Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2020-485-AC1, 2020


Hydrology and Earth System Sciences
Open Access
Discussions

# *Interactive comment on* "Satellite soil moisture data assimilation for improved operational continental water balance prediction" *by* Siyuan Tian et al.

**Siyuan Tian et al.**

siyuan.tian@anu.edu.au

Received and published: 24 November 2020

xcolor



## Response to Anonymous Referee 1

November 24, 2020

The paper "Satellite soil moisture data assimilation for improved operational continental water balance prediction" investigates the application of satellite soil moisture for improving a water balance model. While this study can be useful for modelling objectives there are several issues that need to be addressed.

We would like to thank the reviewer for the overall constructive comments on the manuscript. Below is our response to the issues raised in the review.

**Major comments:**

The paper lacks novelty. The applied methodology that is simplified data assimilation (i.e. nudging approach) does not properly take model and data uncertainties into an account. Several more sophisticated approaches have already been published for soil moisture data assimilation.

We thank reviewer for this comment. We respectfully disagree with the suggestion that the paper lacks originality. We do not claim the methods are original. The novelty of this study is in the application of the satellite soil moisture data assimilation in an existing operational water balance framework. We believe that for an existing operational system, a key consideration when choosing data assimilation method is the minimal disruption/modification to the existing system. This was the basis of the proposed method. The applied approach was chosen based on the tests of other methods not deliberately due to the simplicity. The application of more sophisticated approaches will require significant modification and possibly reinvention of the existing continental operational system. And while our approach is simple, we demonstrated its robustness and usefulness through validation against in-situ/satellite observations including surface soil moisture, root-zone soil moisture, evapotranspiration, streamflow and vegetation greenness.

With regard to the comment about not 'properly' taking model and data uncertainty into account, we disagree. The uncertainties between model and observations were determined through Triple Collocation method which is widely used in error characterisation of soil moisture estimates. Furthermore, we would gratefully add citations in the paper about the any other studies of satellite soil moisture data assimilation in the real-time continental operational system that we used, if the reviewer can provide us some examples.

The term "prediction" does not add much since every model-data integration will affect initial states and correspondingly a few time steps of forecasting. Showing that soil moisture assimilation led to different state estimates than the open-loop results, which is very obvious, does not prove anything. Authors may put more efforts in validating the results against various independent data over the forecasting period. This could more interesting if a calibration scheme was used to improve the model parameters.



We agree with the reviewer that one expects to see the difference in the forecasting results for a few time steps as a result of data assimilation modifying the initial states. However, we believe that quantifying how long these differences persist as a result of data assimilation is not well studied. Showing how long will the impact of initial states last in the system is important for identifying the potential for forecasting the flood and drought impacts and agricultural production. The forecasting driven by rainfall forecasts data will be used in our next study together with the validation against various data over forecasting period. We will ensure in the revised manuscript that this is not about forecast error, rather quantifying the persistence of the constraint.

The model parameters were calibrated offline and an optimal set of parameters based on historical satellite soil moisture from AMSR-E, and in-situ ET and streamflow observations are used in the operational system. Further model calibration is out of the scope of this study, but we acknowledge that this may be required in light of the finding of this study.

The two-step method should be better explained, especially for the second step that deals with the mass conservation constraint. This part is very unclear and requires more details. It is not clear how authors check for water balance after the first step. I am not sure how accurate is to simply distribute the correction (which is not clear how it can be estimated) to other states (and why only these states?).

The analysis increment redistribution is based on the well established tangent linear modelling (TLM). We applied TLM to all model equations. We described only concept of tangent linear modelling in the manuscript since including all the equations in the manuscript is unpractical. All the original model equations can be found in the Van Dijk (2010) and Frost et al. (2018). And if the Editor deems it helpful, we would happily include all the TLM equations as supplementary materials.

The paper lacks thorough background research. There are several other highly related manuscripts to the topic that seem to be missed. These sources could provide a better background and existing knowledge.

We agree with the reviewer that there are plenty of papers related to soil moisture data assimilation. We have included many that are relevant to our arguments and that we are aware of. If however the reviewer can provide us with some specific suggestions regarding existing operational soil moisture data assimilation systems, we would gratefully include them in the revised manuscript.

Line 105-110: Please explain how did you interpolate soil moisture observations into 0.05 degree scale.

We thank the reviewer for pointing this out. In the revised manuscript we will include the following statement:

"Available swath data for each product covering Australia were collated for each 24-hour period approximating the AWRA-L operational time steps and resampled to a regular 0.05-degree grid across the modelling domain using linear interpolation from 2015 to 2019."

Line 115-120: "we derived a set of coefficients for the rescaling by sampling modelled and SSM data from cells surrounding the gaps", How? Details are required.

We thank the reviewer for this comment. In the revised manuscript, we will provide the



following explanation:

"To use SSM products to fill the modelling gap in gauge-sparse region of the continent, we derived a set of coefficients for the observation operator from the cells surrounding the gaps. Specifically, we obtained the maximum SSM values through time and the derived 'slope' and 'intercept' from the observation model for each cell in neighboring region. We applied linear regression to estimate the correspond 'slope' and 'intercept' from the maximum SSM values in the rainfall gaps. This provided a transformation of the SSM into water storage unit (mm) and ensures the assimilation can effectively influence the spatial pattern of soil moisture over the sparsely gauged regions."

Line 135-140: Is not more appropriate to use NDVI to evaluate top layer soil moisture than root-zone? NDVI supposedly better reflects surface soil variations than the rootzone.

We respectfully disagree with the reviewer on this comment. Root-zone soil water availability is the controlling factor for vegetation growth in arid and semi-arid areas. Top-soil (0-5 cm) moisture is not as strongly related to vegetation response as deeper soil water. We chose the modelled root-zone soil moisture (0-1m) over croplands as an indirect evaluation is because the time lag between soil moisture and vegetation response are normally within one month.

Line 155-160: How one can derive Q for different datasets? More details are needed.

We suspect the reviewer is not aware of how triple collocation has been used to infer data error variances. Q here denotes the temporal variance and covariance between three data sets. The triple collocation approach uses these temporal variance $Q_{x,x}$, and covariance$Q_{x,y}$ to infer error variances of the three datasets. We would happily add more references if the reviewer thinks it would help, however we do not think going into further detail about triple collocation is necessary given the wealth of literature on the subject.

Line 170: Very unclear, please revise. Is not S0 the top soil layer? If yes, what do you mean by "soil water storage in S0 for shallow-rooted vegetation and deep-rooted vegetation at surface layer"?

We thank the reviewer for the comment. We mentioned in Section 2.1 Line 92, the soil water storage in each layer is simulated separately for two hydrological response units: shallow-rooted vegetation (grass) and deep-rooted (trees) vegetation. In the revised manuscript, we will clarify it as follow:

"The observation operator ***H*** here is the aggregation of soil water storage estimates in the top-soil layer for two land cover types, i.e. shallow-rooted vegetation and deep-rooted vegetation."

Equation 3 should be better explained when it comes to having more than one observation.

We thank the reviewer for the comment. We can revise the equation with two observation data as below in the revised manuscript as below:

$$k_x = \frac{\frac{1}{\sigma_x^2}}{\frac{1}{\sigma_x^2}+\frac{1}{\sigma_y^2}+\frac{1}{\sigma_z^2}}$$



where x, y, z denotes AWRA-L estimates, SMAP and SMOS soil moisture retrievals.

Line 185-190: Do you mean that instead of calculating and correcting water balance residuals, you distribute S0 increments? I am not sure if this is a correct approach.

We understand the reviewer's confusion here. Effectively what we are calling 're-distribution' is correcting the residuals. The model itself is a water balance model which accounts water balance in the next model step. However, the data assimilation breaks the water balance by reducing the misfit between the model estimates and observations. By distributing S0 increments through the tangent linear modelling, the water balance is maintained after assimilation.

For Section 4.2 authors could use independent evaporation and runoff data to better validate the results.

The independent in-situ ET and streamflow data are used in Section 4.2. The results are shown in Figure 7c and 7d. Section 4.2 focuses on the change in spatial pattern for each grid, since the in-situ ET and runoff observations are limited. The results of independent validation with in-situ data are explained in Section 4.3.

**Minor comments**:

I am not sure whether this is the journal policy or authors' decision but it'd much easier if every line of the manuscript has a line number for the sake of review.

We can include the line number for every line in the revised manuscript.

Lines 225-230: Can you think of any reason behind "missing or underestimated rainfall events", which seems to be large.

As mentioned in Section 2.1 and Line 97, the rainfall forcing used in this operational modelling system is a gridded rainfall derived through interpolating gauge measurements at point scale. The uncertainty of rainfall is limited in regions with insufficient coverage.

Line 255-260: Have you applied any tests of statistical significance?

Yes. The Fig. 1 below demonstrated the change in correlations for surface soil moisture estimates after data assimilation comparing to model open-loop with a 95% confidence level plotted in dashed line.

**Reference**

Frost, A.J., Ramchurn, A. and Smith, A., (2016). The bureau's operational AWRA landscape (AWRA-L) Model. Bureau of Meteorology Technical Report.

van Dijk, A.I.J.M. (2010). AWRA Technical Report 3, Landscape Model (version 0.5) Technical Description, WIRADA, Canberra: CSIRO Water for a Healthy Country Flagship.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2020-485, 2020.



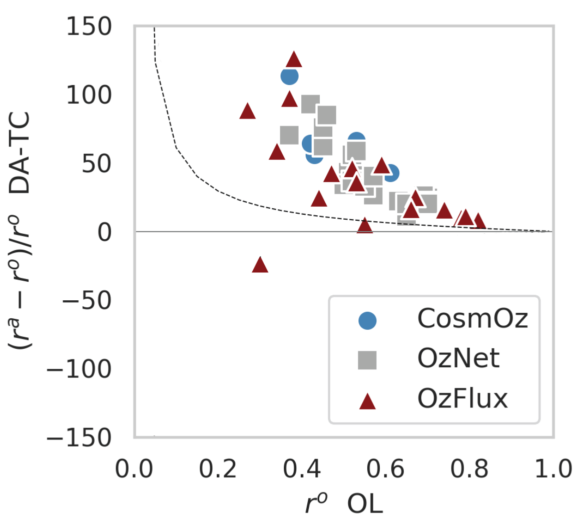


**Fig. 1.** Relative changes in correlations with in-situ surface soil moisture after data assimilation (rˆa) against model open-loop (rˆo)