Author/s:
Alvarez-Garreton, C;Ryu, D;Western, AW;Crow, WT;Su, CH;Robertson, DR

Title:
Dual assimilation of satellite soil moisture to improve streamflow prediction in data-scarce catchments

Date:
2016-07-01

Citation:
Alvarez-Garreton, C., Ryu, D., Western, A. W., Crow, W. T., Su, C. H. & Robertson, D. R. (2016). Dual assimilation of satellite soil moisture to improve streamflow prediction in data-scarce catchments. Water Resources Research, 52 (7), pp.5357-5375. https://doi.org/10.1002/2015WR018429.

Dual assimilation of satellite soil moisture to improve streamflow prediction in data-scarce catchments

Camila Alvarez-Garreton,[1,2,3] Dongryeol Ryu[1], Andrew W. Western[1], Wade T. Crow[4], Chun-Hsu Su[1] and David R. Robertson[5]

Abstract. This paper explores the use of active and passive microwave satellite soil moisture products for improving streamflow prediction within 4 large (>5,000km^2) semi-arid catchments in Australia. We use the probability distributed model (PDM) under a data-scarce scenario and aim at correcting two key controlling factors in the streamflow generation: the rainfall forcing data and the catchment wetness condition. The soil moisture analysis rainfall tool (SMART) is used to correct a near-real time satellite rainfall product (forcing correction scheme) and an ensemble Kalman filter is used to correct the PDM soil moisture state (state correction scheme). These two schemes are combined in a dual correction scheme and we assess the relative improvements of each. Our results demonstrate that the quality of the satellite rainfall product is improved by SMART during moderate-to-high daily rainfall events, which in turn leads to improved streamflow prediction during high flows. When employed individually, the soil moisture state correction scheme generally outperforms the rainfall correction scheme, especially for low flows. Overall, the combined dual correction scheme further improves the streamflow predictions (reduction in root mean square error and false alarm ratio, and increase in correlation coefficient and Nash-Sutcliffe efficiency). Our results provide new evidence of the value of satellite soil moisture observations within data-scarce regions. We also identify a number of challenges and limitations within the schemes.

1. Introduction

Flood prediction in sparsely monitored and ungauged catchments can suffer from large uncertainties given the quality of the data used to force and calibrate the models. Addressing this challenge, a number of studies have explored data assimilation methods to integrate various existing observations from the ground and satellites into streamflow models [e.g., *Moradkhani et al.*, 2005; *Liu and Gupta*, 2007; *Mendoza et al.*, 2012; *Wanders et al.*, 2014].

Within this context, and given the essential role that soil moisture (SM) plays in the runoff generation (*Western et al.* [2002] and references therein), significant attention has been given to satellite SM observations. Microwave retrievals of SM provide near real time estimates of the water content from the top few centimetres of soil, at a global scale every 1-3 days. Moreover, satellite SM estimates have shown good agreement with ground data [*Albergel et al.*, 2009; *Draper et al.*, 2009; *Gruhier et al.*, 2010; *Brocca et al.*, 2011; *Su et al.*, 2013].

A popular approach has been to use satellite SM in a state correction scheme [e.g., *Francois et al.*, 2003; *Brocca et al.*, 2010, 2012; *Alvarez-Garreton et al.*, 2013, 2014; *Chen et al.*, 2014; *Wanders et al.*, 2014; *Alvarez-Garreton et al.*, 2015; *Massari et al.*, 2015]. The rationale is that processed satellite SM can be used to update the SM state of rainfall-runoff models, enabling more accurate prediction of catchment response to precipitation and thus better streamflow. These studies have generally shown positive results for reducing streamflow prediction uncertainty, although important limitations have been identified. The limitations influencing the efficacy of the state update schemes include the limited knowledge and skill gaps in structural and parameter uncertainties, the errors in forcing data, the particular runoff mechanisms within the catchment [*Alvarez-Garreton et al.*, 2015], the experimental setup (e.g., model error quantification, observation error quantification, satellite data processing techniques, data assimilation scheme), and the specific catchment characteristics (e.g., soil type, location and land cover) [*Massari et al.*, 2015]. Since the aim of a state correction scheme is to reduce the errors in the model SM, the reduction in streamflow uncertainty will depend on the error covariance between these two components. This error covariance may be weak when the errors in streamflow come mainly from errors in the rainfall input data [*Crow and Ryu*, 2009]. The latter becomes critical in locations without rain gauges, where the available rainfall data generally comes from satellites.

Satellite rainfall products provide near real-time information with high temporal resolution, which can be used for flood forecasting and monitoring. This information, however, contains bias and errors that are usually corrected by using rain gauges [*Yong et al.*, 2013; *Zhou et al.*, 2014; *Yong et al.*, 2015]. To dispense with the need for weather stations (which are not available in large part of the world), recent studies have shown that these products can potentially be improved by using satellite SM observations [*Pellarin et al.*, 2008; *Crow et al.*, 2009; *Brocca et al.*, 2013, 2014; *Zhan et al.*, 2015]. The argument is that given the information that surface SM contains about antecedent rainfall events, the

[1]Department of Infrastructure Engineering, The University of Melbourne, Parkville, Victoria, Australia.
[2]Laboratorio de Dendrocronologa y Cambio Global, Instituto de Conservacin, Biodiversidad y Territorio, Universidad Austral de Chile, Casilla 567, Valdivia, Chile.
[3]Center for Climate and Resilience Research (CR)^2, Chile.
[4]USDA-ARS Hydrology and Remote Sensing Laboratory, Beltsville, Maryland, USA.
[5]CSIRO Land and Water, Highett, Victoria, Australia.

0043-1397 /16/$9.00

magnitude of these events can be estimated by satellite SM retrievals through water balance models. Although these studies have different approaches, they have all shown the potential improvement of rainfall estimates by using satellite SM.

The potential of SM observations to correct errors in both the model states and the forcing data has motivated recent studies to test these dual forcing/state correction schemes (dual SM-DA). For example, *Massari et al.* [2014] set up a scheme in which in-situ SM observations were used to correct the rainfall (through the SM2RAIN algorithm introduced by *Brocca et al.* [2013]) and to initialise the wetness condition of a simple rainfall-runoff model. Their results showed high potential for SM data to improve flood modelling in a case study.

Using a more complex assimilation scheme and rainfall-runoff model, *Crow and Ryu* [2009] set up a state SM-DA scheme integrated with a rainfall correction scheme (via the antecedent version of the soil moisture analysis rainfall tool, SMART, introduced by *Crow et al.* [2009]) in a series of synthetic twin experiments. To prevent the potential introduction of cross-correlation between observations and forecasting errors coming from the dual use of satellite SM, *Crow and Ryu* [2009] applied the rainfall correction offline (i.e., the corrected rainfall is not used within the analysis cycle used to update SM states). The results of this dual SM-DA scheme were further supported by *Chen et al.* [2014] in a real data application over 13 study catchments in the central United States, with areas ranging between 700 and 10,000 km^2. Both studies showed that the satellite rainfall correction led to improvement in streamflow prediction, especially during high flow periods. Conversely, the soil water state correction mainly led to improvement of the base flow component (low flows periods). The combined state and forcing correction scheme led to improvement of both the high and low flow components of the streamflow; outperforming both the state and forcing correction scheme in isolation. However, it remains unclear how this dual SM-DA scheme performs for different catchment characteristics (such as climate and rainfall-runoff mechanisms) and under different experimental conditions (such as the data assimilation setup, model structure and quality of the forcing data).

In this paper we expand the evaluation of the dual SM-DA proposed by *Crow and Ryu* [2009] by using very distinct catchments and different experimental conditions than *Chen et al.* [2014]. In contrast to previous studies [e.g., *Crow and Ryu*, 2009; *Crow et al.*, 2011; *Chen et al.*, 2014; *Massari et al.*, 2014], we focus on large semi-arid catchments in Australia with a history of relatively frequent flooding. Additionally, these catchments are sparsely instrumented thus streamflow prediction is a great challenge. One of the catchments was previously studied by *Alvarez-Garreton et al.* [2014, 2015] while exploring effective state correction schemes for improving flood prediction. In this paper we expand the state correction scheme proposed by *Alvarez-Garreton et al.* [2014, 2015] by incorporating three other catchments and by combining the data assimilation scheme with a rainfall correction scheme. Also, this is the first work that applies the dual data assimilation scheme to the semi-distributed rainfall-runoff modelling. We devise the dual SM-DA scheme under an scenario without rain gauges (only satellite data is used to force the model) to answer four main questions: 1) Can we improve the quality of an operational satellite rainfall product by the assimilation of satellite soil moisture using SMART? 2) Does this rainfall correction scheme have a positive impact on streamflow predictions? 3) Can we improve streamflow prediction by the assimilation of satellite SM in a state correction scheme? 4) What are the impacts on streamflow prediction of a combined state and forcing correction scheme?

To set up the experiments, we use the probability distributed model (PDM) forced with the Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis (TMPA) rainfall products. We assimilate passive and active satellite SM products to correct the PDM soil moisture state (state correction scheme via an ensemble Kalman filter, EnKF) and to correct the satellite rainfall (forcing correction scheme via SMART).

2. Study Area and Data

The study area consists of four catchments in Queensland, Australia: the Warrego, Comet, Thomson and Barcoo (Figure 1). These catchments were selected for their flooding history, along with their low density of rainfall gauge networks. Some of the main characteristics of the catchments, including the mean annual rainfall (calculated using 3B42 dataset, described below), area and stream gauge at the outlet, are summarised in Table 1. The catchments are located in arid, steppe, hot climatic region [*Peel et al.*, 2007] and feature summer-dominated rainfall (Figure 2). Moreover, since the ground-monitoring network within the catchments is sparse (rainfall gauges are shown in Figure 1), satellite SM data is likely to be more valuable than in well-instrumented catchments.

Streamflow records were collected from the State of Queensland, Department of Natural Resources and Mines (*http://watermonitoring.dnrm.qld.gov.au/*) for each outlet gauge (Table 1). Potential evapotranspiration was obtained from the climatological 0.05° gridded data provided by the Australian Bureau of Meteorology (Australian Data Archive for Meteorology database) and daily values were estimated by assuming a uniform daily distribution within a month.

Satellite rainfall data were obtained from the Tropical Rainfall Measurement Mission (TRMM) Multisatellite Precipitation Analysis (TMPA) [*Huffman et al.*, 2007]. We used the 0.25° resolution corrected TMPA research product (3B42, period 01 January 1998 - 31 December 2013) and the near real-time operational product (3B42-RT, period 01 January 2000 - 28 November 2013), which is derived exclusively from satellite-based observations. A daily averaged time series was calculated for each study sub-catchment (sub-catchments delineation are presented in Figure 3). The 3B42-RT product was corrected using SMART (section 3.2). To evaluate the correction scheme, we used the gauge-interpolated rainfall dataset of the Australian Water Availability Project (AWAP) [*Jones et al.*, 2009] as the benchmark rainfall. The near real-time satellite product was also used to force the rainfall-runoff models. These runs were used as the reference to evaluate the different data assimilation schemes (section 3.5).

Satellite SM products were obtained from one active and two passive sensors. The active sensor product was the TU-WIEN (Vienna University of Technology) Advanced Scatterometer (ASCAT, ASC hereafter) data produced using the change-detection algorithm (Water Retrieval Package, version 5.4) [*Naeimi et al.*, 2009], for the period 04 January 2007 - 14 July 2013. One passive sensor product was the Advanced Microwave Scanning Radiometer - Earth Observing System (AMSR-E, AMS hereafter) version 5 VUA-NASA Land Parameter Retrieval Model, Level 3 gridded product [*Owe et al.*, 2008] for the period 29 July 2002 - 03 October 2011. The second passive sensor product was the Soil Moisture and Ocean Salinity satellite (SMOS, SMO hereafter), version RE02 (Re-processed 1-day global SM) product provided by Centre Aval de Traitement des Donnees for the period 16 January 2010 - 01 January 2014.

A daily averaged SM value was calculated for each product over the study sub-catchments (Figure 3). The areal SM estimate over a catchment was calculated by averaging the values of ascending and descending satellite passes on days when more than 50% of the catchment had valid data. For AMS and SMO, we subtracted the long-term temporal mean of the ascending and descending datasets before areal SM estimation to remove the systematic bias between them [*Brocca et al.*, 2011; *Draper et al.*, 2009].

3. Methods

3.1. Rainfall-Runoff Model

The probability distributed model (PDM) is a parsimonious rainfall-runoff model that has been widely used in hydrologic research and applications [*Moore*, 2007]. PDM belongs to the set of models within the flood forecasting system managed by the Australian Bureau of Meteorology. The model estimates a profile average SM (θ) within the catchment (water content of S_1 in Figure 4) by conceptualising the soil water store S_1 with varying capacities across the catchment. In this study, the spatial heterogeneity of the store capacities was represented by a Pareto distribution function. The SM component and the net rainfall (total rainfall minus evaporation and drainage) define the separation between direct runoff and sub-surface runoff. Direct runoff is transformed into surface runoff by two reservoirs (S_{21} and S_{22} in Figure 4). Subsurface runoff is estimated based on the drainage from S_1 and transformed into baseflow by using a one-storage reservoir (S_3 in Figure 4). Surface runoff and groundwater flow are combined as total runoff (streamflow hereafter). A detailed explanation of the model conceptualisation is presented in *Moore* [2007] and the description of the formulation used in this research is provided in *Alvarez-Garreton et al.* [2015].

A semi-distributed scheme using PDM was set up at a daily time step for each of the study catchments (see Figure 3). The time constant parameters of reservoirs S_{21}, S_{22} and S_3 (k_1, k_2 and k_b, respectively) were scaled by the area of each sub-catchment. Following *Alvarez-Garreton et al.* [2015], the river routing between nodes was represented by a linear Muskingum method [*Gill*, 1978] with a storage time constant k_m. This parameter was scaled by the length of the river channel between consecutive nodes. The rest of the model and routing parameters were treated as homogeneous within each study catchment.

To calibrate the model parameters, we forced the model with 3B42 rainfall dataset and used half of its entire period of record to calculate the objective function, which was based on the Nash-Sutcliffe model efficiency (NSE) [*Nash and Sutcliffe*, 1970]. In particular, we divided the complete period in wet and dry years (based on mean annual rainfall from Table 1) and selected half the wet and half the dry years as calibration period. The calibration was done by using a genetic algorithm [*Chipperfield and Fleming*, 1995].

3.2. Forcing Correction Scheme

Following *Crow et al.* [2011], we implemented the soil moisture analysis rainfall tool (SMART) to correct the 3B42-RT satellite rainfall dataset. This corrected rainfall dataset was used to force the PDM in the forcing correction and the dual correction schemes (see schematic in Figure 5). In general terms, SMART uses the antecedent precipitation index (API) model to estimate a SM proxy. This proxy is corrected by using satellite SM observations via a Kalman filter. The Kalman filter innovations are then used to correct the potential errors in the satellite rainfall data used to force the API model. The API model is used here because SMART was shown to perform better when applied to a linear water balance model lacking saturation (which is a non-linear process represented by rainfall-runoff models). This was one of the key findings in *Crow et al.* [2011], where a more complex land surface model was used that did not enhance the correction of rainfall accumulations. The expected improvement coming from a more realistic modeling of soil moisture (which accounted for the energy balance control on soil water loss, soil saturation and runoff generation) was uncertain given the challenges in evapotranspiration modeling approaches and the multilayered model with finite soil water capacities.

The API model at day t was formulated as

$$\mathrm{API}(t) = \gamma(t)\mathrm{API}(t-1) + P(t), \tag{1}$$

where P is the 3B42-RT rainfall data and $\gamma(t)$ is a dimensionless loss coefficient that varies according to the day of the year (D):

$$\gamma(t) = 0.8 + 0.05(2\pi \mathrm{D(t)}/365). \tag{2}$$

The coefficients in (2) aim to capture radiation and climatological temperature effects. Their values were adopted from *Chen et al.* [2014]. Following *Crow and Ryu* [2009], prior to being assimilated into the API model, satellite SM observations were rescaled into the API space by using a cumulative distribution function. When a rescaled AMS, ASC and/or SMO observation (Θ^{ams}, Θ^{asc} and Θ^{smo}, respectively) was available at time t, API was updated using three sequential steps (the time t was omitted in the following equations when all the terms corresponded to the same time step):

1. If Θ^{ams} was available at time t,

$$\mathrm{API}^{+} = \mathrm{API}^{-} + K^{f}(\Theta^{ams} - \mathrm{API}^{-}). \tag{3}$$

The superscripts minus and plus denote before and after updating, respectively. K^f is the Kalman gain (superscript f refers to forcing scheme) calculated at each time step as

$$K^{f} = \frac{T^{-}}{T^{-} + \Sigma}. \tag{4}$$

Where Σ is the scalar error variance of the observation Θ^{ams}, fixed at 0.04^2 $\mathrm{mm^3 mm^{-3}}$. The use of fixed observation error variance follows previous studies applying SMART [*Chen et al.*, 2014; *Crow and Ryu*, 2009; *Crow et al.*, 2011]. Since λ (10) is calibrated for each sub-catchment and each satellite product, the difference in (relative) SM error estimates is compensated within the calibration process. T^- is the scalar error variance for the API forecast at time t, calculated as

$$T^{-}(t) = \gamma^{2}(t)T^{+}(t-1) + Z + \epsilon P^{2}(t). \tag{5}$$

T^+ is the updated API error variance, calculated whenever API was updated, as

$$T^{+} = (1 - K^{f})T^{-}. \tag{6}$$

The term $Z + \epsilon P^2(t)$ in (5) is the model background uncertainty added in each time step, which assumes greater error in API prediction when P is greater than zero. Following *Crow et al.* [2011] and *Chen et al.* [2014], Z was fixed at 3 $\mathrm{mm^2}$ and ϵ at 5 (dimensionless). If Θ^{ams} was not available at time t, $\mathrm{API}^{+} = \mathrm{API}^{-}$, i.e., no correction was done.

2. If Θ^{asc} was available at time t,

$$\mathrm{API}^{++} = \mathrm{API}^{+} + K^{f}(\Theta^{asc} - \mathrm{API}^{+}). \tag{7}$$

In this step, K^f was calculated by (4) where Σ was the scalar error variance of Θ^{asc}, fixed at 0.04^2 $\mathrm{mm^3 mm^{-3}}$. Similarly to step 1, if API^{+} was updated by (7), T^- was also updated by (6). If Θ^{ams} was not available at time t, $\mathrm{API}^{++} = \mathrm{API}^{+}$.

3. If Θ^{smo} was available at time t,

$$\mathrm{API}^{+++} = \mathrm{API}^{++} + K^f(\Theta^{smo} - \mathrm{API}^{++}). \quad (8)$$

Similarly to the previous steps, K^f was calculated by (4), but now using Σ as the scalar error variance of Θ^{smo}, fixed at 0.04^2 $\mathrm{mm^3mm^{-3}}$. If API^{++} was updated by (8), T^- was also updated by (6). If Θ^{smo} was not available at time t, $\mathrm{API}^{+++} = \mathrm{API}^{++}$.

After the above 3-step updating scheme, the analysis increments δ were defined for each time step t as

$$\delta = \mathrm{API}^{+++} - \mathrm{API}^{-}. \quad (9)$$

Following *Crow et al.* [2011], the rainfall accumulation $[P]$ was corrected by

$$[P]^c = [P] + \lambda[\delta]. \quad (10)$$

The superscript c denotes after correction. The square brackets represent non-overlapping accumulation windows. To ensure that SM observations corrected only past rainfall accumulations, the length of these windows was varied so that the last day of the window had a SM observation. The parameter λ in (10) is constant in time, and was calibrated for each sub-catchment and each satellite SM product by minimising the root-mean-square error between $[P]^c$ and the 3B42 rainfall product. Negative values of $[P]^c$ were reset to zero.

Following *Chen et al.* [2014], we applied a 2 mm threshold value for rainfall correction when $[P]$ was zero (i.e., the correction step in (10) was done only if $\lambda[\delta] > 2$). The latter is done since SMART analysis tends to create spurious very low rainfall due to positive noise in the SM observations, which results in increased false alarm ratios in rainfall events. The consequence of this is that real rainfall values smaller than 2 mm can be discarded, however, these low-intensity rainfall values are unlikely to make significant impacts on the streamflow prediction with the given high evapotranspiration of the study regions.

To get a daily corrected rainfall time series, we redistributed the corrected accumulations in proportion to the original daily rainfall. To remove positive bias due to the resetting-to-zero step, the corrected rainfall was multiplicatively rescaled to match the long-term mean of P. The corrected 3B42-RT dataset using SMART is called $\mathrm{3B42\text{-}RT^C}$.

3.3. State Correction Scheme

To set up the state correction scheme (schematic in Figure 5), we followed the procedure developed by *Alvarez-Garreton et al.* [2015]. The rainfall dataset used in this scheme was the uncorrected 3B42-RT. In summary, the scheme consisted of using satellite SM observations to correct the model SM (θ hereafter) via a stochastic data assimilation framework using an ensemble Kalman filter (EnKF) [*Evensen*, 2003]. Prior to being assimilated, the satellite data was processed to provide consistent information about θ's dynamics. In the following we provide a description of the satellite data processing and the EnKF implementation.

3.3.1. Satellite Soil Moisture Data Processing

Given the active and passive sensors microwave penetration depths, satellite SM observations represent only the top few centimetres of soil. Furthermore, θ is an average profile SM representing a deeper layer. Therefore, the depth that θ represents depends on soil properties and model parameters, and is unique for each sub-catchment and PDM scheme. By assuming a porosity for the study catchments ranging between 0.45 and 0.5 (A-horizon information reported in *McKenzie et al.* [2000]) and S_1 storage capacity ranging from 200 to 280 mm (obtained from calibrated model parameters for the different catchments), θ represents roughly a depth varying between 400 and 600 mm.

To address the depth mismatch between satellite and model, we applied the exponential filter proposed by *Wagner et al.* [1999] to the satellite SM observations and obtained a soil wetness index (SWI) of the root-zone. The use of SWI to characterise the dynamics of the root zone SM based on surface observations has been successfully evaluated in a number of studies [e.g., *Albergel et al.*, 2008; *Brocca et al.*, 2009, 2010; *Ford et al.*, 2013]. We calculated the SWI for each satellite dataset and each sub-catchment by using the following recursive formulation:

$$\mathrm{SWI}(t) = \mathrm{SWI}(t-1) + G(t)\left(\mathrm{SSM}(t) - \mathrm{SWI}(t-1)\right) \quad (11)$$

where t is the daily time step, SSM is the satellite observation (AMS, ASC or SMO) and G is a gain term varying between 0 and 1 calculated as

$$G(t) = \frac{G(t-1)}{G(t-1) + e^{-\left(\frac{t-(t-1)}{T}\right)}}. \quad (12)$$

The parameter T accounts for several physical parameters defining infiltration and percolation processes [*Albergel et al.*, 2008]. T was calibrated for each satellite product and each sub-catchment. The calibration was done by maximising the correlation coefficient between SWI and the model SM (θ) over the entire period of record of the corresponding satellite product.

Once the SWI was calculated for AMS, ASC and SMO, we applied instrumental variables (IV) regression [*Su et al.*, 2014] to remove the systematic (multiplicative and additive) biases between each SWI and θ and to estimate the observation errors. Following *Alvarez-Garreton et al.* [2015], we applied the triple collocation (TC) analysis [*Stoffelen*, 1998; *Yilmaz and Crow*, 2013] to rescale the SWI and estimate its observation error variance. The TC-based method has been used as an optimal rescaling method and error estimator if assumed assumptions are met [*Yilmaz and Crow*, 2013] and it has been increasingly applied in hydrologic data assimilation applications [*Alvarez-Garreton et al.*, 2015; *Chen et al.*, 2014; *Crow and Yilmaz*, 2014]. Data triplets for TC comprised of the model θ and two SWI time series (derived from a passive and an active sensor, respectively). We implemented TC with an imposed threshold sample of 100 [*Scipal et al.*, 2008]. For the periods where only one satellite product was available, or when the threshold for TC triplets was not met, a two-data IV regression was used as a practical substitute. The two-data IV, also known as lagged variables (LV) [*Su et al.*, 2014], was applied to the model θ, a single satellite SWI, and a 1-day lagged variable coming from the model θ.

As a simplification of the seasonal approach proposed by *Alvarez-Garreton et al.* [2015], in this study we applied TC and LV to the complete period of record of AMS, ASC and SMO. This bulk approach provided a scalar observation error variance and constant rescaling factors for each of the SWI datasets. The rescaled SWI datasets for AMS, ASC and SMO were named θ^{ams}, θ^{asc} and θ^{smo}, respectively. Finally, ensembles of 500 members were calculated for the three rescaled datasets ($\boldsymbol{\theta}^{ams}$, $\boldsymbol{\theta}^{asc}$ and $\boldsymbol{\theta}^{smo}$, respectively) by adding a Gaussian noise with mean zero and the error variance obtained from the TC and LV analyses. As discussed in section 5, the adopted bulk estimations may have

implications for the observation error characterisation and data assimilation results.

3.3.2. EnKF Formulation

In the EnKF, the errors in the model and the observations are calculated from Monte-Carlo ensemble realisations. To implement the state-correction scheme, every time there was a 500-member ensemble of observations available ($\boldsymbol{\theta}^{ams}$, $\boldsymbol{\theta}^{asc}$ and/or $\boldsymbol{\theta}^{smo}$ estimated via the satellite data processing described in section 3.3.1), each member of a 500-member ensemble of predictions ($\boldsymbol{\theta}$) at time t was sequentially updated using (t was omitted from the following equations since all the terms corresponded to the same time step):

1. If $\boldsymbol{\theta}^{ams}$ was available at time t,

$$\theta_i^+ = \theta_i^- + K(\theta_i^{ams} - H\theta_i^-). \tag{13}$$

The subscript i indicates a member of the ensemble of predictions and observations. The minus and plus denote θ_i values before and after updating, respectively. The H is an operator that transforms the model state into the measurement space. Given the pre-processing applied to the satellite SM products (section 3.3.1), H reduced to a unit matrix (and therefore was omitted from the following equations). The Kalman gain K was calculated at each time step as

$$K = \frac{C}{C+E}, \tag{14}$$

where E is the θ^{ams} rescaled observation error variance estimated from IV analyses (section 3.3.1) and C is the scalar error covariance of the background prediction $\boldsymbol{\theta}^-$. C was calculated at each time step as

$$C = \frac{1}{N-1}\left(\boldsymbol{\theta}^- - \langle\theta^-\rangle\right)\cdot\left(\boldsymbol{\theta}^- - \langle\theta^-\rangle\right)^T, \tag{15}$$

where $\langle\theta^-\rangle$ is the ensemble mean at time t. If $\boldsymbol{\theta}^{ams}$ was not available, $\boldsymbol{\theta}^+ = \boldsymbol{\theta}^-$, i.e., no correction was done.

2. If $\boldsymbol{\theta}^{asc}$ was available at time t,

$$\theta_i^{++} = \theta_i^+ + K(\theta_i^{asc} - \theta_i^+). \tag{16}$$

To calculate K we applied (14), where E corresponded to the θ^{asc} observation error variance estimated from IV analyses and C was re-calculated by applying (15) to the updated soil moisture $\boldsymbol{\theta}^+$. If $\boldsymbol{\theta}^{asc}$ was not available, $\boldsymbol{\theta}^{++} = \boldsymbol{\theta}^+$.

3. If $\boldsymbol{\theta}^{smo}$ was available at time t,

$$\theta_i^{+++} = \theta_i^{++} + K(\theta_i^{smo} - \theta_i^{++}). \tag{17}$$

Consistently with the previous steps, K was calculated by (14), where E corresponded to the θ^{smo} observation error variance and C was calculated by applying (15) to the updated soil moisture $\boldsymbol{\theta}^{++}$. If $\boldsymbol{\theta}^{smo}$ was not available, $\boldsymbol{\theta}^{+++} = \boldsymbol{\theta}^{++}$.

During the sequence of three updating steps, each subcatchment was treated independently and no spatial cross-correlation in the satellite measurements was considered. The order of the satellite products used in the 3-step sequential assimilation was arbitrary, however different orders were tested with no significant variation in results.

Following *Alvarez-Garreton et al.* [2015], to generate the model background ensemble prediction ($\boldsymbol{\theta}^-$) we applied unbiased perturbations to the rainfall forcing data, the model parameter k_1 and the model SM prediction ($\boldsymbol{\theta}$). These perturbations aimed to represent the main sources of model error, coming from the forcing data, the model parameters and the model structure.

The error models adopted for each perturbation followed a number of previous SM-DA experiments [e.g., *Chen et al.*, 2011; *Brocca et al.*, 2012; *Alvarez-Garreton et al.*, 2014] and consisted in a serially independent log normal multiplicative error (mean 1, standard deviation σ_p) for the rainfall data, a serially independent Gaussian additive error (mean 0, standard deviation σ_k) for parameter k_1, and a serially independent Gaussian additive error (mean 0, standard deviation σ_s) for the model soil moisture. To avoid truncation biases while applying the $\boldsymbol{\theta}$ perturbation, we implemented the bias correction scheme proposed by *Ryu et al.* [2009] to the SM ensemble. This bias correction ensures unbiased state ensembles; however, given the non-linear processes represented by the hydrological model, the perturbation process can still generate biased streamflow ensemble prediction. This can degrade the performance of the EnKF [*Ryu et al.*, 2009]. To remove the biases in streamflow caused by the forcing and state ensemble perturbations, we followed *Alvarez-Garreton et al.* [2015] and also applied the bias correction scheme to the streamflow ensembles (of the subcatchment outlets and routing channels).

The error model parameters (σ_p, σ_s and σ_k) were assumed to be homogeneous within each study catchment and were calibrated using a maximum *a posteriori* likelihood approach (MAP) [*Wang et al.*, 2009] for the period 01 January 1998 - 31 December 2003. MAP has been used as an objective method to estimate reliable model error parameters [*Alvarez-Garreton et al.*, 2015; *Li et al.*, 2014]. With this approach we maximised the likelihood (aggregated over time) of having the observed streamflow within the model streamflow ensemble prediction. In the MAP scheme, the error in the observed streamflow at the outlet of each study catchment was assumed to follow a serially independent multiplicative gaussian error (mean 1, standard deviation 0.2). Further details about model error formulations and MAP calibration can be found in *Alvarez-Garreton et al.* [2015].

3.4. Dual Correction Scheme

The dual correction scheme combined the forcing correction scheme (section 3.2) with the state correction scheme (section 3.3). As illustrated in Figure 5, the streamflow prediction of the dual scheme was obtained by forcing PDM with the corrected rainfall 3B42-RT^C coming from SMART. When a satellite observation was available and $\boldsymbol{\theta}$ from the state correction scheme was updated, the states of the dual run were replaced by the the average of $\boldsymbol{\theta}^+$ ensemble (and the average of the model ensemble states that were not updated) from the state correction scheme. Following *Crow and Ryu* [2009], the state outputs of the dual correction scheme were not fed back to the state correction scheme to avoid cross-correlation between model and observation errors.

3.5. Schemes Evaluation

The reference model run used to evaluate the different correction schemes was the unperturbed model forced with 3B42-RT dataset. The use of the near-real time satellite rainfall product to force the reference run reflects our aim to evaluate the efficacy of the correction schemes under a data scarce scenario.

Given that the streamflow from the reference run, the forcing correction scheme (section 3.2) and the dual correction scheme (section 3.4) are deterministic predictions, the state correction scheme (which provides an ensemble of updated streamflow predictions) was evaluated in terms of its ensemble mean.

The evaluation period was July 2002 to November 2013, which was determined by the availability of the satellite datasets (section 2). The streamflow prediction from the reference run and the 3 correction schemes (forcing correction, state correction and dual correction schemes) were evaluated based on the Nash-Sutcliffe efficiency (NSE) [*Nash and Sutcliffe*, 1970], the root mean square error (RMSE) and the correlation coefficient (R). In particular, we calculated the difference (in percentage) between the statistics of the streamflow from the different correction schemes and the reference run.

To evaluate the improvement from the different schemes during high and low flows periods, the three statistics were calculated in natural space (more sensitive to high flows) and log-transformed space (more sensitive to low flows). Following *Massari et al.* [2015], to avoid exclusion of the zero-flow periods when applying the log transformation, an arbitrary fraction of the mean daily observed streamflow ($\overline{Q}_{obs}/40$) was added to the streamflow time series (observed and simulated) before calculating their logarithm.

Additionally, we estimated the false alarm ratio (FAR) as the number of times (#) the model streamflow prediction exceeded a threshold value (Q^*), while the observed streamflow was less than Q^*:

$$\mathrm{FAR} = \frac{\#(Q_{sim} >= Q^* \ \& \ Q_{obs} < Q^*)}{\#(Q_{obs} < Q^*)}. \qquad (18)$$

Q^* was set as the daily flow rate corresponding to a minor flood classification. The flood classification for the study catchments was provided by the Australian Bureau of Meteorology as river height threshold values. These relate to flood impact rather than recurrence interval. The threshold values for minor floods, expressed as streamflow (mm day^{-1}) for Warrego, Comet, Thomson and Barcoo catchments are 0.06, 0.1, 0.02 and 0.14 , respectively. Similarly, we estimated the probability of detection (POD) of these flow rates as

$$\mathrm{POD} = \frac{\#(Q_{sim} >= Q^* \ \& \ Q_{obs} >= Q^*)}{\#(Q_{obs} >= Q^*)}, \qquad (19)$$

The rainfall correction scheme was further evaluated in terms of the corrected rainfall dataset. For this we used the gauge-interpolated AWAP rainfall as a benchmark dataset and calculated 5 statistics for the 3B42-RT dataset before and after SMART correction. The statistics used here were the mean daily bias, the coefficient of determination R^2, the RMSE, the FAR and the POD. In this case, FAR and POD were calculated based on daily rainfall threshold values specified in section 4.1.

4. Results

4.1. Rainfall Correction

To categorise the evaluation of SMART into meaningful ranges, we analysed the histograms of daily rainfall for the benchmark rainfall dataset (AWAP). Figure 6 shows the frequency of daily rainfall accumulations within four representative sub-catchments (each from one of the four study catchments). To calculate the histograms we filtered daily accumulations greater than 2 mm, which resulted in 532, 658, 444 and 553 daily records for the four sub-catchments, respectively (panels a to d in Figure 6). These plots reveal that for all the sub-catchments more than 50% of the daily rainfall values are within the first histogram bin, which corresponds to 2 to 7 mm (each histogram bin corresponds to a 5 mm increment). Almost 20% (slight variation across the catchments) of the daily records range between 7 and 12 mm and the rest is distributed above 12 mm. Based on this, the evaluation statistics of the forcing correction scheme (section 3.5) were calculated within these three ranges. FAR and POD in particular were calculated with (18) and (19), using the upper and lower ranges bounds as threshold values, respectively.

To illustrate the tendency of the satellite rainfall estimates to over or under predict daily rainfall events, in Figure 7 we present the mean daily bias of the rainfall before and after SMART correction (using AWAP as the benchmark dataset) for each sub-catchment within the 4 study catchments (29 sub-catchments in total, see Figure 3). The plots in Figure 7 do not represent long-term biases, but rather the mean daily bias within specific evaluation ranges. Before SMART correction, the near real-time satellite product generally over predicted daily accumulations of rainfall for the low-to-mean ranges (panels a and b). For the high rainfall events (panel c), the behaviour of the satellite product before correction was the opposite, there was a fairly consistent under-prediction of daily accumulations. This over-prediction of low rainfall events and under-prediction of high rainfall events is consistent with the literature [*Ebert et al.*, 2007].

To interpret the plots in Figure 7, we should recall that SMART is formulated to reduce the random component of the error in the satellite rainfall data, and thus a reduction in long-term biases should not be expected. However, since the plots illustrate the mean biases within specific ranges, some variation of the general tendency to over or under predict rainfall events at some locations was identified after SMART correction. For the lowest rainfall range (panel a), the over-prediction of daily accumulations was reduced within 90% of the catchments after applying SMART (up to 15% of reduction in positive biases). For the second rainfall range (panel b) the impacts of SMART correction were not so consistent. Almost half of the cases where there was over prediction of rainfall estimates (20 in total) were improved after applying SMART. Similarly, almost half of the under prediction cases (8 in total) were improved after SMART correction (e.g., 2 sub-catchments within Warrego, 1 sub-catchment within Comet and 1 sub-catchment within Thomson). For the high rainfall range (panel c) there was consistent negative bias in 3B42-RT, which was not reduced after applying SMART. The latter could be due to the limited information that SM provides when the surface soil gets saturated, and to the tendency of under-estimate peak rainfall events coming from SMART's core formulation (further discussion in section 5).

Figure 8 presents the statistics calculated for 3B42-RT before (x-axes) and after SMART correction (y-axes), for the 3 rainfall ranges defined above and for the 29 study sub-catchments. Additionally, the R^2 and the RMSE were calculated for the complete range of daily rainfall accumulations. The coefficient of determination R^2 between 3B42-RT and AWAP for the 3 rainfall ranges (panels a, b and c in Figure 8) and for the complete rainfall accumulation range (panel d) increased after SMART correction. The improvement was moderate, but consistent throughout most catchments (it should be noted that different scales in x- and y-axes were used for the different rainfall ranges). The low R^2 values in panel a are consistent to the larger errors found in 3B42-RT product for daily accumulations below 10 mm [*Pipunic et al.*, 2015]. In terms of the RMSE, the forcing correction scheme reduced the satellite rainfall data error, across all catchments and all daily rainfall ranges (panels e, f, g and h in Figure 8). The contrasting results between reduced RMSE (panel g in Figure 8) and increased mean biases (panel c in Figure 7) for the high rainfall range is consistent with the SMART core formulation. SMART was

effective at reducing the error variance of the rainfall estimates (random component of the satellite rainfall error), which is reflected by a reduction in RMSE; however, the existing biases in the original dataset were not significantly impacted by the scheme. Positive results after SMART correction were also observed in the increased POD (panels i, j and k) and decreased FAR (panels l and m) statistics, across all catchments and all daily rainfall ranges.

Based on the complete range of daily rainfall events, improvement in R^2 and reduction in RMSE (panel d and h, Figure 8) are consistent, and similar in magnitude, with the evaluation of SMART made by *Chen et al.* [2014] within 13 U.S. basins.

In summary, although SMART led to increased rainfall mean biases at some locations for specific ranges of rainfall events (Figure 7), overall the scheme improved the quality of the satellite rainfall data (Figure 8). In particular, the positive impacts within high rainfall events (increased R^2 and POD, decreased RMSE) suggest that this could be a suitable scheme to improve the prediction of high streamflow events.

4.2. Satellite Data Processing

Figure 9 summarises some of the satellite data processing results, panel a shows the parameter T from equation 12 that maximised the correlation coefficient (presented in panel b) between the model SM and SWI for each sub-catchment and each satellite product. The parameter T varied within a range of 3 to 42 days, which is consistent with the range of values found in previous studies [e.g., *Albergel et al.*, 2008; *Brocca et al.*, 2009; *Ford et al.*, 2013]. These variations in T could be due to a series of factors, including the particular sub-catchment physical processes, the retrieval method of the satellite product, the quality of the SM predicted by the model, and the different periods of time used for the calibration. Across all sub-catchments, and similarly to previous findings [*Alvarez-Garreton et al.*, 2015], T values were larger for the SMO product, which would be inconsistent with L-band having a deeper penetration than AMS C-band (to limit the comparison within passive retrievals). This might be due to factors including the different retrieval methods (which have quite different assumptions pertaining to spatial heterogeneity) and the influence of radio-frequency interference noise.

The observation error variances for SWI derived from AMS, ASC and SMO, respectively, are presented in panel c, Figure 9. SMO-derived SWI generally outperformed the others two sensors, which is consistent with its higher correlation with the model (panel b). The passive AMS product showed the largest error across the study sub-catchments. It should be noted that the errors presented in Figure 9 come from TC analyses, the results from LV procedure (applied when there was only one satellite product available or when the sample threshold for TC was not met) maintained a similar comparative relationship among sensors; however, the magnitude of the error was consistently higher. This over-estimation of the observation errors by LV is consistent to previous studies [*Su et al.*, 2014; *Alvarez-Garreton et al.*, 2015] and it is likely to be explained by error autocorrelation of the lagged variables used in the triplets. This will have the impact of giving greater weights to the model predictions in the assimilation.

4.3. Streamflow Prediction Evaluation

The statistics of the reference runs used to evaluate the different assimilation schemes are presented in Table 2. It can be seen from this table that the quality of the streamflow prediction in most of these catchments was poor, with low NSE and R values (calculated using both the raw and the log-transformed streamflow values). The only catchment that showed good quality statistics is the Comet. The poor performance of the model in the study catchments was mainly due to the low quality of the forcing rainfall data (the near real-time 3B42-RT product). This was confirmed by the higher statistics obtained after calibrating the models with the (higher quality) gauged-interpolated AWAP dataset (presented as reference in Table 3). The relevance of using these reference runs is that they represent the data scarce scenario within most areas in the world. The calibrated model error parameters (σ_p, σ_s and σ_k) for the study catchments are presented in Table 4. It can be seen from this table that Comet catchment presents the lowest error in rainfall, which is consistent with the better performance in streamflow prediction.

The results of the different assimilation schemes are presented in Figure 10. Overall, the use of (processed) satellite SM to correct the model SM and/or the forcing rainfall led to an improvement over the reference model runs. In terms of NSE and RMSE, there was a wide range of improvement for both the high flows (panels a, e) and the low flows (panels b, f). In 3 out of 4 high flow cases and all the low flow cases, the state correction scheme outperformed the forcing correction scheme. The combined dual scheme further improved the results, irrespective of wether the high flows or low flows were emphasised in the evaluation. The only case where this relation changed was for the high flows in Comet catchment (panel a), where the forcing scheme consistently showed a greater positive impact in streamflow prediction. Based on the improvements in R, the dual correction scheme in general outperformed the other two schemes. This is the true for 2 out of 4 high flow cases (panel c) and for 3 out of 4 low flow cases (panel d).

The dual correction scheme also led to a consistent decrease of the FAR (panel g in Figure 10) within all the study catchments. By applying this dual correction scheme, the number of incorrectly predicted minor floods was reduced by 10 to 30%. If these predictions were to be applied to feed operational flood alert systems, this improvement in FAR would have a significant impact. In terms of POD (panel h), the data assimilation schemes had a much lower (negative and positive) impact on the streamflow prediction (less than 10% of POD variation). The POD is only improved in Comet catchment, where the dual scheme showed the highest effect. For the other catchments there was a decrease of POD after the state and dual correction schemes.

5. Discussion

The results presented here demonstrate that active and passive satellite SM retrievals have the potential to improve an operational satellite rainfall product (3B42-RT). We also showed that assimilating the satellite SM observations into the streamflow modelling generally had positive impacts in the quality of the streamflow prediction. Finally, by combining the forcing and state correction schemes we further improved the streamflow predictions for most study catchments.

We implemented the dual correction scheme with some key methodological configurations that differ from previous studies [*Crow and Ryu*, 2009; *Chen et al.*, 2014; *Massari et al.*, 2014]. These included the rainfall-runoff model used, the semi-distributed model configuration, and the MAP error model estimation procedure. Notwithstanding these methodological differences, the outcomes are consistent with previous studies [*Crow and Ryu*, 2009; *Chen et al.*, 2014; *Massari et al.*, 2014] where the combined dual correction scheme enhanced the benefits of the individual state and forcing correction schemes. This suggests that using passive satellite SM retrievals would a robust tool for improving an

operational satellite rainfall production and streamflow prediction.

Regarding SMART implementation, we used a fixed observation error variance (section 3.2). The representativity issue of this simplification was partially compensated by the calibration of λ (10), which reduces the difference in (relative) observation error estimates. The correction parameter λ was calibrated for product separately to account for the difference in the microwave soil moisture retrieval errors.

Overall, the assimilation of SM retrievals via SMART improved the rainfall estimates over the study catchments, with a decrease in RMSE and FAR, and an increase in R^2 and POD within most sub-catchments (Figure 8). Similarly to previous studies [*Crow et al.*, 2011; *Chen et al.*, 2014], SMART showed limitations during wet conditions. This resulted in the under-prediction of some high intensity rainfall events (increased negative biases in panel c, Figure 7). As mentioned in section 4.1, this could be due to the limited information about rainfall that SM provides when the surface soil is wet. This key issue affects not only SMART, but also other correction schemes aiming to estimate (or reduce the error in) rainfall based on SM information [e.g., *Brocca et al.*, 2013; *Zhan et al.*, 2015]. Another reason for the under-prediction of rainfall peaks could be the precipitation error variance minimisation approach used by SMART (the Kalman filter). It has been shown that an error variance minimisation algorithm increases the conditional bias of the rainfall estimates, which is manifested by an underestimation of strong rainfall [*Ciach et al.*, 2000].

When the SM was near the lower limit of the volumetric water content (dry conditions), SMART consistently reduced the over-prediction of small rainfall events (decreased positive bias in panel a, Figure 7). This suggests that, in contrast to previous studies [e.g., *Crow et al.*, 2011; *Chen et al.*, 2014], the noise in the SM retrieval signal was not misinterpreted by SMART as rainfall. The better performance of SMART during low intensity rainfall events could be explained by the different climatology of our study region and by the different quality of the satellite products (rainfall and SM products).

The impact of SMART correction in the streamflow modelling was assessed in section 4.3. The overall improvement of the forcing data via SMART was successfully transferred into the streamflow modelling. Similar to previous studies [*Crow and Ryu*, 2009; *Chen et al.*, 2014], during high flow periods SMART led to a consistent positive impact on the streamflow modelling across all catchments (increased NSE, R and a reduced RSME in Figure 10). Therefore, even when some peak rainfall values were under predicted at some locations (negative mean biases in Figure 7), the corrected rainfall featured consistently lower errors than the near real-time satellite rainfall (Figure 8). This resulted in an improved overall performance of the rainfall-runoff model after SMART rainfall correction. The low flow estimations were also improved after SMART; however, the improvement was less significant than during high flows. This was expected given the higher control that rainfall exerts in the streamflow generation during intense events.

The correction of the model SM state by the assimilation of satellite SM led to a significant improvement in the prediction of low flows (panels b, d and f in Figure 10). The improvement during high flows was less for most cases (panels a, c and e in Figure 10), which is consistent with the higher control that the catchment wetness condition has in the streamflow generation during low flow periods. Our results agree with various studies demonstrating the potential of these observations for enhancing streamflow modelling [e.g., *Brocca et al.*, 2010; *Alvarez-Garreton et al.*, 2014; *Wanders et al.*, 2014; *Alvarez-Garreton et al.*, 2015; *Massari et al.*, 2015]. Notwithstanding this evidence, there are key choices to be made in setting up the SM data assimilation schemes that can have significant impacts on their results. As clearly described by *Massari et al.* [2015], these schemes are highly influenced by local conditions and methodological issues. The latter should be carefully taken into consideration before drawing general conclusions.

Regarding our methodology, the first key step to set up the state correction scheme was the satellite data processing (section 3.3.1). The use of the exponential filter to estimate the SWI of the root zone based on surface observations was a simple solution that has shown positive results in several studies [e.g., *Albergel et al.*, 2008; *Brocca et al.*, 2009; *Ford et al.*, 2013]. However, there are some issues related to the autocorrelation in the observation errors and the potential cross-correlation with the model SM errors that have been highlighted when SWI is used within a data assimilation scheme [*Brocca et al.*, 2010; *Alvarez-Garreton et al.*, 2015]. There is an important research gap here since the implications of the latter issues have not yet been assessed, and the use of other profile SM estimation methods [e.g., *Richards*, 1931; *Manfreda et al.*, 2014] have not been tested within this data assimilation context. The rescaling of the observations and the quantification of their errors was performed here by a triple collocation based approach (TC and LV detailed in section 3.3.1), which has been assessed as an optimal rescaling procedure if assumptions are met [*Yilmaz and Crow*, 2013]. In particular, we applied TC and LV to the complete observation period (bulk estimation of rescaling parameters and observation errors), which does not consider the temporal variability (e.g., seasonality) in the observation errors [*Draper and Reichle*, 2015; *Su and Ryu*, 2015]. This simplification could potentially lead to overcorrection of the model state if the actual error is higher, and vice versa. To address this, some studies have applied seasonal rescaling and error estimation [*Alvarez-Garreton et al.*, 2015] or have separately treated anomalies and seasonality within the TC implementation [*Chen et al.*, 2014]. Despite these attempts to address the temporal variation in the observation errors, further investigation is required to assess the impacts of rescaling assumptions and simplifications in satellite SM data assimilation.

A practical implication of the highlighted limitations within the satellite data processing is that lower errors estimated for a particular dataset (e.g. SMO-derived SWI in Figure 9) do not necessarily imply a better performance of the product in the data assimilation schemes. Therefore, a comparative assessment of the skill of the different satellite SM products to improve streamflow prediction in the proposed schemes cannot be drawn from these results (such comparison would require to run the assimilation schemes independently for the different satellite products). Acknowledging this limitation, the benefit of sequentially assimilating the three satellite products is that since we are using a statistically optimal updater, integrating multiple observations should provide better results. Additionally, given the different period of record of the SM products, using the three products enables a longer evaluation period for the assimilation schemes.

The second key step in the state correction scheme was the representation and quantification of the model errors, which has a direct impact in the data assimilation results [*Massari et al.*, 2015]. There are a number of methods to quantify model errors such as the assumption of arbitrary error parameter values [*Chen et al.*, 2014], the maximisation of ensemble verification criteria [*De Lannoy et al.*, 2006; *Brocca et al.*, 2010; *Massari et al.*, 2015], the auto-tuned land data assimilation system proposed by *Crow and Yilmaz* [2014] and the maximisation of the likelihood of having the streamflow observations within the streamflow ensemble prediction [*Alvarez-Garreton et al.*, 2015; *Li et al.*, 2014] (adopted in this study). The evaluation of these techniques and their

impacts on the assimilation of SM into rainfall-runoff models has not been studied deeply. In particular, the reliability and quality of the generated open-loop ensembles (in Monte Carlo-based applications) used to evaluate the data assimilation results are usually not assessed. In our case, since we based the evaluation on deterministic predictions (as explained in section 3.5), the skill of the stochastic state correction scheme in terms of ensemble prediction characteristics was not assessed.

Despite the highlighted limitations and challenges within the forcing and state correction schemes, our experiments demonstrated that the streamflow prediction for these sparsely gauged locations is improved by the assimilation of satellite soil moisture. The state correction scheme generally showed higher positive impact on the streamflow prediction than the forcing correction scheme, for both high flows and low flows. This larger improvement during high flows contrasts with previous studies [*Crow and Ryu*, 2009; *Chen et al.*, 2014] where the forcing correction scheme generally outperformed the state correction for high flows. This could be due to several factors including differences in rainfall-runoff mechanisms between catchments (streamflow in these catchments features long periods of zero- flow, a negligible base flow component and sharp flow peaks after rainfall events, when the catchments have reached a threshold level of wetness [*Alvarez-Garreton*, 2015]). Other factors may include a different quality of the forcing data before correction, different quality of the satellite SM products and the different experimental methodologies. Finally, we showed that the combination of a better representation of the catchment wetness condition (via state correction scheme) with higher quality forcing data (via forcing correction scheme) in most cases outperformed the results of separately applying either data assimilation scheme (Figure 10).

It should be noted that while the proposed schemes were able to improve the quality of an operational satellite rainfall product and the PDM SM state, which in turn led to better streamflow predictions, the streamflow predictions after the dual SM-DA scheme still did not outperform the case where the gauge-based rainfall data was used as input forcing (evaluation scores presented in Table 3). This implies that while satellite SM may be useful for improving satellite rainfall products and SM states of hydrological models within data scarce regions, it is more critical to have a higher quality forcing data for accurate streamflow prediction in the study regions.

6. Conclusions

We explored the use of active and passive satellite SM products for improving the streamflow prediction of a rainfall-runoff model (PDM) within 4 large semi-arid catchments. We set up our experiments under a scenario without rain gauges to represent the data scarcity in most areas worldwide, which led to poor streamflow predictions before assimilation. Within this context, two key variables controlling the runoff generation were corrected by the assimilation of the surface SM observations: the satellite rainfall forcing data and the PDM soil moisture state.

The forcing correction used SMART [*Crow et al.*, 2011] and the results showed a consistent improvement in the the operational satellite rainfall (increased R and reduced RMSE and mean bias). In general, the use of the corrected rainfall data to force the rainfall-runoff models improved the streamflow prediction (increased NSE, R^2 and decreased RMSE), especially during high flow periods.

The state correction scheme generally showed a higher positive impact on the streamflow prediction compared with the forcing correction scheme, especially for low flows. The combined dual correction scheme enhanced the benefits of the individual schemes, which led to an improved prediction of both low and high flows. Notwithstanding this, the updated streamflow predictions did not outperform the case where the gauge-based rainfall data was used as input forcing.

We have highlighted a number of limitations within the forcing and state correction schemes that should be addressed to advance towards a robust data assimilation framework. Although our results are case specific and depend on the catchment characteristics, degree of instrumentation and the experimental set up, they provide new evidence of the value of satellite SM for improving both an operational satellite rainfall product and the streamflow prediction within data scarce regions.

Acknowledgments. This research was conducted with financial support from the Australian Research Council (ARC Linkage Project No. LP110200520) and the Bureau of Meteorology, Australia. C. Alvarez-Garreton was supported by a Becas Chile scholarship and by CONICYT/FONDAP/1511000. We are grateful to all who contributed to the data sets used in this study. We thank Chris Leahy and Soori Sooriyakumaran from the Australian Bureau of Meteorology for providing catchment data and gratefully acknowledge their advice. AWAP rainfall data was downloaded from `http://www.bom.gov.au/jsp/awap/rain/archive.jsp`. AMSR-E data was downloaded from `ftp://hydro1.sci.gsfc.nasa.gov/data/s4pa/WAOB/LPRM_AMSRE_A_SOILM3.002/` (see also `http://gcmd.nasa.gov`). ASCAT data can be downloaded from `http://hsaf.meteoam.it/soil-moisture.php`. The SMOS data was downloaded from `http://www.catds.fr/Products/Available-products-from-CPDC`. The TMPA data was downloaded from `http://disc2.nascom.nasa.gov/opendap/TRMM_L3/TRMM_3B42/`. We also thank the anonymous reviewers and associate editor for their comments which have improved the quality of this paper.

References

Albergel, C., C. Rüdiger, T. Pellarin, J.-C. Calvet, N. Fritz, F. Froissard, D. Suquia, A. Petitpa, B. Piguet, E. Martin, et al. (2008), 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.

Albergel, C., C. Rüdiger, D. Carrer, J.-C. Calvet, N. Fritz, V. Naeimi, Z. Bartalis, and S. Hasenauer (2009), An evaluation of ascat surface soil moisture products with in-situ observations in southwestern france, *Hydrology and Earth System Sciences*, *13*(2), 115–124.

Alvarez-Garreton, C., D. Ryu, A. W. Western, W. T. Crow, and D. E. Robertson (2013), Impact of observation error structure on satellite soil moisture assimilation into a rainfall-runoff model, in *MODSIM2013, 20th International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand*, edited by J. Piantadosi, R. Anderssen, and J. Boland, pp. 3071–3077.

Alvarez-Garreton, C., D. Ryu, A. W. Western, W. T. Crow, and D. E. Robertson (2014), The impacts of assimilating satellite soil moisture into a rainfall-runoff model in a semi-arid catchment, *Journal of Hydrology*, *519*, 2763–2774.

Alvarez-Garreton, C., D. Ryu, A. W. Western, C.-H. Su, W. T. Crow, D. E. Robertson, and C. Leahy (2015), Improving operational flood ensemble prediction by the assimilation of satellite soil moisture: comparison between lumped and semi-distributed schemes, *Hydrology and Earth System Sciences*, *19*(4), 1659–1676.

Alvarez-Garreton, C. (2015), Improving flood prediction in sparsely gauged catchments by the assimilation of satellite soil moisture into a rainfall-runoff model.

Brocca, L., F. Melone, T. Moramarco, and R. Morbidelli (2009), Antecedent wetness conditions based on ERS scatterometer data, *Journal of Hydrology*, *364*(1), 73–87.

Brocca, L., F. Melone, T. Moramarco, W. Wagner, V. Naeimi, Z. Bartalis, and S. Hasenauer (2010), Improving runoff prediction through the assimilation of the ASCAT soil moisture product, *Hydrology and Earth System Sciences*, *14*(10), 1881–1893.

Brocca, L., S. Hasenauer, T. Lacava, F. Melone, T. Moramarco, W. Wagner, W. Dorigo, P. Matgen, J. Martínez-Fernández, P. Llorens, et al. (2011), Soil moisture estimation through ASCAT and AMSR-E sensors: an intercomparison and validation study across europe, *Remote Sensing of Environment*, *115*(12), 3390–3408.

Brocca, L., T. Moramarco, F. Melone, W. Wagner, S. Hasenauer, and S. Hahn (2012), Assimilation of surface-and root-zone ASCAT soil moisture products into rainfall–runoff modeling, *Geoscience and Remote Sensing, IEEE Transactions on*, *50*(7), 2542–2555.

Brocca, L., T. Moramarco, F. Melone, and W. Wagner (2013), A new method for rainfall estimation through soil moisture observations, *Geophysical Research Letters*, *40*(5), 853–858.

Brocca, L., L. Ciabatta, C. Massari, T. Moramarco, S. Hahn, S. Hasenauer, R. Kidd, W. Dorigo, W. Wagner, and V. Levizzani (2014), Soil as a natural rain gauge: Estimating global rainfall from satellite soil moisture data, *Journal of Geophysical Research: Atmospheres*, *119*(9), 5128–5141.

Chen, F., W. T. Crow, P. J. Starks, and D. N. Moriasi (2011), Improving hydrologic predictions of a catchment model via assimilation of surface soil moisture, *Advances in Water Resources*, *34*, 526–536.

Chen, F., W. T. Crow, and D. Ryu (2014), Dual forcing and state correction via soil moisture assimilation for improved rainfall runoff modelling, *Journal of Hydrometeorology, accepted*.

Chipperfield, A., and P. Fleming (1995), The matlab genetic algorithm toolbox, in *Applied Control Techniques Using MATLAB, IEE Colloquium on*, pp. 10/1–10/4.

Ciach, G. J., M. L. Morrissey, and W. F. Krajewski (2000), Conditional bias in radar rainfall estimation, *Journal of Applied Meteorology*, *39*(11), 1941–1946.

Crow, W., M. Van Den Berg, G. Huffman, and T. Pellarin (2011), Correcting rainfall using satellite-based surface soil moisture retrievals: The Soil Moisture Analysis Rainfall Tool (SMART), *Water Resources Research*, *47*(8), W08,521.

Crow, W. T., and D. Ryu (2009), A new data assimilation approach for improving runoff prediction using remotely-sensed soil moisture retrievals, *Hydrology and Earth System Sciences*, *13*(1), 1–16.

Crow, W. T., and M. T. Yilmaz (2014), The auto-tuned land data assimilation system (ATLAS), *Water Resources Research*, *50*(1), 371–385.

Crow, W. T., G. J. Huffman, R. Bindlish, and T. J. Jackson (2009), Improving Satellite-Based Rainfall Accumulation Estimates Using Spaceborne Surface Soil Moisture Retrievals, *Journal of Hydrometeorology*, *10*, 199–212.

De Lannoy, G. J., P. R. Houser, V. Pauwels, and N. E. Verhoest (2006), Assessment of model uncertainty for soil moisture through ensemble verification, *Journal of Geophysical Research: Atmospheres (1984–2012)*, *111*(D10).

Draper, C., and R. Reichle (2015), The impact of near-surface soil moisture assimilation at subseasonal, seasonal, and interannual time scales, *Hydrology and Earth System Sciences Discussions*, *12*(8).

Draper, C. S., J. P. Walker, P. J. Steinle, R. A. de Jeu, and T. R. Holmes (2009), An evaluation of AMSR–E derived soil moisture over Australia, *Remote Sensing of Environment*, *113*(4), 703–710.

Ebert, E. E., J. E. Janowiak, and C. Kidd (2007), Comparison of near-real-time precipitation estimates from satellite observations and numerical models, *Bulletin of the American Meteorological Society*, *88*(1), 47–64.

Evensen, G. (2003), The ensemble kalman filter: Theoretical formulation and practical implementation, *Ocean dynamics*, *53*(4), 343–367.

Ford, T. W., E. Harris, and S. M. Quiring (2013), Estimating root zone soil moisture using near-surface observations from SMOS, *Hydrology and Earth System Sciences Discussions*, *10*(6), 8325–8364.

Francois, C., A. Quesney, and C. Ottlé (2003), Sequential assimilation of ers-1 sar data into a coupled land surface-hydrological model using an extended Kalman filter, *Journal of Hydrometeorology*, *4*(2), 473–487.

Gill, M. A. (1978), Flood routing by the Muskingum method, *Journal of Hydrology*, *36*(3), 353–363.

Gruhier, C., P. De Rosnay, S. Hasenauer, T. R. Holmes, R. A. De Jeu, Y. H. Kerr, E. Mougin, E. Njoku, F. Timouk, W. Wagner, et al. (2010), Soil moisture active and passive microwave products: intercomparison and evaluation over a Sahelian site, *Hydrology and Earth System Sciences*.

Huffman, G., D. Bolvin, E. Nelkin, D. Wolff, R. Adler, G. Gu, Y. Hong, K. Bowman, and E. Stocker (2007), The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales, *Journal of Hydrometeorology*, *8*(1), 38–55.

Jones, D. A., W. Wang, and R. Fawcett (2009), High-quality spatial climate data-sets for australia, *Australian Meteorological and Oceanographic Journal*, *58*(4), 233.

Li, Y., D. Ryu, A. W. Western, Q. J. Wang, D. R. Robertson, and W. T. Crow (2014), An integrated error parameter estimation and lag-aware data assimilation scheme for real-time flood forecasting, *Journal of Hydrology*, *519*, 2722–2736.

Liu, Y. Q., and H. V. Gupta (2007), Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework, *Water Resources Research*, *43*.

Manfreda, S., L. Brocca, T. Moramarco, F. Melone, and J. Sheffield (2014), A physically based approach for the estimation of root-zone soil moisture from surface measurements, *Hydrology and Earth System Sciences*, *18*(3), 1199–1212.

Massari, C., L. Brocca, T. Moramarco, Y. Tramblay, and J.-F. Didon Lescot (2014), Potential of soil moisture observations in flood modelling: estimating initial conditions and correcting rainfall, *Advances in Water Resources*.

Massari, C., L. Brocca, A. Tarpanelli, and T. Moramarco (2015), Data assimilation of satellite soil moisture into rainfall-runoff modelling: A complex recipe?, *Remote Sensing*, *7*(9), 11,403–11,433.

McKenzie, N. J., D. Jacquier, L. Ashton, and H. Cresswell (2000), *Estimation of soil properties using the Atlas of Australian Soils*, CSIRO Land and Water Canberra.

Mendoza, P. A., J. McPhee, and X. Vargas (2012), Uncertainty in flood forecasting: A distributed modeling approach in a sparse data catchment, *Water Resources Research*, *48*(9).

Moore, R. J. (2007), The PDM rainfall-runoff model, *Hydrology and Earth System Sciences*, *11*(1), 483–499.

Moradkhani, H., S. Sorooshian, H. Gupta, and P. Houser (2005), Dual state–parameter estimation of hydrological models using ensemble Kalman filter, *Advances in Water Resources*, *28*(2), 135–147.

Naeimi, V., K. Scipal, Z. Bartalis, S. Hasenauer, and W. Wagner (2009), An improved soil moisture retrieval algorithm for ERS and Metop scatterometer observations, *Geoscience and Remote Sensing, IEEE Transactions on*, *47*(7), 1999–2013.

Nash, J., and J. Sutcliffe (1970), River flow forecasting through conceptual models part i: A discussion of principles, *Journal of Hydrology*, *10*(3), 282–290.

Owe, M., R. de Jeu, and T. Holmes (2008), Multisensor historical climatology of satellite-derived global land surface moisture, *Journal of Geophysical Research: Earth Surface (2003–2012)*, *113*(F1).

Peel, M. C., B. L. Finlayson, and T. A. McMahon (2007), Updated world map of the köppen-geiger climate classification, *Hydrology and earth system sciences discussions*, *4*(2), 439–473.

Pellarin, T., A. Ali, F. Chopin, I. Jobard, and J.-C. Bergès (2008), Using spaceborne surface soil moisture to constrain satellite precipitation estimates over west Africa, *Geophysical Research Letters*, *35*(2).

Pipunic, R. C, D. Ryu, J. F. Costelloe, and C. -H. Su (2015), An evaluation and regional error modeling methodology for near-real-time satellite rainfall data over Australia, *Journal of Geophysical Research: Atmospheres*, *120*(20).

Richards, L. A. (1931), Capillary conduction of liquids through porous mediums, *Physics*, *1*(5), 318–333.

Ryu, D., W. T. Crow, X. Zhan, and T. J. Jackson (2009), Correcting Unintended Perturbation Biases in Hydrologic Data Assimilation, *Journal of Hydrometeorology*, *10*, 734–750.

Scipal, K., T. Holmes, R. De Jeu, V. Naeimi, and W. Wagner (2008), A possible solution for the problem of estimating the error structure of global soil moisture data sets, *Geophysical Research Letters*, *35*(24).

Stoffelen, A. (1998), Toward the true near-surface wind speed: Error modeling and calibration using triple collocation, *Journal of Geophysical Research: Oceans (1978–2012)*, *103*(C4), 7755–7766.

Su, C., D. Ryu, W. T. Crow, and A. W. Western (2014), Beyond triple collocation: Applications to soil moisture monitoring, *Journal of Geophysical Research - Atmospheres*, *119(11)*, 6416–6439.

Su, C.-H., and D. Ryu (2015), Multi-scale analysis of bias correction of soil moisture, *Hydrology and Earth System Sciences*, *19*(1), 17–31.

Su, C.-H., D. Ryu, R. I. Young, A. W. Western, and W. Wagner (2013), Inter-comparison of microwave satellite soil moisture retrievals over the murrumbidgee basin, southeast australia, *Remote Sensing of Environment*, *134*, 1–11.

Wagner, W., G. Lemoine, and H. Rott (1999), A method for estimating soil moisture from ers scatterometer and soil data, *Remote Sensing of Environment*, *70*(2), 191–207.

Wanders, N., D. Karssenberg, A. d. Roo, S. de Jong, and M. Bierkens (2014), The suitability of remotely sensed soil moisture for improving operational flood forecasting, *Hydrology and Earth System Sciences*, *18*(6), 2343–2357.

Wang, Q., D. Robertson, and F. Chiew (2009), A bayesian joint probability modeling approach for seasonal forecasting of streamflows at multiple sites, *Water Resources Research*, *45*(5), W05,407.

Western, A. W., R. B. Grayson, and G. Blöschl (2002), Scaling of soil moisture: A hydrologic perspective, *Annual Review of Earth and Planetary Sciences*, *30*(1), 149–180.

Yilmaz, M. T., and W. T. Crow (2013), The optimality of potential rescaling approaches in land data assimilation, *Journal of Hydrometeorology*, *14*(2), 650–660.

Yong, B., L. Ren, Y. Hong, J. J. Gourley, Y. Tian, G. J. Huffman, X. Chen, W. Wang, and Y. Wen (2013), First evaluation of the climatological calibration algorithm in the real-time tmpa precipitation estimates over two basins at high and low latitudes, *Water Resources Research*, *49*(5), 2461–2472.

Yong, B., D. Liu, J. J. Gourley, Y. Tian, G. J. Huffman, L. Ren, and Y. Hong (2015), Global view of real-time trmm multisatellite precipitation analysis: Implications for its successor global precipitation measurement mission, *Bulletin of the American Meteorological Society*, *96*(2), 283–296.

Zhan, W., M. Pan, N. Wanders, and E. Wood (2015), Correction of real-time satellite precipitation with satellite soil moisture observations, *Hydrology & Earth System Sciences Discussions*, *12*(6).

Zhou, T., B. Nijssen, G. Huffman, and D. Lettenmaier (2014), Evaluation of the trmm real-time multi-satellite precipitation analysis version 7 for macro scale hydrologic prediction, *J. Hydrometeor*, *15*(4), 1651–1660.

Corresponding author: Camila Alvarez-Garreton, Center for Climate and Resilience Research $(CR)^2$, Chile. (calvarezgarreton@gmail.com)

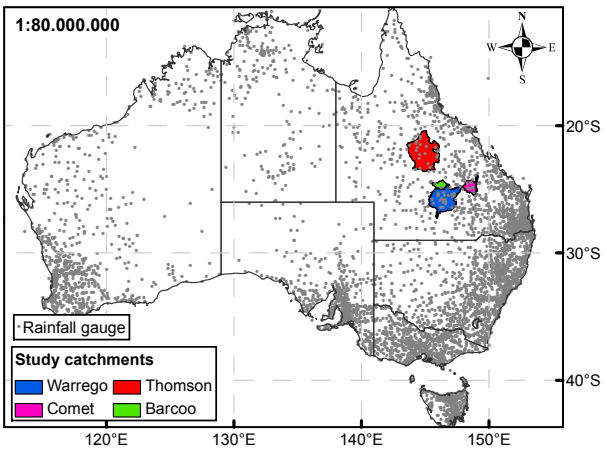

Figure 1. Study catchments and rainfall gauges.

Table 1. Study catchments characteristics.

Catchment	Outlet stream gauge	Record initial year	Mean annual rainfall (mm)	Area (km^2)
Warrego	Warrego River at Wyandra	1967	537	42,870
Comet	Comet River at The Lake	1972	723	10,470
Thomson	Thomson River at Longreach	1969	516	57,734
Barcoo	Barcoo River at Blackall	1969	570	5,758

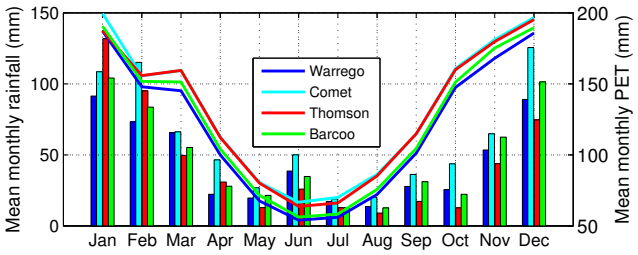

Figure 2. Seasonal rainfall (bars) and potential evapotranspiration (lines) of the study catchments, calculated for the period January 1998 to August 2013.

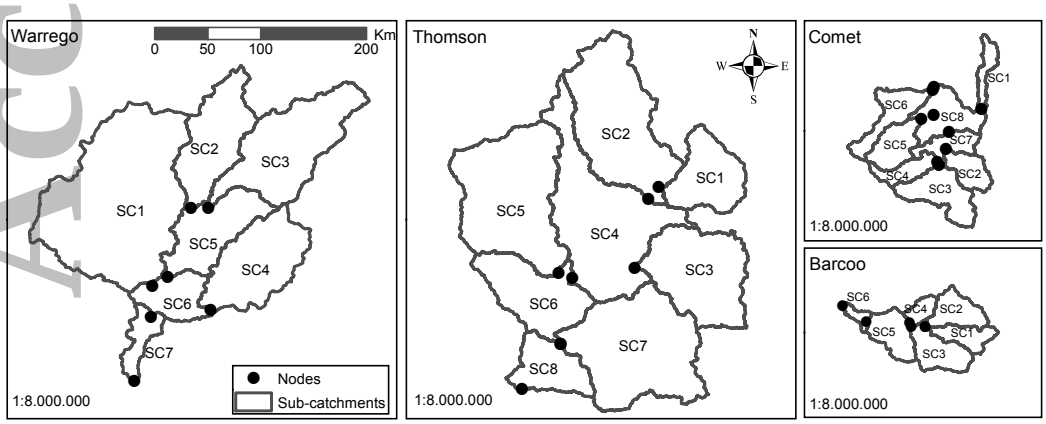

Figure 3. Semi-distributed schemes within the study catchments.

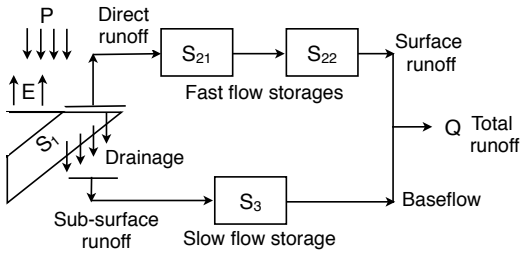

Figure 4. The PDM scheme.

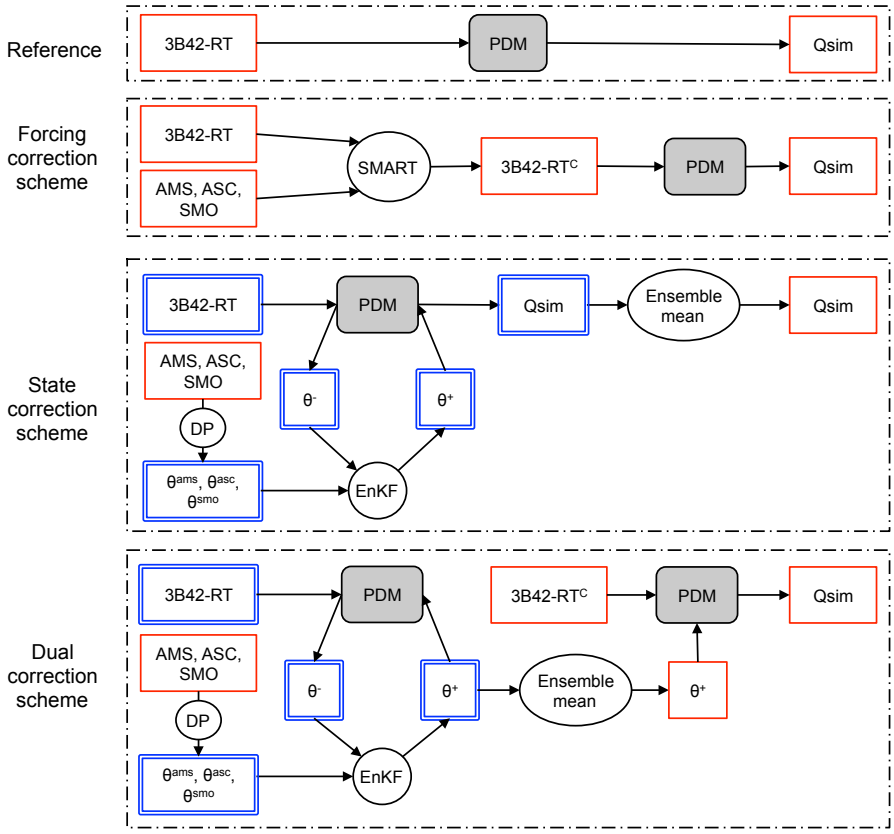

Figure 5. Diagram of the reference evaluation run and the 3 correction schemes. The red single-lined boxes correspond to deterministic variables while the blue double-lined boxes correspond to stochastic variables (ensembles). The circle labelled DP indicates the data processing of the satellite SM observations detailed in section 3.3.1

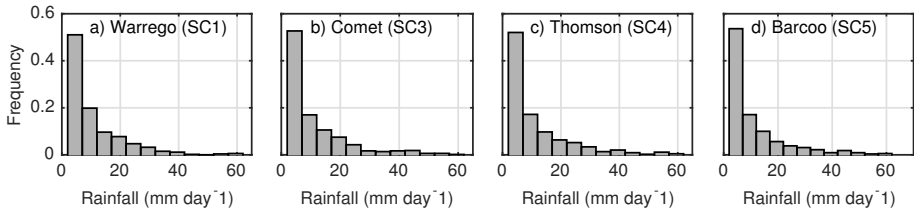

Figure 6. Histograms of the benchmark daily AWAP rainfall accumulations larger than 2 mm over 4 representative sub-catchments.

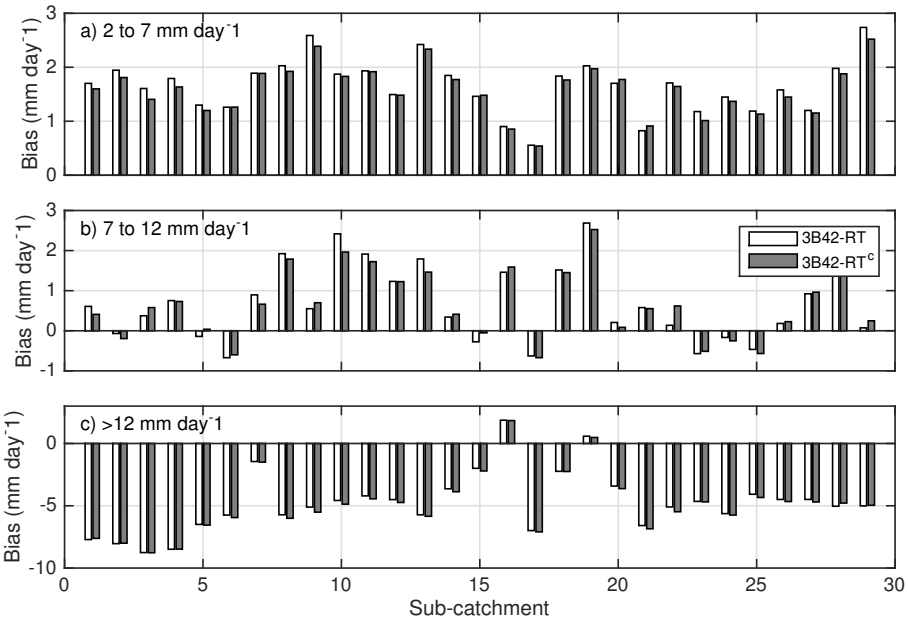

Figure 7. Mean bias in daily rainfall before and after SMART correction within the study catchments: Warrego (sub-catchments 1 to 7), Comet (8 to 15), Thomson (16 to 23) and Barcoo (24 to 29). The mean bias was calculating for daily accumulations of the benchmark dataset varying between 2 to 7 mm (panel a), 7 to 12 mm (panel b) and above 12 mm (panel c).

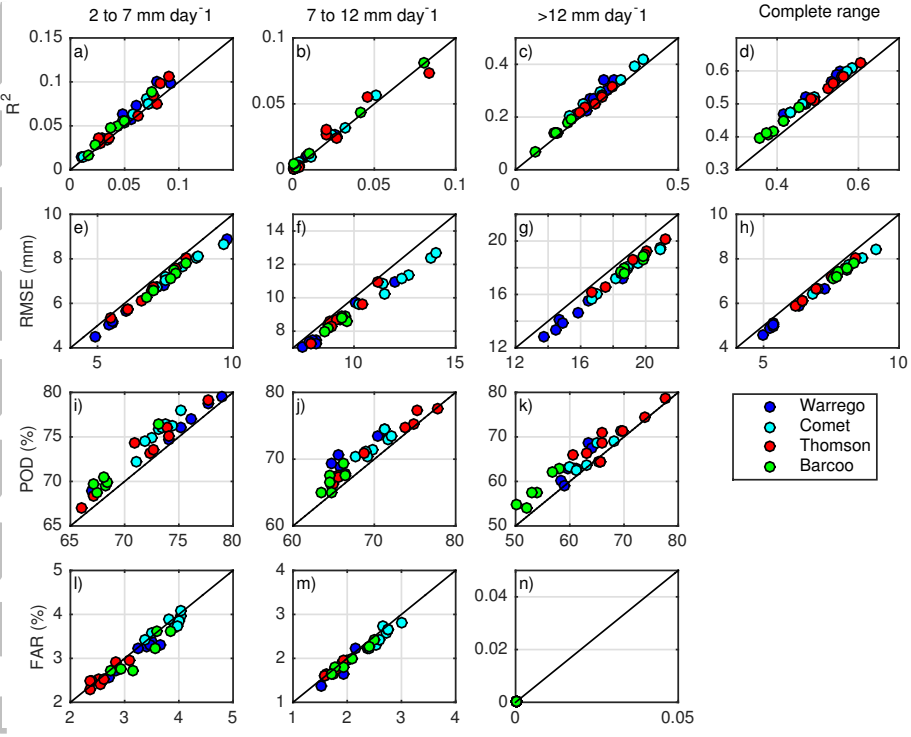

Figure 8. Sub-catchment wise evaluation of SMART analyses using AWAP as the benchmark dataset. Y-axes presents the corrected 3B42-RT^C statistics and x-axes the uncorrected 3B42-RT statistics. The 3 columns show the results for the indicated daily rainfall ranges. The 4 rows present the results for R^2, RMSE, POD and FAR statistics, respectively.

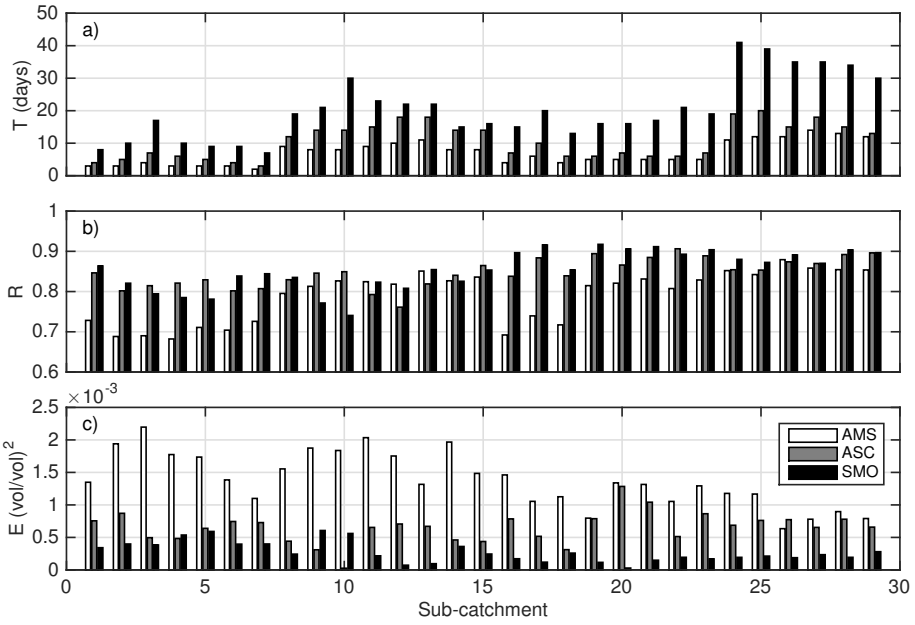

Figure 9. Satellite data processing results for the Warrego (sub-catchments 1 to 7), Comet (8 to 15), Thomson (16 to 23) and Barcoo (24 to 29). Panel a shows the calibrated parameter T used in the SWI estimates. Panel b presents the correlation coefficient between AMS, ASC and SMO-derived SWI and the model soil moisture. Panel c presents the observation error variance in the observation space.

Table 2. Streamflow prediction statistics for the reference runs (forced with satellite rainfall data) calculated using the raw streamflow values (r) and for the log-transformed values (l).

Catchment	NSE		R		RMSE (mm)		FAR	POD
	(r)	(l)	(r)	(l)	(r)	(l)	(r)	(r)
Warrego	0.30	0.40	0.58	0.74	0.34	1.11	0.07	0.85
Comet	0.70	-0.30	0.87	0.56	0.78	1.49	0.22	0.81
Thomson	0.28	0.10	0.55	0.75	0.21	1.51	0.23	0.95
Barcoo	0.24	-0.03	0.50	0.62	0.53	1.21	0.10	0.73

Table 3. Streamflow prediction statistics from the models forced with gauged-based rainfall data. (r) refers to the raw streamflow values and (l) to the log-transformed values.

Catchment	NSE		R		RMSE (mm)		FAR	POD
	(r)	(l)	(r)	(l)	(r)	(l)	(r)	(r)
Warrego	0.85	0.69	0.92	0.87	0.15	0.80	0.05	0.92
Comet	0.74	0.21	0.89	0.71	0.73	1.16	0.15	0.81
Thomson	0.54	0.56	0.77	0.86	0.17	1.06	0.17	0.96
Barcoo	0.43	0.25	0.66	0.73	0.46	1.04	0.06	0.79

Table 4. Model error parameters calibrated with MAP

Catchment	σ_p	σ_s	σ_k
Warrego	0.98	0.03	0.10
Comet	0.70	0.03	0.05
Thomson	0.85	0.02	0.08
Barcoo	0.89	0.02	0.03

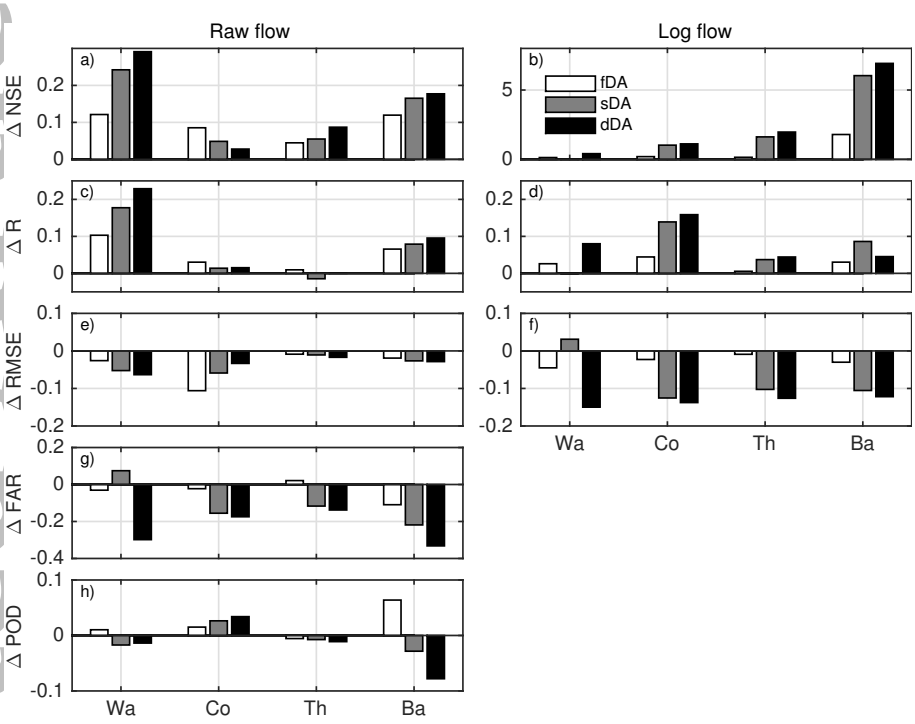

Figure 10. Data assimilation results for the Warrego (Wa), Comet (Co), Thomson (Th) and Barcoo (Ba). The bars show the forcing correction scheme (fDA), the state correction scheme (sDA) and the dual correction scheme (dDA). The statistics in the left panels (a, c, e, g and h) used the raw streamflow values. The statistics in the right panels (b, d and f) used log-transformed streamflow values.