Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2020-485-AC2, 2021

Hydrology and Earth System Sciences
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: 8 January 2021

xcolor

Response to Anonymous Referee 2

January 8, 2021

The authors present in their manuscript an application of assimilating SMAP and SMOS soil moisture into the AWRA-L hydrological model. The innovation of this manuscript lies in the development of a two-step data assimilation approach. In the first step, model states are updated using a Kalman filter type approach whereby error covariances are obtained through triple collocation. The second step is to mitigate the mass balance error created by the data assimilation through what the authors named the Analysis Increment Redistribution approach. The topic is relevant for reader of HESS. The manuscript is generally well written and methodology and results are well explained. I believe the manuscript can be considered for publication after consideration of the following comments .

We would like to thank the reviewer for the thoughtful comments and suggestions. We will revise the manuscript based on the reviewer's comments. Please see below for our detailed responses to all the comments.

General comments:

Even though the manuscript is well written general, I found still a number of grammar mistakes. Several of them I have indicated in the specific comments below, but I would recommend the authors to check the manuscript carefully again.

Thank you. In the revise manuscript, we will check the manuscript thoroughly.

* In their DA approach the author assume that the error (co-)variance are temporally constant, while there is ample evidence that this is reality not the case. For instance, due varying sensing depths as a function of the soil moisture content itself. In the discussion section the author mention this as point of improvement for the future, but I would appreciate if the authors could introduce this assumption early in the manuscript.

We agree with the reviewer that the assumption of temporally constant error variance is a simplification. Ideally we would have calculated seasonally varying error variances to account for the variations in surface soil moisture. However, the derived variances would have been based on too few data for the TC approach to yield good quality statistics given the relatively short time period (2016-2019). As the number of remotely sensed data increase with time, a temporally varying error is certainly a consideration in future refinements to the method. We will add the following statement in the method section in L181 in the revised manuscript.

"Here, we applied the triple collocation approach (Section 3.1) to characterise the temporal error variances of the model estimates and the two satellite observations for each grid cell across Australia. Given the relatively short time series (small number) of observations, however, a single set of error variances is calculated for all time. This results in spatially varying but temporally static error variances (and thus gain weights) for each of the three sources (Fig. 2). We acknowledge the limitations of assuming a

temporally constant error variances and future refinements to the assimilation method will consider introducing seasonally error variances."

Section 4: Results * When presenting your assimilation results figure 4 and onwards, do you only present the results with assimilation of SMAP observations? It would be interesting to see also the results for the assimilation of SMOS to get an idea about the what the effect of observation uncertainty is on the analysis results.

We thank the reviewer for the opportunity to clarify. Figures 4 onwards display the results of assimilating both SMAP and SMOS. We will clarify this in the Figure caption and abstract as follow:

Figure 4: Comparison of daily average surface soil water storage estimates (S_0) for December 2019 from (a) model open-loop (OL), (b) joint assimilation of SMAP and SMOS with Triple Collocation (DA-TC) and (c) difference between estimates DA-TC and OL.)

Abstract: *"In this study, we assimilate satellite soil moisture retrievals from both SMAP and SMOS missions simultaneously into the Australian Water Resources Assessment Landscape model (AWRA-L) using the proposed framework and evaluate its impact on the model's accuracy against in-situ observations across water balance components."*

Section 4.3: * Differences in root zone soil moisture, ET and streamflow after DA are actually quite small, while in figure 8 there is still as substantial difference between the observed and simulation streamflow. I would expect more discussion here on how this gap in streamflow between model and observation can be closed. Can this be done with soil moisture assimilation?

Thank you for this comment. We agree with the reviewer that the change in the root-zone soil moisture, ET and streamflow after DA appear marginal for the locations of in-situ monitoring sites, as mentioned in L260. One reason is due to the limited numbers of in-situ monitoring sites (as shown in Figure 1d). Changes in those components across the continent can be seen in Figure 6 a-c. The soil moisture assimilation alone cannot address the disparity between modelled and observed streamflow. The difference suggest a need for re-calibration of AWRA-L against streamflow for this catchment. Nevertheless it is encouraging to see (in Fig. 8) that soil moisture data assimilation, particularly after AIR, does 'nudge' the model estimates towards the streamflow observations. We will add the following discussion in L273 when revising the manuscript:

"The negative streamflow analysis increment (Fig. 8) indicates that water is removed from the surface water store after the assimilation of SSM and application of AIR, which is appears to compensate for the overall overestimate of OL simulations, in this example. Although the change in streamflow due to the soil moisture data assimilation is small compared to the disparity between model and observed streamflow, the adjustment in the direction towards observations highlights the importance of accurate antecedent soil moisture conditions in the simulation of runoff response. The joint assimilation of gauge-measured streamflow and satellite soil moisture retrievals into AWRA-L is expected to improve the streamflow simulation."

* How do you explain that the correlation between the AWRA-L root zone soil moisture and NDVI improves, while the correlation with the root zone soil moisture measurements do not improve (see box plots)?

Thank you for the question. We believe the reasons that the improvements in root-zone moisture are better illustrated in NDVI compared to the in-situ data are due to (i) the limited number of in-situ sites (less than 30 root-zone soil moisture monitoring sites available across Australia); (ii) the scale disparity of the point measurements and modelling grid cell; and (iii) the model open-loop already performs reasonable well, with average correlation > 0.8. Thus, we proposed the indirect verification of root-zone soil moisture with NDVI at grid cell scale. We will further clarify this in L277 as below:

"The limited number of root-zone soil moisture monitoring sites and the large spatial disparity between in-situ measurements at point scale and modelling resolution (5km grid cell) scale are the substantial limitations for wide-area evaluation of root-zone soil moisture estimates."

Section 4.4 * The authors evaluate the persistence of data assimilation through comparison of the open loop and DA-TCAIR. Could the authors also include the DA-TC in this analysis? I would be interested to see what AIR in itself does to the persistence of the soil moisture data assimilation. This would potentially also support the use of DA-TCAIR over DA- TC.

Thank you for the suggestion. The results of DA-TC are shown in Fig.1 below. In comparison with manuscript Figure 10, you will note that the results of DA-TCAIR and DA-TC for the upper layer soil water storage are the same (perhaps obvious since AIR does not change top-soil layer). However, DA-TCAIR does change Ss, ET and Qtot, which can be seen to be quite different to the DA-TC results below. To save space, we suggest including the figure below as supplementary materials, however we will leave the final decision to the discretion of the Editor.

Specific comments: Abstract: I would suggest to specify the following in the abstract * the name of the soil moisture product assimilated * the method of state updating

Thank you for the suggestion. We will revise the abstract as below:

A simple and effective two-step data assimilation framework was developed to improve soil moisture representation in an operational large-scale water balance model. The first step is the Kalman filter type sequential state updating process that exploits temporal covariance statistics between modelled and satellite-derived soil moisture to produce analysed estimates. The second step is to use analysed surface moisture estimates to impart mass conservation constraints (mass redistribution) on related states and fluxes of the model using tangent linear modeling theory in a post-analysis adjustment after the state updating at each time step. In this study, we assimilate satellite soil moisture retrievals from SMAP and SMOS missions simultaneously to the Australian Water Resources Assessment Landscape model (AWRA-L) using the proposed framework and evaluate its impact on the model's accuracy against in-situ observations across water balance components. We show that the correlation between simulated surface soil moisture and in-situ observation increases from 0.54 (open-loop) to 0.77 (data assimilation). Furthermore, indirect verification of root-zone soil moisture using remotely sensed vegetation time series across cropland areas results in significant improvements of 0.11 correlation units. The improvements gained from data assimilation can persist for more than one week in surface soil moisture estimates and one month in root-zone soil moisture estimates, thus demonstrating the efficacy of this data assimilation framework.

L15: Could the authors provide also correlation coefficients for the comparison of the root zone soil moisture and vegetation time series? Instead of only the increment.

Thank you for the suggestion. In L287, we included that the correlation of root zone soil moisture from DA-TC with NDVI is 0.55 on average, while the correlation of DA-TCAIR is 0.66. We will include a averaged time series plot for a small region (15km x 20km, origin:(-36.25, 142.15)) as an example in the revised manuscript, since taking the average of all the cropland grid cells across the continent is not appropriate.

The revised figure is shown in Fig.2. The correlations of the region are 0.47, 0.68 and 0.74 for the open-loop, DA-TC and DA-TCAIR respectively.

L41: 'As the assimilation .. ' Sentence seems incomplete.

Thank you. We will revise the sentence as follows:

The assimilation of remotely sensed soil moisture or total water storage data may lead to undesired impacts on groundwater or evapotranspiration simulations due to the mass imbalance or random error covariances (Girotto et al., 2017;Tangdamrongsub et al., 2020;Tian et al., 2017).

P2L45: check sentence.

Thank you. We will revise the sentence as below:

"However, studies considering mass conservation in data assimilation often require extra data sources such as evapotranspiration and runoff as constraints or without considering the fluxes in the redistribution (Li et al., 2012;Pan and Wood, 2006)."

L61: replace 'has' by 'have'

Done.

L65: Could the authors explain why this limits the operational use?

Thank you. We will clarify this sentence when revising the manuscript as below:

"However, unlike the aforementioned systems where data assimilation is inherent in the system design, many operational water balance models, or catchment hydrology models, are calibrated to observations a priori. *Including data assimilation as an afterthought restrains the flexibility of the system, thereby limiting the complexity of the data assimilation scheme for operational use."*

L85-87: Please add references in support of this statement

Thank you. We will add references for this statement as below:

"The outputs from the operational AWRA-L has been widely used in various agricultural applications and natural resources risk assessment and planning, including commodity forecasting, irrigation scheduling, flood and drought risk analysis, as well as flood forecasting (Frost et al., 2018; Hafeez et al., 2015; Nguyen et al., 2019; van Dijk and Renzullo, 2011; van Dijk et al., 2013)."

L111: change Figure1 to Figure 1

Done.

L115: What do the authors mean by 'dynamics' and it is unclear why this would flatten to zero as a result of mean and variance matching.

Thank you for this comment. We meant that the mean and variance values of modelled soil moisture over gauge-sparse areas such as Western Australia are zeros or close to zero. Matching the satellite soil moisture (SSM) to model mean and variance for those regions will make the SSM to zero. We will clarify this as below in the revised manuscript:

"For regions with sparse rain-gauge coverage such as central Western Australia (Fig1.c), AWRA-L modeled S0 persists as zeros or very low values for the experiment period, reflecting a deficiency in the gauge-based analysis of daily rainfall used to drive model simulations.The result of mean and variance matching in these gauge-sparse areas will flatten the variability of SSM time series to zero when using values of the modelled S0 for these areas directly."

L116: The coefficients of what are derived? More explanation is needed here.

We will include the following explanation about the coefficients following the above mentioned response.

"To resolve this problem, and fully leverage the information available in the SSM products to fill the gaps in modelled outputs across the continent, we derived a set of coefficients for the mean and variance matching over the gauge sparse regions by

sampling modelled and SSM data from cells surrounding the gaps."

L139: change 'were' to 'was'

Done

Eq. 1. The letter Q is used for the variance while sigma2 also indicate this. Could the authors explain this? Should the reader interpret this both as variances?

Thank you for pointing out this potential confusion. However, we mentioned in L154-155 that Q denotes the temporal variance of the time series, while σ^2 refers to the error variance in the data.

L165: Why do the authors make this statement because they apply mean and variance matching to suppress the systematic differences between the observations and simulations.

The reviewer is correct. We did apply mean and variance matching as a way of suppressing systematic differences. It so happens that this transformation also provides the state space-to-observation mapping required for data assimilation. Here we simply explained the role observation operator here to the readers as a key component of data assimilation.

L182: Why do they authors refer to Crow and Van den Berg (2010) here? If they have used TC as method to derive uncertainty levels I would have expected the reference earlier in the manuscript.

We have the citation in the Introduction (L73) along with others. However we agree with the reviewer that the citation is warranted earlier and we will move it to the beginning of the method section in L146 as below:

"It was first applied to near-surface wind data (Stoffelen, 1998) and later extensively applied to soil moisture (Chen et al., 2018;Crow and Yilmaz, 2014;Crow and Van den Berg, 2010; Dorigo et al., 2017;McColl et al., 2014;Scipal et al., 2008;Su et al.,2014b;Yilmaz and Crow, 2014;Zwieback et al., 2013) and rainfall (Alemohammad et al., 2015;Massari et al., 2017)."

L195-205: How do the authors obtain dM/dx? Is this a fixed value or a quantity that is updated every time step?

The dM/dx is derived from on the AWRA model equations for each state variable that related to the S0. This is standard approach of tangent linear and adjoint modelling. The value of delta x (analysis increment) is updated every time step. The equations of dM/dx are fixed however their values change with every analysis increment.

Figure 3: Could the authors add a time series of the measured soil moisture to this figure.

Yes. We will include the in-situ measurements in this figure as shown in Fig.3.

L226: Could the author indicate where the Murray-Darling Basin is? Readers not familiar to the continent may not know where it is.

Yes. We will include the boundary of the Murray-Darling as shown in Fig.4.

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

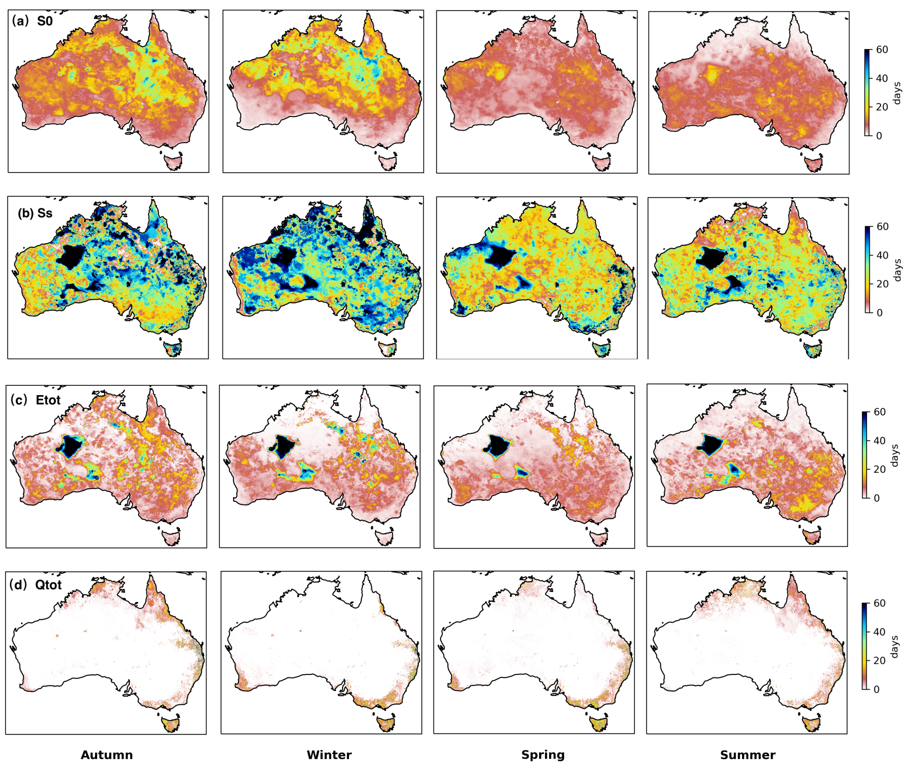

Fig. 1. Quantified impacts of data assimilation on forecasting AWRA-L state variables using the initial condition from DA-TC:

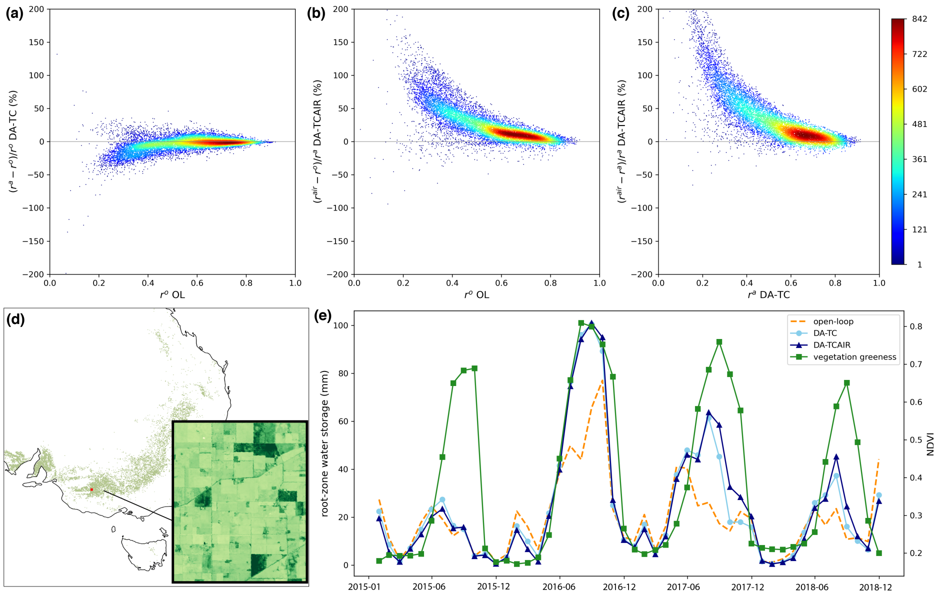

Fig. 2. (a-c) Comparison of correlations between vegetation greenness (NDVI) with AWRA-L modelled root-zone soil moisture over cropland (a-c). (e) Time series of NDVI and root-zone soil water storage in (d)

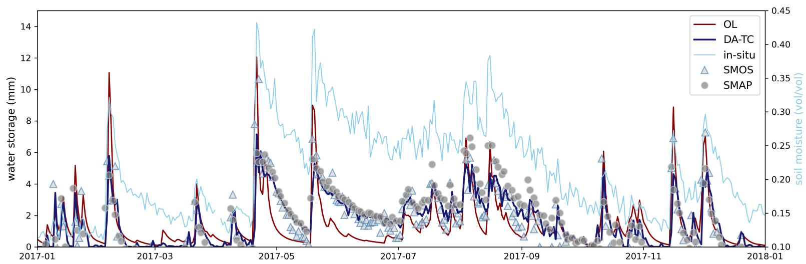

Fig. 3. Time series of AWRA-L surface soil water storage estimates from open-loop (OL) compared to estimates after data assimilation (DA-TC) of SMAP and SMOS soil moisture retrievals at CosmOz monitoring site

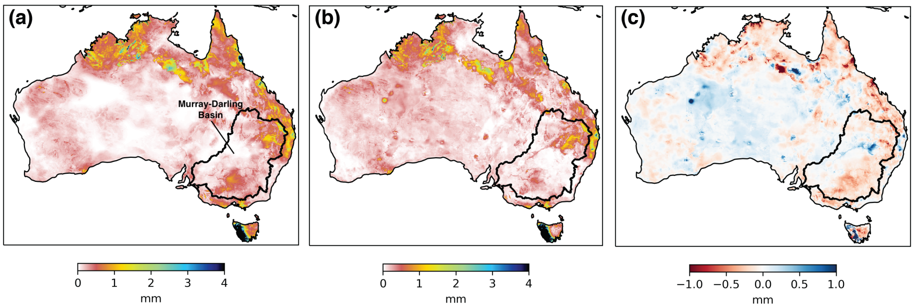

Fig. 4. Comparison of daily average surface soil water storage estimates (S0) for December 2019 from (a) OL, (b) DA-TC and (c) difference between DA-TC and OL