



Dual Forcing and State Correction via Soil Moisture Assimilation for Improved Rainfall–Runoff Modeling

FAN CHEN

*Science Systems and Applications, Inc., and Hydrology and Remote Sensing Laboratory,
Agricultural Research Service, U.S. Department of Agriculture, Beltsville, Maryland*

WADE T. CROW

*Hydrology and Remote Sensing Laboratory, Agricultural Research Service, U.S. Department of Agriculture,
Beltsville, Maryland*

DONGRYEOL RYU

Department of Infrastructure Engineering, University of Melbourne, Melbourne, Victoria, Australia

(Manuscript received 16 December 2013, in final form 14 April 2014)

ABSTRACT

Uncertainties in precipitation forcing and prestorm soil moisture states represent important sources of error in streamflow predictions obtained from a hydrologic model. An earlier synthetic twin experiment has demonstrated that error in both antecedent soil moisture states and rainfall forcing can be filtered by assimilating remotely sensed surface soil moisture retrievals. This opens up the possibility of applying satellite soil moisture estimates to address both key sources of error in hydrologic model predictions. Here, in an attempt to extend the synthetic analysis into a real-data environment, two satellite-based surface soil moisture products—based on both passive and active microwave remote sensing—are assimilated using the same dual forcing/state correction approach. A bias correction scheme is implemented to remove bias in background forecasts caused by synthetic perturbations in the ensemble filtering routines, and a triple collocation–based technique is adopted to derive rescaled observations and observation error variances. Results are largely in agreement with the earlier synthetic analysis. That is, the correction of satellite-derived rainfall forcing is able to improve streamflow prediction, especially during relatively high-flow periods. In contrast, prestorm soil moisture state correction is more efficient in improving the base flow component of streamflow. When rainfall and soil moisture state corrections are combined, the RMSE of both the high- and low-flow components of streamflow can be reduced by ~40% and ~30%, respectively. However, an unresolved issue is that soil moisture data assimilation also leads to underprediction of very intense precipitation/high-flow events.

1. Introduction

Surface soil moisture plays a key role in determining the partitioning of surface-incident rainfall between infiltration and surface runoff. As a result, the characterization of prestorm soil moisture states is an important component of most hydrologic prediction systems. With

the growing availability of satellite-derived surface soil moisture retrievals (Naeimi et al. 2009; Entekhabi et al. 2010; Kerr et al. 2010), the application of data assimilation techniques in hydrological modeling has become increasingly popular. In particular, the assimilation of surface soil moisture retrievals into rainfall–runoff modeling has been demonstrated to benefit streamflow predictions via the improved constraint of antecedent soil moisture states that determine the infiltration capacity of the land surface (e.g., Aubert et al. 2003; Crow and Ryu 2009, hereafter CR09; Draper et al. 2011; Lee et al. 2011; Brocca et al. 2012; Li et al. 2013). Practical challenges in application, such as

Corresponding author address: Fan Chen, USDA ARS Hydrology and Remote Sensing Laboratory, Rm. 104, Bldg. 007, BARC-W, Beltsville, MD 20705.
E-mail: fan.chen@ars.usda.gov

insufficient model coupling strength (Chen et al. 2011) as well as biased precipitation data (Draper et al. 2011), have also been discussed.

In addition to prestorm soil moisture level, precipitation uncertainty is another important source of error in streamflow predictions. Because of its global coverage at high temporal frequency (as high as 3 hourly), satellite precipitation measurements have become increasingly utilized in forcing hydrologic models (Hong et al. 2007; Wu et al. 2012) and are often the only source of precipitation data available for ungauged basins. However, real-time products with relatively low latency (less than daily) typically contain significant bias and errors that can propagate into other model variables during operational streamflow forecasting. Left uncorrected, such errors can significantly impact subsequent hydrologic model predictions (Harris et al. 2007; Li et al. 2009; Pan et al. 2010; Gebregiorgis et al. 2012). Recent work has also clarified that remotely sensed soil moisture has a potential role in reducing uncertainty in precipitation accumulation products (Crow et al. 2009; Pellarin et al. 2008; Brocca et al. 2013). In particular, Crow et al. (2011) introduced the Soil Moisture Analysis Rainfall Tool (SMART) to correct short-term (i.e., from daily to pentad scale) precipitation accumulation estimates using satellite-based surface soil moisture retrievals. As a result, remotely sensed surface soil moisture retrievals have the potential to address two separate sources of error in rainfall–runoff modeling: 1) uncertainty in prestorm soil moisture states and 2) error in within-storm precipitation accumulation totals. Recognizing this potential, CR09 presented a set of synthetic twin data assimilation experiments that verified that a dual-correction strategy (tasked with correcting both antecedent soil moisture states and incident rainfall forcing) is superior to competing techniques that apply data assimilation to correct either precipitation forcing or prestorm model states in isolation.

While encouraging, results in CR09 are based solely on synthetic experiments that do not capture the entire range of challenges involved in applying actual satellite-based surface soil moisture retrievals to hydrologic prediction applications. As a result, this paper attempts to build upon CR09 by testing its dual-correction soil moisture data assimilation strategy using real satellite data products. With a combination of passive microwave data from the Soil Moisture Ocean Salinity (SMOS) satellite and active microwave observation from the Advanced Scatterometer (ASCAT), near-daily soil moisture retrieval temporal frequency can be achieved. Through enhancement of both precipitation and antecedent soil moisture, this dual-assimilation system aims to simultaneously correct random errors in both soil moisture

TABLE 1. Outlet location, basin size, and long-term runoff ratios for basins used in the analysis.

Basin	USGS No.	Outlet		Size (km ²)	Runoff ratio
		Lat (°N)	Lon (°W)		
1	07144780	37.844	97.935	1967	0.073
2	07152000	36.811	97.277	4647	0.211
3	07144200	37.832	97.387	3317	0.136
4	07147070	37.795	97.012	1065	0.185
5	07147800	37.224	96.994	4700	0.260
6	07172000	37.003	96.315	1112	0.232
7	07197000	35.921	94.838	767	0.270
8	06908000	38.992	93.196	2800	0.294
9	07056000	35.983	92.745	2072	0.283
10	07052500	36.805	93.461	2456	0.256
11	06933500	37.929	91.997	7100	0.229
12	07057500	36.622	92.248	1402	0.336
13	07019000	38.505	90.591	9470	0.253

and rainfall precipitation inputs (via SMART) and therefore maximize the utility of existing remotely sensed soil moisture data products for hydrologic prediction applications in ground data–poor regions. In addition, the combined use of active and passive surface soil moisture retrieval provides an important analogy for the eventual availability of simultaneous active–passive soil moisture products from the National Aeronautics and Space Administration (NASA) Soil Moisture Active Passive (SMAP) mission (Entekhabi et al. 2010).

This paper is organized as follows. Section 2 describes the study basins and data used in the analysis, and section 3 describes the hydrologic modeling and data assimilation methodology. Results are presented in section 4, discussed in section 5, and summarized in section 6.

2. Study site and data

The analysis is focused on 13 basins located in the central United States (Table 1, Fig. 1) that have been screened for adequate rain gauge and low levels of anthropogenic stream regulation (Schaake et al. 2001). The chosen basins are from small to intermediate in size (700–10 000 km²) and demonstrate a general westward trend of decreasing annual precipitation and vegetation density. Climatic normal potential evapotranspiration (PET) data as well as parameter values for the Sacramento Soil Moisture Accounting (SAC-SMA, hereafter just SAC) model within each basin are obtained from the Model Parameter Estimation Experiment (MOPEX) datasets (Schaake et al. 2001). Daily mean discharge data measured at basin outlets are obtained from the U.S. Geological Survey (USGS).

The ASCAT on board the Meteorological Operational (MetOp) satellite measures radar backscatter in

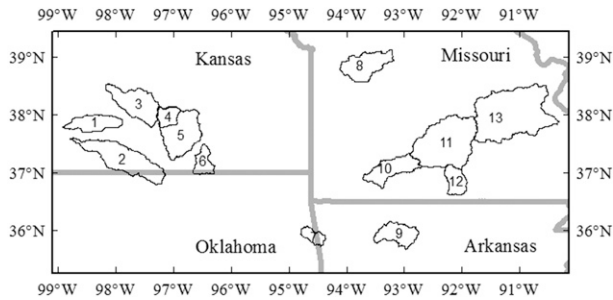


FIG. 1. Locations of the 13 basins used in the analysis, numbered from west to east.

C band (5.255 GHz) with approximately 25-km resolution. MetOp was launched on 19 October 2006 and is in a sun-synchronous orbit, with equatorial crossing times of approximately 0930 (descending overpass) and 2130 (ascending overpass) local solar time. ASCAT surface soil moisture retrievals on 0.25° spatial grid were obtained from Naeimi et al. (2009) by applying the Water Retrieval Package 5 (WARP5) algorithm. For details regarding retrieval of ASCAT surface soil moisture data, refer to Wagner et al. (1999) and Naeimi et al. (2009). Retrievals of ascending and descending overpasses are combined into a daily dataset. In cases where both ascending and descending retrievals for a grid cell are available on a single day, the two retrievals are averaged to obtain a single daily value.

The SMOS satellite, designated for soil moisture observation, was launched on 2 November 2009 by the European Space Agency (ESA). The Microwave Imaging Radiometer using Aperture Synthesis (MIRAS) on board SMOS senses L-band microwave emission (1.400–1.427 GHz) that penetrates the top 5 cm of the soil column. Data used in this study are the SMOS operational Soil Moisture User Data Product 2, version 5.01 (SMUDP2; <http://earth.esa.int/smos>), retrieved using the Dobson dielectric model and processed at the National Oceanic and Atmospheric Administration (NOAA) National Environmental Satellite, Data, and Information Service (NESDIS) into a daily 0.25° gridded product as part of the Soil Moisture Operational Product System (SMOPS; a full description of SMOPS is available at www.ospo.noaa.gov/Products/land/smops/). As with ASCAT, soil moisture retrievals obtained from ascending and descending SMOS overpasses (at 0600 and 1800 local solar time) of SMOS are combined into a single daily dataset. When merged in this way, the combination of ASCAT and SMOS soil moisture datasets provides near-daily coverage of the study sites.

SMART has generally been applied to the correction of real-time precipitation products derived exclusively from satellite-based observations (Crow et al. 2009,

2011). Therefore, results presented here are meant to reflect data availability restrictions commonly faced by hydrologic prediction systems operating in ground data-poor regions. Therefore, the precipitation dataset used to force the SAC model, and subject to correction via SMART, is the daily accumulation from Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis, version 7 (TMPA), 3B40RT (Ostrenga et al. 2013). This is a real-time, microwave-only product generated without any ground-based rain gauge observations. The NOAA Climate Prediction Center's (CPC) 0.25°-grid unified gauge-based retrospective analysis of daily precipitation (Xie et al. 2010) is used as an independent source of verifying data for satellite-derived (and/or SMART corrected) rainfall accumulation estimates.

The products described above are spatially resampled to provide watershed-scale averages for all basins shown in Fig. 1. Daily precipitation is defined as the total depth of precipitation between 1200 and 1200 UTC. Accordingly, observed daily mean streamflow is obtained from hourly discharge data between 1200 and 1200 UTC measured at USGS gauges located at watershed outlets and converted to water-equivalent depth. However, the defined day for soil moisture products is shifted 12 h forward (i.e., from 0000 to 2400 UTC) in order to capture the delay between possible rainfall and resulting soil moisture.

3. Modeling and data assimilation methodology

a. Hydrologic modeling

All hydrologic modeling was based on application of the SAC hydrologic model to the 13 basins described in section 2 and Table 1. The SAC model has been used extensively for operational streamflow forecasting within medium-sized (~1000 km²) river basins in the United States (Burnash et al. 1973; Georgakakos 2006). Soil moisture accounting in the model is based on the estimation of six interdependent soil water states: upper-zone free water content (UZFWC), upper-zone tension water content (UZTWC), lower-zone tension water content (LZTWC), lower-zone free primary water content (LZFPC), lower-zone free supplemental water content (LZFSC), and basin-saturated fraction [additional impervious area (ADIMP)]. The movement of water between these states is calculated using the SAC model parameterization described in Sorooshian et al. (1993). Forced by PET and precipitation, the SAC model produces four separate runoff processes: surface infiltration excess runoff (SER) occurring when rainfall accumulation within a given time step is large enough to fill available upper-zone tension and free water storage

TABLE 2. List of calibrated SAC model parameters.

Parameter	Description	Range of value	Unit
UZTWM	Max capacity of the upper-zone tension water storage	1–300	mm
UZFWM	Max capacity of the upper-zone free water storage	5–150	mm
LZTWM	Max capacity of the lower-zone tension water storage	10–500	mm
LZFPM	Max capacity of the lower-zone primary free water storage	10–1000	mm
LZFSM	Max capacity of the lower-zone primary free water storage	5–400	mm
ADIMP	Additional impervious area	0–1	decimal fraction
UZK	Upper-zone free water lateral depletion rate	0.1–0.75	day ⁻¹
LZPK	Lower-zone primary free water depletion rate	0.001–0.05	day ⁻¹
LZSK	Lower-zone supplemental free water depletion rate	0.01–0.35	day ⁻¹
ZPERC	Max percolation rate	5–350	—
REXP	Exponent of the percolation equation	1–5	—
PCTIM	Impervious fraction of the watershed area	0–1	decimal fraction
RIVA	Riparian vegetation area	0–1	decimal fraction
PFREE	Fraction of water percolating from upper-zone which goes directly to lower-zone free water storage	0–0.8	—
SIDE	Ratio of deep recharge to channel base flow	0–1	—
RSERV	Fraction of lower-zone free water not transferable to lower-zone tension water	0–1	decimal fraction

capacity; surface saturation runoff (SSR) occurring when rainfall falls on saturated portions of the basin (as defined by ADIMP); shallow subsurface interflow (SIF) expressed as a direct function of UZFWC; and deep base flow (BF) expressed as a direct function of LZFWC and LZFC. The SAC model was run on a daily time step for a 2.5-yr period between 12 January 2010 and 31 May 2012.

Table 2 provides a list of key SAC model parameters calibrated for the period of simulation (from 12 January 2010 to 31 May 2012). Calibrated parameter values for each of the 13 study basins shown in Fig. 1 were obtained via optimized calibration against USGS streamflow observations using the shuffled complex evolution method developed at the University of Arizona (SCE-UA) algorithm (Duan et al. 1994). Calibrated parameter sets in all thirteen basins met the baseline criteria of producing a Nash–Sutcliffe efficiency of 0.50 (or higher) in daily streamflow when SAC is forced using the NOAA CPC rainfall product. Based on these calibrated parameters, Fig. 2 shows comparisons between observed streamflow and SAC predictions (forced by CPC and TMPA rainfall, respectively) for three representative basins.

b. Data assimilation

Two separate data assimilation approaches are applied to integrate remotely sensed soil moisture into SAC streamflow modeling. The first approach, described in section 3b(1), is via the SMART algorithm (Crow et al. 2011), which aims to correct precipitation accumulation data before it is used to force the SAC model. The second approach is correcting the SAC soil moisture states using either an ensemble Kalman filter (EnKF) or smoother (EnKS) embedded in the SAC

modeling process, described in section 3b(2). Results will be presented for the application of each approach in isolation and both modules combined (to correct both antecedent soil moisture and incident rainfall accumulations). A triple collocation (TC) procedure used to derive observation error variances and to scale observations to model space is described in section 3b(3), and the application of a bias correction technique for the EnKF/EnKS analysis is described in section 3b(4).

1) RAINFALL CORRECTION USING THE SMART

The SMART algorithm was first described in Crow et al. (2009) and later updated in Crow et al. (2011) and Chen et al. (2012). Its basic function is the correction of satellite-based rainfall products using filtering innovations calculated during the assimilation of remotely sensed surface soil moisture products into a simple water balance model. Past work with SMART has generally focused on the assimilation of a single soil moisture retrieval product, which restricted it to the correction of precipitation accumulations within a >3-day accumulation window (Crow et al. 2009). However, here we will apply SMART to simultaneously integrate both (passive) SMOS and (active) ASCAT retrievals in an attempt to correct daily rainfall inputs required by a hydrologic model.

At its core, SMART is based on the application of an Antecedent Precipitation Index (API) model:

$$API_i = \gamma_i API_{i-1} + P'_i, \quad (1)$$

where P' is a (potentially erroneous) rainfall accumulation estimate obtained on day i and γ_i is a dimensionless

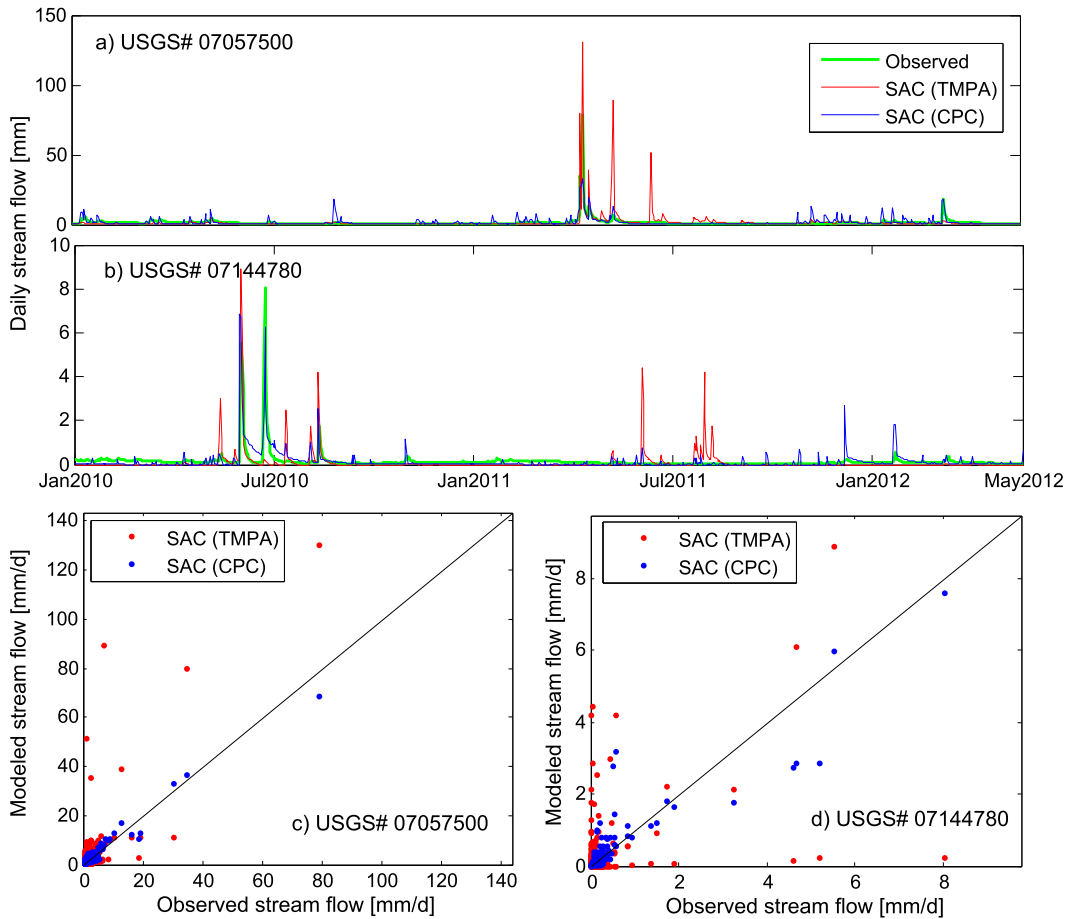


FIG. 2. Comparison of observed discharge (green) for two representative basins and SAC-predicted streamflow forced by TMPA 3B40RT precipitation (red) or CPC gauge analysis (blue) in both (a),(b) time series and (c),(d) scatterplot form. Note that SAC is calibrated to produce acceptable streamflow predictions when forced by CPC precipitation.

API loss coefficient that varies according to day of year D :

$$\gamma_i = 0.8 + 0.05(2\pi D_i/365). \quad (2)$$

The API model described by Eqs. (1) and (2) obviously represents a simple linear modeling strategy that fails to capture both surface saturation and the impact of radiation and meteorology on γ_i . Nevertheless, Crow et al. (2011) demonstrate that, for the specific case of applying SMART at relatively coarse spatial scales, it performs as well as more complex modeling approaches. In particular, Crow et al. (2011) demonstrated that attempts to modify Eq. (1) to account for nonlinear surface saturation generally led to degraded SMART results.

When available, remotely sensed surface soil moisture retrieval(s) are used to update Eq. (1) via a Kalman filter:

$$\text{API}_i^+ = \text{API}_i^- + \mathbf{K}_i(\Theta_i - \mathbf{H}_i \text{API}_i^-), \quad (3)$$

where the plus and minus superscripts denote values after and before Kalman filter updating on day i . Since SMOS and ASCAT data are simultaneously assimilated, the observation vector Θ_i is expressed as

$$\Theta_i = \begin{pmatrix} \theta_i^{\text{SMOS}} \\ \theta_i^{\text{ASCAT}} \end{pmatrix}, \quad (4)$$

where θ_i^{SMOS} and θ_i^{ASCAT} are SMOS and ASCAT surface soil moisture retrievals that have been transformed from raw surface soil moisture retrievals to the reference space of API via a TC-based preprocessing step [see section 3b(3) for details]. Variable \mathbf{H} is a dynamic observation operator defined as

$$\mathbf{H}_i = \begin{pmatrix} a_{\text{SMOS}} \\ a_{\text{ASCAT}} \end{pmatrix}, \quad (5)$$

where a_{SMOS} and a_{ASCAT} are logical numbers that have the value of one when the corresponding observation is available on day i and zero otherwise.

The Kalman gain \mathbf{K} in Eq. (3) is defined as

$$\mathbf{K}_i = \frac{T_i^- \mathbf{H}_i^T}{\mathbf{H}_i T_i^- \mathbf{H}_i^T + \mathbf{R}}, \quad (6)$$

where T is the scalar error variance for API forecasts, superscript T indicates transpose, and \mathbf{R} is the soil moisture observation error covariance matrix defined as

$$\mathbf{R} = \begin{pmatrix} \sigma_{\text{SMOS}}^{*2} & 0 \\ 0 & \sigma_{\text{ASCAT}}^{*2} \end{pmatrix}. \quad (7)$$

Note that \mathbf{R} is transformed to the error space of API via a TC preprocessing step [see section 3b(3)].

At times when API is updated via Eq. (3), T is also updated following

$$T_i^+ = (1 - \mathbf{K}_i \mathbf{H}_i) T_i^-. \quad (8)$$

Between observations, API is forecasted forward in time using Eq. (1) and T is estimated as

$$T_i^- = \gamma_i^2 T_{i-1}^+ + Z + \xi P_i^2, \quad (9)$$

where Z is the model background uncertainty added during every daily forecast time step and ξP_i^2 represents a conditioning of Z by assuming greater uncertainty in API forecasts when $P' > 0$ (Crow et al. 2011). Following Crow et al. (2011), the value of ξ and Z are fixed at 5.0 (dimensionless) and 3 mm^2 , respectively.

Rainfall correction operates on the basis of analysis increments δ defined by

$$\delta_i = \text{API}_i^+ - \text{API}_i^- = \mathbf{K}_i (\Theta_i - \mathbf{H}_i \text{API}_i^-). \quad (10)$$

As δ should correlate with recent random errors in P' , it can be leveraged to linearly correct multiday rainfall accumulations $[P']$ via

$$[\tilde{P}']_l = [P']_l + \lambda [\delta]_l, \quad (11)$$

where λ is a scaling factor and l is an index of non-overlapping accumulation windows. The square brackets represent accumulation over a multiday window. Note that the corrected rainfall is not dynamically fed back into the API model. In practice, the length of the correction window is allowed to vary according to the availability of soil moisture observations (Chen et al. 2012). In other words, each correction window is defined so that a single satellite observation is available on the

last day of the window. Therefore, this approach ensures that soil moisture retrievals are intuitively leveraged to improve only past, instead of future, rainfall accumulations. The corrected rainfall accumulations are then redistributed to daily values in proportion to the original daily rainfall within each correction window. When $[P']_l$ is zero and $\lambda [\delta]_l$ is positive, the addition is applied to the last day of the correction window.

The constant λ in Eq. (11) can be calibrated at each watershed so that the root-mean-square error (RMSE) of $[P']$ against a benchmark precipitation product (e.g., CPC) is minimized. However, temporally and spatially fixed λ values have also been shown to yield acceptable results (Crow et al. 2009, 2011; Chen et al. 2012). Here, following earlier results present in Crow et al. (2009), a fixed value of $\lambda = 0.50$ is applied. Negative values of $[P']$ produced are reset to zero. Since such resetting may introduce a positive overall bias in \tilde{P}' , the entire corrected time series is multiplicatively rescaled to match the long-term mean of P' .

SMART has been found to overestimate the occurrence of low-intensity rainfall events when positive noise in soil moisture due to retrieval error is misidentified as rainfall signals (Crow et al. 2009, 2011). Here a modification is proposed to reduce this tendency: when $[P']_l$ is zero, rainfall correction is applied to the last day of the window only when $\lambda [\delta]_l$ is greater than or equal to 2 mm. This 2-mm threshold represents an arbitrary rainfall intensity over which the signal from soil moisture is considered an indication of real rainfall, with full awareness of a trade-off that real rainfall signals under 2 mm may be discarded. Likewise, saturation of near-surface soil moisture can lead to conditions where soil moisture dynamics do not fully capture accumulation from large rainfall events. This limitation is discussed at length later in the paper (see section 5).

2) STATE CORRECTION USING ENSEMBLE KALMAN FILTER/SMOOTHER

A 35-member EnKF is employed to assimilate satellite-based surface soil moisture retrievals into the SAC model. The choice of 35 as ensemble size is arbitrary; however, larger ensemble sizes failed to produce significant changes to presented results. The EnKF is based on the generation and propagation of a Monte Carlo ensemble of model replicates to provide the error covariance information required by the Kalman filter update equation. An ensemble of model realizations is generated by perturbing key SAC model forcings (i.e., PET and P) and by the application of direct perturbations to the six SAC model soil moisture states. Additive PET perturbations are randomly sampled from a Gaussian distribution with a mean of 0 mm and a standard

deviation of 1.0 mm (negative values are reset to zero). Precipitation is perturbed by multiplying a random factor sampled from a mean-one, lognormal distribution with a dimensionless standard deviation of 1.0. Daily additive noise applied to individual soil moisture states are assumed to be serially uncorrelated and mutually independent random variables sampled from a mean-zero

Gaussian distribution with a standard deviation equal to 2.0% of the total capacity of each state. Note that the integration of these perturbations in time leads to the generation of realistic auto- and cross-correlated error in SAC soil moisture states.

At time i , the vector of SAC model states associated with the j th Monte Carlo replicate is given as

$$\mathbf{S}_{ij} = (\text{UZFWC}_{ij}, \text{UZTWC}_{ij}, \text{LZTWC}_{ij}, \text{LZPFC}_{ij}, \text{LZSFC}_{ij}, \text{ADIMP}_{ij})^T. \quad (12)$$

This vector can be transformed into an estimate of volumetric surface soil moisture using a linear observation operator:

$$\mathbf{H}_i = \begin{pmatrix} a_{\text{SMOS}}\varphi, a_{\text{SMOS}}\varphi, 0, 0, 0, 0 \\ a_{\text{ASCAT}}\varphi, a_{\text{ASCAT}}\varphi, 0, 0, 0, 0 \end{pmatrix}, \quad (13)$$

where a_{SMOS} or a_{ASCAT} are logical numbers that equal one when the corresponding observation is available on day i and zero otherwise. The constant φ is used to convert SAC upper-zone soil water content to volumetric soil moisture ($\text{m}^3 \text{m}^{-3}$):

$$\varphi = \frac{\rho}{\text{UZFWC}_{\text{max}} + \text{UZTWC}_{\text{max}}}, \quad (14)$$

where ρ is soil porosity, $\text{UZFWC}_{\text{max}}$ (mm) is the maximum capacity of free water in the surface zone, and $\text{UZTWC}_{\text{max}}$ (mm) is the maximum capacity of tension water in the surface zone. When remotely sensed surface soil moisture observation is available, replicates of \mathbf{S} are updated via

$$\mathbf{S}_{ij}^+ = \mathbf{S}_{ij}^- + \mathbf{K}_i(\boldsymbol{\Theta}_i + \mathbf{v}_{ij} - \mathbf{H}_i\mathbf{S}_{ij}^-), \quad (15)$$

where the observation vector $\boldsymbol{\Theta}_i$ is defined as in section 3b(1), but is transformed to match the reference space of the SAC volumetric surface soil moisture using a TC method [see section 3b(3)]. The perturbation term \mathbf{v} is a 2×1 vector of mean-zero Gaussian random number with variance of

$$\begin{pmatrix} \sigma_{\text{SMOS}}^{*2} \\ \sigma_{\text{ASCAT}}^{*2} \end{pmatrix}.$$

The gain \mathbf{K} in Eq. (15) is calculated as

$$\mathbf{K}_i = \frac{\mathbf{CH}_i^T}{\mathbf{H}_i\mathbf{CH}_i^T + \mathbf{R}}, \quad (16)$$

where \mathbf{C} is a 6×6 covariance matrix of forecast error calculated from the 35-member ensemble of background

\mathbf{S} predictions, and \mathbf{R} is the observation error covariance matrix defined as

$$\mathbf{R} = \begin{pmatrix} \sigma_{\text{SMOS}}^{*2} & 0 \\ 0 & \sigma_{\text{ASCAT}}^{*2} \end{pmatrix}, \quad (17)$$

which has been transformed to the error space of SAC-modeled volumetric surface soil moisture [see section 3b(3)]. Final EnKF state analyses are obtained by averaging across the ensemble mean of \mathbf{S} .

In contrast to the EnKF, where model states are only updated using concurrent observations, a 1-day-lag EnKS is also employed to use the observation at time $j + 1$ to update the model states between time j and $j + 1$ (inclusive of end points). While the SAC model is run on a daily time step, the three free water states (i.e., UZFWC, LZPFC, and LZSFC) and ADIMP actually evolve on a 3-hourly basis. Thus, with an observation at time i , the EnKS updates an augmented, 40-element state vector $\mathbf{S}_{j-1 \rightarrow j}^*$, which includes the six SAC soil moisture state variables at times $j - 1$ and j as well as the subdaily states of the four variables (i.e., UZFWC, LZPFC, LZSFC, and ADIMP) between times $j - 1$ and j . The matrix \mathbf{C}^* is the new covariance matrix for this 40-element augmented state vector \mathbf{S}^* . As in the EnKF, components of this augmented covariance matrix are sampled directly from the SAC model ensemble and updated with an expression analogous to Eq. (15):

$$\mathbf{S}_{i-1 \rightarrow ij}^{*+} = \mathbf{S}_{i-1 \rightarrow ij}^{*-} + \mathbf{K}_i(\boldsymbol{\Theta}_i + \mathbf{v}_{ij} - \mathbf{H}_i^*\mathbf{S}_{i-1 \rightarrow ij}^{*-}), \quad (18)$$

where

$$\mathbf{K}_i^* = \frac{\mathbf{CH}_i^{*T}}{\mathbf{H}_i^*\mathbf{C}^*\mathbf{H}_i^{*T} + \mathbf{R}} \quad (19)$$

and \mathbf{H}_i^* is a 2×40 matrix similar to \mathbf{H}_i (i.e., elements corresponding to UZFWC and UZTWC are either $a_{\text{SMOS}}\varphi$ or $a_{\text{ASCAT}}\varphi$, all others being zero). As in EnKF,

final EnKS state analyses are obtained by averaging across the updated soil moisture ensemble.

3) OBSERVATION RESCALING AND ERROR ESTIMATION

Satellite soil moisture retrievals typically reflect a different long-term mean and dynamic range than comparable soil moisture time series generated by the assimilation model that they must be integrated into (Reichle et al. 2004). Therefore, following Yilmaz and Crow (2013), the TC-based rescaling method proposed by Stoffelen (1998) is applied to 1) rescale ASCAT and SMOS soil moisture retrievals into a climatology consistent with the SAC and/or API model prior to data assimilation and 2) obtain error covariance information for the subsequent assimilation of ASCAT and SMOS in both the API model for SMART and the SAC model for the EnKF/EnKS implementations. Note that the surface layer in the SAC model (i.e., the upper zone) is implicitly defined by the calibrated capacity of free and tension water contents (i.e., parameters UZFWM and UZTWM; Table 2) and typically does not match the penetration depth of satellite retrievals (0–5 cm). However, by scaling the SMOS and ASCAT data to the climatology and approximate dynamic range of the model surface-layer soil moisture, the impact caused by different vertical representativeness can be minimized.

This approach is based on the assumed availability of three independent soil moisture datasets: θ_A , θ_B , and θ_C . Each time series is decomposed into a corresponding climatological anomaly time series (θ_A^* , θ_B^* , and θ_C^*) by subtracting the long-term 31-day moving average μ_D from the raw data, where μ_D is sampled from a multi-year dataset using a 31-day sampling window centered on the day of the year (i.e., D). This is done for two reasons. First, the observational error variance applied in a Kalman filter should reflect random error sources and not systematic errors due to, for example, differences in seasonal soil moisture climatologies (Crow and van den Berg 2010). Second, past work has demonstrated that TC-derived error variances are more accurate when calculated for anomalies relative to a seasonally varying climatology relative to absolute soil moisture products (Miralles et al. 2010; Draper et al. 2012).

Here, θ_C^* will refer to soil moisture products acquired from the offline model integration (without any actual data assimilation). During application of TC to data assimilation activities, this represents the appropriate reference for rescaling the other two soil moisture products. Therefore, using θ_C^* as the model reference (API in the case of SMART rainfall correction and SAC in the case of EnKF and EnKS state correction) and θ_A

and θ_B representing SMOS and ASCAT products, θ_A^* is scaled to the reference space of θ_C^* via

$$\theta_{A \rightarrow C}^* = \mu_{\theta_C^*} + (\theta_A^* - \mu_{\theta_A^*}) \cdot \varphi_{A \rightarrow C}. \quad (20)$$

Note that $\theta_{A \rightarrow C}^*$ is the rescaled anomaly of θ_A in the reference space of θ_C^* , and $\mu_{\theta_C^*}$ and $\mu_{\theta_A^*}$ are the mean of θ_C^* and θ_A^* . Following Stoffelen (1998), the scaling factor $\varphi_{A \rightarrow C}$ is calculated from

$$\varphi_{A \rightarrow C} = \frac{\overline{\theta_B^* \cdot \theta_C^*}}{\overline{\theta_A^* \cdot \theta_B^*}}, \quad (21)$$

where the overbar denotes temporal averaging. Similarly, $\varphi_{B \rightarrow C}$ can be calculated for θ_B . Then, based on a typical TC analysis, the error variances of θ_A^* and θ_B^* can be estimated as

$$\sigma_A^{*2} = \overline{(\theta_{B \rightarrow C}^* - \theta_{A \rightarrow C}^*) \cdot (\theta_C^* - \theta_{A \rightarrow C}^*)} \quad (22)$$

and

$$\sigma_B^{*2} = \overline{(\theta_{A \rightarrow C}^* - \theta_{B \rightarrow C}^*) \cdot (\theta_C^* - \theta_{B \rightarrow C}^*)}, \quad (23)$$

provided that an assumption of mutually independent errors is met. These values of σ_A^{*2} and σ_B^{*2} are used to formulate the observation error covariance matrix \mathbf{R} and ensure that the relative uncertainties in SMOS and ASCAT soil moisture retrieval products are properly considered during data assimilation.

Finally, θ_A and θ_B are transformed to the mean reference space of θ_C via

$$\theta_{A \rightarrow Cj} = \mu_{C,D_j} + \theta_{A \rightarrow Cj}^* \quad (24)$$

and

$$\theta_{B \rightarrow Cj} = \mu_{C,D_j} + \theta_{B \rightarrow Cj}^*, \quad (25)$$

where j is the temporal index (day) and D_j is the corresponding Julian day D (or day of year), and μ_{C,D_j} is the long-term mean of θ_C on D sampled from the multiyear θ_C time series using a 31-day sampling window centered on D . Note that the use of TC to constrain modeling (as opposed to observation) errors is complicated by the nonlinear nature of the SAC model (Crow and Yilmaz 2014). As a result, TC is used only to estimate the relative size of active and passive soil moisture retrieval errors.

The separate treatment of seasonality and anomalies (i.e., decomposing–combining) ensures that the final rescaled datasets have similar dynamic range as the reference dataset and removes potential seasonal differences between them. Prior to the data assimilation runs, the

above procedure is performed twice to provide one set of scaled error variances and observations for SMART rainfall correction (based on the API model) and a separate set for EnKF/EnKS SAC model state correction.

4) CORRECTION OF PERTURBATION BIAS

Because of the nonlinear nature of most hydrologic models, an ensemble of model forecasts perturbed by mean-zero Gaussian noise can produce biased background predictions. During the application of sequential data assimilation, such bias can lead to significant cumulative water balance errors (Ryu et al. 2009). Here, the simple bias correction method introduced in Ryu et al. (2009) is applied during the EnKF/EnKS state correction analysis. Specifically, the mean bias of perturbed state variable y of an n -member ensemble at time t is sampled as

$$\delta_t = \frac{1}{n} \sum_{i=1}^n (y_t^i - y_t^o), \quad (26)$$

where y^o is a model forecast generated from a single open loop run without perturbing forcing and states. The perturbation bias sampled in Eq. (26) is then subtracted from each (perturbed) y_t^i to ensure a bias-zero forecast ensemble. Past work suggests that this correction can significantly improve EnKF/EnKS profile soil moisture estimates (Ryu et al. 2009). Because of the linear nature of the API model, no comparable bias correction is applied within SMART.

c. Data assimilation strategies

Using real satellite-based surface soil moisture retrievals (obtained from both the passive microwave SMOS sensor and the active microwave ASCAT instrument), this study examines five different data assimilation strategies for improving SAC streamflow predictions.

- (i) A rainfall-correction-only (RC) strategy based on the application of the SMART algorithm to correct rainfall forcing data prior to its input into the SAC model. Note that this approach, unlike all the other strategies introduced below, does not involve the assimilation of active–passive soil moisture retrievals into the SAC model (and thus no direct state correction).
- (ii) Two separate state-correction-only (EnKF and EnKS) strategies where active–passive surface soil moisture retrievals are directly assimilated into the SAC model using EnKF or EnKS to update SAC soil moisture states. An SAC streamflow prediction is then obtained by initializing a separate SAC

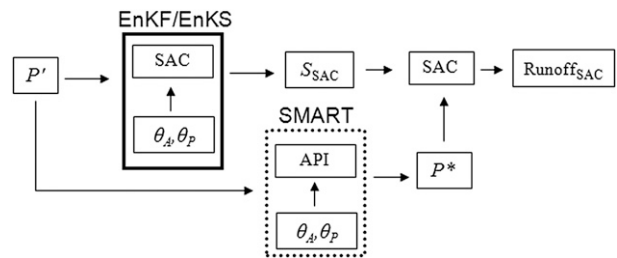


FIG. 3. Schematic for dual rainfall/state correction strategy for leveraging active–passive soil moisture (θ_A , θ_P) products to improve SAC streamflow predictions. The variables P^o and P^* represent observed and corrected rainfall, respectively. The solid box indicates EnKF methods applied to incorporate θ_A , θ_P into SAC soil moisture states S_{SAC} , while the dashed box indicates sequential Kalman filtering that assimilates both θ_A and θ_P into the API model for rainfall correction.

realization with updated model states and forcing it with uncorrected TMPA 3B42RT rainfall. In contrast with the RC strategy introduced above, no rainfall correction is attempted.

- (iii) Two separate dual-correction strategies (RC–EnKF and RC–EnKS), in which both state and rainfall correction is attempted. An illustration of these strategies is provided in Fig. 3. Here, active–passive soil moisture is assimilated into an ensemble of SAC realizations—forced by uncorrected TMPA 3B40RT rainfall data—using either an EnKF or EnKS, and a single set of corrected soil moisture states (i.e., the soil moisture analysis) are obtained by averaging the updated ensemble. As the data assimilation system marches forward in time, a new 1-day SAC model simulation is initialized every day using the previous day’s EnKF/EnKS soil moisture analysis and forced by corrected rainfall (obtained from a separate application of SMART) to produce the dual-correction streamflow estimates. Following the generation of streamflow, this 1-day SAC simulation is terminated and its state predictions are discarded. Note that feeding these state predictions back into the EnKF/EnKS data assimilation system would result in the cross correlation of modeling and observational errors and, therefore, the degradation of EnKF/EnKS state estimates (CR09).

4. Results

Prior to consideration of SAC streamflow results, the improvement of TMPA 3B40RT rainfall via SMART application is evaluated in section 4a. This is followed by an assessment of the bias correction scheme described

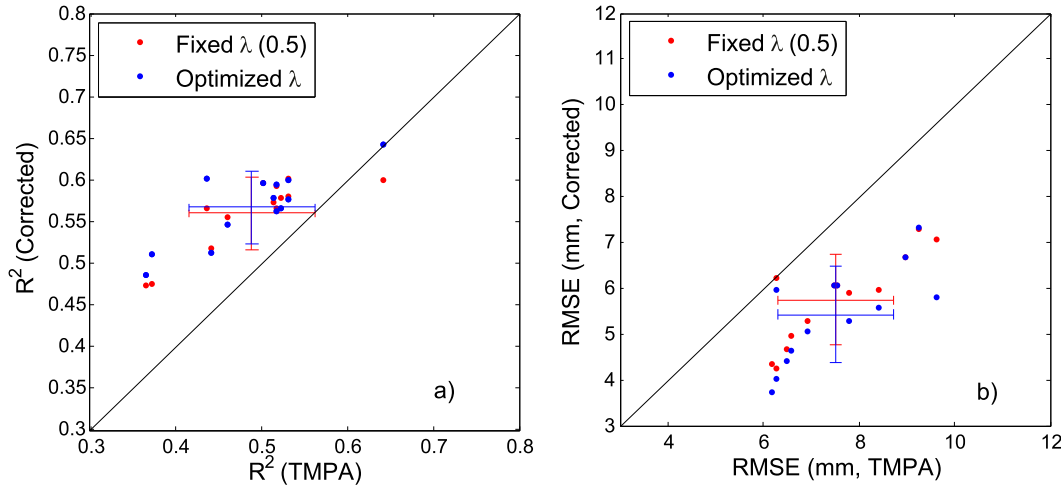


FIG. 4. Basinwise (a) R^2 and (b) RMSE of TMPA 3B40RT rainfall (compared to CPC gauge analysis) before and after SMART correction using both fixed (red) and optimized (blue) λ . Corresponding error bars of one std dev are located at mean R^2 or RMSE values sampled across all basins.

above on SAC streamflow results in section 4b. Finally, streamflow data assimilation results for the five separate cases described in section 3c are presented in section 4c.

a. SMART rainfall correction

Based on application of the SMART approach with the proposed modifications outlined in section 3b(1), daily TMPA 3B40RT rainfall accumulations are corrected using ASCAT and SMOS soil moisture observations. These corrections are verified based on comparisons against daily rainfall estimates obtained from the CPC rain gauge analysis. The application of SMART using ASCAT and SMOS surface soil moisture retrievals substantially improves the coefficient of determination R^2 and RMSE performance of daily TMPA accumulations relative to the CPC benchmark for the case of optimized λ (on a basin by basin case) to minimize RMSE differences relative to CPC (Fig. 4). Moreover, replacing these optimized values with a globally fixed value of $\lambda = 0.50$ results in only a slight reduction in the performance of the SMART algorithm (Fig. 4). As a result, all subsequent SMART results are based on using a globally fixed value of $\lambda = 0.50$.

In addition to R^2 and RMSE results shown in Fig. 4, changes in three categorical skill metrics are also assessed: false alarm ratio (FAR), probability of detection (POD), and threat score (TS). These metrics are calculated based on the number of events (defined as a daily rainfall accumulation in excess of a given threshold) falsely predicted, successfully captured, or missed:

$$FAR = \frac{F}{H + F}, \tag{27}$$

$$POD = \frac{H}{H + M}, \text{ and} \tag{28}$$

$$TS = \frac{H}{H + F + M}, \tag{29}$$

where H is the number of events successfully predicted by a given rainfall product, F is the number of nonevents (i.e., actual rainfall below event threshold) erroneously predicted to occur, and M is the number of actual events that are missed. As a result, FAR refers to the fraction of predicted events (at a given percentile threshold) that are actually nonevents and POD refers to the fraction of all qualifying events correctly predicted. A perfect rainfall time series would yield $FAR = 0$, $POD = 1$, and $TS = 1$ for any given threshold value. Assuming the CPC rain gauge analysis as truth, average changes in the 13 basins in FAR, POD, and TS after SMART correction are calculated for 1-, 3-, and 5-day accumulations (Fig. 5). Given the known tendency of the CPC gauge analysis to overpredict frequency of light events due to spatial interpolation (Janowiak et al. 2004), “true” events of less than 2.0mm are excluded from the categorical metrics analysis.

In contrast to the almost uniformly good R^2 and RMSE results shown in Fig. 4, a number of problems emerge in Fig. 5. First, SMART frequently misinterprets soil moisture retrieval noise as low-intensity rainfall. As a result, the implementation of SMART tends to increase FAR at low-intensity thresholds. Second, SMART has a tendency to underpredict large events and therefore decrease their POD. As discussed in Crow et al. (2011) and Chen et al. (2012), this can be

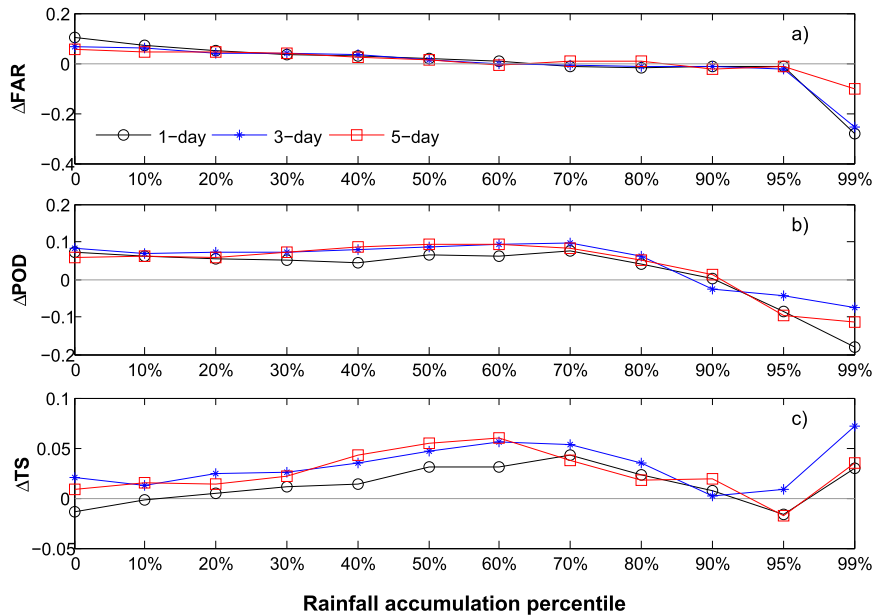


FIG. 5. Change in (a) FAR, (b) POD, and (c) TS for 1-, 3-, and 5-day rainfall accumulations after SMART correction of TMPA 3B40RT data. Data points represent averaged results for all 13 basins used in this analysis.

attributed to the saturation of the surface layer during periods of intense precipitation and the resulting difficulty of inverting a saturated soil moisture signal into an accumulation amount. Additional discussion of difficulties encountered during high precipitation/streamflow events is presented in section 5.

Nevertheless, as an integrated measure of performance that is sensitive to both FAR and POD, TS results in Fig. 5 suggest an improvement in SMART-corrected accumulations over a broad range of intensity thresholds. Although updating at near-daily frequency is possible, better categorical skill obtained in the 3- and 5-day scales (relative to daily) implies that SMART may ultimately be better suitable for application in larger basins with longer streamflow concentration times.

b. Ensemble perturbation bias correction

One issue facing the EnKF approach is the biased background predictions arising from the nonlinear model physics, which could degrade the subsequent data assimilation analysis (De Lannoy et al. 2006; Ryu et al. 2009). Here, we examine the impact of the Ryu et al. (2009) bias removal procedure described in section 3b(4) on EnKF-predicted streamflow. Note that results are based on the use of uncorrected TMPA 3B40RT rainfall (as opposed to corrected output from the SMART algorithm).

For all 13 basins described in Table 1, Fig. 6a shows the impact of ensemble bias correction on the RMSE between an SAC–EnKF analysis (based on simultaneously

assimilating both active ASCAT and passive SMOS surface soil moisture retrievals) and observed streamflow. Application of the approach results in reduced streamflow RMSE (versus USGS streamflow gauges) in 9 out of the 13 basins examined. RMSE results in Fig. 6a are highly sensitive to performance during peak streamflow events. To examine the impact of data assimilation on the low-flow (or base flow) component of the hydrograph, changes in RMSE (Δ RMSE) for log-transformed streamflow is also shown in Fig. 6b. Since this procedure is found more effective in removing bias in deep-layer soil moisture (Ryu et al. 2009), it is expected to have more impact on EnKF/EnKS base flow prediction. Indeed, this trend is reflected on the larger magnitude of Δ RMSE in log flow (Fig. 6b) versus raw flow values (Fig. 6a). In addition, forecast bias correction has greater positive impact on the grass-/cropland-dominated western basins examined, while less improvement is noted in the wetter and more heavily vegetated eastern basins (Fig. 6).

Streamflow time series for two representative basins (basins 3 and 7 in Table 1) are shown in Fig. 7. In both basins (as well as in all other basins), EnKF-predicted low flow is persistently reduced by an almost constant volume throughout the simulation period, after bias correction. This indicates a persistent positive bias generated in the EnKF perturbation process, likely more significant in the deep-layer storages, that is adjusted for via the bias correction. The cause of this

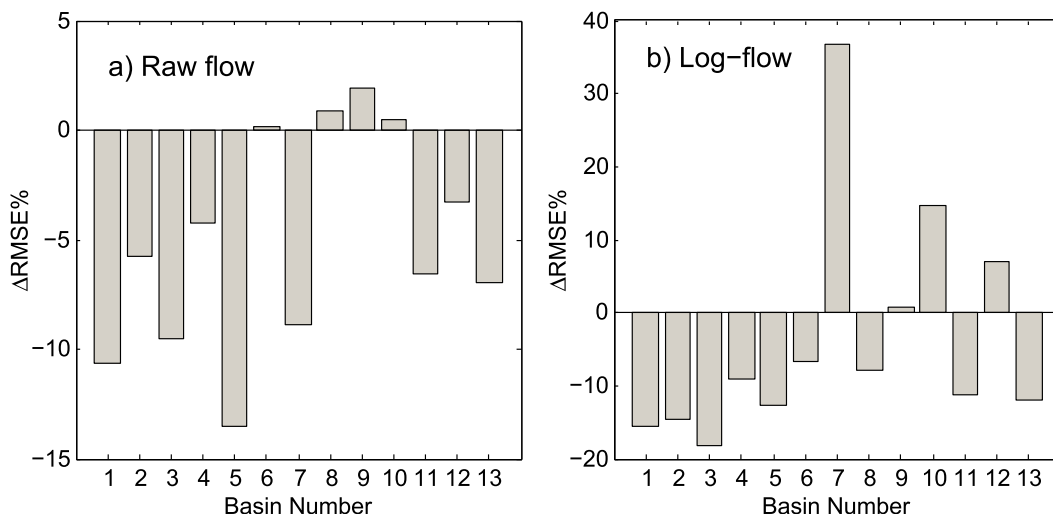


FIG. 6. Change in streamflow RMSE (Δ RMSE; %) after bias correction (relative to the no-bias-correction EnKF run) in study basins. Negative Δ RMSE indicates improvement in streamflow prediction. Basins are arranged from west to east, as annual rainfall increases and dominant land cover transitions from grassland/crop to deciduous forest.

positive bias is that, in this relatively arid region, deep soil layers rarely reach saturation but are often near depletion, so that positive perturbations are usually kept while negative values are frequently truncated when they result in negative soil moisture flows. During observed low-flow periods, SAC predictions are often underestimated, or even intermittent (e.g., zero flow). In the case of basin 7, EnKF-generated low flow is not

enough to make up the missing volume (Fig. 7d). Although it is only decreased by a marginal amount (on the order of $\sim 0.04 \text{ mm day}^{-1}$) after bias correction, the proportional reduction is significant, and it leads to the sharp increase in proportional RMSE of log flow (i.e., highest positive Δ RMSE in Fig. 6b). However, the degrading effect on high flow is small, and overall RMSE is improved (Fig. 6a). In basin 3 (Figs. 7a,b), the EnKF run

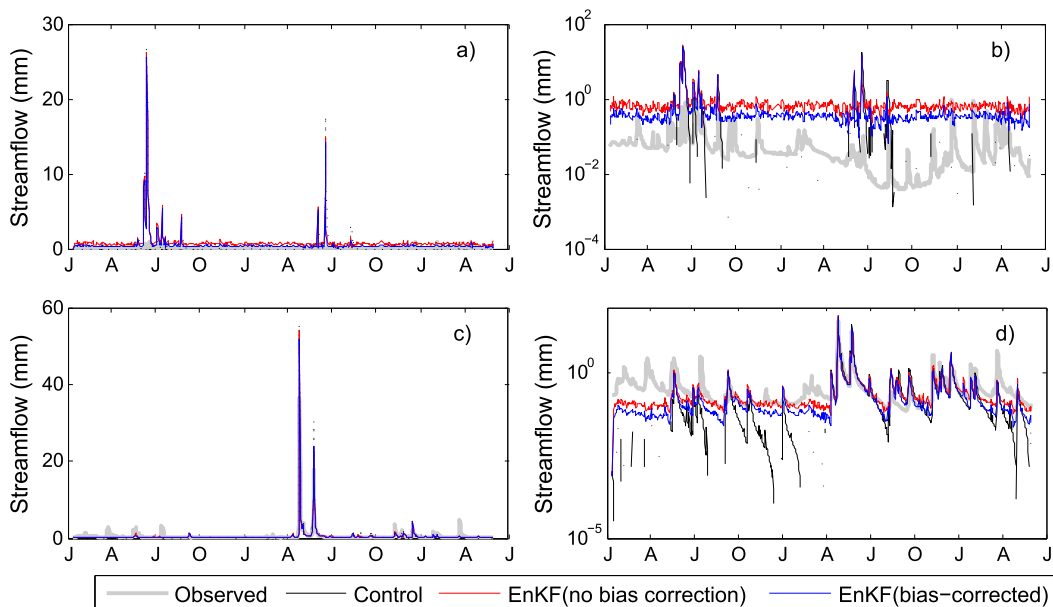


FIG. 7. Observed and simulated discharge in basins (a),(b) 3 and (c),(d) 7. Streamflow values are shown in linear ordinates in (a) and (c) and logarithmic ordinates in (b) and (d). The plotted log flow line of the control run is broken when predicted streamflow values are zero. Labels of horizontal axis are in repetitions of “J A J O,” referring to the months of January, April, July, and October, respectively.

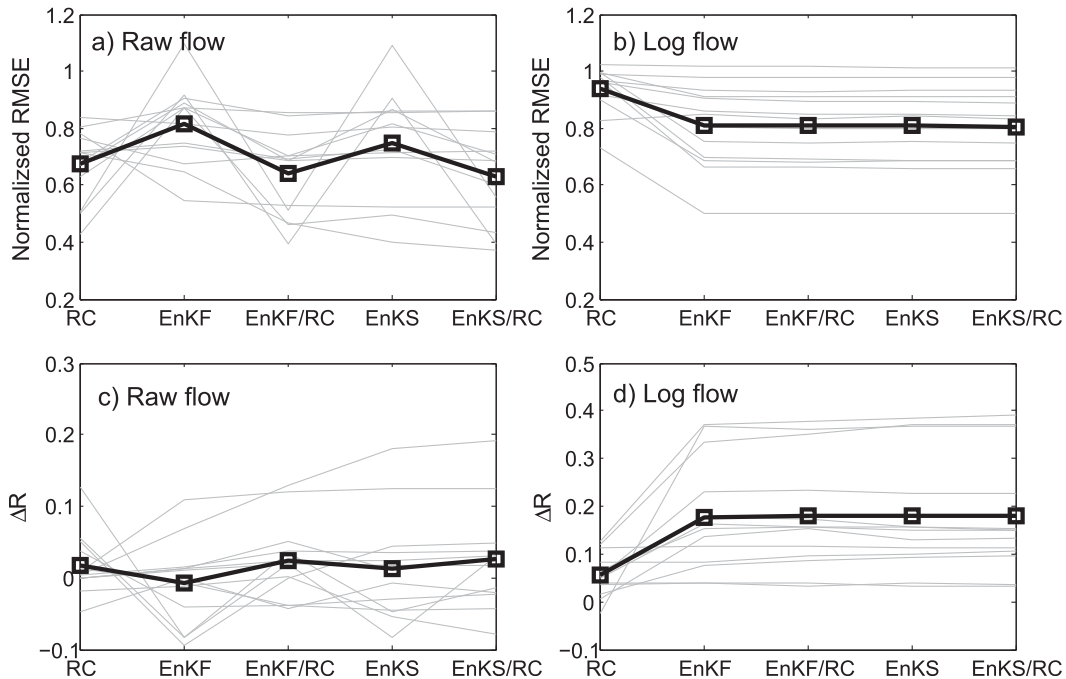


FIG. 8. Changes in streamflow prediction skills (compared to CPC data) after data assimilation in terms of (a) normalized raw flow RMSE (divided by RMSE of SAC control run), (b) normalized log flow RMSE, (c) raw flow correlation coefficient difference (between analysis and SAC control run; ΔR), and (d) log flow correlation coefficient difference. Individual basins are shown as gray lines, and boldface lines with square symbols are averages across all basins.

produces a consistently positive bias during the low-flow periods, which is reduced by $\sim 40\%$ after bias correction is applied. Overall, Figs. 5 and 6 support the ability of the Ryu et al. (2009) bias correction procedure to improve EnKF streamflow estimates. As a result, it is applied to all subsequent EnKF and EnKS analyses.

c. Streamflow results

Following the SMART-based correction of rainfall (section 4a) and the bias correction of EnKF/EnKS (section 4b), we are now ready to examine streamflow predictions associated with all five data assimilation cases introduced in section 3c. In particular, Fig. 8 summarizes the average normalized RMSE (i.e., ratio of RMSE after data assimilation to the RMSE of the SAC control run) across all 13 basins for application of 1) the RC case, where SAC is forced with SMART-corrected rainfall and no state correction; 2) the ENKF state-correction-only case, where antecedent states are updated using an EnKF but SMART is not applied; 3) the EnKS state-correction-only case, where antecedent states are updated using an EnKS but SMART is not applied; and 4) two dual (RC–EnKS and RC–EnKF) cases where simultaneous rainfall and antecedent soil moisture state correction are attempted.

Overall, results in Fig. 8 demonstrate a reduction in streamflow prediction errors (i.e., a normalized RMSE of less than one) for all five cases. However, the SMART-based RC appears to provide the most benefit for reducing the normalized RMSE of the raw flows (Fig. 8a). This is consistent with earlier synthetic results in CR09, where average normalized streamflow RMSE of ~ 0.5 (RC), 1.0 (EnKF), 0.75 (EnKS), and 0.45 (EnKS–RC) were obtained for 97 individual MOPEX basins. On the other hand, when low-flow components are emphasized by examining log flows in Figs. 8b and 8d, the average skill obtained by the EnKF/EnKS cases is higher than that of the RC cases, suggesting that state correction has more impact on flow estimation during interstorm, dry-down periods. As in the CR09 synthetic experiment, the dual-correction cases demonstrate virtually the same amount of overall skill during low-flow conditions as the state-correction-only EnKF and EnKS cases.

However, closer examination reveals several noteworthy differences between real-data results presented here and the earlier synthetic results in CR09. First, unlike CR09 results in which EnKF state correction alone had little positive impact on runoff prediction, the real-data EnKF case proved capable of improving

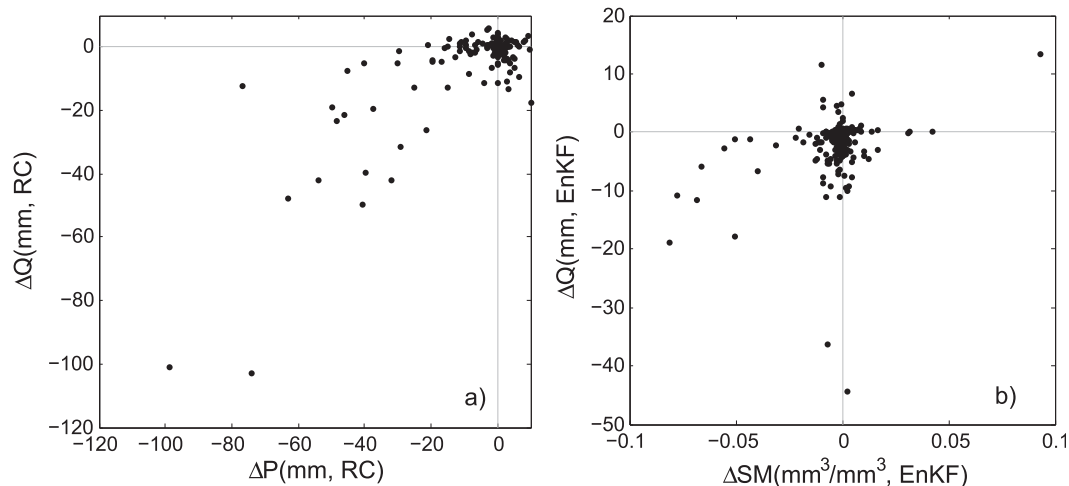


FIG. 9. Changes of rainfall (ΔP ; mm), antecedent upper-layer soil moisture (ΔSM ; $\text{m}^3 \text{m}^{-3}$), and predicted streamflow (ΔQ ; mm) after data assimilation for high-flow events (i.e., observed streamflow ≥ 5 mm). Parenthetical text on axes labels indicate the data assimilation case applied.

RMSE in all but one basin (see Fig. 8a). This may be partially attributed to the benefit of the bias correction approach applied here (but not in CR09). Secondly, the normalized RMSE for the real-data EnKS case is only marginally lower than comparable EnKF case results, whereas the relative advantage of the EnKS over an EnKF was much more significant in CR09. Finally, counter to the results in CR09, the dual-correction cases (EnKF-RC and EnKS-RC) do not clearly outperform the RC case in most basins for flow peaks (manifested in RMSE; Fig. 8a). This can be largely attributed to the synergism of RC and EnKF/EnKS that tends to produce underestimated flow estimations (see discussion in section 5).

5. Discussion

Overall, results presented here are encouraging for efforts to improve operational streamflow forecasts using remotely sensed surface soil moisture. Rainfall correction using the SMART algorithm is able to improve the flow prediction, especially the high-flow component, effectively, although at the cost of underestimating some of the largest flow events (discussed below). Although less as effective, state (soil moisture) correction can also improve the flow prediction and contribute more to the low-flow component. When rainfall and state corrections are combined, RMSE of both the high- and low-flow components can be significantly reduced by $\sim 40\%$ and $\sim 30\%$, respectively.

However, a common theme running through this analysis, as well as earlier SMART validation work (see, e.g., Chen et al. 2012), is the challenge of applying soil

moisture-based assimilation strategies to improve the characterization of intense precipitation/streamflow events.

The reliable prediction of high-flow episodes is, of course, a vital goal for most hydrologic prediction systems. Given known shortcomings in SMART-corrected rainfall with regards to the detection of intense precipitation (Crow et al. 2011; Chen et al. 2012), it is important to specifically assess the impact of soil moisture data assimilation on high-flow events. To demonstrate the response of high-flow events to changes in forcing (precipitation) and antecedent state (soil moisture) due to the data assimilation (DA) procedures, Fig. 9 plots the post-DA increment of streamflow Q , rainfall (i.e., P), and antecedent surface-layer soil moisture SM of “high flow” events from all basins in the RC and EnKF cases. High-flow events are defined as daily observed streamflow volume of ≥ 5 mm (converted to water-depth unit for each basin). Based on this threshold, a total of 233 high-flow events (accounted for 2% of total daily flow events) are identified for the 13 basins studied. As in the rainfall correction case, data assimilation (both rainfall and state correction alone) often leads to a negative skew in ΔQ for large-flow events, and the largest negative ΔQ events are primarily associated with decreased rainfall (Figs. 9a,b). Figure 9a reveals a clear tendency for SMART to reduce P (and therefore Q) for many of the high-flow events. As noted in Fig. 5, this strong skeptical bias toward high P events has a deleterious effect of the POD performance of SMART during intense rainfall events.

There appears to be at least two sources for this bias. First, the Kalman filter at the core of the

TABLE 3. Change in FAR, POD, and TS (see section 4a) of daily streamflow events (meeting certain depth threshold requirements) associated with various data assimilation cases. All statistics are based on qualified events from all 13 basins. Numbers in parentheses indicate the total number of events in all basins.

Event threshold (mm)		RC	EnKF	EnKF-RC	EnKS	EnKS-RC
2 (908)	Δ FAR	-0.01	-0.06	-0.07	-0.08	-0.08
	Δ POD	0.04	-0.06	-0.07	-0.07	-0.07
	Δ TS	0.02	-0.01	-0.01	-0.01	-0.01
5 (233)	Δ FAR	-0.04	-0.10	-0.10	-0.12	-0.12
	Δ POD	0.03	-0.05	-0.07	-0.06	-0.06
	Δ TS	0.03	0.03	0.03	0.04	0.04
10 (60)	Δ FAR	-0.03	-0.07	-0.09	-0.10	-0.11
	Δ POD	-0.07	-0.10	-0.15	-0.10	-0.13
	Δ TS	0.02	0.03	0.04	0.06	0.05

SMART approach has a tendency to produce conditionally biased results. That is, in an attempt to minimize the error variance of its P predictions, SMART will systematically underpredict peak values. This problem is well known in variance-minimized rainfall estimates (Ciach et al. 2000) but has likely not received the attention it deserves in land data assimilation. This limitation is obviously serious for hydrologic applications tasked with high-flow (i.e., flood) detection and forecasting. It is worth noting that other techniques exist for correcting precipitation via soil moisture exist that do not utilize a variance-minimizing strategy (e.g., Brocca et al. 2013). Therefore, it is possible that other rainfall correction techniques may be better suited to estimating very large rainfall/streamflow events than SMART.

Second, during periods of intense precipitation, surface soil moisture will reach saturation and can no longer be reliably inverted into a rainfall estimate. In practice, this makes it difficult for SMART to accurately identify the occurrence of intense rainfall events. Crow et al. (2011) attempted to address this POD problem for large events through specific modifications to the SMART rainfall correction approach—in particular, the addition of new terms into Eq. (9). However, the fact that POD problems persist here (in a follow-on analysis using different data products) implies that the changes suggested by Crow et al. (2011) to improve the large-event POD were ad hoc and not generally effective. Furthermore, because of the strong coupling between ΔP and changes in ΔQ for large events (Fig. 9), SMART POD problems (for rainfall) are likely to be carried over into streamflow results generated from the SAC hydrologic model (shown in Table 3).

In addition to rainfall correction, updating model soil moisture states is expected to have a positive impact on detecting high-flow events since it is a key

factor that determines the partitioning between surface runoff and infiltration (e.g., Li et al. 2013; Marchi et al. 2010). Unfortunately, EnKF and EnKS state-correction-only cases are found to result in a net decrease of POD for high-flow events here (see Table 3). In addition to the inherent conditional bias issue discussed above, the mismatched timing of satellite overpass and rainfall event(s) may also contribute to the underestimation in predicted high-flow events by the EnKF and EnKS state-correction-only cases. When satellite overpasses occur prior to a rainfall event, observations will indicate a drier condition, while observations at a much later time relative to rainfall may not adequately capture the signal in soil moisture. Both situations will falsely introduce a negative bias into model states and subsequent streamflow conditions. In such cases, the degrading effects can be more serious for EnKS, because the instantaneous satellite measurement is used to update all 3-hourly subdaily states. Thus, the advantage of EnKS (i.e., its ability to improve runoff prediction of time j as well as $j + 1, j + 2, \dots$, with observation at time j) turns into a disadvantage; this explains the greater number of underestimated moderate events for the EnKS case relative to the EnKF case (Table 3). Such impacts can potentially be mitigated by calculating API at a finer temporal scale (e.g., 3 or 6 hourly) and by assimilating remotely sensed soil moisture retrievals, which are more accurately matched to the timing of individual precipitation events.

6. Summary

The design of appropriate data assimilation systems to simultaneously ingest both active and passive microwave-based surface soil moisture retrievals represents a critical goal for facilitating the development of hydrologic applications for SMAP surface products.

Here, satellite surface soil moisture retrievals are used in a suite of data assimilation procedures in an attempt to improve streamflow predictions obtained from a hydrologic model. This attempt is based on applying data assimilation strategies that address errors in both prestorm soil moisture conditions and/or intrastorm rainfall accumulation estimates.

Unlike the earlier synthetic analysis in CR09, this analysis is based on real surface soil moisture retrievals derived from real, satellite-based active (i.e., ASCAT) and passive (i.e., SMOS) microwave sensors. By simultaneously assimilating both products, a near-daily average update frequency is obtainable, which is important for flow prediction in relatively small watersheds with short streamflow concentration temporal scales. The ensemble bias correction procedure outlined in Ryu et al. (2009) is also validated and exercised in the EnKF/EnKS routines. Another major methodological advance is our use of a TC approach for rescaling remote sensing observations and obtaining accurate estimates of their error variance statistics.

In general, results presented here are consistent with earlier synthetic results presented in CR09, which demonstrate the potential of combined rainfall–state correction (based on the assimilation of satellite-derived surface soil moisture retrievals) to improve estimation of both the peak and base flow components of total streamflow. In particular, peak flow estimation generally benefits the most from (SMART based) rainfall correction while improved low-flow predictions are primarily aided by state correction via an EnKF/EnKS. On the other hand, a number of notable differences between the results here and CR09 are also found. Here, the real-data, EnKF-only case proved capable of improving streamflow RMSE in most basins studied—a capability that was not noted in the earlier CR09 synthetic analysis. This improvement may be related to the bias correction procedure adopted here that largely removed the soil moisture bias caused by ensemble perturbation. Furthermore, unlike the CR09 results, the real-data EnKS case has little advantage over the EnKF case, and the dual-correction cases (state and rainfall) do not outperform the RC case in most basins in terms of flow peaks, suggesting that the advantage associated with the application of a smoother in the CR09 synthetic analysis may be difficult to realize in a real-data environment.

Moreover, a major unresolved challenge is the underprediction of intense precipitation/high-flow events (see section 5). Difficulties in predicting large-flow events represent, of course, a major shortcoming for any hydrologic prediction system, and future research is required for minimizing this tendency.

Nevertheless, the results of this analysis illustrate the future potential of utilizing simultaneous active–passive surface soil moisture retrievals to correct state and forcing errors degrading hydrologic modeling predictions. Naturally, these efforts will receive a significant boost with the late 2014 launch of the NASA SMAP mission.

Acknowledgments. Research was funded via a grant from the NASA Precipitation Measurement Mission (PMM) program (W. T. Crow, principal investigator).

REFERENCES

- Aubert, D., C. Loumagne, and L. Oudin, 2003: Sequential assimilation of soil moisture and streamflow data in a conceptual rainfall-runoff model. *J. Hydrol.*, **280**, 145–161, doi:10.1016/S0022-1694(03)00229-4.
- 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 modelling. *IEEE Trans. Geosci. Remote Sens.*, **50**, 2542–2555, doi:10.1109/TGRS.2011.2177468.
- , F. Melone, T. Moramarco, and W. Wagner, 2013: A new method for rainfall estimation through soil moisture observations. *Geophys. Res. Lett.*, **40**, 853–858, doi:10.1002/grl.50173.
- Burnash, R. J. C., R. L. Ferral, and R. A. McGuire, 1973: A generalized streamflow simulation system—Conceptual modeling for digital computers. Tech. Rep., Joint Federal and State River Forecast Center, Sacramento, CA, 204 pp.
- 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. *Adv. Water Resour.*, **34**, 526–536, doi:10.1016/j.advwatres.2011.01.011.
- , —, and T. Holmes, 2012: Improving long-term, retrospective precipitation datasets using satellite-based surface soil moisture retrievals and the Soil Moisture Analysis Rainfall Tool (SMART). *J. Appl. Remote Sens.*, **6**, 063604, doi:10.1117/1.JRS.6.063604.
- Ciach, G. J., M. L. Morrissey, and W. F. Krajewski, 2000: Conditional bias in radar rainfall estimation. *J. Appl. Meteor.*, **39**, 1941–1946, doi:10.1175/1520-0450(2000)039<1941:CBIRRE>2.0.CO;2.
- Crow, W. T., and D. Ryu, 2009: A new data assimilation approach for improving runoff prediction using remotely-sensed soil moisture retrievals. *Hydrol. Earth Syst. Sci.*, **13**, 1–16, doi:10.5194/hess-13-1-2009.
- , and M. J. van den Berg, 2010: An improved approach for estimating observation and model error parameters in soil moisture data assimilation. *Water Resour. Res.*, **46**, W12519, doi:10.1029/2010WR009402.
- , and M. T. Yilmaz, 2014: The Auto-Tuned Land Assimilation System (ATLAS). *Water Resour. Res.*, **50**, 371–385, doi:10.1002/2013WR014550.
- , G. F. Huffman, R. Bindlish, and T. J. Jackson, 2009: Improving satellite-based rainfall accumulation estimates using spaceborne soil moisture retrievals. *J. Hydrometeorol.*, **10**, 199–212, doi:10.1175/2008JHM986.1.

- , M. J. van Den Berg, G. F. Huffman, and T. Pellarin, 2011: Correcting rainfall using satellite-based surface soil moisture retrievals: The soil moisture analysis rainfall tool (SMART). *Water Resour. Res.*, **47**, W08521, doi:10.1029/2011WR010576.
- De Lannoy, G. J. M., P. R. Houser, V. R. N. Pauwels, and N. E. C. Verhoest, 2006: Assessment of model uncertainty for soil moisture through ensemble verification. *J. Geophys. Res.*, **111**, D10101, doi:10.1029/2005JD006367.
- Draper, C. S., J.-F. Mahfouf, J.-C. Calvet, E. Martin, and W. Wagner, 2011: Assimilation of ASCAT near-surface soil moisture into the SIM hydrological model over France. *Hydrol. Earth Syst. Sci.*, **15**, 3829–3841, doi:10.5194/hess-15-3829-2011.
- , R. H. Reichle, G. J. M. De Lannoy, and Q. Liu, 2012: Assimilation of passive and active microwave soil moisture retrievals. *Geophys. Res. Lett.*, **39**, L04401, doi:10.1029/2011GL050655.
- Duan, Q. Y., S. Sorooshian, and V. K. Gupta, 1994: Optimal use of the SCE-UA global optimization method for calibrating watershed models. *J. Hydrol.*, **158**, 265–284, doi:10.1016/0022-1694(94)90057-4.
- Entekhabi, D., and Coauthors, 2010: The Soil Moisture Active and Passive (SMAP) Mission. *Proc. IEEE*, **98**, 704–716, doi:10.1109/JPROC.2010.2043918.
- Gebregiorgis, A. S., Y. Tian, C. D. Peters-Lidard, and F. Hossain, 2012: Tracing hydrologic model simulation error as a function of satellite rainfall estimation bias components and land use and land cover conditions. *Water Resour. Res.*, **48**, W11509, doi:10.1029/2011WR011643.
- Georgakakos, K. P., 2006: Analytical results for operational flash flood guidance. *J. Hydrol.*, **317**, 81–103, doi:10.1016/j.jhydrol.2005.05.009.
- Harris, A., S. Rahman, F. Hossain, L. Yarborough, A. C. Bagtzoglou, and G. Eason, 2007: Satellite-based flood modeling using TRMM-based rainfall products. *Sensors*, **7**, 3416–3427, doi:10.3390/s7123416.
- Hong, Y., R. F. Adler, F. Hossain, S. Curtis, and G. J. Huffman, 2007: A first approach to global runoff simulation using satellite rainfall estimation. *Water Resour. Res.*, **43**, W08502, doi:10.1029/2006WR005739.
- Janowiak, J. E., P. Xie, R. J. Joyce, M. Chen, and Y. Yarosh, 2004: Validation of satellite-derived rainfall estimates and numerical model forecasts of precipitation over the United States. *Proc. 29th Annual Climate Diagnostics and Prediction Workshop*, Madison, WI, NWS, 6 pp. [Available online at www.cpc.ncep.noaa.gov/products/outreach/proceedings/cdw29_proceedings/Janowiak.pdf.]
- Kerr, Y. H., and Coauthors, 2010: The SMOS mission: New tool for monitoring key elements of the global water cycle. *Proc. IEEE*, **98**, 666–687, doi:10.1109/JPROC.2010.2043032.
- Lee, H., D. Seo, and V. Koren, 2011: Assimilation of streamflow and in situ soil moisture data into operational distributed hydrologic models: Effects of uncertainties in the data and initial model soil moisture states. *Adv. Water Resour.*, **34**, 1597–1615, doi:10.1016/j.advwatres.2011.08.012.
- Li, L., and Coauthors, 2009: Evaluation of the real-time TRMM-based multi-satellite precipitation analysis for an operational Flood Prediction System in Nzoia Basin, Lake Victoria, Africa. *Nat. Hazards*, **50**, 109–123, doi:10.1007/s11069-008-9324-5.
- Li, Y., D. Ryu, A. W. Western, and Q. J. Wang, 2013: Assimilation of stream discharge for flood forecasting: The benefits of accounting for routing time lags. *Water Resour. Res.*, **49**, 1887–1900, doi:10.1002/wrcr.20169.
- Marchi, L., M. Borga, E. Preciso, and E. Gaume, 2010: Characterisation of selected extreme flash floods in Europe and implications for flood risk management. *J. Hydrol.*, **394**, 118–133, doi:10.1016/j.jhydrol.2010.07.017.
- Miralles, D. G., W. T. Crow, and M. H. Cosh, 2010: Estimating spatial sampling errors in coarse-scale soil moisture estimates derived from point-scale observations. *J. Hydrometeorol.*, **11**, 1423–1429, doi:10.1175/2010JHM1285.1.
- Naeimi, V., K. Scipal, Z. Bartalis, S. Hasenauer, and W. Wagner, 2009: An improved soil moisture retrieval algorithm for ERS and METOP scatterometer observations. *IEEE Trans. Geosci. Remote Sens.*, **47**, 1999–2013, doi:10.1109/TGRS.2008.2011617.
- Ostrega, D., Z. Liu, W. Teng, and S. Kempler, 2013: Newly released version 7 TRMM Multi-satellite Precipitation Analysis (TMPA) products and data services at NASA GES DISC. *29th Conf. on Environmental Information Processing Technologies*, Austin, TX, Amer. Meteor. Soc., 13. [Available online at <https://ams.confex.com/ams/93Annual/webprogram/Paper220338.html>.]
- Pan, M., H. Li, and E. Wood, 2010: Assessing the skill of satellite-based precipitation estimates in hydrologic applications. *Water Resour. Res.*, **46**, W09535, doi:10.1029/2009WR008290.
- Pellarin, T., A. Ali, F. Chopin, I. Jobard, and J.-C. Bergs, 2008: Using spaceborne surface soil moisture to constrain satellite precipitation estimates over West Africa. *Geophys. Res. Lett.*, **35**, L02813, doi:10.1029/2007GL032243.
- Reichle, R. H., R. D. Koster, J. Dong, and A. Berg, 2004: Global soil moisture from satellite observations, land surface models, and ground data: Implications for data assimilation. *J. Hydrometeorol.*, **5**, 430–442, doi:10.1175/1525-7541(2004)005<0430:GSMFSO>2.0.CO;2.
- Ryu, D., W. T. Crow, X. Zhan, and T. J. Jackson, 2009: Correcting unintended perturbation biases in hydrologic data assimilation. *J. Hydrometeorol.*, **10**, 734–750, doi:10.1175/2008JHM1038.1.
- Schaake, J., Q. Duan, V. Koren, and A. Hall, 2001: Toward improved parameter estimation of land surface hydrology models through the Model Parameter Estimation Experiment (MOPEX). *IAHS Publ.*, **270**, 91–98. [Available online at http://iahs.info/uploads/dms/iahs_270_091.pdf.]
- Sorooshian, S., Q. Duan, and V. K. Gupta, 1993: Calibration of rainfall-runoff models: Application of global optimization to the Sacramento Soil Moisture Accounting Model. *Water Resour. Res.*, **29**, 1185–1194, doi:10.1029/92WR02617.
- Stoffelen, A., 1998: Towards the true near-surface wind speed: Error modeling and calibration using triple collocation. *J. Geophys. Res.*, **103**, 7755–7766, doi:10.1029/97JC03180.
- Wagner, W., G. Lemoine, and H. Rott, 1999: A method for estimating soil moisture from ERS scatterometer and soil data. *Remote Sens. Environ.*, **70**, 191–207, doi:10.1016/S0034-4257(99)00036-X.
- Wu, H., R. F. Adler, Y. Hong, Y. Tian, and F. Policelli, 2012: Evaluation of global flood detection using satellite-based rainfall and a hydrologic model. *J. Hydrometeorol.*, **13**, 1268–1284, doi:10.1175/JHM-D-11-087.1.
- Xie, P., M. Chen, and W. Shi, 2010: CPC unified gauge-based analysis of global daily precipitation. *24th Conf. on Hydrology*, Atlanta, GA, Amer. Meteor. Soc., 2.3A. [Available online at <https://ams.confex.com/ams/90annual/webprogram/Paper163676.html>.]
- Yilmaz, M. T., and W. T. Crow, 2013: The optimality of potential rescaling approaches in land data assimilation. *J. Hydrometeorol.*, **14**, 650–660, doi:10.1175/JHM-D-12-052.1.