

# Operational soil moisture data assimilation for improved continental water balance prediction

Siyuan Tian<sup>1</sup>, Luigi John Renzullo<sup>1</sup>, Robert C. Pipunic<sup>2</sup>, Julien Lerat<sup>2</sup>, Wendy Sharples<sup>2</sup>, and Chantal Donnelly<sup>2</sup>

<sup>1</sup>Australian National University

<sup>2</sup>Bureau of Meteorology

November 21, 2022

## Abstract

A simple and efficient method was developed to improve soil moisture representation in an operational water balance model through satellite data assimilation. The proposed method exploits temporal covariance statistics between modelled and satellite-derived soil moisture to produce analysed estimates, as a weighted combination of all data sources. We demonstrate the application of the method to the Australian Water Resources Assessment (AWRA) model and evaluate the accuracy of the approach against in-situ observations across the water balance. The correlation between simulated surface soil moisture and in-situ observation is increased from 0.54 (open-loop) to 0.77 (data assimilation). We suggest an approach to use analysed surface moisture estimates to impart mass conservation constraints on related states and fluxes of the AWRA model in a post-analysis adjustment. 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.

1 **Operational soil moisture data assimilation for improved continental water balance**  
2 **prediction**

3  
4

5 Siyuan Tian<sup>1\*</sup>, Luigi J. Renzullo<sup>1</sup>, Robert C. Pipunic<sup>2</sup>, Julien Lerat<sup>2</sup>, Wendy Sharples<sup>2</sup>, Chantal  
6 Donnelly<sup>2</sup>

7 <sup>1</sup> Fenner School of Environment & Society, The Australian National University, ACT, 2601, Australia

8 <sup>2</sup> Water Program, Bureau of Meteorology, Australia

9

10 Corresponding author: Siyuan Tian ([siyuan.tian@anu.edu.au](mailto:siyuan.tian@anu.edu.au))

11

12 **Key Points:**

- 13 • We develop a simple and efficient method to improve operational soil water balance  
14 model through satellite data assimilation.
- 15 • We suggest an approach to use analyzed surface soil moisture estimates to impart mass  
16 conservation constraints on related states and fluxes
- 17 • The impact of the satellite data assimilation on model estimates of soil moisture can  
18 persist for several weeks.

19

20

21

22

23 **Abstract**

24 A simple and efficient method was developed to improve soil moisture representation in an  
25 operational water balance model through satellite data assimilation. The proposed method  
26 exploits temporal covariance statistics between modelled and satellite-derived soil moisture to  
27 produce analysed estimates, as a weighted combination of all data sources. We demonstrate the  
28 application of the method to the Australian Water Resources Assessment (AWRA) model and  
29 evaluate the accuracy of the approach against in-situ observations across the water balance. The  
30 correlation between simulated surface soil moisture and in-situ observation is increased from  
31 0.54 (open-loop) to 0.77 (data assimilation). We suggest an approach to use analysed surface  
32 moisture estimates to impart mass conservation constraints on related states and fluxes of the  
33 AWRA model in a post-analysis adjustment. The improvements gained from data assimilation  
34 can persist for more than one week in surface soil moisture estimates and one month in root-zone  
35 soil moisture estimates.

36

37 **Plain Language Summary**

38 The access to accurate daily continental soil water balance predictions is valuable for water  
39 management practitioners, policy makers and researchers in support of water resources  
40 assessment and agriculture planning. This study develops a simple and robust method for an  
41 operational water balance model to incorporate satellite soil moisture products for improved  
42 accuracy and spatial representation of soil water storage predictions. The integration of satellite  
43 soil moisture products can provide persistent constraints in model predictions for up to several  
44 weeks.

45

## 46 **1 Introduction**

47 Accurate estimation of soil moisture is fundamental to monitoring and forecasting water  
48 availability and land surface conditions under extreme events such as droughts, heatwaves and  
49 floods (Ines et al., 2013; Sheffield and Wood, 2007; Tian et al., 2019a; Wanders et al., 2013).  
50 The assimilation of satellite soil moisture into land surface and hydrology models has been  
51 repeatedly demonstrated to improve model representation of soil water dynamics (Draper et al.,  
52 2012; Kumar et al., 2009; Pipunic et al., 2008; Reichle and Koster, 2005; Renzullo et al., 2014;  
53 Tian et al., 2017; Tian et al., 2019b). Soil moisture is the linchpin between atmospheric fluxes,  
54 surface- and ground-water hydrology, thus it is important that any changes in modelled estimates  
55 are not detrimental to other components of the water balance.

56 Soil moisture anomalies can persist for months (Vinnikov et al., 1996), but the spatial pattern can  
57 vary significantly due to the heterogeneous spatial distribution of rainfall and variability in soil  
58 properties, land cover type and topography. Due to this large spatial variability of soil moisture,  
59 the utility of ground-based, point-scale measurements is limited. Soil moisture estimates from  
60 land surface models are adversely affected by the uncertainties of atmospheric forcing, model  
61 dynamics and model parameterization. Remotely sensed data can provide spatially and  
62 temporally varying constraints on the modelling of biophysical landscape variables that are often  
63 superior to that achieved by a single static set of model parameters. Data assimilation merges  
64 models and observations in a way that compensates for the deficiencies in each (e.g. uncertainty,  
65 coverage), resulting in improved accuracy, coverage, and ultimately forecasting capability.

66 Methods of assimilation are many and varied, however commonalities exist between them. These  
67 commonalities are such, that for any time step, the time integrated first guess (the forecast) of  
68 soil moisture states are adjusted by an amount determined by the difference between observed  
69 and modelled soil moisture (the innovation), which is weighted by the respective error variances  
70 of modelled and observed quantities (the gain), to generate revised soil moisture states (the  
71 analysis). At this point, the model soil moisture states are out of balance with the other stores and

72 fluxes, until the model integrates forward to the next time step, whereupon water balance is  
73 restored through model physics.

74 In addition to water balance closure, from an operational perspective, is it important that the  
75 method of data assimilation be: computationally efficient for routine, automated simulation over  
76 the whole model domain; robust to data gaps; and make lasting positive improvements to future  
77 predictions of soil water stores and fluxes. Currently, there are very few operational continental  
78 land surface modelling systems that provide high-resolution near-real time soil moisture  
79 estimates that have been constrained through the assimilation of satellite observations. Some  
80 recent examples include surface soil wetness observations from Advanced Scatterometer  
81 (ASCAT) active radar system, on the meteorological operational satellite (MetOp), been  
82 assimilated into Unified Model (Davies et al., 2005) through nudging to provide soil moisture  
83 analysis at 40 km globally (Dharssi et al., 2011). Additionally, ASCAT data are used in the  
84 ECMWF (European Centre for Medium-Range Weather Forecasts) Land Data Assimilation  
85 System through a simplified Extended Kalman Filter approach (De Rosnay et al., 2013) to  
86 provide near-real time surface soil moisture and root-zone soil moisture at 25-km resolution  
87 globally. However, soil moisture products from a passive radiometer system such as SMOS (Soil  
88 Moisture and Ocean Salinity) mission (Kerr et al., 2001) or the SMAP mission (Entekhabi et al.,  
89 2010) have not been fully explored in an operational data assimilation system.

90 In this study, we develop a simple, computationally efficient, and effective data assimilation  
91 approach for assimilating satellite soil moisture products into an operational national water  
92 balance model. We demonstrate the application of the method to the Australian Water Resources  
93 Assessment Community Modelling system (AWRA-CMS), which provides daily water balance  
94 estimates at 5-km resolution across Australia, with the assimilation of satellite surface soil  
95 moisture (SSM) from both SMOS and SMAP. A post-analysis adjustment is proposed to impart  
96 mass conservation constraints on related states and fluxes such as root-zone soil water storage,  
97 evapotranspiration and streamflow thus improving the accuracy of the water balance post  
98 assimilation. The impacts of data assimilation on model predictions is assessed by quantifying  
99 the persistence of the correction to key model components with respect to open-loop simulations.

## 100 **2 Materials and Methods**

### 101 2.1 Australian Water Resources Assessment Community Modelling system (AWRA-CMS)

102 The Australian Water Resources Assessment (AWRA) Community Modelling system (AWRA-  
103 CMS) is a freely available version of the AWRA Landscape model (Van Dijk, 2010) which  
104 simulates the water balance in the Australian landscape ([https://github.com/awracms/awra\\_cms](https://github.com/awracms/awra_cms)).  
105 The operational implementation of the AWRA-CMS by the Australian Bureau of Meteorology  
106 provides daily 0.05 degree (approximately 5 km) national gridded soil moisture, runoff,  
107 evapotranspiration and deep drainage estimates, and underpins the annual national water  
108 resource assessments and water use accounts (Frost et al., 2018) as well as providing situational  
109 soil moisture for flood forecasting, agriculture and other applications. AWRA is a one-  
110 dimensional distributed model that simulates the water balance for each grid cell across the  
111 modelling domain by distributing rainfall into plant-accessible water, soil moisture and  
112 groundwater stores, and removing water through evapotranspiration, runoff and deep drainage.  
113 The soil water column has been partitioned into three layers (upper: 0–10 cm, lower: 10–100 cm,  
114 and deep: 1–6 m) simulated separately for deep- and shallow-rooted vegetation. In addition to  
115 the modelling of soil columns, the model includes a surface water and a groundwater storage that  
116 are simulated at each grid cell and conceptualized as if operating within a small unimpaired  
117 catchment. In this study, we used daily precipitation and air temperature from the gridded  
118 climate data services (Jones et al., 2009), daily solar exposure produced from geostationary  
119 satellites (Grant et al., 2008), and interpolated site-based wind speed (McVicar et al., 2008) as  
120 model forcing inputs.

### 121 2.2 Satellite soil moisture (SSM)

122 To optimize the daily spatial coverage, we used two satellite soil moisture products derived from  
123 passive L-band systems: the Soil Moisture Active-Passive (SMAP) product from NASA  
124 (Entekhabi et al., 2010); and the product from the European Space Agency's (ESA's) Soil  
125 Moisture and Ocean Salinity (SMOS) mission (Kerr et al., 2001). The SMAP product is the  
126 level-2 enhanced radiometer half-orbit 9-km EASE-grid soil moisture (Chan et al., 2018). The  
127 SMOS product is the level-2 soil moisture product on ~ 25-km grid (Rahmoune et al., 2013).  
128 Both SMAP and SMOS produce volumetric soil moisture estimates (units  $\text{m}^3/\text{m}^3$ ) of

129 approximately the upper 5 cm of soil. Available swath data for each product covering Australia  
 130 were sourced and collated for 24-hour period approximating the AWRA-CMS operational time  
 131 steps and interpolated to a regular 0.05-degree grid across the modelling domain from 2015 to  
 132 2019. This provided maximum possible spatial coverage for each data product in representing  
 133 surface soil moisture at the end of the model's time step integration each day.

### 134 2.3 Data assimilation approach through triple collocation (TC)

135 The data assimilation method used here is a time sequential updating of model state(s) given  
 136 observations of relevant model variables (Reichle, 2008). Two key modelling components in  
 137 data assimilation: the *dynamics operator*, which describes the time integration of the system  
 138 states and fluxes, which in this study is the AWRA-CMS; and the *observation operator*, which  
 139 provides the mathematical mapping from state to observation space (or vice versa). The role of  
 140 the observation operator is to perform a mapping between observation and state space, as often  
 141 observations are not directly comparable to model states.

142 The state updating equation for sequential data assimilation is written as:

$$143 \quad X_t^a = X_t^f + K_t[Y_t - H(X_t^f)] \quad (1)$$

144 which says that the best estimate of model state, known as analysis ( $X_t^a$ ), is equal to the first  
 145 guess or forecast ( $X_t^f$ ) plus a weighted difference between observations,  $Y_t$ , and the model  
 146 equivalent to the observation,  $H(X_t^f)$ , for that time step. The multiplier,  $K_t$ , is known as the *gain*  
 147 *factor* which contains uncertainty expressed as error variance for both model estimates ( $\sigma_B^2$ ) and  
 148 observations  $\sigma_R^2$ . For a unity observation operator and assuming independence between model  
 149 estimates and observations, gain factor typically assumes the form:

$$150 \quad K = \frac{\sigma_B^2}{\sigma_B^2 + \sigma_R^2}. \quad (2)$$

151 In this study, the state variable of focus is the moisture storage in AWRA's upper soil layer,  $S_0$ .  
 152 Satellite surface soil moisture (SSM) products from both SMOS and SMAP are used as the  
 153 observations to update the model simulation. Satellite soil moisture estimates are provided in  
 154 volumetric units ( $\text{m}^3/\text{m}^3$ ), whereas modeled upper-layer soil moisture is given in terms of

155 storage of water (i.e. units mm). The observation operator used here is a linear transformation  
 156 which matches the mean and variance between model and observation time series (Tian et al.,  
 157 2017). As such, the observation operator also simultaneously removes systematic bias between  
 158 model estimates and satellite observations. In addition, for region with sparse rain-gauge  
 159 coverage such as central Western Australia, the linear transformation of the satellite soil moisture  
 160 products draws on data sampled from neighboring cells with similar soil moisture conditions, to  
 161 account for known poor model estimates from consistent underestimation of rainfall (S1).

162 The gain factor,  $K$ , contains information on the error variances of the model and observations.  
 163 Observation error variance is often estimated through field campaigns (Draper et al., 2009;  
 164 Panciera et al., 2014), but these rarely represent the spatial and temporal variability of errors in  
 165 gridded satellite products. Alternatively, data providers often specify error estimates, but their  
 166 magnitude can be overly optimistic. Triple collocation (TC) was developed as a method of  
 167 quantifying error characteristics in geophysical variables when the true error structure is elusive.  
 168 It was first applied to near-surface wind data (Stoffelen, 1998) and later extensively applied to  
 169 soil moisture (Dorigo et al., 2017; McColl et al., 2014; Scipal et al., 2008; Su et al., 2014) and  
 170 rainfall (Massari et al., 2017). The assumption of this approach is that three independent data sets  
 171 of the same geophysical variable can be used to infer the error variances in each. Here we use TC  
 172 as a way of inferring error variances from our three independent estimates of surface soil  
 173 moisture, AWRA  $S_0$ , SMAP, and SMOS. McColl et al. (2014) shows that the error variances of  
 174 each data set can be calculated from the temporal variance and covariance between data sets  
 175 respectively as:

$$176 \quad \sigma_x^2 = \left( Q_{x,x} - \frac{Q_{x,y}Q_{x,z}}{Q_{y,z}} \right), \quad \sigma_y^2 = \left( Q_{y,y} - \frac{Q_{x,y}Q_{y,z}}{Q_{x,z}} \right) \quad \text{and} \quad \sigma_z^2 = \left( Q_{z,z} - \frac{Q_{z,y}Q_{x,z}}{Q_{x,y}} \right) \quad (3)$$

177 where  $x, y$  and  $z$  denote AWRA, SMAP or SMOS soil moisture estimates respectively and  $Q$   
 178 denotes temporal variance and covariance between the data sets. These estimates of the error

179 variances are then used in the determination of gain factors (Eq. 2) for the three estimates of soil  
 180 moisture, thus recasting Equation (1) as:

$$181 \quad X_t^a = K_{AWRA} X_t^f + K_{SMAP} Y_t^{SMAP} + K_{SMOS} Y_t^{SMOS} \quad (4)$$

182 The analysed soil water state derived from Eq. (4) represents an optimal blending of AWRA  $S_0$ ,  
 183 SMAP and SMOS (S2).

#### 184 2.4 Analysis increment redistribution (AIR)

185 The assimilation of satellite soil moisture often violates mass conservation in the model through  
 186 the analysis update (Eq. 4). The difference between the analysis,  $X_t^a$ , and the forecast,  $X_t^f$ , (known  
 187 as the *analysis increment*) represents an amount of water that has been added or subtracted from  
 188 the system that was not present at the start of model integration for the given time step. In this  
 189 study, we use the concept of tangent linear modelling (Errico, 1997; Giering, 2000) to  
 190 redistribute the analysis increment of  $S_0$  to all the relevant model states and fluxes (e.g. lower  
 191 layer and deep layer soil water storage, evapotranspiration and runoff). This was considered as a  
 192 way of maintaining mass (i.e. water) balance within a model time step, which data assimilation is  
 193 known to break. We refer to this approach as analysis increment redistribution, or simply as AIR  
 194 hereafter. To illustrate, Equation 5 gives an example of what should the resulting changes ( $\Delta$ ) in  
 195 drainage ( $D_0$ ) be, given the analysis increment in  $S_0$  ( $S_0^a - S_0^f$ ):

$$196 \quad \Delta D_0 = (1 - \beta_0) k_{0sat} \left[ \left( \frac{S_0^a}{S_{0max}} \right)^2 - \left( \frac{S_0^f}{S_{0max}} \right)^2 \right], \quad (5)$$

197 where the  $k_{0sat}$  and  $S_{0max}$  are model parameters representing the saturated hydraulic  
 198 conductivity and maximum storage of the upper soil layer, respectively, and  $\beta_0$  is the  
 199 proportion of upper soil layer lateral drainage (S3 for more detail). Corresponding adjustments  
 200 of total lateral interflow for both upper and lower soil layer are then propagated to the river  
 201 water storage and total runoff. In addition, the analysis increments of  $S_0$  and change in lower  
 202 soil layer water storage after application of the AIR are used to revise the total

203 evapotranspiration. The adjustments to the relevant states and fluxes are derived from AWRA  
204 model formulation (S3).

## 205 2.5 In-situ measurements

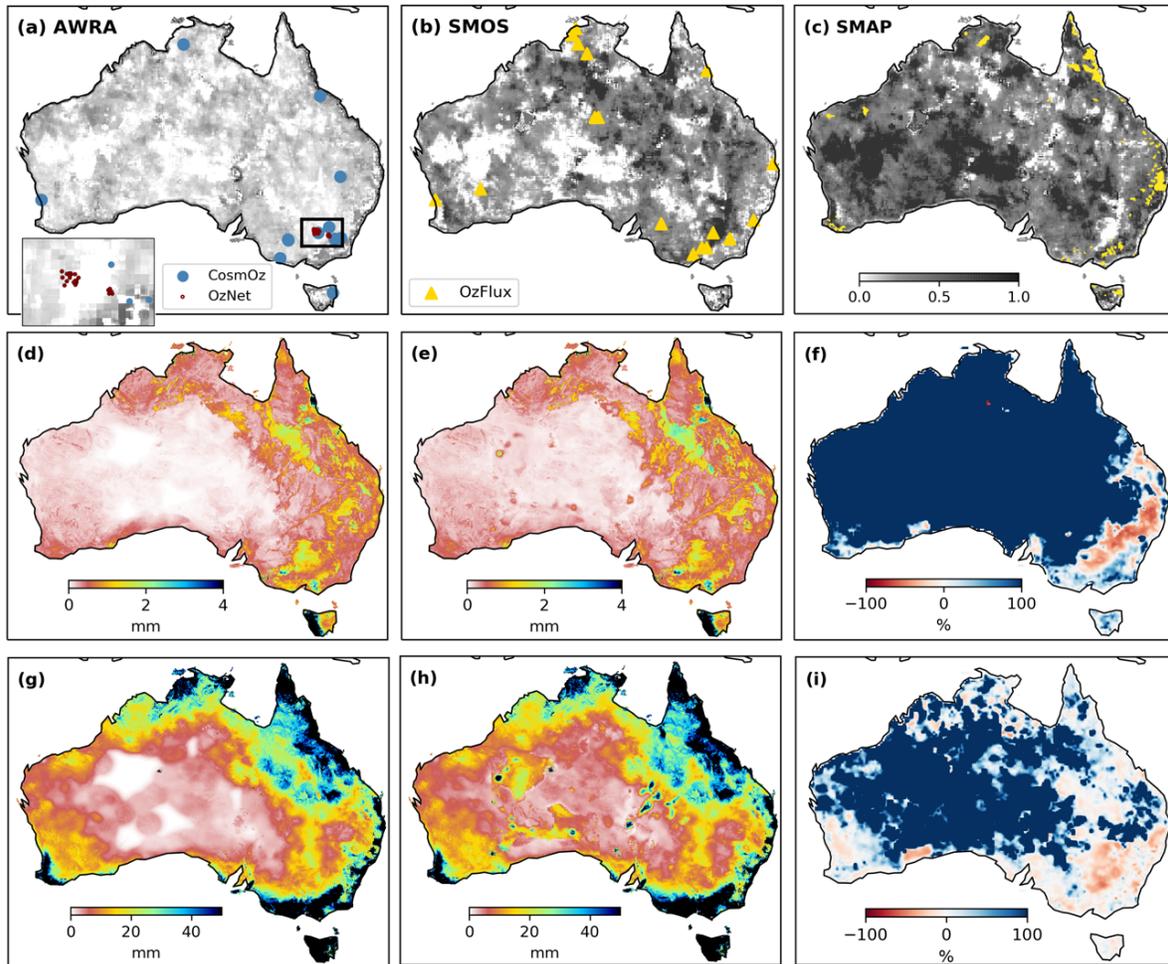
206 Evaluation of the modelled soil water storages was made against measurements from three soil  
207 moisture monitoring networks in Australia from 2016 to 2018, namely OzNet (Smith et al.,  
208 2012), CosmOz (Hawdon et al., 2014) (Fig. 1a) and OzFlux (Fig. 1b). AWRA model estimates  
209 of upper layer soil water storage were compared against in situ measurements from the top 10 cm  
210 of soil across all three networks. In situ measurements of root-zone moisture varied across  
211 networks from 0-30 cm to 0-1 m. As such, AWRA soil water storages over the root-zone were  
212 constructed accordingly by combining upper- and lower-layer soil water storage in the  
213 appropriate proportions. OzFlux sites are primarily used for the evaluation of AWRA  
214 evapotranspiration estimates, which were calculated from accumulated latent heat flux  
215 measurements at each location. In total, there are 45 sites for soil moisture validation and 14 sites  
216 for evapotranspiration validation. Streamflow observations for 100 catchments across Australia  
217 have been used in the validation based on the quality and data availability (Fig. 1c).

## 218 2.6 Vegetation index

219 In water-limited regions like Australia, shallow-rooted vegetation normally responds quickly  
220 with soil water availability, typically within a month. Consistency between root-zone soil water  
221 storage and vegetation greenness may be considered as an indirect independent verification of  
222 the simulation of root-zone soil water dynamic (Tian et al., 2019a; Tian et al., 2019b). The 0.05-  
223 degree monthly Normalized Difference Vegetation Index (NDVI) from Moderate Resolution  
224 Imaging Spectroradiometer (MODIS, MYD13C2 v6) is used to evaluate estimates of root-zone  
225 soil moisture over cropland and grassland regions of the continent. The 250m land cover  
226 classification map from Geoscience Australia (Lymburner et al. 2015) is resampled to 0.05  
227 degree over model domain and used in the identification of crop and grassland cells.

228

229



230

231 **Figure 1** Study area showing gain factors of the TC data assimilation method rescaled to highlight relative contribution of the  
 232 respective estimate: (a) AWRA-simulated  $S_0$ , (b) SMOS soil moisture, and (c) SMAP soil moisture. Also displayed are the  
 233 locations of in-situ monitoring stations from (a) CosmOz and OzNet networks, (b) OzFlux network, and (c) catchments for  
 234 streamflow validation. Subfigures (d) and (e) are the average  $S_0$  simulations for 2019 from AWRA open-loop (OL) and TC  
 235 assimilation of SMOS and SMAP data (DA-TC). Subfigure (f) shows the average relative change of analysed  $S_0$  (TC)  
 236 compared to OL simulations in 2019. Subfigures (g) and (h) are the average  $S_s$  simulations for 2019 from AWRA OL and DA-TC.  
 237 Subfigure (i) shows the average relative change of analysed  $S_s$  after analysis increment redistribution (TC-AIR)  
 238 compared to OL simulations in 2019.

239

## 240 **3 Results and Discussion**

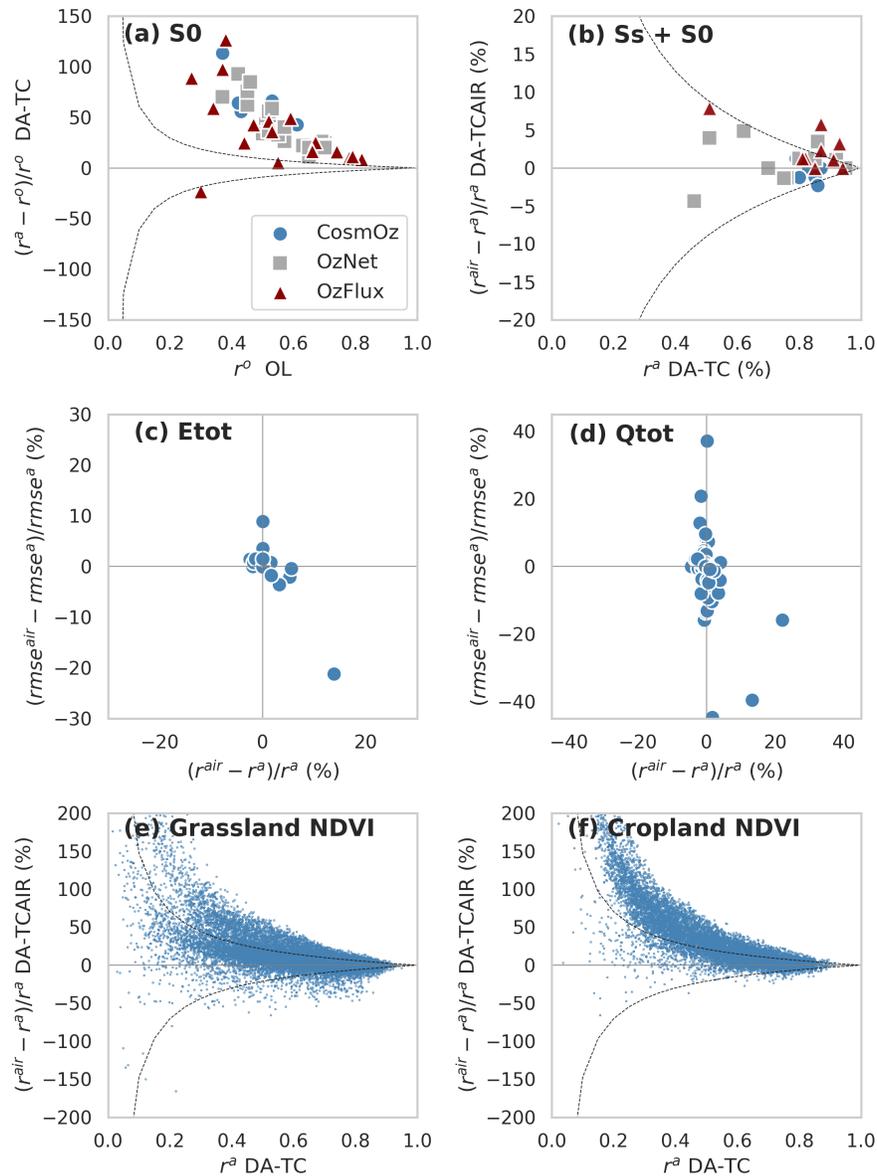
### 241 3.1 Improved spatial representation of soil moisture variability

242 The analysed upper layer soil water storage estimates receive a greater contribution from SSM  
243 products, in particular SMAP observations, compared to model simulations (Figures 1a-c).  
244 AWRA model simulations are driven by gauge-based rainfall analyses. As such they have  
245 difficulty in adequately simulating soil moisture patterns over regions lacking in rain gauge  
246 coverage, such as Western Australia and central Australia (Fig. 1d). Water storage simulations  
247 over these regions default to zero, thus very little or no weight was given to the AWRA estimates  
248 in these regions (Fig. 1a). In contrast, SMAP SSM data is heavily weighted in the assimilation  
249 due to the smaller error variance derived from TC (Fig. 1c). This is expected since SMAP is the  
250 best-performing satellite soil moisture product over the majority of applicable global land pixels  
251 (Chen et al., 2018). AWRA simulations of  $S_0$  are dominated by the satellite SSM data as a result  
252 of TC data assimilation in the region which largely eliminates the erroneous artefacts associated  
253 with deficient rainfall data forcing (Fig. 1e).

254 Moreover, the SSM data assimilation has the effect of adding moisture to AWRA  $S_0$  simulations  
255 over most of Australia, with predictions on average often in excess of 100% of those from the  
256 OL simulations (Fig. 1f). The notable exception to this is in the southeast of Australia,  
257 particularly within the Murray-Darling Basin, where SSM data assimilation reduced AWRA  $S_0$   
258 by more than 50%. This suggests that AWRA simulations underestimated the severity of the  
259 drought experienced in the region in 2019.

260 TC assimilation only updates  $S_0$  directly with satellite SSM, thus the  $S_s$  and other water storage  
261 receives the impact from assimilation once the model integrates forward to the next time step  
262 from the analysed  $S_0$  as initial conditions. The AIR method adjusts  $S_s$  and other relevant states  
263 and fluxes as a post-correction according to the change in  $S_0$  to maintain water balance. The  
264 average  $S_s$  with the correction in drainage and lateral interflow from the change in  $S_0$  after TC  
265 assimilation shows significant different spatial pattern with a relative change more than 100%

266 over those regions with sparse rain-gauge coverage against OL simulations (Fig. 1g-i, i.e. see the  
 267 white regions in Fig1g in particular).



268

269 **Figure 2** Evaluation of AWRA-estimated upper soil water storage  $S_0$ , root-zone soil water storage  $S_s + S_0$ , total  
 270 evapotranspiration  $E_{tot}$  and runoff  $Q_{tot}$ : (a) relative change in correlation of  $S_0$  from TC assimilation (DA-TC) compared to  
 271 model open-loop (with dots above the zero line showing improved performance); (b) relative change in correlation of root-zone  
 272 soil water storage from TC-AIR to DA-TC without mass redistribution; (c) – (d) relative correlation and RMSE changes in  
 273  $E_{tot}$  and  $Q_{tot}$  compared to DA-TC without AIR (with dots in the bottom right quadrant showing both improved correlation and  
 274 reduced RMSE); (e) – (f) relative change in correlation between monthly root-zone soil water storage from TC-AIR with NDVI  
 275 compared to DA-TC for all grid cells classified as grassland and cropland. Note that dashed curves delineate a 95% level of  
 276 statistical significance.

## 277 3.2 Improved water balance estimates

278 Comparisons of AWRA simulations with and without SSM data assimilation were made against  
279 in-situ measurements networks from 2016 to 2018. Consistent, statistically significant  
280 improvement in modelled upper layer soil water storage estimates ( $S_0$ ) was observed across all  
281 sites (Fig. 2a) with the exception of a single OzFlux site located in a tropical rainforest, where  
282 microwave SSM retrievals are typically poor in areas of dense vegetation (Njoku and Entekhabi,  
283 1996). TC-based assimilation (Section 2.2) increases the correlation between in-situ surface SM  
284 measurements from 0.47 to 0.72 on average for CosmOz sites, 0.54 to 0.69 for OzFlux sites, and  
285 0.56 to 0.77 for OzNet sites compared to OL. This is a significant improvement in AWRA  
286 simulations of surface soil moisture dynamics. Compared to ensemble methods of data  
287 assimilation (e.g. Tian et al. 2017; 2019b) which rely on an initial guess of the error variance and  
288 post hoc correction (e.g. inflation factors, Anderson, 2009), this proposed method based on TC is  
289 simple, effective and computationally efficient, thus well suited to an operational system  
290 simulating large-scale hydrology. Overall subtle improvements were observed across the AWRA  
291 estimates of root-zone soil water storage, evapotranspiration and streamflow (results not shown,  
292 see Tian et al., 2019c). The level of improvement is not surprising since those variables were not  
293 directly updated with the TC assimilation and are only influenced through the integration of the  
294 model to the next time step (Tian et al., 2019c).

295 The lack of water balance closure is arguably a weak point in data assimilation (Pan and Wood,  
296 2006). Hence, we applied an analysis increments redistribution (AIR, Section 2.3) as a post-  
297 correction to all relevant model states and fluxes to enforce mass conservation (water balance).  
298 Although the absolute change  $S_0$  is small relative to the volume of  $S_s$ , the corresponding change  
299 in lower layer soil water storage is allocated through AIR based on model physics (S3). The  
300 adjusted root-zone soil water storage ( $S_0+S_s$ ) shows better agreement with in-situ measurements  
301 by up to 10% compared to TC estimates without AIR (Fig. 2b). Improvements are found over the  
302 majority of sites from OzFlux and OzNet with measurements. Improvements in correlation  
303 together with reduced RMSE (Root-Mean-Squares Error) with in-situ measurements for  $E_{tot}$  are  
304 found more than 10% relative to the TC estimates without AIR for some sites (Fig. 2c). Further  
305 improvements in  $Q_{tot}$  simulations are found for some sites, with up to 40% reduction in RMSE  
306 (Fig. 2d). Improvements in runoff simulation are due to, first, the SSM assimilation improving

307 pre-storm soil moisture status (Pauwels et al., 2001; Crow and Ryu, 2009), and then AIR adjusts  
308 the interflow and river storage accordingly. This indicates the importance of accurate antecedent  
309 soil moisture condition in the simulation of runoff response to subsequent rainfall.

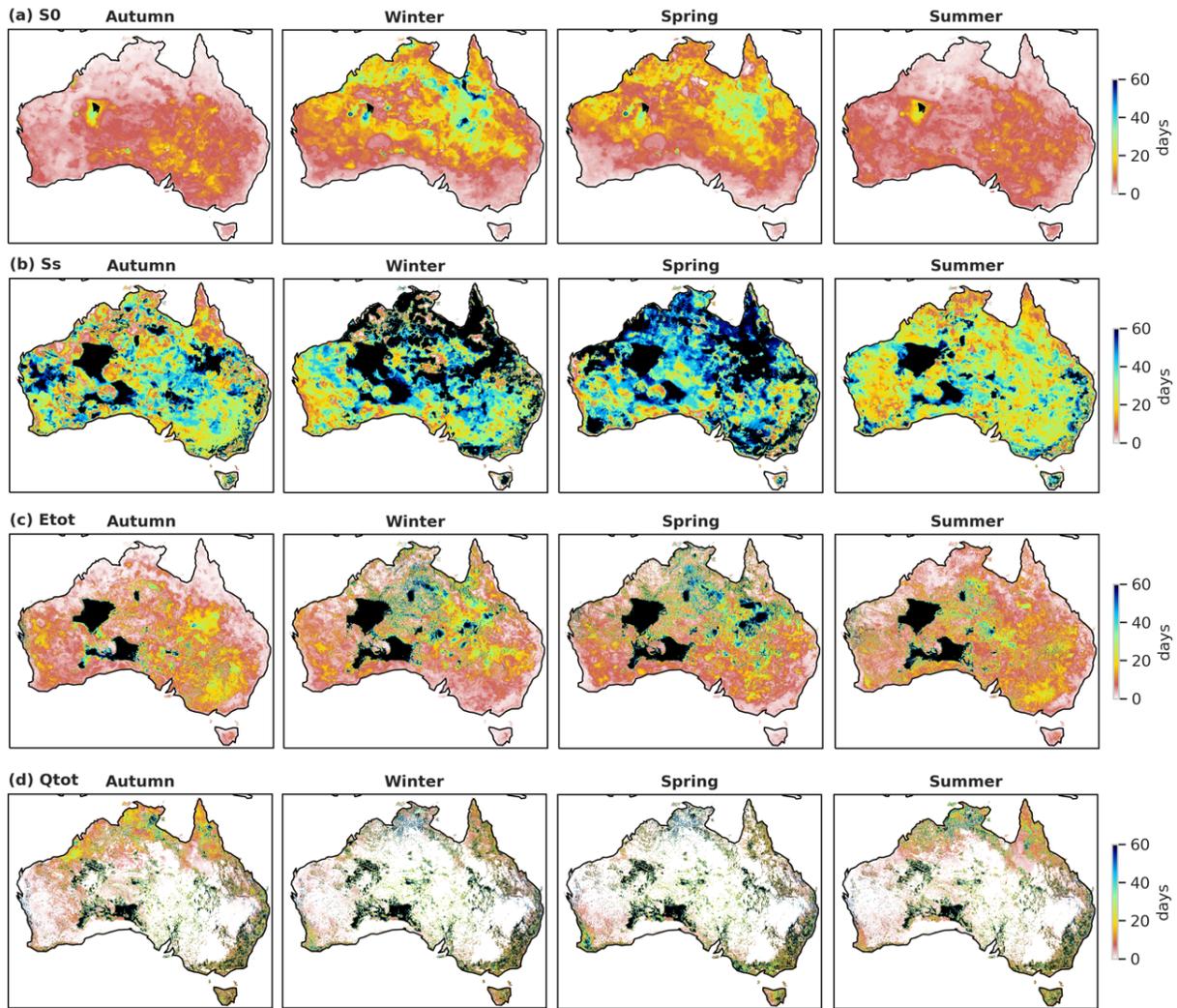
310 The inadequate distribution of in-situ observations as well as the large spatial disparity between  
311 ground measurement and modelling scales are a great limitation for the evaluation of root-zone  
312 soil moisture and evapotranspiration. AWRA simulation of root-zone soil moisture are compared  
313 against satellite-derived NDVI in an indirect verification of model performance and as a way of  
314 evaluating the impact of data assimilation. We calculated the correlation between time series of  
315 monthly average AWRA root-zone soil moisture from OL, DA-TC and TC-AIR simulations  
316 against NDVI for cropland and grassland of Australia over the period 2015 to 2018. These cover  
317 types we selected as their rooting depths are commensurate with the combined soil depths of the  
318 upper- and lower-soil water storages in AWRA. Figure 2e-f show the relative change in  
319 correlation between root-zone simulations from DA-TC and those from TC-AIR data against  
320 NDVI data for grassland and cropland areas of Australia. The figure shows that for the vast  
321 majority of model grids, TC-AIR shows statistically significant increase in correlation with  
322 NDVI compared to DA-TC alone, with an average increase in correlation with NDVI from 0.64  
323 to 0.67 for grassland and 0.55 to 0.66 for cropland compared to OL. This demonstrates that  
324 enforcing mass balances as part of the SSM data assimilation each time step is essential to  
325 improving the simulation of root-zone soil water balance. The improved consistency with NDVI  
326 also illuminates the potential of improving agricultural planning with more accurate information  
327 of root-zone soil water availability.

### 328 3.3 Impacts on model predictions

329 Accurate soil water estimates can provide initial conditions for improved flood forecasting and  
330 groundwater forecasting (Getirana et al., 2020a; Getirana et al., 2020b; Wanders et al., 2013).  
331 Few studies quantify how long the impacts of data assimilation persist in the model system's  
332 memory. In this study we used 100-day model simulations from initial states provided by the  
333 AWRA OL and DA-TC with AIR. We calculated the number of days it took for the simulation  
334 from the analysed DA-TCAIR states to converge to within +/- 5% of those from OL. The  
335 experiments were run for one year from 1 March 2018 to 28 February 2019. Results show that  
336 data assimilation can impact on model states and fluxes for weeks and sometimes up to 2-3

337 months (Fig. 3). The impacts of DA-TC with AIR can persist in simulated  $S_0$  for as long as a  
338 week over coastal regions, and longer in central Western Australia and Northern Australia with  
339 up to 1 month persistence in winter and spring (Fig. 3a). There is less impact on  $S_0$  simulations  
340 during wet season since the  $S_0$  can saturate rapidly due to the heavy rainfall. Overall, the longest  
341 persistence is found in winter with a continental average of 13 days; the shortest persistence is 6  
342 days on average in autumn and summer. The memory of initial conditions in simulations of  $S_s$   
343 can persist even longer due to the slower response to rainfall variability and higher field capacity.  
344 Summer persistence for  $S_s$  is the least with a continental average of 30 days; in winter, this is  
345 increased to 45 days.

346 Evapotranspiration estimates, however, do not feedback into the system and are highly variable  
347 in time and space. On average, the impact of the antecedent soil moisture conditions on  
348 evapotranspiration simulations can persist for 1 week over coastal areas, but up to months in  
349 central Western Australia. The continental average varies from 13 to 20 days for each season.  
350 The areas with the longest persistence are those areas with artefacts of zero rainfall in the  
351 forcing. This demonstrates that improvements in AWRA estimates after SSM assimilation over  
352 regions with sparse rain-gauge coverage can persist in the system for more than 2 months. The  
353 impact on runoff varies from 1 week to 3 months over the continent. The majority of areas  
354 impacted for more than 2 months are in locations of little rainfall and runoff. However, there  
355 remains between 1-2 week impacts over north-eastern areas with heavy runoff.



356

357

358

359

360

361

**Figure 3** Quantified impacts of data assimilation on forecasting AWRA state variables through the forecast of states for 100 days using the initial condition from DA-TCAIR: average time period that the impact of data assimilation can persist in autumn (2018.03-2018.05), Winter (2018.06-2018.08), Spring (2018.09-2018.11) and Summer (2018.12-2019.02) on (a) upper-layer soil water storage  $S_0$ , (b) lower-layer soil water storage  $S_s$ , (c) total evapotranspiration  $E_{tot}$  and (d) total runoff  $Q_{tot}$ .

#### 362 **4 Conclusion**

363 In this study, we proposed a simple and robust method for assimilating SMAP and SMOS soil  
364 moisture products into the operational Australian Water Resources Assessment (AWRA) model.  
365 The method involves the sequential (daily) updating of the model's upper layer soil water storage  
366 with satellite soil moisture observations through a linear combination with weights determined  
367 through triple collocation (DA-TC). Evaluation against in-situ measurements showed that  
368 simulations of surface soil moisture dynamics is improved significantly after TC data  
369 assimilation with an average increase of 0.23 correlation units compared with open-loop  
370 simulations. Furthermore, we proposed an additional component to the data assimilation  
371 whereby the analysis increment of the upper layer soil water storage is propagated into relevant  
372 model states and fluxes as a way of maintaining mass balance (TC-AIR). An evaluation of the  
373 root-zone soil moisture, evapotranspiration and streamflow estimates showed that the TC-AIR  
374 appeared to only provide marginal, yet positive, improvement over the TC data assimilation  
375 method alone. However, in an indirect verification of modelled root-zone soil moisture against  
376 satellite-derived NDVI, TC-AIR was seen to provide significant improvement on TC method  
377 alone. This demonstrates that by enforcing mass balances as part of the SSM data assimilation  
378 each time step, AWRA can better represent soil water dynamics with greater consistency with  
379 vegetation response.

380  
381 The assimilation of satellite soil moisture estimates together with the mass redistribution reduces  
382 the uncertainties in model estimates resulting mainly from uncertain forcing and model physics,  
383 and provides temporally and spatially varying constraints on model water balance estimates. For  
384 example, the assimilation resolves the gaps in rainfall forcing, and the underestimate of drought  
385 condition over south-eastern areas in 2019. We demonstrate that the impacts of data assimilation  
386 can persist in the model system for more than a week for surface soil water storage and more  
387 than a month for root-zone soil water storage. This highlights the importance of accurate initial  
388 hydrological states for improving forecast skill over longer lead times. Hence, an operational  
389 water balance modelling system, with satellite data assimilation, has strong potential to add value  
390 for assessing and predicting water availability for a range of decisions across industries and  
391 sectors.

## 392 Acknowledgments

393 This project is supported by collaborative research agreement between the Australian Bureau of  
394 Meteorology and Australian National University. We would like to thank Stuart Baron-Hay from  
395 the Bureau of Meteorology for his help with implementation of the in AWRA-CMS. This  
396 research was undertaken with the assistance of resources and services from the National  
397 Computational Infrastructure (NCI), which is supported by the Australian Government through  
398 the National Collaborative Research Infrastructure Strategy.

## 399 Data Availability

400 The AWRA-CMS code is accessible from github ([https://github.com/awracms/awra\\_cms](https://github.com/awracms/awra_cms)).  
401 SMAP product used here is the level-2 enhanced radiometer half-orbit 9-km EASE-grid soil  
402 moisture from the US National Snow and Ice Data Center (<https://nsidc.org>). SMOS level-2 soil  
403 moisture product is available from ESA's SMOS online dissemination service ([https://smos-  
404 diss.eo.esa.int/oads/access/](https://smos-diss.eo.esa.int/oads/access/)). The MYD13C2 NDVI data is accessible through Land Processes  
405 Distributed Active Archive Centre (<https://lpdaac.usgs.gov>). The National Dynamic Land Cover  
406 Dataset of Australia is available from Geoscience Australia (<https://www.ga.gov.au>).

## 407 References

- 408 Chan, S. K., Bindlish, R., O'Neill, P., Jackson, T., Njoku, E., Dunbar, S., et al. (2018).  
409 Development and assessment of the SMAP enhanced passive soil moisture product. *Remote*  
410 *Sensing of Environment*, 204, 931-941.
- 411 Chen, F., Crow, W. T., Bindlish, R., Colliander, A., Burgin, M. S., Asanuma, J., & Aida, K.  
412 (2018). Global-scale evaluation of SMAP, SMOS and ASCAT soil moisture products using  
413 triple collocation. *Remote Sensing of Environment*, 214, 1-13
- 414 Crow, W. T., & Ryu, D. (2009). A new data assimilation approach for improving runoff  
415 prediction using remotely-sensed soil moisture retrievals. *Hydrology and Earth System Sciences*,  
416 13(1), 1-16.
- 417 Davies, T., Cullen, M. J. P., Malcolm, A. J., Mawson, M. H., Staniforth, A., White, A. A., &  
418 Wood, N. (2005). A new dynamical core for the Met Office's global and regional modelling of  
419 the atmosphere. *Quarterly Journal of the Royal Meteorological Society*, 131(608), 1759-1782.
- 420 de Rosnay, P., Drusch, M., Vasiljevic, D., Balsamo, G., Albergel, C., & Isaksen, L. (2013). A  
421 simplified Extended Kalman Filter for the global operational soil moisture analysis at ECMWF.  
422 *Quarterly Journal of the Royal Meteorological Society*, 139(674), 1199-1213.
- 423 Dharssi, I., Bovis, K. J., Macpherson, B., & Jones, C. P. (2011). Operational assimilation of  
424 ASCAT surface soil wetness at the Met Office. *Hydrology and Earth System Sciences*, 15(8),  
425 2729-2746.

- 426 Dorigo, W., Wagner, W., Albergel, C., Albrecht, F., Balsamo, G., Brocca, L., et al. (2017). ESA  
427 CCI Soil Moisture for improved Earth system understanding: State-of-the art and future  
428 directions. *Remote Sensing of Environment*, 203, 185-215.
- 429 Draper, C. S., Reichle, R. H., De Lannoy, G. J. M., & Liu, Q. (2012). Assimilation of passive  
430 and active microwave soil moisture retrievals. *Geophysical Research Letters*, 39.
- 431 Draper, C. S., Walker, J. P., Steinle, P. J., de Jeu, R. A. M., & Holmes, T. R. H. (2009). An  
432 evaluation of AMSR-E derived soil moisture over Australia. *Remote Sensing of Environment*,  
433 113(4), 703-710.
- 434 Lymburner, L., Tan, P., McIntyre, A., Thankappan, M., Sixsmith, J. (2015). Dynamic Land  
435 Cover Dataset Version 2.1. Geoscience Australia, Canberra.  
436 <http://pid.geoscience.gov.au/dataset/ga/83868>
- 437 Entekhabi, D., Njoku, E. G., O'Neill, P. E., Kellogg, K. H., Crow, W. T., Edelstein, W. N., et al.  
438 (2010). The Soil Moisture Active Passive (SMAP) Mission. *Proceedings of the Ieee*, 98(5), 704-  
439 716.
- 440 Errico, R. M. (1997). What is an adjoint model?. *Bulletin of the American Meteorological*  
441 *Society*, 78(11), 2577-2592.
- 442 Frost, A., A. Ramchurn, and A. Smith (2018). The Australian Landscape Water Balance model  
443 (AWRA-L v6), Technical Description of the Australian Water Resources Assessment Landscape  
444 model version 6. *Bureau of Meteorology Technical Report*, Canberra: Bureau of Meteorology.
- 445 Getirana, A., Rodell, M., Kumar, S., Beaudoin, H. K., Arsenault, K., Zaitchik, B., et al. (2020).  
446 GRACE Improves Seasonal Groundwater Forecast Initialization over the United States. *Journal*  
447 *of Hydrometeorology*, 21(1), 59-71.
- 448 Getirana, A., Jung, H. C., Arsenault, K., Shukla, S., Kumar, S., Peters-Lidard, C., et al. (2020).  
449 Satellite Gravimetry Improves Seasonal Streamflow Forecast Initialization in Africa. *Water*  
450 *Resources Research*, 56(2), doi: 10.1029/2019wr026259.
- 451 Giering, R. (2000). Tangent linear and adjoint biogeochemical models. *Inverse Methods in*  
452 *Global Biogeochemical Cycles*, 114, 33-48.
- 453 Grant, I., Jones, D., Wang, W., Fawcett, R., and Barratt, D. (2008). Meteorological and remotely  
454 sensed datasets for hydrological modelling: A contribution to the Australian Water Availability  
455 Project, paper presented at Catchment-scale Hydrological Modelling and Data Assimilation  
456 (CAHMDA-3). *International Workshop on Hydrological Prediction: Modelling, Observation*  
457 *and Data Assimilation*, Melbourne, Citeseer.
- 458 Hawdon, A., McJannet, D., & Wallace, J. (2014). Calibration and correction procedures for  
459 cosmic-ray neutron soil moisture probes located across Australia. *Water Resources Research*,  
460 50(6), 5029-5043.
- 461 Ines, A. V. M., Das, N. N., Hansen, J. W., & Njoku, E. G. (2013). Assimilation of remotely  
462 sensed soil moisture and vegetation with a crop simulation model for maize yield prediction.  
463 *Remote Sensing of Environment*, 138, 149-164.
- 464 Jones, D. A., Wang, W., & Fawcett, R. (2009). High-quality spatial climate data-sets for  
465 Australia. *Australian Meteorological and Oceanographic Journal*, 58(4), 233-248.

- 466 Kerr, Y. H., Waldteufel, P., Wigneron, J. P., Martinuzzi, J. M., Font, J., & Berger, M. (2001).  
467 Soil moisture retrieval from space: The Soil Moisture and Ocean Salinity (SMOS) mission. *Ieee*  
468 *Transactions on Geoscience and Remote Sensing*, 39(8), 1729-1735.
- 469 Koster, R. D., Dirmeyer, P. A., Guo, Z. C., Bonan, G., Chan, E., Cox, P., et al. (2004). Regions  
470 of strong coupling between soil moisture and precipitation. *Science*, 305(5687), 1138-1140.
- 471 Massari, C., Crow, W., & Brocca, L. (2017). An assessment of the performance of global rainfall  
472 estimates without ground-based observations. *Hydrology and Earth System Sciences*, 21(9),  
473 4347-4361.
- 474 McColl, K. A., Vogelzang, J., Konings, A. G., Entekhabi, D., Piles, M., & Stoffelen, A. (2014).  
475 Extended triple collocation: Estimating errors and correlation coefficients with respect to an  
476 unknown target. *Geophysical Research Letters*, 41(17), 6229-6236.
- 477 McVicar, T. R., Van Niel, T. G., Li, L. T., Roderick, M. L., Rayner, D. P., Ricciardulli, L., &  
478 Donohue, R. J. (2008). Wind speed climatology and trends for Australia, 1975-2006: Capturing  
479 the stilling phenomenon and comparison with near-surface reanalysis output. *Geophysical*  
480 *Research Letters*, 35(20)
- 481 Njoku, E. G., & Entekhabi, D. (1996). Passive microwave remote sensing of soil moisture.  
482 *Journal of Hydrology*, 184(1-2), 101-129.
- 483 Pan, M., & Wood, E. F. (2006). Data assimilation for estimating the terrestrial water budget  
484 using a constrained ensemble Kalman filter. *Journal of Hydrometeorology*, 7(3), 534-547.
- 485 Panciera, R., Walker, J. P., Jackson, T. J., Gray, D. A., Tanase, M. A., Ryu, D., et al. (2014). The  
486 Soil Moisture Active Passive Experiments (SMAPEX): Toward Soil Moisture Retrieval From the  
487 SMAP Mission. *Ieee Transactions on Geoscience and Remote Sensing*, 52(1), 490-507.
- 488 Pauwels, V. R. N., Hoeben, R., Verhoest, N. E. C., & De Troch, F. P. (2001). The importance of  
489 the spatial patterns of remotely sensed soil moisture in the improvement of discharge predictions  
490 for small-scale basins through data assimilation. *Journal of Hydrology*, 251(1-2), 88-102.
- 491 Pipunic, R. C., Walker, J. P., & Western, A. (2008). Assimilation of remotely sensed data for  
492 improved latent and sensible heat flux prediction: A comparative synthetic study. *Remote*  
493 *Sensing of Environment*, 112(4), 1295-1305.
- 494 Rahmoune, R., Ferrazzoli, P., Kerr, Y. H., & Richaume, P. (2013). SMOS Level 2 Retrieval  
495 Algorithm Over Forests: Description and Generation of Global Maps. *Ieee Journal of Selected*  
496 *Topics in Applied Earth Observations and Remote Sensing*, 6(3), 1430-1439.
- 497 Reichle, R. H., & Koster, R. D. (2005). Global assimilation of satellite surface soil moisture  
498 retrievals into the NASA Catchment land surface model. *Geophysical Research Letters*, 32(2).  
499 <https://doi.org/10.1016/j.advwatres.2008.01.001>
- 500 Reichle, R. H. (2008). Data assimilation methods in the Earth sciences. *Advances in Water*  
501 *Resources*, 31(11), 1411-1418.
- 502 Renzullo, L. J., van Dijk, A. I. J. M., Perraud, J. M., Collins, D., Henderson, B., Jin, H., et al.  
503 (2014). Continental satellite soil moisture data assimilation improves root-zone moisture analysis  
504 for water resources assessment. *Journal of Hydrology*, 519, 2747-2762.

- 505 Scipal, K., Drusch, M. & Wagner, W., (2008). Assimilation of a ERS scatterometer derived soil  
506 moisture index in the ECMWF numerical weather prediction system, *Advances in water*  
507 *resources*, 31(8), 1101-1112.
- 508 Sheffield, J., & Wood, E. F. (2007). Characteristics of global and regional drought, 1950-2000:  
509 Analysis of soil moisture data from off-line simulation of the terrestrial hydrologic cycle.  
510 *Journal of Geophysical Research-Atmospheres*, 112(D17).
- 511 Smith, A. B., Walker, J. P., Western, A. W., Young, R. I., Ellett, K. M., Pipunic, R. C., et al.  
512 (2012). The Murrumbidgee soil moisture monitoring network data set. *Water Resources*  
513 *Research*, 48.
- 514 Stoffelen, A. (1998). Toward the true near-surface wind speed: Error modeling and calibration  
515 using triple collocation. *Journal of Geophysical Research-Oceans*, 103(C4), 7755-7766.
- 516 Su, C. H., Ryu, D., Crow, W. T., & Western, A. W. (2014). Beyond triple collocation:  
517 Applications to soil moisture monitoring. *Journal of Geophysical Research-Atmospheres*,  
518 119(11), 6419-6439.
- 519 Tian, S., van Dijk, A. I. J. M., Tregoning, P., & Renzullo, L. J. (2019). Forecasting dryland  
520 vegetation condition months in advance through satellite data assimilation. *Nature*  
521 *Communications*, 10. (1), 1-7.
- 522 Tian, S., Tregoning, P., Renzullo, L. J., van Dijk, A. I. J. M., Walker, J. P., Pauwels, V. R. N., &  
523 Allgeyer, S. (2017). Improved water balance component estimates through joint assimilation of  
524 GRACE water storage and SMOS soil moisture retrievals. *Water Resources Research*, 53(3),  
525 1820-1840.
- 526 Tian, S., Renzullo, L. J., van Dijk, A. I. J. M., Tregoning, P., & Walker, J. P. (2019). Global joint  
527 assimilation of GRACE and SMOS for improved estimation of root-zone soil moisture and  
528 vegetation response. *Hydrology and Earth System Sciences*, 23(2), 1067-1081.
- 529 Tian, S., Renzullo, L. J., Pipunic, R., Sharples, W., Lerat, J., & Donnelly, C.(2019),.  
530 Assimilating satellite soil moisture retrievals to improve operational water balance modelling.  
531 *Proceedings of the 23rd International Congress on Modelling and Simulation*, Canberra,  
532 Australia. <https://doi.org/10.36334/modsim.2019.H6.tian>
- 533 van Dijk, A.I.J.M. (2010). AWRA Technical Report 3, Landscape Model (version 0.5) Technical  
534 Description, WIRADA, Canberra: CSIRO Water for a Healthy Country Flagship.
- 535 Vinnikov, K. Y., Robock, A., Speranskaya, N. A., & Schlosser, A. (1996). Scales of temporal  
536 and spatial variability of midlatitude soil moisture. *Journal of Geophysical Research-*  
537 *Atmospheres*, 101(D3), 7163-7174.
- 538 Wanders, N., Karssenber, D., de Roo, A., de Jong, S. M., & Bierkens, M. F. P. (2014). The  
539 suitability of remotely sensed soil moisture for improving operational flood forecasting.  
540 *Hydrology and Earth System Sciences*, 18(6), 2343-2357.

**Operational soil moisture data assimilation for improved continental water balance prediction**

Siyuan Tian<sup>1\*</sup>, Luigi J. Renzullo<sup>1</sup>, Robert Pipunic<sup>2</sup>, Julien Lerat<sup>2</sup>, Wendy Sharples<sup>2</sup>, Chantal Donnelly<sup>2</sup>

<sup>1</sup> Fenner School of Environment & Society, The Australian National University, ACT, 2601, Australia

<sup>2</sup> Water Program, Bureau of Meteorology, Australia

**Contents of this file**

Text S1 to S3

**Introduction**

This document expands the description of data assimilation method and analysis increment mass redistribution (AIR). Specifically it covers:

- the method of linear transformation of the satellite soil moisture products over gauge-sparse region;
- the derivation of equation 4 from equation 1; and
- all the equations for AIR.

**Text S1.**

The observation operator that links AWRA model upper-layer soil water storage ( $S_o$ ) state with the satellite soil moisture (SSM) is a linear transformation derived from temporal mean and variance matching between the two estimates. Mean and variance matching is an accepted practice of correcting systematic bias between model estimates and observations, and in our case map observations into state space for data assimilation. However, for regions of Australia with little, or no, rain gauge coverage, AWRA model  $S_o$  persist as zeros or very low values, 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 be to flatten the dynamics of SSM time series to zero.

To resolve this problem, and make full use of the 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. We obtained the maximum SSM values through time and the derived 'slope' and 'intercept' from the observation model for each cell in neighboring region. Then we applied linear regression to estimate the correspond slope and intercept from the maximum SSM values in the rainfall gaps. This provided an observation model to transform the SSM in into water storage unit (mm) and ensures the assimilation can effectively impart spatial pattern of soil moisture over the sparsely gauged regions.

**Text S2.**

The generic form of the state updating equation for sequential data assimilation is given as:

$$X_t^a = X_t^f + K_t[Y_t - H(X_t^f)],$$

where the terms are defined as for Eq. (1). In this study, there are two satellite soil moisture observations, transformed into model space (i.e. water storage) through the observation operator, denotes as  $Y_t^{SMAP}$  and  $Y_t^{SMOS}$ . Since the error variances of the SSM products are independent, we can therefore write the above as:

$$\begin{aligned} X_t^a &= X_t^f + [K_{SMAP}, K_{SMOS}] \begin{bmatrix} Y_t^{SMAP} - X_t^f \\ Y_t^{SMOS} - X_t^f \end{bmatrix} \\ &= (1 - K_{SMAP} - K_{SMOS})X_t^f + K_{SMAP}Y_t^{SMAP} + K_{SMOS}Y_t^{SMOS} \\ &= K_{AWRA}X_t^f + K_{SMAP}Y_t^{SMAP} + K_{SMOS}Y_t^{SMOS}, \end{aligned}$$

where the gain factors are calculated as:

$$K_{AWRA} = \frac{\frac{1}{\sigma_x^2}}{\frac{1}{\sigma_x^2} + \frac{1}{\sigma_y^2} + \frac{1}{\sigma_z^2}}, K_{SMAP} = \frac{\frac{1}{\sigma_y^2}}{\frac{1}{\sigma_x^2} + \frac{1}{\sigma_y^2} + \frac{1}{\sigma_z^2}}, K_{SMOS} = \frac{\frac{1}{\sigma_z^2}}{\frac{1}{\sigma_x^2} + \frac{1}{\sigma_y^2} + \frac{1}{\sigma_z^2}}$$

respectively. The error variance  $\sigma^2$  for each data set are obtained through the triple collocation (TC) methods, Eq (3).

### Text S3.

The influence of the improved, or analysed, upper-layer soil water,  $S_0^a$  is only realized in the TC data assimilation once the model integrates to the next time step when the water balance is restored between model components. We proposed an analysis increment redistribution (AIR) modification to the TC data assimilation method (TC-AIR) as a way of maintaining water balance at each time step. The idea borrows from tangent linear modelling (TLM), where only relevant model components are modified to accommodate the increment. The following are the specific components of AWRA model which are relevant here (for greater detail see Frost et al., 2016) and are used in the AIR approach, and they include the modifications necessary to impart the water balance constraint.

The analysis increments after the data assimilation can be calculated as:

$$\Delta S_0 = S_0^a - S_0^f,$$

where  $S_0^a$  denotes the analysed upper-layer soil water storage and  $S_0^f$  denotes the forecast, or initial estimate. The change in  $S_0$  affects the drainage to the lower-layer soil water storage ( $D_0$ ) and interflow draining laterally from the top soil layer ( $Q_{I0}$ ). The corresponding change in drainage to lower-layer soil water storage from the increment  $\Delta S_0$  is calculated as:

$$\Delta D_0 = (1 - \beta_0)k_{0sat} \left[ \left( \frac{S_0^a}{S0max} \right)^2 - \left( \frac{S_0^f}{S0max} \right)^2 \right],$$

$$\Delta Q_{I0} = \beta_0 k_{0sat} \left[ \left( \frac{S_0^a}{S0max} \right)^2 - \left( \frac{S_0^f}{S0max} \right)^2 \right],$$

where the  $k_{0sat}$  and  $S0max$  are model parameters representing the saturated hydraulic conductivity and maximum storage of the upper soil layer, respectively. The proportion of overall top layer drainage that is lateral drainage ( $\beta_0$ ) given as:

$$\beta_0 = \tanh \left( k_\beta \beta \frac{S_0^a}{S0max} \right) \tanh \left( k_\zeta \left( \frac{k_{0sat}}{k_{ssat}} - 1 \right) \frac{S_0^a}{S0max} \right),$$

where  $\beta$  and  $k_\beta$  are the slope radians and scaling factor, and  $k_\zeta$  is a scaling factor for the ratio of saturated hydraulic conductivity. The revised lower-layer soil water storage  $S_s^a$  is then determined as:

$$S_s^a = S_s^f + \Delta D_0.$$

The change in  $S_s$  will lead to the change in the shallow soil water storage ( $D_s$ ) and lateral interflow ( $Q_{Is}$ ). The soil water storage at lower layer is thus updated as:

$$S_d^a = S_s^a + \Delta D_s.$$

Similarly, the groundwater storage  $S_g$  will be adjusted with the increment of deep soil layer drainage.

The total runoff ( $Q_{tot}^a$ ) should be updated as:

$$Q_{tot}^a = (1 - e^{-k_r})(S_r^f + Q_{tot}^f + \Delta Q_{Is} + \Delta Q_{I0}),$$

where  $k_r$  is a routing delay factor.

The surface water storage  $S_r$  should be updated accordingly as:

$$S_r^a = S_r^f + \Delta Q_{Is} + \Delta Q_{I0} - \Delta Q_{tot}.$$

The total evapotranspiration change ( $\Delta E_{tot}$ ) caused by the changes in  $S_0$  and  $S_s$  can be updated as follow:

$$\Delta E_{tot} = \delta E_s * \Delta S_0 + \delta E_t * \Delta S_s,$$

where the  $E_s$  is the evaporation flux from the surface soil store ( $S_0$ ) and  $E_t$  is the total actual plant transpiration. The term  $\delta E_s$  is given as

$$\delta E_s = (1 - f_{sat})E_{t\_rem}\delta f_{soile},$$

where  $f_{soile}$  is relative soil evaporation and  $f_{sat}$  is the fraction of the grid cell that is saturated, and

$$E_{t\_rem} = E_0 - (E_t - \delta E_t),$$

The term  $\delta E_t$  is from the changes in root-water uptake from shallow and deep soil layers as

$$\delta E_t = \delta U_s + \delta U_d,$$

with

$$\delta U_s = \delta U_{smax} \frac{\max(\delta U_{smax}, \delta U_{dmax})}{\delta U_{smax} + \delta U_{dmax}}$$

$$\delta U_d = \delta U_{dmax} \frac{\max(\delta U_{smax}, \delta U_{dmax})}{\delta U_{smax} + \delta U_{dmax}}$$

$\delta U_{smax} = \frac{U_{s0}}{w_{slim}} \delta w_s$ ,  $\delta U_{dmax} = \frac{U_{d0}}{w_{dlim}} \delta w_d$ , where  $U_{smax}$  and  $U_{dmax}$  are the maximum root water uptake from the shallow soil store and from deep soil store.  $w_{slim}$  and  $w_{dlim}$  is the water-limiting relative water content from the *shallow and deep* soil layer.

Finally,

$\delta f_{soile} = \frac{f_{soilmax}}{w_{olim}} \delta w_0$ , where  $f_{soilmax}$  is the scaling factor corresponding to unlimited soil water supply, with

$$\delta w_0 = \frac{1}{S_{0max}}, \quad \delta w_s = \frac{1}{S_{smax}}, \quad \text{and} \quad \delta w_d = \frac{1}{S_{dmax}},$$

where the  $w_z$  is the relative soil wetness of layer  $z$ , *i.e.* either 0, s or d.

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