J., Kumar, S., Moradkhani, H., Seo, D.-J., Schwanenberg, D., Smith, P., van Dijk, A. I. J. M., van Velzen, N., He, M., Lee, H., Noh, S. J., Rakovec, O., and Restrepo, P.: Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities, Hydrology and Earth System Sciences, 16, 3863–3887, 2012.

Liu, Y., Liu, Y., and Wang, W.: Inter-comparison of satellite-retrieved and Global Land Data Assimilation System-simulated soil moisture datasets for global drought analysis, Remote Sensing of Environment, 220, 1–18, https://doi.org/10.1016/j.rse.2018.10.026, 2019.

Loew, A., Dall'Amico, J. T., Schlenz, F., and Mauser, W.: The Upper Danube soil moisture validation site: measurements and activities, in: Proceedings of the Symposium Earth Observation and Water Cycle Science, vol. 674, p. 56, Frascati, Italy, 2009.

Long, D. and Ulaby, F. T.: Microwave radar and radiometric remote sensing, Artech House, 2015.

López López, P., Wanders, N., Schellekens, J., Renzullo, L. J., Sutanudjaja, E. H., and Bierkens, M. F. P.: Improved large-scale hydrological modelling through the assimilation of streamflow and downscaled satellite soil moisture observations, Hydrology and Earth System Sciences, 20, 3059–3076, 2016.

Lü, H., Crow, W. T., Zhu, Y., Ouyang, F., and Su, J.: Improving streamflow prediction using remotely-sensed soil moisture and snow depth, Remote Sensing, 8, 2016.

Lv, S., Zeng, Y., Wen, J., Zhao, H., and Su, Z.: Estimation of penetration depth from soil effective temperature in microwave radiometry, Remote Sensing, 10, https://doi.org/10.3390/rs10040519, 2018.

Ma, H., Zeng, J., Chen, N., Zhang, X., Cosh, M. H., and Wang, W.: Satellite surface soil moisture from SMAP, SMOS, AMSR2 and ESA CCI: A comprehensive assessment using global ground-based observations, Remote Sensing of Environment, 231, 111 215, https://doi.org/10.1016/j.rse.2019.111215, 2019.

Marczewski, W., Slominski, J., Slominska, E., Usowicz, B., Usowicz, J., Romanov, S., Maryskevych, O., Nastula, J., and Zawadzki, J.: Strategies for validating and directions for employing SMOS data, in the Cal-Val project SWEX (3275) for wetlands, Hydrology and Earth System Sciences Discussions, 7, 7007–7057, https://doi.org/10.5194/hessd-7-7007-2010, 2010.

Martens, B., Miralles, D., Lievens, H., Fernández-Prieto, D., and Verhoest, N.: Improving terrestrial evaporation estimates over continental Australia through assimilation of SMOS soil moisture, International Journal of Applied Earth Observation and Geoinformation, 48, 146–162, https://doi.org/10.1016/j.jag.2015.09.012, 2016.

Martens, B., Miralles, D. G., Lievens, H., van der Schalie, R., de Jeu, R. A. M., Férnandez-Prieto, D., Beck, H. E., Dorigo, W., and Verhoest, N. E. C.: GLEAM v3: satellite-based land evaporation and root-zone soil moisture, Geoscientific Model Development, 10, 1903–1925, 2017.

Massari, C., Crow, W., and Brocca, L.: An assessment of the accuracy of global rainfall estimates without ground-based observations, Hydrology and Earth System Sciences, 21, 4347–4361, https://doi.org/10.5194/hess-2017-163, 2017.

Mattar, C., Santamaría-Artigas, A., Durán-Alarcón, C., Olivera-Guerra, L., and Fuster, R.: LAB-net the First Chilean soil moisture network for remote sensing applications, in: Quantitative Remote Sensing Symposium (RAQRS), Valencia, Spain, 2014.

McCabe, M. F., Ershadi, A., Jimenez, C., Miralles, D. G., Michel, D., and Wood, E. F.: The GEWEX LandFlux project: evaluation of model evaporation using tower-based and globally-gridded forcing data, Geoscientific Model Development, 9, 283–305, https://doi.org/10.5194/gmd-9-283-2016, 2016.

McColl, K. A., Kaighin, A., Vogelzang, J., Konings, A. G., Entekhabi, D., Piles, M., and Stoffelen, A.: Extended triple collocation: Estimating errors and correlation coefficients with respect to an unknown target, Geophysical Research Letters, 41, 6229–6236, 2014.

McKay, M. D., Conover, W. J., and Beckman, R. J.: A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, Technometrics, 21, 239–245, 1979.

Mendoza, P. A., Clark, M. P., Barlage, M., Rajagopalan, B., Samaniego, L., Abramowitz, G., and Gupta, H.: Are we unnecessarily constraining the agility of complex process-based models?, Water Resources Research, 51, 716–728, https://doi.org/10.1002/2014WR015820, 2015.

Mialon, A., Coret, L., Kerr, Y. H., Secherre, F., and Wigneron, J.: Flagging the topographic impact on the SMOS signal, IEEE Transactions on Geoscience and Remote Sensing, 46, 689–694, 2008.

Miralles, D. G., Crow, W. T., and Cosh, M. H.: Estimating spatial sampling errors in coarse-scale soil moisture estimates derived from point-scale observations, Journal of Hydrometeorology, 11, 1423–1429, 2010.

Miralles, D. G., Gentine, P., Seneviratne, S. I., and Teuling, A. J.: Land-atmospheric feedbacks during droughts and heatwaves: state of the science and current challenges, Annals of the New York Academy of Sciences, 1436, https://doi.org/10.1111/nyas.13912, 2019.

Miyaoka, K., Gruber, A., Ticconi, F., Hahn, S., Wagner, W., Salda na, J. F., and Anderson, C.: Triple collocation analysis of soil moisture from Metop-A ASCAT and SMOS against JRA-55 and ERA-Interim, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 10, 2274–2284, 2017.

Moghaddam, M., Entekhabi, D., Goykhman, Y., Li, K., Liu, M., Mahajan, A., Nayyar, A., Shuman, D., and Teneketzis, D.: A wireless soil moisture smart sensor web using physics-based optimal control: concept and initial demonstrations, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 3, 522–535, 2010.

Moghaddam, M., Silva, A., Clewley, D., Akbar, R., Hussaini, S., Whitcomb, J., Devarakonda, R., Shrestha, R., Cook, R., Prakash, G., Santhana Vannan, S., and Boyer, A.: Soil Moisture Profiles and Temperature Data from SoilSCAPE Sites, USA, https://doi.org/10.3334/ORNLDAAC/1339, http://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=1339, 2016.

Montero, R. A., Schwanenberg, D., Krahe, P., Lisniak, D., Sensoy, A., Sorman, A. A., and Akkol, B.: Moving horizon estimation for assimilating H-SAF remote sensing data into the HBV hydrological model, Advances in Water Resources, 92, 248–257, 2016.

Moradkhani, H.: Hydrologic remote sensing and land surface data assimilation, Sensors, 8, 2986–3004, 2008.

Morbidelli, R., Saltalippi, C., Flammini, A., Rossi, E., and Corradini, C.: Soil water content vertical profiles under natural conditions: matching of experiments and simulations by a conceptual model, Hydrological Processes, 28, 4732–4742, 2014.

Muñoz Sabater, J., Lawrence, H., Albergel, C., Rosnay, P., Isaksen, L., Mecklenburg, S., Kerr, Y., and Drusch, M.: Assimilation of SMOS brightness temperatures in the ECMWF Integrated Forecasting System, Quarterly Journal of the Royal Meteorological Society, 145, 2524–2548, 2019.

Myeni, L., Moeletsi, M. E., and Clulow, A. D.: Present status of soil moisture estimation over the African continent, Journal of Hydrology: Regional Studies, 21, 14–24, https://doi.org/10.1016/j.ejrh.2018.11.004, 2019.

Naz, B. S., Kurtz, W., Montzka, C., Sharples, W., Goergen, K., Keune, J., Gao, H., Springer, A., Hendricks Franssen, H.-J., and Kollet, S.: Improving soil moisture and runoff simulations at 3 km over Europe using land surface data assimilation, Hydrology and Earth System Sciences, 23, 277–301, 2019.

Njoku, E. G., Ashcroft, P., Chan, T. K., and Li, L.: Global survey and statistics of radio-frequency interference in AMSR-E land observations, IEEE Transactions Geoscience and Remote Sensing, 43, 938–947, 2005.

Ochsner, T. E., Cosh, M. H., Cuenca, R. H., Dorigo, W. A., Draper, C. S., Hagimoto, Y., Kerr, Y. H., Larson, K. M., Njoku, E. G., Small, E. E., and Zreda, M.: State of the art in large-scale soil moisture monitoring, Soil Science Society of America Journal, 77, 1888–1919, 2013.

Ojo, E. R., Bullock, P. R., L'Heureux, J., Powers, J., McNairn, H., and Pacheco, A.: Calibration and evaluation of a frequency domain reflectometry sensor for real-time soil moisture monitoring, Vadose Zone Journal, 14, 2015.

Oliva, R., Daganzo, E., Kerr, Y. H., Mecklenburg, S., Nieto, S., Richaume, P., and Gruhier, C.: SMOS radio frequency interference scenario: status and actions taken to improve the RFI environment in the 1400–1427-MHz passive band, IEEE Transactions on Geoscience and Remote Sensing, 50, 1427–1439, 2012.

O'Neill, P. E., Chan, S., Njoku, E. G., Jackson, T., and Bindlish, R.: SMAP Enhanced L3 radiometer global daily 9 km EASE-grid soil moisture, version 3, https://doi.org/10.5067/T90W6VRLCBHI, 2019.

Orth, R., Staudinger, M., Seneviratne, S. I., Seibert, J., and Zappa, M.: Does model performance improve with complexity? A case study with three hydrological models, Journal of Hydrology, 523, 147–159, https://doi.org/10.1016/j.jhydrol.2015.01.044, 2015.

Osenga, E. C., Arnott, J. C., Endsley, K. A., and Katzenberger, J. W.: Bioclimatic and soil moisture monitoring across elevation in a mountain watershed: opportunities for research and resource management, Water Resources Research, 55, 2493–2503, 2019.

Pablos, M., González-Zamora, A., Sánchez, N., and Martínez-Fernández, J.: Assessment of root zone soil moisture estimations from SMAP, SMOS and MODIS Observations, Remote Sensing, 10, 2018.

Pan, M. and Wood, E. F.: Impact of accuracy, spatial availability, and revisit time of satellite-derived surface soil moisture in a multiscale ensemble data assimilation system, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 3, 49–56, 2010.

Pan, M., Wood, E. F., McLaughlin, D. B., Entekhabi, D., and Luo, L.: A multiscale ensemble filtering system for hydrologic data assimilation. Part I: Implementation and synthetic experiment, Journal of Hydrometeorology, 10, 794–806, 2009.

Parajka, J., Naeimi, V., Blöschl, G., Wagner, W., Merz, R., and Scipal, K.: Assimilating scatterometer soil moisture data into conceptual hydrologic models at the regional scale, Hydrology and Earth System Sciences, 10, 353–368, 2006.

Parinussa, R. M., Holmes, T. R. H., Yilmaz, M. T., and Crow, W. T.: The impact of land surface temperature on soil moisture anomaly detection from passive microwave observations, Hydrology and Earth System Sciences, 15, 3135–3151, 2011.

Parinussa, R. M., Holmes, T. R. H., Wanders, N., Dorigo, W. A., and de Jeu, R. A. M.: A preliminary study toward consistent soil moisture from AMSR2, Journal of Hydrometeorology, 16, 932–947, 2015.

Paulik, C., Dorigo, W., Wagner, W., and Kidd, R.: Validation of the ASCAT Soil Water Index using in situ data from the International Soil Moisture Network, International Journal of Applied Earth Observation and Geoinformation, 30, 1–8, https://doi.org/10.1016/j.jag.2014.01.007, 2014.

Pellarin, T., Calvet, J.-C., and Wagner, W.: Evaluation of ERS scatterometer soil moisture products over a half-degree region in southwestern France, Geophysical Research Letters, 33, https://doi.org/10.1029/2006GL027231, 2006.

Petropoulos, G. P. and McCalmont, J. P.: An operational in situ soil moisture & soil temperature monitoring network for West Wales, UK: the WSMN network, Sensors, 17, 7, https://doi.org/10.3390/s17071481, 2017.

Petropoulos, G. P., Ireland, G., and Barrett, B.: Surface soil moisture retrievals from remote sensing: Current status, products & future trends, Physics and Chemistry of the Earth, Parts A/B/C, 83–84, 36–56, https://doi.org/10.1016/j.pce.2015.02.009, 2015.

Piepmeier, J. R., Johnson, J. T., Mohammed, P. N., Bradley, D., Ruf, C., Aksoy, M., Garcia, R., Hudson, D., Miles, L., and Wong, M.: Radio-frequency interference mitigation for the Soil Moisture Active Passive microwave radiometer, IEEE Transactions on Geoscience and Remote Sensing, 52, 761–775, 2014.

Pulvirenti, L., Pierdicca, N., and Marzano, F. S.: Prediction of the error induced by topography in satellite microwave radiometric observations, IEEE Transactions on Geoscience and Remote Sensing, 49, 3180–3188, 2011.

Reichle, R., De Lannoy, G., Koster, R. D., Crow, W. T., Kimball, J. S., and Liu, Q.: SMAP L4 global 3-hourly 9 km EASE-grid surface and root zone soil moisture geophysical data, version 4, https://doi.org/10.5067/KPJNN2GI1DQR, 2019a.

Reichle, R. H. and Koster, R. D.: Bias reduction in short records of satellite soil moisture, Geophysical Research Letters, 31, https://doi.org/10.1029/2004GL020938, 2004.

Reichle, R. H., De Lannoy, G. J. M., Liu, Q., Ardizzone, J. V., Colliander, A., Conaty, A., Crow, W., Jackson, T. J., Jones, L. A., Kimball, J. S., Koster, R. D., Mahanama, S. P., Smith, E. B., Berg, A., Bircher, S., Bosch, D., Caldwell, T. G., Cosh, M., González-Zamora, l., Collins, C. D. H., Jensen, K. H., Livingston, S., Lopez-Baeza, E., Martínez-Fernández, J., McNairn, H., Moghaddam, M., Pacheco, A., Pellarin, T., Prueger, J., Rowlandson, T., Seyfried, M., Starks, P., Su, Z., Thibeault, M., van der Velde, R., Walker, J., Wu, X., Zeng, Y., Reichle, R. H., De Lannoy, G. J. M., Liu, Q., Ardizzone, J. V., Colliander, A., Conaty, A., Crow, W., Jackson, T. J., Jones, L. A., Kimball, J. S., Koster, R. D., Mahanama, S. P., Smith, E. B., Berg, A., Bircher, S., Bosch, D., Caldwell, T. G., Cosh, M., González-Zamora, A., Collins, C. D. H., Jensen, K. H., Livingston, S., Lopez-Baeza, E., Martínez-Fernández, J., McNairn, H., Moghaddam, M., Pacheco, A., Pellarin, T., Prueger, J., Rowlandson, T., Seyfried, M., Starks, P., Su, Z., Thibeault, M., van der Velde, R., Walker, J., Wu, X., and Zeng, Y.: Assessment of the SMAP Level-4 surface and root-zone soil moisture product using in situ measurements, Journal of Hydrometeorology, 18, 2621–2645, 2017.

Reichle, R. H., Liu, Q., Koster, R. D., Crow, W. T., De Lannoy, G. J. M., Kimball, J. S., Ardizzone, J. V., Bosch, D., Colliander, A., Cosh, M., Kolassa, J., Mahanama, S. P., Prueger, J., Starks, P., and Walker, J. P.: Version 4 of the SMAP level-4 soil moisture algorithm and data product, Journal of Advances in Modeling Earth Systems, 11, 3106–3130, 2019b.

Rodell, M., Houser, P., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J., Walker, J., Lohmann, D., , and Toll, D.: The Global Land Data Assimilation System, Bulletin of the American Meteorological Society, 85, 381–394, 2004.

Rondinelli, W. J., Hornbuckle, B. K., Patton, J. C., Cosh, M. H., Walker, V. A., Carr, B. D., and Logsdon, S. D.: Different rates of soil drying after rainfall are observed by the SMOS satellite and the South Fork in situ soil moisture network, Journal of Hydrometeorology, 16, 889–903, 2015.

Rui, H., Beaudoing, H., and Loeser, C.: README document for NASA GLDAS version 2 data products, NASA Goddard Earth Science Data Information and Services Center (GES DISC), Greenbelt, Maryland, https://hydro1.gesdisc.eosdis.nasa.gov/data/GLDAS/README_GLDAS2.pdf, 2020.

Sahlu, D., Moges, S. A., Nikolopoulos, E. I., Anagnostou, E. N., and Hailu, D.: Evaluation of high-resolution multisatellite and reanalysis rainfall products over East Africa, Advances in Meteorology, 2017, https://doi.org/10.1155/2017/4957960, 2017.

Samaniego, L., Kumar, R., and Attinger, S.: Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale, Water Resources Research, 46, https://doi.org/10.1029/2008WR007327, 2010.

Satgé, F., Ruelland, D., Bonnet, M.-P., Molina, J., and Pillco, R.: Consistency of satellite-based precipitation products in space and over time compared with gauge observations and snow-hydrological modelling in the Lake Titicaca region, Hydrology and Earth System Sciences, 23, 595–619, 2019.

Scipal, K., Holmes, T., de Jeu, R., Naeimi, V., and Wagner, W.: A possible solution for the problem of estimating the error structure of global soil moisture data sets, Geophysical Research Letters, 35, https://doi.org/10.1029/2008GL035599, 2008.

Sheffield, J., Goteti, G., and Wood, E. F.: Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling, Journal of Climate, 19, 3088–3111, 2006.

Shellito, P. J., Small, E. E., Colliander, A., Bindlish, R., Cosh, M. H., Berg, A. A., Bosch, D. D., Caldwell, T. G., Goodrich, D. C., McNairn, H., Prueger, J. H., Starks, P. J., van der Velde, R., and Walker, J. P.: SMAP soil moisture drying more rapid than observed in situ following rainfall events, Geophysical Research Letters, 43, 8068–8075, 2016a.

Shellito, P. J., Small, E. E., and Cosh, M. H.: Calibration of Noah soil hydraulic property parameters using surface soil moisture from SMOS and basinwide in situ observations, Journal of Hydrometeorology, 17, 2275–2292, 2016b.

Singh, G., Das, N. N., Panda, R. K., Colliander, A., Jackson, T. J., Mohanty, B. P., Entekhabi, D., and Yueh, S. H.: Validation of SMAP soil moisture products using ground-based observations for the paddy dominated tropical region of India, IEEE Transactions on Geoscience and Remote Sensing, 57, 8479–8491, 2019.

Smith, A. B., Walker, J. P., Western, A. W., Young, R. I., Ellett, K. M., Pipunic, R. C., Grayson, R. B., Siriwardena, L., Chiew, F. H. S., and Richter, H.: The Murrumbidgee soil moisture monitoring network data set, Water Resources Research, 48, https://doi.org/10.1029/2012WR011976, 2012.

Steele-Dunne, S., Lynch, P., McGrath, R., Semmler, T., Wang, S., Hanafin, J., and Nolan, P.: The impacts of climate change on hydrology in Ireland, Journal of Hydrology, 356, 28–45, 2008.

Su, C.-H., Narsey, S. Y., Gruber, A., Xaver, A., Chung, D., Ryu, D., and Wagner, W.: Evaluation of post-retrieval de-noising of active and passive microwave satellite soil moisture, Remote Sensing of Environment, 163, 127–139, https://doi.org/10.1016/j.rse.2015.03.010, 2015.

Su, C.-H., Zhang, J., Gruber, A., Parinussa, R., Ryu, D., Crow, W. T., and Wagner, W.: Error decomposition of nine passive and active microwave satellite soil moisture data sets over Australia, Remote Sensing of Environment, 182, 128–140, https://doi.org/10.1016/j.rse.2016.05.008, 2016.

Tagesson, T., Fensholt, R., Guiro, I., Rasmussen, M. O., Huber, S., Mbow, C., Garcia, M., Horion, S., Sandholt, I., Holm-Rasmussen, B., Göttsche, F. M., Ridler, M.-E., Olén, N., Lundegard Olsen, J., Ehammer, A., Madsen, M., Olesen, F. S., and Ardö, J.: Ecosystem properties of semiarid savanna grassland in West Africa and its relationship with environmental variability, Global Change Biology, 21, 250–264, 2015.

Tavakol, A., Rahmani, V., Quiring, S. M., and Kumar, S. V.: Evaluation analysis of NASA SMAP L3 and L4 and SPoRT-LIS soil moisture data in the United States, Remote Sensing of Environment, 229, 234–246, https://doi.org/https://doi.org/10.1016/j.rse.2019.05.006, 2019.

Teuling, A. J., Uijlenhoet, R., van den Hurk, B., and Seneviratne, S. I.: Parameter sensitivity in LSMs: An analysis using stochastic soil moisture models and ELDAS soil parameters, Journal of Hydrometeorology, 10, 751–765, 2009.

Thorstensen, A., Nguyen, P., Hsu, K., and Sorooshian, S.: Using densely distributed soil moisture observations for calibration of a hydrologic model, Journal of Hydrometeorology, 17, 571–590, 2016.

Tian, S., Renzullo, L. J., van Dijk, A. I. J. M., Tregoning, P., and Walker, J. P.: Global joint assimilation of GRACE and SMOS for improved estimation of root-zone soil moisture and vegetation response, Hydrology and Earth System Sciences, 23, 1067–1081, 2019.

Ticconi, F., Anderson, C., Figa-Salda na, J., Wilson, J. J. W., and Bauch, H.: Analysis of radio frequency interference in Metop ASCAT backscatter measurements, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 10, 2360–2371, 2017.

Tobin, K. J., Crow, W. T., Dong, J., and Bennett, M. E.: Validation of a new root-zone soil moisture product: Soil MERGE, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 12, 3351–3365, 2019.

Van Cleve, K., Chapin, F. S., Stuart, R., and Roger, W.: Bonanza Creek long term ecological research project climate database, www.lter.uaf.edu, 2015.

Vereecken, H., Huisman, J. A., Bogena, H., Vanderborght, J., Vrugt, J. A., and Hopmans, J. W.: On the value of soil moisture measurements in vadose zone hydrology: A review, Water Resources Research, 44, https://doi.org/10.1029/2008WR006829, 2008.

Vergopolan, N., Chaney, N. W., Beck, H. E., Pan, M., Sheffield, J., Chan, S., and Wood, E. F.: Combining hyper-resolution land surface modeling with SMAP brightness temperatures to obtain 30-m soil moisture estimates, Remote Sensing of Environment, 242, 111 740, https://doi.org/10.1016/j.rse.2020.111740, 2020.

Vetter, T., Huang, S., Aich, V., Yang, T., Wang, X., Krysanova, V., and Hattermann, F.: Multi-model climate impact assessment and intercomparison for three large-scale river basins on three continents, Earth System Dynamics, 6, 17–43, 2015.

Wagner, W., Lemoine, G., and Rott, H.: A method for estimating soil moisture from ERS scatterometer and soil data, Remote Sensing of Environment, 70, 191–207, 1999.

Wagner, W., Blöschl, G., Pampaloni, P., Calvet, J.-C., Bizzarri, B., Wigneron, J.-P., and Kerr, Y.: Operational readiness of microwave remote sensing of soil moisture for hydrologic applications, Hydrology Research, 38, 1–20, 2007.

Wagner, W., Hahn, S., Kidd, R., Melzer, T., Bartalis, Z., Hasenauer, S., Salda na, J. F., de Rosnay, P., Jann, A., Schneider, S., Komma, J., Kubu, G., Gerhard, B., Katharina, A., Aubrecht, C., Züger, J., Gangkofner, U., Kienberger, S., Brocca, L., Wang, Y., Blöschl, G., Eitzinger, J., Steinnocher, K., Zeil, P., and Rubel, F.: The ASCAT soil moisture product: a review of its specifications, validation results, and emerging applications, Meteorologische Zeitschrift, 22, 5–33, 2013.

Wanders, N., Bierkens, M. F. P., de Jong, S. M., de Roo, A., and Karssenberg, D.: The benefits of using remotely sensed soil moisture in parameter identification of large-scale hydrological models, Water Resources Research, 50, 6874–6891, 2014.

Wang, H. and Yong, B.: Quasi-global evaluation of IMERG and GSMaP precipitation products over land using gauge observations, Water, 12, 243, https://doi.org/10.3390/w12010243, 2020.

Wu, Q., Liu, H., Wang, L., and Deng, C.: Evaluation of AMSR2 soil moisture products over the contiguous United States using in situ data from the International Soil Moisture Network, International Journal of Applied Earth Observation and Geoinformation, 45, 187–199, 2016.

Xia, Y., Hao, Z., Shi, C., Li, Y., Meng, J., Xu, T., Wu, X., and Zhang, B.: Regional and global land data assimilation systems: Innovations, challenges, and prospects, Journal of Meteorological Research, 33, 159–189, 2019.

Yamazaki, D., Ikeshima, D., Tawatari, R., Yamaguchi, T., O'Loughlin, F., Neal, J. C., Sampson, C. C., Kanae, S., and Bates, P. D.: A high-accuracy map of global terrain elevations, Geophysical Research Letters, 44, 5844–5853, 2017.

Yang, K., Qin, J., Zhao, L., Chen, Y., Tang, W., Han, M., Lazhu, Chen, Z., Lv, N., Ding, B., Wu, H., and Lin, C.: A multiscale soil moisture and freeze-thaw monitoring network on the third pole, Bulletin of the American Meteorological Society, 94, 1907–1916, 2013.

Ye, N., Walker, J., Guerschman, J., Ryu, D., and Gurney, R.: Standing water effect on soil moisture retrieval from L-band passive microwave observations, Remote Sensing of Environment, 169, 232–242, https://doi.org/10.1016/j.rse.2015.08.013, 2015.

Yilmaz, M. T. and Crow, W. T.: Evaluation of assumptions in soil moisture triple collocation analysis, Journal of Hydrometeorology, 15, 1293–1302, 2014.

Yilmaz, M. T., Crow, W. T., Anderson, M. C., and Hain, C.: An objective methodology for merging satellite- and model-based soil moisture products, Water Resources Research, 48, https://doi.org/10.1029/2011WR011682, 2012.

Zacharias, S., Bogena, H., Samaniego, L., Mauder, M., Fuß, R., Pütz, T., Frenzel, M., Schwank, M., Baessler, C., Butterbach-Bahl, K., Bens, O., Borg, E., Brauer, A., Dietrich, P., Hajnsek, I., Helle, G., Kiese, R., Kunstmann, H., Klotz, S., Munch, J. C., Papen, H., Priesack, E., Schmid, H. P., Steinbrecher, R., Rosenbaum, U., Teutsch, G., and Vereecken, H.: A network of terrestrial environmental observatories in Germany, Vadose Zone Journal, 10, 955–973, 2011.

Zhang, D. and Zhou, G.: Estimation of soil moisture from optical and thermal remote sensing: A review, Sensors, 16, 2016.

Zhang, D., Liu, X., Bai, P., and Li, X.-H.: Suitability of satellite-based precipitation products for water balance simulations using multiple observations in a humid catchment, Remote Sensing, 11, 2019a.

Zhang, R., Kim, S., and Sharma, A.: A comprehensive validation of the SMAP Enhanced Level-3 Soil Moisture product using ground measurements over varied climates and landscapes, Remote Sensing of Environment, 223, 82–94, https://doi.org/10.1016/j.rse.2019.01.015, 2019b.

Zhang, X., Zhang, T., Zhou, P., Shao, Y., and Gao, S.: Validation analysis of SMAP and AMSR2 soil moisture products over the United States using ground-based measurements, Remote Sensing, 9, 2017.

Zhang, Y., Viney, N. R., Chiew, F. H. S., van Dijk, A. I. J. M., and Liu, Y. Y.: Improving hydrological and vegetation modelling using regional model calibration schemes together with remote sensing data, in: 19th International Congress on Modelling and Simulation, 12–16 December, 2011, pp. 3448–3454, Perth, Australia, 2011.

Zreda, M., Desilets, D., Ferré, T. P. A., and Scott, R. L.: Measuring soil moisture content non-invasively at intermediate spatial scale using cosmic-ray neutrons, Geophysical Research Letters, 35, https://doi.org/10.1029/2008GL035655, 2008.

Zreda, M., Shuttleworth, W. J., Zeng, X., Zweck, C., Desilets, D., Franz, T., and Rosolem, R.: COSMOS: the COsmic-ray Soil Moisture Observing System, Hydrology and Earth System Sciences, 16, 4079–4099, 2012.

Table 1. The 18 soil moisture products evaluated in this study. For the single-sensor satellite products, the spatial ~~resolution~~ sampling represents the footprint size and the temporal ~~resolution~~ sampling the average revisit time. Acronyms: A = ascending; D = descending; PMW = passive microwave; AMW = active microwave; P = precipitation; DA = data assimilation.

Acronym	Details	Spatial ~~resolution~~ sampling	Temporal ~~resolution~~ sampling	Temporal coverage	Latency	Refer
Satellite products						
AMSR2[a]	AMSR2/GCOM-W1 LPRM L3 V001 (soil_moisture_x); single-sensor PMW product; only D passes	∼47 km	~~Daily~~ 1–3 days	2012–present	~~Several hours~~ ∼1.5 days	Parinu
ASCAT[a]	Combination of H115 and H116; single-sensor AMW product; A and D passes	∼30 km	1–2 days	2007–present	~~Several hours~~ 2–4 months	Wagn b)
SMAPL3E[a]	SPL3SMP_E.003 L3 Enhanced Radiometer EASE-Grid V3; single-sensor PMW product; A and D passes	∼30 km	1–3 days	2015–present	~~Several hours~~ ∼2 days	Entek (2018
SMOS[a]	L2 User Data Product (MIR_SMUDP2) V650; single-sensor PMW product; A and D passes	∼40 km	1–3 days	2010–present	~~Several~~ ∼12 hours	Kerr e
ESA-CCI[a]	ESA-CCI SM V04.4 COMBINED; multi-sensor merged AMW- and PMW-based product derived from AMSR2, ASCAT, and SMOS	0.25°	Daily	1978–2018	About a year	Dorig (2019
MeMo	Multi-sensor merged PMW product derived from ~~ASCAT~~AMSR2, SMAPL3E, and SMOS with SWI filter	0.1°	3-hourly	2015–present	~~Several~~ ∼12 hours	This s
Open-loop models (i.e., without data assimilation)						
ERA5-Land	Volumetric soil water layer 1 (0–7 cm); H-TESSEL model; forced with ERA5 P (Hersbach et al., 2020)	0.1°	Hourly	1979–2020	~~Several~~ 2–3 months	C3S (
GLDAS-Noah	GLDAS_NOAH025_3H.2.1 (SoilMoi0_10cm_inst) forced with GPCP V1.3 Daily Analysis P (Huffman et al., 2001)	0.25°	3-hourly	1948–2020	~~2–3~~ ∼4 months	Rodel
HBV-ERA5	HBV forced with ERA5 P (Hersbach et al., 2020)	0.28°	3-hourly	1979–2020	~~Several months~~ ∼6 days	This s
HBV-IMERG	HBV forced with IMERGHHE V06 P (Huffman et al., 2014, 2018)	0.1°	3-hourly	2000–present	~~Several~~ ∼3 hours	This s
HBV-MSWEP	HBV forced with MSWEP V2.4 P ~~(?)~~ (Beck et al., 2019b)	0.1°	3-hourly	2000–present	~~Several~~ ∼3 hours[b]	This s
VIC-PGF	Layer 1 (0–30 cm) of VIC forced with PGF (Sheffield et al., 2006)	0.25°	Daily	1950–2016	Several years	He et
Models with satellite data assimilation						
ERA5	ECMWF ERA5-HRES reanalysis layer 1 (0–7 cm); ASCAT soil moisture DA	0.28°	Hourly	1979–2020	~~Several months~~ ∼6 days	Hersb
GLEAM	GLEAM V3.3a surface layer (0–10 cm); MSWEP V2.2 P forcing; ESA-CCI DA	0.25°	Daily	1980–2018	6–12 months	Marte
HBV-ERA5+SMAPL3E	HBV forced with ERA5 P; SMAPL3E DA	0.1°	3-hourly	~~1979–2020~~ 2015–2020	~~Several months~~ ∼6 days	This s
HBV-IMERG+SMAPL3E	HBV forced with IMERG P; SMAPL3E DA	0.1°	3-hourly	~~2000–present~~2015–present	~~Several hours~~ ∼2 days	This s
HBV-MSWEP+SMAPL3E	HBV forced with MSWEP P; SMAPL3E DA	0.1°	3-hourly	~~1979–present~~2015–present	~~Several hours[b]~~ ∼2 days	This s
SMAPL4	SMAP L4 V4 surface layer (0–5 cm); NASA Catchment model forced with GEOS P corrected using CPC Unified (Chen et al., 2008); SMAP brightness temperature DA	9 km	3-hourly	2015–present	~~2–3~~ ∼2 days	Reich (2019

[a] We also evaluated versions of these products with Soil Wetness Index (SWI) filter (Wagner et al., 1999; Albergel et al., 2008) with the time lag constant T set to 5 days.
[b] At a latency of hours, MSWEP does not include daily gauge corrections and is therefore of lower quality. The data evaluated here have an effective latency of several days.

Table 2. Median Pearson correlations (R) between *in situ* measurements and retrievals from ascending and descending overpasses for the single-sensor soil moisture products (Table 1). The approximate local solar ~~times~~ time (LST) of the overpasses ~~are~~ is reported ~~between~~ in parentheses. Probability (p) values were determined using the Kruskal and Wallis (1952) test. A small p-value indicates that the difference in median R is unlikely to be due to chance.

	Correlation (R)		
Product	Ascending (LST)	Descending (LST)	p-value
AMSR2	0.40 (13:30)	0.50 (01:30)	0.000
ASCAT	0.41 (21:30)	0.47 (09:30)	0.000
SMAPL3E	0.65 (18:00)	0.65 (06:00)	0.643
SMOS	0.49 (06:00)	0.48 (18:00)	0.271

Table A1. The measurement networks part of the ISMN archive from which we have used *in situ* soil moisture data.

Network	Reference(s) or website
ARM	www.arm.gov
BIEBRZA	www.igik.edu.pl
BNZ-LTER	Van Cleve et al. (2015)
COSMOS	Zreda et al. (2008, 2012)
CTP	Yang et al. (2013)
DAHRA	Tagesson et al. (2015)
FMI	http://fmiarc.fmi.fi
FR	www.inrae.fr
HOBE	Kang et al. (2014); Jin et al. (2014)
HYDROL-NET	Morbidelli et al. (2014)
iRON	Osenga et al. (2019)
LAB-net	Mattar et al. (2014)
MySMNet	Kang et al. (2016)
ORACLE	https://gisoracle.inrae.fr
OZNET	Smith et al. (2012)
REMEDHUS	http://campus.usal.es/~hidrus/
RISMA	Ojo et al. (2015)
RSMN	http://assimo.meteoromania.ro
SCAN	www.wcc.nrcs.usda.gov
SMOSMANIA	Calvet et al. (2007); Albergel et al. (2008)
SNOTEL	www.wcc.nrcs.usda.gov
SOILSCAPE	Moghaddam et al. (2010); Moghaddam et al. (2016)
SWEX	Marczewski et al. (2010)
TERENO	Zacharias et al. (2011)
UDC	Loew et al. (2009)
USCRN	Bell et al. (2013)
VAS	http://nimbus.uv.es
WSMN	Petropoulos and McCalmont (2017)