

# A new data assimilation approach for improving runoff prediction using remotely-sensed soil moisture retrievals

W. T. Crow<sup>1</sup> and D. Ryu<sup>2</sup>

<sup>1</sup>USDA ARS Hydrology and Remote Sensing Laboratory, Beltsville, MD, USA

<sup>2</sup>Department of Civil and Environmental Engineering, University of Melbourne, Melbourne, Australia

Received: 27 June 2008 – Published in Hydrol. Earth Syst. Sci. Discuss.: 22 July 2008

Revised: 18 November 2008 – Accepted: 9 December 2008 – Published: 7 January 2009

**Abstract.** A number of recent studies have focused on enhancing runoff prediction via the assimilation of remotely-sensed surface soil moisture retrievals into a hydrologic model. The majority of these approaches have viewed the problem from purely a state or parameter estimation perspective in which remotely-sensed soil moisture estimates are assimilated to improve the characterization of pre-storm soil moisture conditions in a hydrologic model, and consequently, its simulation of runoff response to subsequent rainfall. However, recent work has demonstrated that soil moisture retrievals can also be used to filter errors present in satellite-based rainfall accumulation products. This result implies that soil moisture retrievals have potential benefit for characterizing both antecedent moisture conditions (required to estimate sub-surface flow intensities and subsequent surface runoff efficiencies) and storm-scale rainfall totals (required to estimate the total surface runoff volume). In response, this work presents a new sequential data assimilation system that exploits remotely-sensed surface soil moisture retrievals to simultaneously improve estimates of both pre-storm soil moisture conditions and storm-scale rainfall accumulations. Preliminary testing of the system, via a synthetic twin data assimilation experiment based on the Sacramento hydrologic model and data collected from the Model Parameterization Experiment, suggests that the new approach is more efficient at improving stream flow predictions than data assimilation techniques focusing solely on the constraint of antecedent soil moisture conditions.

## 1 Introduction

Enhancement of runoff and/or flood forecasts is frequently cited as a key benefit of satellite-based surface soil moisture retrievals (Entekhabi et al., 2003; Lakshmi 2004; NRC 2007). This potential is likely to receive greater attention in the next decade as attempts are made to demonstrate operational applications for soil moisture data products emerging from both current and next-generation satellite missions. Of particular importance are upcoming launches of the first two dedicated soil moisture missions: the ESA Soil Moisture and Ocean Salinity (SMOS) mission in 2009 (Kerr et al., 2001) and the NASA Soil Moisture Active/Passive (SMAP) mission in 2012 (NRC, 2007).

As represented in traditional hydrologic models, surface runoff prediction is a dual estimation problem requiring information describing both the volume of rainfall occurring within a storm and the ability of a watershed to infiltrate such rainfall. This infiltration capacity is largely determined by prevailing soil moisture conditions. Therefore, to date, most strategies for integrating remotely-sensed soil moisture into the rainfall/runoff prediction (or forecasting) problem have focused solely on improving the estimation of antecedent soil moisture conditions. A variety of methodologies have been applied to this goal including the direct use of remotely-sensed soil moisture fields to initialize a hydrologic model (Goodrich et al., 1994; Jacobs et al., 2003; Weisling et al., 2007), the calibration of hydrologic model soil moisture predictions using remotely-sensed soil moisture retrievals (Parajka et al., 2006) and the optimal merging of modeled and remotely-sensed soil moisture using sequential data assimilation techniques (Pauwels et al., 2002; Aubert et al., 2003; Francois et al., 2003; Crow et al., 2005; Kantamneni et al., 2005).



Correspondence to: W. T. Crow  
(wade.crow@ars.usda.gov)

To date, results from such experiments have been mixed and there is currently little compelling evidence that remotely-sensed soil moisture retrievals can aid runoff prediction in ungauged basins (Parajka et al., 2006). Somewhat typical is Crow et al. (2005) who found an improved correlation between antecedent precipitation index (API) values and subsequent storm-scale runoff ratios when soil moisture retrievals from a passive microwave radiometer were sequentially assimilated into the API model. However, the marginal advantage of assimilating soil moisture disappeared when the API model was modified slightly to incorporate air temperature observations into estimates of soil water loss due to evapotranspiration. Other studies were able to identify improvement (upon the integration of remotely-sensed soil moisture) in only a subset of the total basins examined (Pauwels et al., 2002; Parajka et al., 2006).

The above-mentioned approaches are all based on the assumption that an improved representation of antecedent soil moisture conditions in hydrologic models will ensure improved runoff prediction. However, a number of important cases exist where antecedent soil moisture conditions are of relatively minor importance for determining eventual basin response to rainfall. For example, theoretical arguments suggest that the role of antecedent soil moisture is diminished for very intense runoff events that are of primary importance for flood forecasting (Wood et al., 1990). In addition, for basins lacking adequate rain-gauge coverage, constraining antecedent soil moisture represents only a fraction of the overall stream flow prediction problem – the larger fraction of uncertainty being due to error in observed rainfall (Oki et al., 1999). Finally, the relationship between antecedent soil moisture and runoff is strongly nonlinear and characterized by sharp thresholds which are ill-suited for the application of data assimilation techniques designed for linear models.

These difficulties suggest that some merit exists in efforts to reformulate the basis for integrating remote sensing retrievals into hydrologic models. For example, Crow et al. (2009) demonstrates that remotely-sensed surface soil moisture retrievals can also be used to directly improve the accuracy of satellite-based rainfall accumulation estimates. At least in data-poor areas of the world heavily reliant on satellite-based rainfall retrievals, this result broadens the basis of attempts to enhance runoff prediction via surface soil moisture retrievals. Specifically, it presents an opportunity to simultaneously reduce the impact of antecedent soil moisture and rainfall accumulation uncertainty on hydrologic model predictions.

This paper attempts to realize this potential by reframing the remotely-sensed soil moisture/hydrologic forecasting problem in such a way that potential benefits of remotely-sensed soil moisture on both state (i.e. antecedent soil moisture) and flux (i.e. observed rainfall) estimation are captured. Given the dual use of remotely-sensed soil moisture retrievals in this framework, special emphasis will be placed on designing a system that avoids the potentially deleterious

effect of correlated errors between hydrologic model forecasts and assimilated observations.

## 2 Modeling and data

All hydrologic modeling here is based on application of the Sacramento (SAC) hydrologic model. In the United States, the SAC model has been used extensively for operational stream flow forecasting within medium-sized ( $\sim 1000 \text{ km}^2$ ) river basins (Burnash et al., 1973; Geogakakos, 2005). 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 (ADIMP). The movement of water between these states is based on the SAC model parameterization described in Sorooshian et al. (1993).

Combined with measurements of rainfall accumulation, these six states are used to predict 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 capacity, surface saturation runoff (SSR) occurring when rainfall falls on saturated portions of the basin (as defined by ADIMP), shallow sub-surface interflow (SIF) expressed as a direct function of UZFWC, and deep base flow (BF) expressed as a direct function of LZFSC and LZFPC. Here, we will make a distinction between “direct” surface runoff components (SER and SSR) that are driven primarily by incident rainfall and exhibit only a secondary dependence on antecedent soil moisture conditions and “indirect” sub-surface runoff generating processes (SIF and BF) that are wholly a function of soil moisture and do not require the simultaneous presence of non-zero rainfall to generate runoff.

Potential evapotranspiration (PET), daily rainfall ( $P$ ), and stream flow time series data are acquired for specific basins from data sets prepared as part of the Model Parameterization Experiment (MOPEX) (Schaake et al., 2001). Inclusion into the United States portion of the MOPEX experiment was predicated on individual basins meeting threshold requirements related to a lack of anthropogenic stream flow impoundment and/or diversion and possessing adequate spatial rain gauge coverage. Here, we additionally subset the original United States MOPEX datasets to include only basins located below  $36^\circ \text{ N}$  latitude (to minimize snow effects) with an area greater than  $100 \text{ km}^2$  (to eliminate basins smaller than the resolution of soil moisture products expected from next-generation satellite sensors). Of the 438 United States MOPEX basins, 97 meet these two additional criteria. Figure 1 plots long-term runoff ratios (mean annual stream

flow divided by mean annual rainfall) and drainage area for each of these 97 basins. Note the wide range of basic climatic conditions and basins scales considered in the analysis.

Based on MOPEX  $P$  and PET forcing data, the SAC model was run on a daily time step over each of the basins in Fig. 1 during the 55-year period between 1 January 1949 and 31 December 2003. Basin specific model parameters are obtained from SAC model stream flow calibration performed as part of the MOPEX experiment. Based on these calibrated parameters, Figure 2 provides representative examples of observed and predicted stream flow for five of the US MOPEX basins considered here. Stream flow routing is based on convoluting runoff using a simple exponentially decaying unit hydrograph with a folding length varied between 1 and 5 days (depending on basin size). The reasonable performance of the SAC model over a range of climate and basin size conditions suggests that it forms a reliable basis for the synthetic data assimilation experiments to follow.

### 3 Data assimilation

Here, two separate data assimilation approaches are considered for the integration of remotely-sensed soil moisture information into the SAC model. First, the use of a simplified Kalman filtering methodology to correct rainfall input fed into the SAC model. Second, the application of either an Ensemble Kalman filter (EnKF) or smoother (EnKS) to correct SAC soil moisture states based on the availability of remotely-sensed surface soil moisture retrievals. The data assimilation approach utilized for both correction strategies are described in the following two sub-sections (Sect. 3.1 and 3.2). As noted in Sect. 1, the central theme of this paper is unifying these two methodologies and developing a data assimilation system capable of simultaneously correcting both SAC model soil moisture states and rainfall inputs.

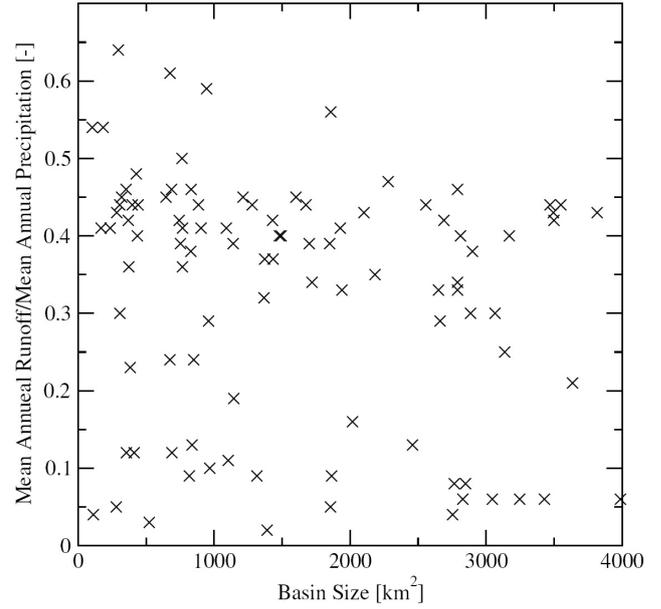
#### 3.1 Rainfall correction using the Kalman filter

Using remotely-sensed soil moisture retrievals from the Advanced Microwave Scanning Radiometer (AMSR-E) aboard the NASA Aqua satellite, Crow et al. (2009) demonstrated the feasibility of correcting uncertain short-term rainfall accumulation estimates using remotely-sensed surface soil moisture retrievals. Their approach is based the assimilation of surface soil moisture retrievals into a simple Antecedent Precipitation Index (API) model

$$\text{API}_j = \gamma_j \text{API}_{j-1} + P'_j \quad (1)$$

where  $j$  is a daily time index,  $P'$  an (uncertain) estimate of daily rainfall accumulation [mm], and  $\gamma$  varies according to day-of-year ( $d$ ) as

$$\gamma_j = \alpha + \beta \cos(2\pi d_j/365). \quad (2)$$



**Fig. 1.** Drainage size and long-term runoff ratio (mean annual runoff/mean annual rainfall) at the outlet of the 97 MOPEX basins used in the study.

Here, the dimensionless parameters  $\alpha$  and  $\beta$  are held constant at values of 0.85 and 0.05. Remotely-sensed surface soil moisture estimates  $\theta$  are used to update Eq. (1) via a Kalman filter

$$\text{API}_j^+ = \text{API}_j^- + K_j(\theta_j - \text{API}_j^-), \quad (3)$$

and “-” and “+” denote API values before and after Kalman filter updating, respectively. Following Reichle and Koster (2005), daily  $\theta$  estimates are obtained by rescaling raw volumetric soil moisture retrievals  $\theta^\circ$  [ $\text{m}^3\text{m}^{-3}$ ] following

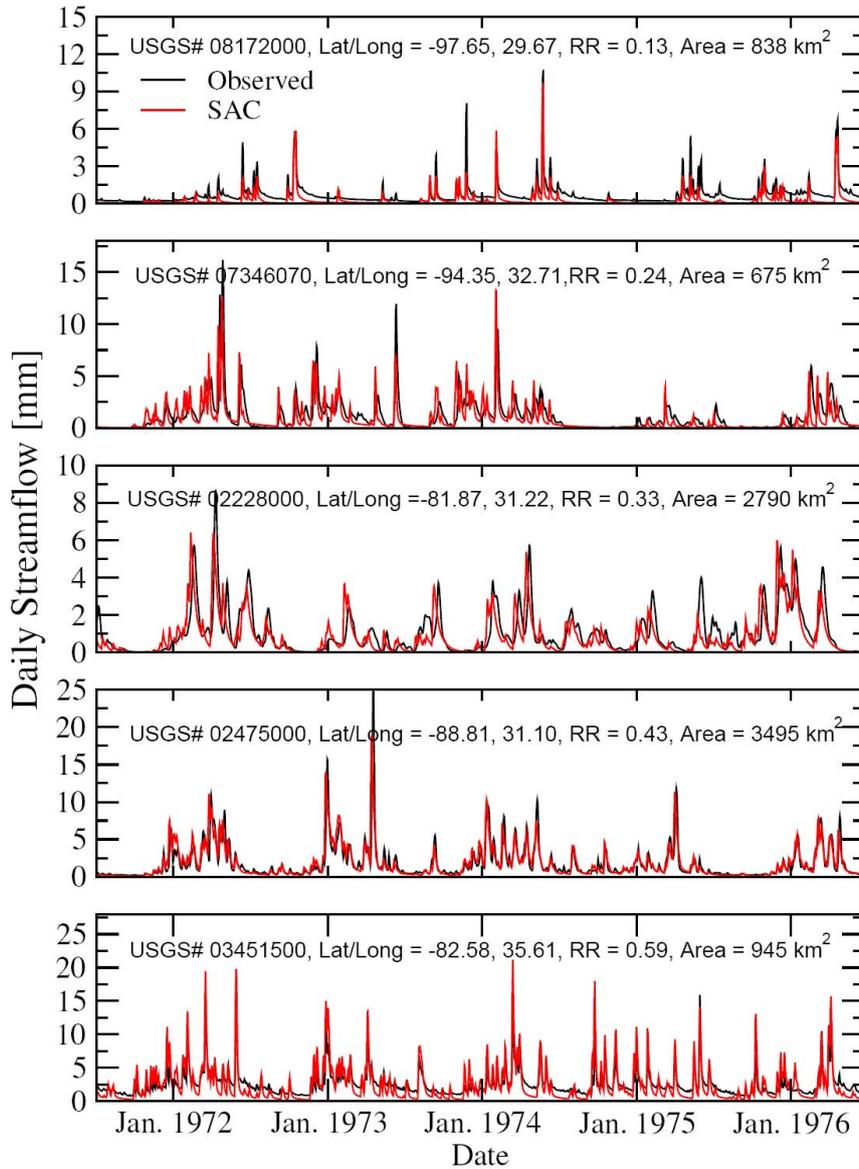
$$\theta_j = (\theta_j^\circ - \mu^\theta)(\sigma^{\text{API}}/\sigma^\theta) + \mu^{\text{API}} \quad (4)$$

to match the API model in expressing soil moisture in water depth units [mm] and ensure that rescaled retrievals possess a long-term mean ( $\mu$ ) and standard deviation ( $\sigma$ ) matching those derived from a multi-year integration of API for the same pixel. Soil moisture retrieval mean ( $\mu^\theta$ ) and standard deviation ( $\sigma^\theta$ ) estimates are obtained by sampling a long-term time series of  $\theta^\circ$ . Likewise, the API mean ( $\mu^{\text{API}}$ ) and standard deviation ( $\sigma^{\text{API}}$ ) statistics in Eq. (4) are sampled from an API time series generated using Eq. (1) and no Kalman filter updating. The Kalman gain  $K$  in Eq. (3) is then given by

$$K_j = T_j^- / (T_j^- + R) \quad (5)$$

where  $T^-$  is the scalar error variance in API forecasts and  $R$  is the error variance of a rescaled  $\theta$  retrieval. At measurement times,  $T^-$  is updated via

$$T_j^+ = (1 - K_j)T_j^-. \quad (6)$$



**Fig. 2.** Comparison of SAC model stream flow predictions (in red) with observed hydrographs (in black) for five representative MOPEX basins. United States Geologic Survey (USGS) basin identification number, latitude/longitude coordinates, long-term runoff ratio (RR) and drained area at basin outlet are listed for each basin.

Between soil moisture retrievals, and the adjustment of API and  $T$  via (3) and (6), API is forecasted in time using observed  $P'$  and (1). In parallel,  $T^+$  is updated in time as

$$T_j^- = \gamma_j^2 T_{j-1}^+ + Q \quad (7)$$

where  $Q$  relates the forecast uncertainty added to an API estimate during propagation between times  $j-1$  and  $j$ . Here temporally constant values of  $R$  and  $Q$  are calibrated on a pixel-by-pixel basis using the innovation tuning procedure described in Crow and Bolten (2007).

To correct rainfall, Crow et al. (2009) utilize analysis increments  $\delta$  calculated during the updating of API with  $\theta$  via (3)

$$\delta_j = \text{API}_j^+ - \text{API}_j^- = K_j(\theta_j - \text{API}_j^-) \quad (8)$$

Values of  $\delta$  reflect the depth of water [mm] added to an API forecast in response to information contained in surface soil moisture retrievals. As such, it contains information concerning errors in near-past  $P'$  estimates used to forecast API. To this end, Crow et al. (2009) propose a simple additive correction which utilizes  $\delta$  to correct errors in uncertain  $P'$  estimates

$$P_j^* = P_j' + \lambda \delta_j. \quad (9)$$

The rescaling parameter  $\lambda$  is required to capture the impact of processes which may lead to differences between  $\delta$  and rainfall errors. Foremost of which is the near certainty that not all errors in API predictions are directly attributable to rainfall uncertainty. Some portion of  $\delta$  will almost certainly be associated with our simplistic representation of soil water loss (i.e. the combined effect of soil drainage and evapotranspiration) in (1). This implies a  $\lambda$  value less than one is required to filter the impact of such error before it can be misattributed to rainfall. Likewise some portion of the original rainfall error is damped via either runoff or infiltration beyond the shallow surface zone prior to the acquisition of a  $\theta$  retrieval used to calculate  $\delta$ . Such processes will require an increase in  $\lambda$  to compensate for the volume of rainfall error that is not directly detectable by the remote sensing observations.

As a practical solution, Crow et al. (2009) propose estimating temporally constant values of  $\lambda$  via the minimization of the root-mean-square difference between corrected rainfall  $P^*$  and some additional estimate of rainfall accumulation. Here, such tuning is performed relative to the benchmark  $P$  obtained from dense rain gauges within each MOPEX basin. Such tuning against high-quality rain gauge data will not be feasible in many data-poor settings; however, Crow et al. (2009) demonstrates that  $\lambda$  can also be accurately specified using an additional, independently-acquired, satellite-based rainfall product.

An additional concern is the possibility that the application of (9) will lead to non-physical negative values of  $P^*$ . Simply resetting such values to zero creates a long-term bias in  $P^*$  values relative to  $P'$ . As an alternative we define a positive threshold  $\tau$  such that  $P_j^* = 0$  for  $P_j^* < \tau$  and  $P_j^* = P_j^* - \tau$  for  $P_j^* \geq \tau$ . The value of  $\tau$  is then iteratively varied until the application of these rules leads to a resulting  $P^*$  time series which is unbiased with respect to  $P'$ .

### 3.2 State correction using the Ensemble Kalman filter or smoother

The Ensemble Kalman filter (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 to update state estimates based on the availability of observations. Here, this ensemble is generated using a combination of noise applied to both SAC model forcing (i.e. PET and  $P$ ) and SAC model soil moisture states (see Sect. 4 for details). At time  $j$ , the vector of SAC model states associated with the  $i$ th Monte Carlo replicate is

$$S_{i,j} = [\text{UZFWC}_{i,j}, \text{UZTWC}_{i,j}, \text{LZTWC}_{i,j}, \text{LZPFW}_{i,j}, \text{LZSFW}_{i,j}, \text{ADIMP}_{i,j}]^T \quad (10)$$

This vector can be transformed into an estimate of volumetric surface soil moisture (assumed to correspond to a remote sensing observation) via the application of the linear observation operator

$$H = [\rho / (\text{UZFWC}_{\max} + \text{UZTWC}_{\max}), \rho / (\text{UZFWC}_{\max} + \text{UZTWC}_{\max}), 0, 0, 0, 0] \quad (11)$$

where  $\rho$  is soil porosity,  $\text{UZFWC}_{\max}$  [m] the maximum capacity of free water in the surface zone and  $\text{UZTWC}_{\max}$  [m] the maximum capacity of tension water in the surface zone. Given the concurrent availability of a remotely-sensed surface soil moisture observation  $\theta^\circ$  with error variance  $R^\circ$ , replicates of  $\mathbf{S}$  are updated following

$$\mathbf{S}_{i,j}^+ = \mathbf{S}_{i,j}^- + \mathbf{K}_j(\theta_j^\circ + v_{i,j} - \mathbf{H}\mathbf{S}_{i,j}^-) \quad (12)$$

where the perturbation term  $v$  is a mean-zero Gaussian random variable with scalar variance  $R^\circ$  and  $\mathbf{K}$  is

$$\mathbf{K}_j = \mathbf{H}\mathbf{C}_j / (\mathbf{H}\mathbf{C}_j\mathbf{H}^T + R^\circ). \quad (13)$$

Here, the forecast error covariance matrix  $\mathbf{C}$  is sampled from a 35-member Monte Carlo ensemble of background SAC model  $\mathbf{S}$  predictions. Final EnKF state predictions are obtained by averaging replicates across the entire ensemble.

The EnKF is designed to update model-forecasted state predictions at the same time an observation is acquired. No attempt is made to reanalyze previous model predictions in response to a particular observation. In contrast, the Ensemble Kalman Smoother (EnKS) can be used to update all model states predictions within a fixed lag of past time (Dunne and Entekhabi, 2005). While the SAC model is run on a daily time step, variations in the three free water states (i.e. UZFWC, LZPFW, and LZSFW) and ADIMP are actually calculated on a three-hourly basis using an sub-daily model time loop. For our application of the EnKS, an augmented  $\mathbf{S}_j$  vector is created ( $\mathbf{S}_{j-1 \rightarrow j}^*$ ) which contains not only the six SAC model soil moisture state variables at time  $j$  but also all SAC model state predictions between times  $j-1$  and  $j$  (inclusive of end points) and including 3-hourly water balance calculations of UZFWC, LZPFW, LZSFW and ADIMP. 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 (10)

$$\mathbf{S}_{i,j-1 \rightarrow j}^{*,+} = \mathbf{S}_{i,j-1 \rightarrow j}^{*,-} + \mathbf{K}_j^*(\theta_j^\circ + v_{i,j} - \mathbf{H}^*\mathbf{S}_{i,j-1 \rightarrow j}^{*,-}) \quad (14)$$

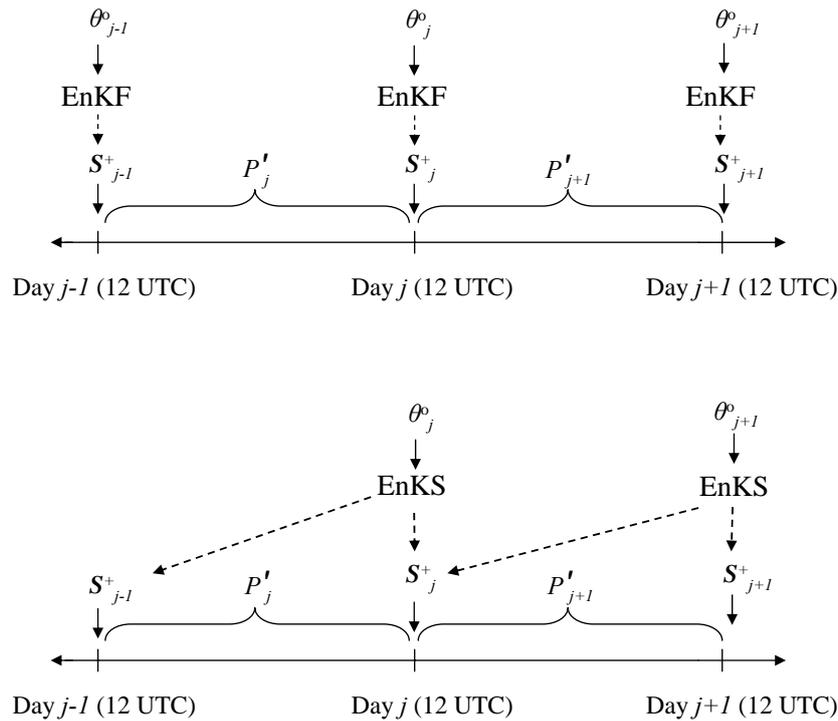
where

$$\mathbf{K}_j^* = \mathbf{H}^*\mathbf{C}_j^* / (\mathbf{H}^*\mathbf{C}_j^*\mathbf{H}^{*T} + R^\circ) \quad (15)$$

and  $\mathbf{H}^*$  is a 40-element vector of the form

$$\mathbf{H}^* = [\rho / (\text{UZFWC}_{\max} + \text{UZTWC}_{\max}), \rho / (\text{UZFWC}_{\max} + \text{UZTWC}_{\max}), 0, \dots, 0]. \quad (16)$$

As in the EnKF, final EnKS state predictions are obtained by averaging across the updated soil moisture ensemble.



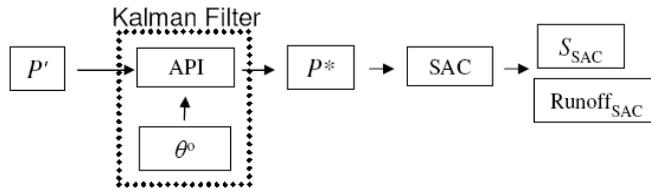
**Fig. 3.** Schematics for the assimilation of remotely-sensed soil moisture retrievals  $\theta^o$  into the SAC model (to improve its internal soil moisture states  $S$ ) using both an Ensemble Kalman filtering (EnKF; top) and fixed-lag Ensemble Kalman Smoothing (EnKS; bottom) approach.

Figure 3 provides a brief illustration of differences between the EnKF and a fixed-lag EnKS approach. For a real-time filtering problem (Fig. 3a), a soil moisture observation at time  $j$  is used to update concurrent SAC model state replicates at time  $j$  using an EnKF. These updated forecasts, and an estimation of total rainfall accumulation occurring between time  $j$  and  $j+1$ , are then used to initiate a SAC model ensemble of states predictions between times  $j$  and  $j+1$ . Alternatively, the entire analysis could be delayed until a soil moisture observation is obtained at time  $j+1$ . In this formulation, the one-day, fixed-lag EnKS is employed to update all SAC model state replicates between  $j$  and  $j+1$  using the soil moisture observation at time  $j+1$  (Fig. 3b). Note that, unlike the EnKF, the EnKS allows for SAC model states between  $j$  and  $j+1$  to be corrected based on the observation obtained at time  $j+1$ . The key advantage of the EnKS is that state estimates at time  $j$  (as well as intermediate free water states calculated between  $j$  and  $j+1$ ) are constrained via information gleaned from the subsequent observation at time  $j+1$ . In contrast, the EnKF is only forward propagating in the sense that EnKF estimates at any particular time are not impacted by subsequent observations. Consequently, flux and state predictions obtained from the EnKS should be relatively more accurate than comparable predictions by the EnKF (Dunne and Entekhabi, 2005).

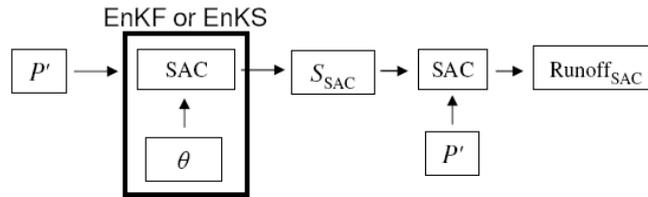
#### 4 Synthetic experiment methodology

Our overall approach is based on the application of the Sacramento (SAC) hydrologic model to 97 MOPEX study basins along the southern tier of the US. A series of synthetic data assimilation experiments are individually conducted for each basin. All such experiments are based on the designation of output from a single SAC model realization as “truth”. The approximate realism of these truth simulations is supported by comparisons between their stream flow predictions and long-term hydrographs obtained from stream flow observations taken at the outlet of MOPEX basins (Fig. 2). Runoff and soil moisture predictions from the truth SAC runs are withheld to serve as a benchmark for future runs and surface soil moisture predictions (perturbed by a suitable amount of additive Gaussian noise) are assumed to represent remotely-sensed surface soil moisture retrievals. Using either an EnKF or EnKS approach (see Fig. 3), these retrievals are subsequently assimilated back into a perturbed representation of the SAC model to examine the degree to which their integration can correct the perturbed SAC model simulation back to benchmark results obtained in the “truth” SAC model simulation. Results obtained directly from the perturbed representation of the SAC model (prior to the implementation of any data assimilation technique) are referred to as “open loop” results which define the baseline by which the relative improvement in subsequent data assimilation results is quantified.

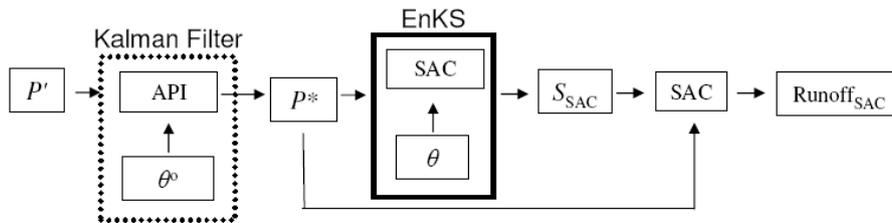
**Case 1 - Rainfall Correction only**



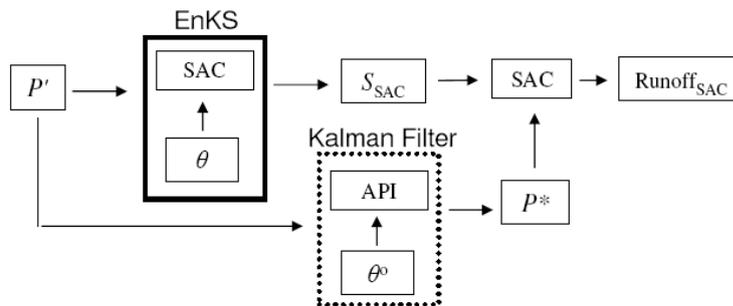
**Case 2 (EnKF) / Case3 (EnKS) - State Correction Only**



**Case 4 - Sub-Optimal State/Rainfall Correction**



**Case 5 - Dual State/Rainfall Correction**



**Fig. 4.** Schematics for five cases of incorporating remotely-sensed surface soil moisture retrievals ( $\theta$  or  $\theta^o$ ) into SAC model runoff ( $Runoff_{SAC}$ ) and soil moisture ( $S_{SAC}$ ) predictions. The dashed box in Case 1, 4 and 5 represents the observed rainfall ( $P'$ ) rainfall correction procedure outlined in Sect. 3a. The solid box in Cases 2, 3, 4 and 5 represents either the EnKF or EnKS-based assimilation of surface soil moisture retrievals into the SAC model.

Perturbations to the SAC model are based on additive noise applied directly to SAC water balance states in  $\mathbf{S}$  and the daily PET input time series. Daily perturbations applied to individual 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 5% of the total capacity of each state. Additive PET perturbations are similarly uncorrelated and sampled from a mean-zero, Gaussian distribution with a standard deviation

of 1 mm. Negative PET values resulting from such perturbations are simply reset to zero. In addition to internal model and PET errors, uncertainty in rainfall is captured through the multiplicative scaling of observed rainfall  $P$  with a random factor  $\chi$  sampled from a mean-one, log-normal distribution with a dimensionless standard deviation of one

$$P' = \chi P, \chi \sim \text{Log}_N(1, 1). \tag{17}$$

For our particular representation of a synthetic twin experiment, all model perturbations (presented above) are actually applied twice. During their first application, they are applied to degrade the SAC model truth simulation and create a perturbed open-loop SAC model simulation. Subsequently, they are re-applied to the open model simulation (on top of the original set of perturbations) to create an ensemble of SAC model runs (calculated around the perturbed SAC model simulation) during the application of an EnKF or EnKS to correct the perturbed SAC model simulation back to the truth simulation. In addition, the same set of synthetically-generated soil moisture retrievals assimilated into the SAC model are also assimilated into an API model (see Sect. 3.1) in an attempt to correct for precipitation error introduced into SAC precipitation forcing via (17). In this way, the synthetic experiment accounts for the possibility of correcting both SAC model state and rainfall forcing error.

Remotely-sensed surface soil moisture retrievals are assumed to be available at a daily frequency with a root-mean-squared (RMS) accuracy of  $0.03 \text{ m}^3 \text{ m}^{-3}$  (defined as the fraction of total soil volume occupied by water).  $R^\circ$  is the square of this value and  $R = R^\circ (\sigma^{\text{API}}/\sigma^\theta)^2$ . During the application of the EnKS and EnKF within the synthetic experiment, all model and observational error covariances are assumed to be known. However, the sensitivity of key experimental results to the magnitude of these covariance values is examined in Sect. 6.3.

## 5 State and/or Rainfall Correction strategies

Our primary analysis will focus on comparing soil moisture and runoff results derived from the five separate data assimilation strategies outlined in Fig. 4. The first “Rainfall Correction” strategy (Case 1) is based on the application of the Crow et al. (2009) procedure reviewed in Sect. 3.1 to correct rainfall forcing data prior to its use as a SAC model forcing variable. Note that this approach does not involve the actual assimilation of soil moisture retrievals into the SAC model. Instead, the “Rainfall Correction” approach attempts to improve runoff prediction solely through the correction of SAC rainfall forcing. Conversely, the “State Correction Only – EnKF/EnKS” approach (comprising Cases 2 and 3 in Fig. 4) employs the assimilation of surface soil moisture retrievals into the SAC model using an EnKF (or EnKS) without attempting to correct model rainfall input. Starting with Case 2, we reference the SAC model twice in the schematic for each case. The first reference occurs as part of an ensemble created to run the EnKF or EnKS and predict SAC model soil moisture states in  $\mathbf{S}$  (or  $\mathbf{S}^*$ ). The second occurrence is during a post-processing step in which the ensemble-mean of these state predictions are directly inserted into a single realization of the SAC model for the sole purpose of predicting runoff ( $\text{Runoff}_{\text{SAC}}$  in Fig. 4). Note that the ensemble-mean soil moisture prediction made by this post-processing run is

not used to initialize any subsequent SAC model forecast. At least for the Case 2 implementation of the EnKF, it is also possible to neglect this post-processing stage and simply average SAC/EnKF runoff predictions across the ensemble to obtain a single EnKF runoff prediction. However, we found that the inclusion of the post-processing stage had a generally beneficial impact on EnKF runoff predictions relative to this alternative approach. Consequently, we retained the use of a post-processing step for all EnKF-based data assimilation results.

The “State Correction Only – EnKS” approach (Case 3) is identical to Case 2 except that estimation of the augmented SAC model state vector  $\mathbf{S}^*$  is based on implementation of a one-day, fixed-lag EnKS – rather than an EnKF – to update SAC model soil moisture states (Fig. 3). Both the EnKF and EnKS are applied to produce Cases 2 and 3, respectively. However, to reduce the proliferation of cases, only the EnKS is employed for Cases 4 and 5 described below.

None of the first three cases in Fig. 4 take the next step of simultaneously attempting both rainfall and state correction based on remotely-sensed surface soil moisture retrievals. This possibility is first examined in Case 4 where corrected rainfall is used to both force an EnKS state correction procedure and during the post-processing calculation of runoff. This type of approach is potentially problematic in that surface soil moisture retrievals are used both to modify forcing data for SAC model forecasts and as observations which are subsequently assimilated into the SAC model via the EnKS. Such dual use of soil moisture retrievals can conceivably lead to correlation between forecasting and observations errors within the EnKF, and, consequently, sub-optimal filter performance. A final potential strategy (Case 5) tries to mitigate this possibility by utilizing corrected rainfall only in the post-processing calculation of runoff (Fig. 4) and using uncorrected rainfall ( $P'$ ) for generation of the SAC model forecast ensemble in the EnKS. Since soil moisture predictions made during the post-processing stage are not fed back into the EnKS, this strategy avoids the potential for cross-correlated errors within the EnKS while still allowing for the dual correction of errors present in both antecedent soil moisture and rainfall.

A possible simplified structure for Case 4 and 5 schematics in Fig. 4 is to eliminate the API model and instead use SAC analysis increments (produced during the EnKS-based state correction procedure) to correct rainfall. However, it is currently unclear whether the API-based approach in Sect. 3.1 can be successfully applied to the SAC model. In particular, adaption of the rainfall correction scheme to a multi-state hydrologic model is complicated by large variations in water storage capacity existing between various soil water model states. These variations imply that analysis increments applied to each soil water state will respond to antecedent rainfall errors occurring over different time scales. Unless an arbitrary decision is made to exclude analysis increments applied to certain states, such variation requires a

more complex form of (9) (currently limited to operating at a single time scale) and the specification of additional  $\lambda$  scaling factors. In addition, nonlinear, multi-state land surface models like the SAC model can badly confound the innovation-based tuning procedure required to implement the procedure (Crow and Van Loon, 2006). While these problems are potentially resolvable, real-data verification of a rainfall correction procedure has been limited to the API rainfall correction approach presented in Sect. 3.1, and current prospects for basing the approach on more complex models are unclear. Consequently, in an attempt to maximize the realism of the synthetic experiments presented here, our analysis will follow Crow et al. (2009) and retain the use of an API model for the rainfall correction procedure. Further discussion of this point is presented in Sects. 7 and 8.

## 6 Results

Figure 4 lays out a number of possible approaches for integrating remotely-sensed surface soil moisture retrievals into runoff estimates produced by a hydrologic model. To date, most data assimilation studies focusing on this goal have followed Case 2 by formulating the problem purely in a state estimation framework and applying a sequential filtering algorithm to improve the estimation of pre-storm antecedent soil moisture conditions in the hope that this will aid in the subsequent estimation of storm-scale runoff. As stated above, our primary focus is on evaluating the added benefit of reformulating the runoff estimation problem as a smoothing reanalysis problem (e.g. Case 3) and attempting the simultaneous correction of both model soil moisture states and the rainfall forcing used to drive the model (e.g. Cases 4 and 5). Figure 5 shows sample time-series results for a single MOPEX basin. Given the availability of remotely-sensed surface soil moisture retrievals, one can correct a time series of daily rainfall accumulations (Fig. 5a) and/or implement an EnKF (or EnKS) to correct SAC model soil moisture predictions (Fig. 5b). Both types of corrections should aid in the subsequent calculations of runoff by the SAC model. Cases 1, 2 and 3 explore the application of one type of correction (antecedent soil moisture or rainfall) in isolation. However, Cases 4 and 5 explore the possibility of obtaining better SAC model runoff estimates by simultaneously implementing both corrections (Fig. 5c).

### 6.1 MOPEX basin results

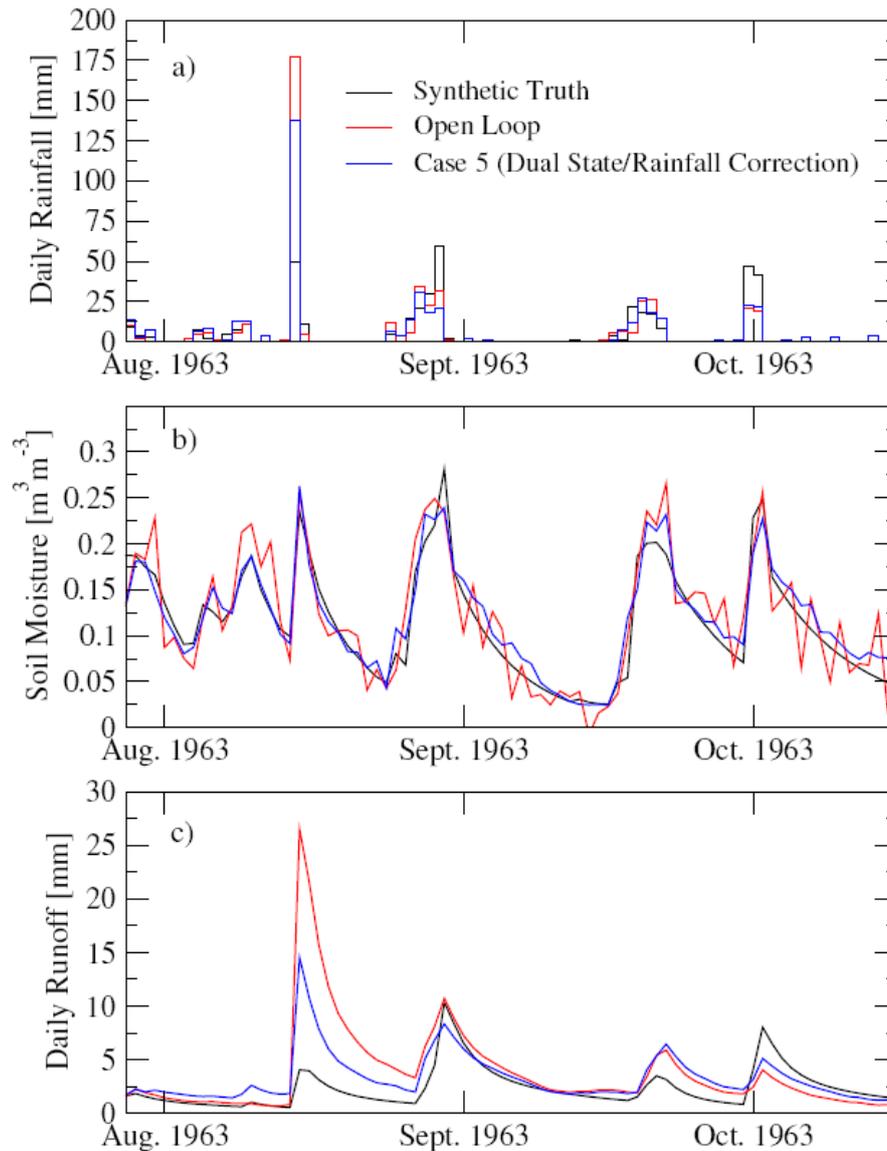
Based on the synthetic twin experimental methodology introduced in Sect. 4, Fig. 6 compares runoff and upper-zone soil moisture root-mean-square error (RMSE) results calculated for all 97 MOPEX basins and the five separate data assimilation cases described in Fig. 4. Unless otherwise noted, all subsequent results are presented as normalized RMSE in which open loop SAC model RMSE results are used to nor-

malize RMSE results obtained after the implementation of various data assimilation techniques. Since normalized values reflect the fraction of modeling error that is addressed by a particular technique, an improvement in performance relative to the uncorrected open loop case is captured by a normalized RMSE value less than one (see dotted line in Fig. 6). All RMSE results are based on daily SAC model predictions made during the 55-year period between 1 January 1949 and 31 December 2003. Symbols in Fig. 6 represent the mean for all basins and error bars reflect the one-standard deviation spread of normalized RMSE across all 97 basins.

In Fig. 6, results for the case of rainfall correction only (Case 1) and of EnKF-based state correction (Case 2) are diametrically opposed in that Case 1 reduces daily rainfall RMSE relative to the open loop case, but provides little, if any, net improvement to upper-zone soil moisture predictions – defined as the product of  $\mathbf{H}$  in (11) and  $\mathbf{S}$  in (12). In contrast, application of the EnKF to correct antecedent soil moisture predictions yields a significant improvement to upper-zone soil moisture estimates but leads to no net improvement in daily runoff. Modifying the state-estimation technique to be based on a fixed-lag EnKS reanalysis (Case 3) clearly enhances the accuracy of both runoff predictions and soil moisture estimates relative to the analogous EnKF-based case (Case 2).

Despite this improvement, Case 3 results are still based solely on the application of a state-correction strategy. Cases 4 and 5 results in Figure 6 demonstrate how optimal aspects of Case 1, 2 and 3 runoff and soil moisture results can be combined, and even enhanced, by reformulating the estimation problem using either of the dual state/rainfall strategies (Cases 4 and 5) outlined in Fig. 4. In particular, Case 5 is able to match the high soil moisture accuracy of Case 3 while providing runoff results which are even slightly better than already good Case 1 results.

As noted in Sect. 1, a danger in our strategy for simultaneously correcting both rainfall and internal soil moisture states is that information contained in surface soil moisture retrievals will be overexploited – leading to the possibility of degenerate runoff predictions. Case 4 results in Fig. 6 illustrate such an example. Here, surface soil moisture retrievals are used both to correct rainfall amounts used to forecast the SAC model ensemble and as the observation assimilated into the ensemble via an EnKS. This leads to cross-correlation between SAC model forecasting error and observation error in remotely-sensed soil moisture retrievals assimilated into the SAC model by the EnKS. Such correlation violates a key Kalman filtering assumption and degrades Case 4 soil moisture and runoff results relative to their Case 5 equivalents (Fig. 6). By withholding the use of corrected rainfall until the post-processing calculation of runoff (and discarding soil moisture predictions made by the SAC model during this calculation), Case 5 avoids the negative impact of cross-correlated errors and produces superior runoff and soil moisture predictions.



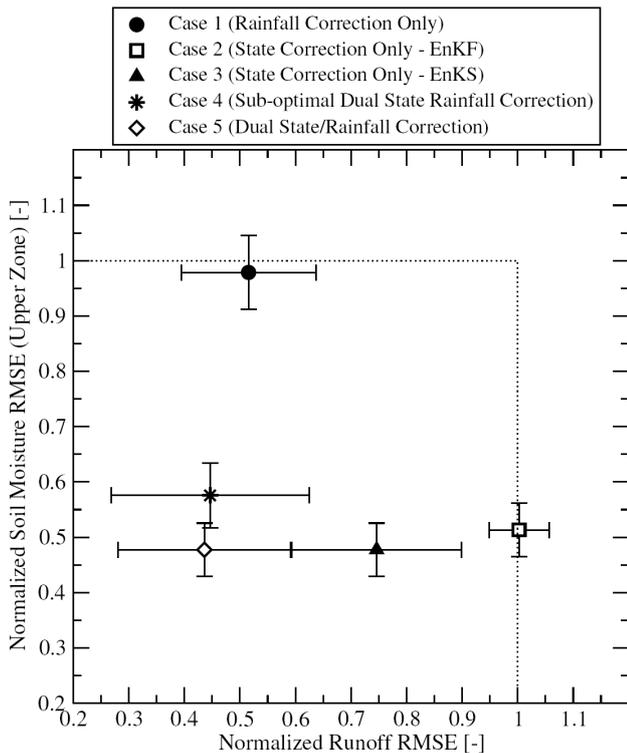
**Fig. 5.** For USGS basin #02228000, example time series of truth, open case and corrected (Case 5) (a) rainfall, (b) SAC upper-zone soil moisture and (c) SAC runoff results.

While mildly degraded results are noted in Fig. 6, the full effect of this degeneracy appears only in SAC lower-zone soil moisture results. Figure 7 is identical to Figure 6 except the y-axis is re-plotted as normalized lower-zone soil moisture RMSE (instead of daily runoff). Lower-zone soil moisture is defined as

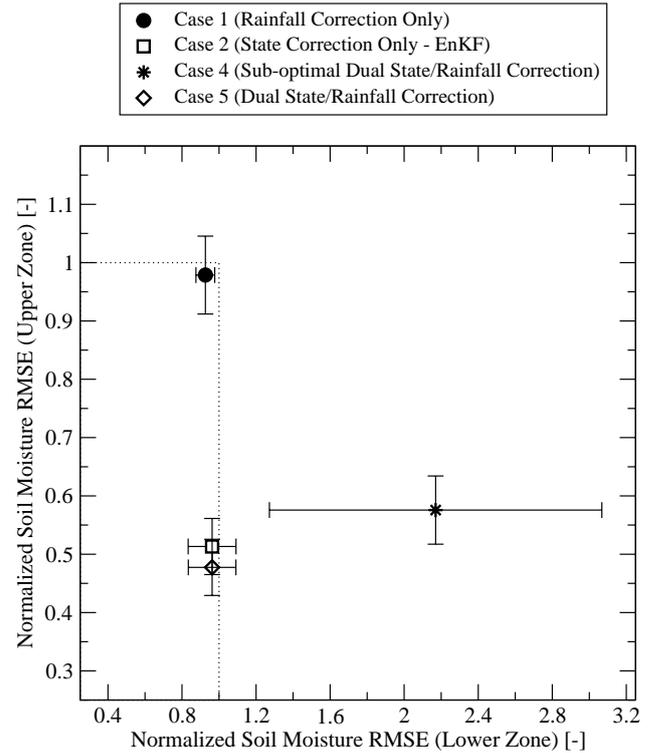
$$\theta_{\text{zone2}} = \rho(\text{LZTWC} + \text{LZPFC} + \text{LZSFC}) / (\text{LZTWC}_{\text{max}} + \text{LZPFC}_{\text{max}} + \text{LZSFC}_{\text{max}}) \quad (18)$$

where  $\text{LZTWC}_{\text{max}}$ ,  $\text{LZPFC}_{\text{max}}$  and  $\text{LZSFC}_{\text{max}}$  are maximum capacities of SAC model states LZTWC, LZPFC and LZSFC, respectively. Because the states underlying lower-zone soil moisture are not directly observed via  $\mathbf{H}^*$  in (16), and the SAC model predicts relatively little vertical cou-

pling between its upper and lower soil zones, all cases in Fig. 4 provide only a modest correction relative to the open loop. However, Case 4 results cannot even meet this minimal threshold and instead clearly degrade lower-zone soil moisture predictions relative to the open loop. The source of this degradation is the cross-correlation of modeling and observational error induced by using corrected rainfall accumulations during the EnKS forecast step. The long-term effects of this correlation are particularly pernicious for lower-zone soil moisture estimates since these values cannot be directly constrained via surface observations and can therefore accumulate unchecked over long time periods. As a result of this problem, Case 4 results will be dropped from the remainder of the analysis.



**Fig. 6.** Upper-zone soil moisture and runoff results for the five cases outlined in Fig. 4. Plotted symbols represent the mean of RMSE results (normalized by open loop RMSE results) for all basins. Error bars represent the one-standard deviation spread of normalized RMSE results across all 97 MOPEX basins.



**Fig. 7.** Upper- and lower-zone soil moisture results for the five cases outlined in Fig. 3. Plotted symbols represent the mean of RMSE results (normalized by open loop RMSE results) for all basins. Error bars represent the one-standard deviation spread of results across all 97 MOPEX basins. Case 3 results (not shown) correspond exactly to shown Case 5 results.

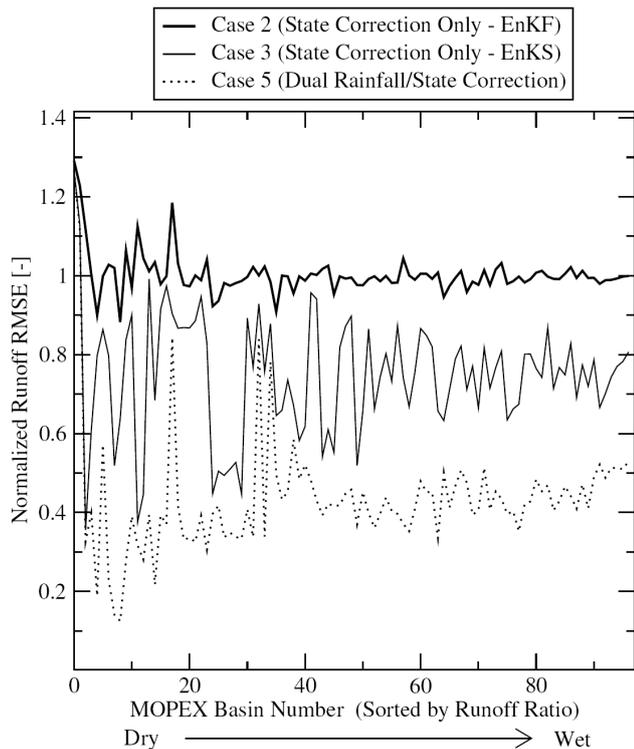
### 6.2 Sensitivity of results to climate and runoff processes

As demonstrated in Fig. 1, MOPEX basins selected for this study capture a wide range of long-term runoff ratio values. Such variability is lost upon the averaging performed to construct Figs. 6 and 7. In order to examine any possible trends with regards to climate, Figure 8 re-plots normalized daily runoff RMSE results as a function of long-term basin runoff ratio (sorted from the driest to the wettest of the 97 MOPEX basins). Despite a large range of overall basin wetness, little variation is observed when moving from drier to wetter basins. For all basins, regardless of long-term climate characteristics, Case 2 provides little or no added skill to runoff predictions; however roughly equal added skill is obtained upon reformulating the problem using a smoothing approach (Case 3) and, subsequently, adding a rainfall correction component (Case 5).

Despite a lack of strong variation of results with climate, insight into Figs. 6 and 8 can be obtained by decomposing total runoff results into various individual runoff processes captured by the SAC model. Here, total SAC model runoff consists of four separate components: surface infiltration-excess runoff (SER), surface saturation runoff (SSR), shal-

low sub-surface interflow (SIF) and deep sub-surface base flow (BF). A useful classification is to divide these four separate processes into “direct” and “indirect” runoff generation processes. Indirect runoff components SIF and BF are runoff processes in which the rainwater path to channel flow proceeds through one (or more) of the SAC model soil moisture states. The rate at which these processes operate is therefore a direct function of soil moisture and only indirectly linked to antecedent rainfall. Consequently, they can be adequately constrained by state estimation techniques. The impact of this is seen in Fig. 9, where no added advantage (in terms of RMSE accuracy) is associated with adding our rainfall correction approach on top of EnKS state estimation results (i.e. equivalent results for SIF and BF are obtained in Cases 3 and 5). Overall better correction results for SIF relative to BF can be attributed to the sensitivity of SIF to upper-zone soil moisture states that are assumed to be directly observed by remotely-sensed surface soil moisture retrievals.

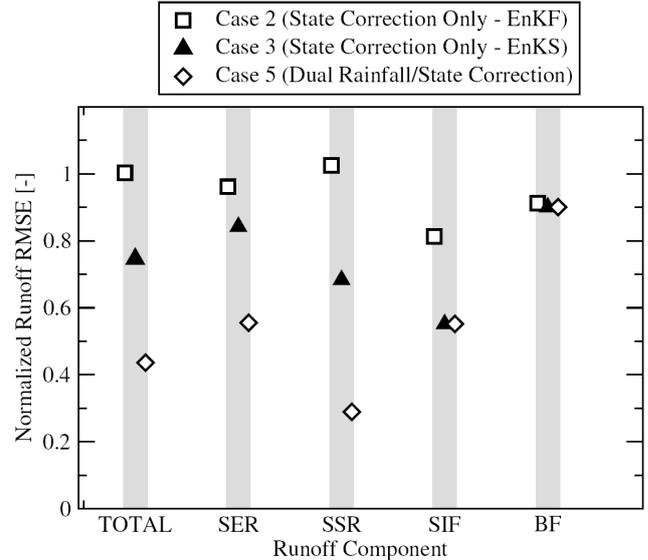
In contrast, direct runoff processes are those in which – during saturated surface conditions – rainfall is directly routed to runoff without first transitioning through an intermediate soil moisture state. Consequently, antecedent soil



**Fig. 8.** The impact of basin runoff ratio (mean annual runoff/mean annual rainfall) on Case 2, 3 and 5 runoff results.

moisture impacts these processes only indirectly through the specification of a pre-storm infiltration capacity or the extent of saturated contributing areas. Improved specification of these soil moisture states via application of the EnKS leads to improved SER and SSR results relative to the EnKF baseline (compare Cases 2 and 3 in Fig. 9). However, because of their direct link to rainfall, SER and SSR estimates can be further enhanced through the application of our dual rainfall/state correction procedure (compare Cases 3 and 5 in Fig. 9). Therefore the relative advantage of Case 5 (noted in Figs. 6 and 8) is based solely on the improved constraint of direct, surface runoff processes captured by the SAC model.

The importance of direct runoff processes can also be observed when varying the performance metric by which SAC runoff predictions are evaluated in Fig. 6. Qualitatively similar results are obtained when regenerating Fig. 6 using mean absolute error (MAE), as opposed to RMSE, as the performance metric for SAC runoff predictions (not shown). However, the relative magnitude of correction observed in Case 5 results is reduced. For instance, defining the relative fraction of open loop error in terms of MAE (i.e. assimilation MAE/open loop MAE) as opposed to RMSE, increases the fraction of open loop error for Case 5 results from 0.44 to 0.75 and decreases the marginal advantage of Case 5 results versus Case 3 results from 0.31 to 0.11. This reduction is associated with the reduced weight that MAE applies to large



**Fig. 9.** Case 2, 3 and 5 total runoff results decomposed into surface excess runoff (SER), surface saturation runoff (SSR), surface interflow (SIF) and base flow (BF) components.

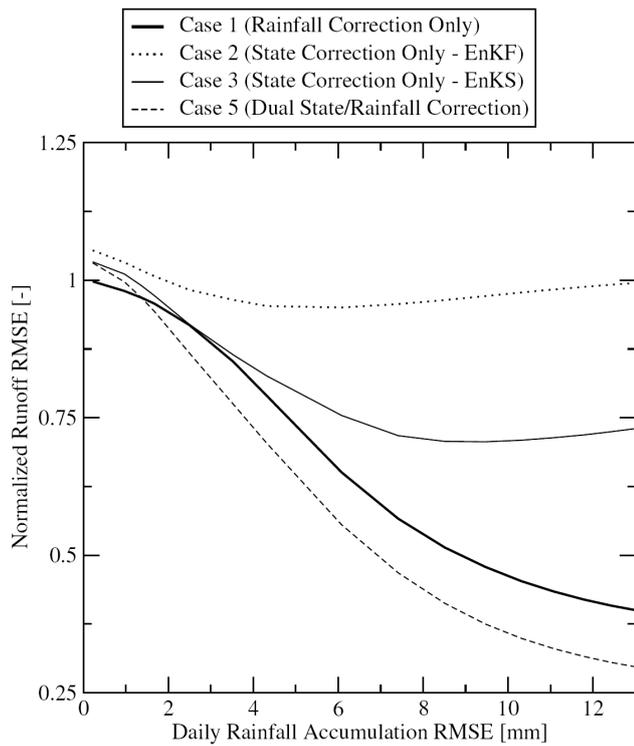
runoff events relative to RMSE and would seem to indicate that the marginal benefit of our dual state/rainfall correction procedure (as expressed by the difference between Case 5 and Case 3 results) lies primarily in constraining relatively high flow events dominated by direct surface runoff.

### 6.3 Sensitivity of results to error assumptions

A large number of assumptions underlie synthetic data assimilation results presented in Figs. 5 to 9. Perhaps most critically, the magnitude of synthetic noise, introduced to represent observational and modeling uncertainty in the synthetic experiment, is specified in a somewhat arbitrary manner. Here we examine the sensitivity of key results to these values.

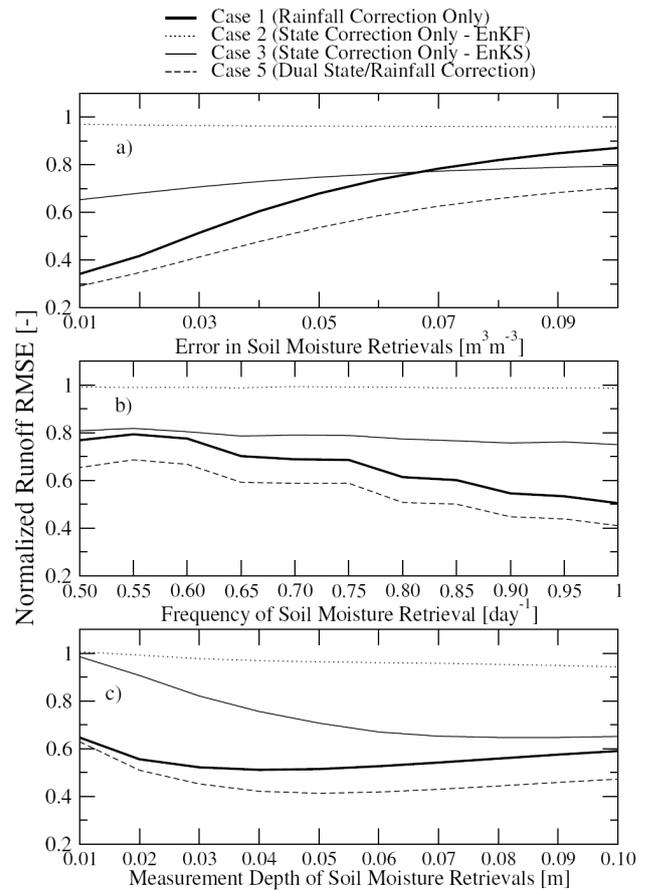
The introduction of error in rainfall observations is based on the multiplicative rescaling of daily rainfall values by a random variable sampled from a mean-one, log-normal distribution. By varying the standard deviation of this distribution, various levels of RMSE error in estimates of daily rainfall accumulation can be obtained. For instance, the default choice of one for the standard deviation of  $\chi$  in (17) produces an average daily rainfall RMSE of about 8.5 mm. Figure 10 recalculates Case 1, 2, 3 and 5 results for a range of specified standard deviations, and thus long-term RMSE, in daily rainfall accumulations. For computational reasons, these sensitivity results are derived for only the sub-set of 5 MOPEX basins shown in Fig. 2.

For small rainfall errors, Fig. 10 demonstrates minor runoff corrections relative to the open loop. This suggests that, for well-instrumented basins in which highly accurate



**Fig. 10.** Sensitivity of Case 1, 2, 3 and 5 runoff results to the magnitude of rainfall error.

rainfall accumulation estimates can be obtained, none of our proposed strategies for integrating surface soil moisture retrievals are effective for correcting SAC model runoff predictions relative to the open loop. However, as rainfall error increases, substantial improvement is noted for Case 1 (“Rainfall Correction Only”), Case 3 (“EnKS State Correction Only”) and Case 5 (“Dual State/Rainfall Correction”) results. Of particular relevance is the relative difference between the best state correction-only case (clearly Case 3) and the dual state/rainfall correction case (Case 5). A substantial difference between the two cases does not appear until a moderate (>4 mm) level of rainfall accumulation RMSE is reached. Above this point, however, the relative advantage of Case 5 is clear and a substantial relative advantage is associated with the implementation of our rainfall correction scheme. Over continental areas, levels of daily RMSE between 6 and 10 mm are common in many satellite rainfall accumulation products lacking rain gauge correction (see e.g. Crow and Bolten, 2007). Consequently, it appears that the largest applicability of our approach will be for regions in which operational hydrologic forecasting applications depend heavily on uncorrected satellite retrievals for real-time rainfall information. The procedure will be of substantially less value for heavily instrumented regions in which accurate real-time rainfall accumulation information is available from ground-based instrumentation.



**Fig. 11.** The sensitivity of Case 1, 2, 3 and 5 runoff results to (a) the accuracy, (b) the frequency and (c) the vertical measurement depth of remotely-sensed surface soil moisture retrievals.

Conversely, one might expect a reverse trend when varying the magnitude of perturbations applied directly to internal model states and/or SAC PET inputs (see Sect. 4). Since these perturbations are not tied to rainfall uncertainty, an increase in their magnitude will increase the fraction of total modeling error that cannot be addressed through our rainfall correction scheme. Consequently, the additional advantage of the dual correction strategy in Case 5 might be lessened relative to the application of the state-correction only approach in Case 3. However, this tendency is not noted in sensitivity results in which the magnitude of these perturbations is increased. Such results (not shown) demonstrate little variation in the performance of Case 3 and Case 5 relative to the open loop. One potential reason for this lack of sensitivity may be known bias problems encountered when propagating mean-zero model state perturbations (as required by the Monte Carlo nature of the EnKS and EnKF) through a nonlinear model (Ryu et al., 2009). These biases limit the effectiveness of EnKF or EnKS state correction techniques when applied to models with higher levels of internal uncertainty. This tendency may counter the relative advantage

enjoyed by state-correction techniques when internal modeling errors are large compared to external rainfall forcing errors. Regardless of the specific cause, the relative advantage of Case 5 versus Case 3 seen in Figs. 6 and 8 is essentially maintained for a wide range of error variances assumed for perturbations to internal SAC model states and PET input.

## 7 Sensitivity of results to observation characteristics

In addition to assumptions concerning modeling uncertainties, a series of attributes are also assumed for remotely-sensed surface soil moisture retrievals. Specifically, they are assumed to be available on a daily frequency, measure approximately the top 10 centimeters of the soil column and have a RMSE accuracy of  $0.03 \text{ [cm}^3 \text{ cm}^{-3}]$  volumetric. In general, these assumptions are optimistic reflections of expectations for next-generation satellite retrievals and the impact of less ideal retrieval conditions must be considered.

Figure 11 displays Case 1, 2, 3 and 5 results for a series of synthetic data assimilation experiments in which the accuracy, frequency and measurement depth of surface soil moisture retrievals have been systematically varied. With regards to retrieval accuracy (Figure 11a) and frequency (Fig. 11b), there exists a systematic narrowing of the difference between Case 3 and Case 5 as retrieval error increases and/or frequency decreases. This suggests that benefits of our rainfall correction approach are relative more sensitive to limitations in the accuracy and frequency of retrievals than EnKF/EnKS-based state correction approaches. Given the need to correct daily rainfall accumulation amounts, the reduction in accuracy observed in Fig. 11b for retrieval frequencies of less than once per day is not surprising. However, it is worth noting that from the mid-latitudes to the poles, combining ascending and descending overpass data from passive microwave sensors (e.g. AMSR-E) typically provides measurements for at least 4 out of every 5 days.

Of all the assumptions underlying the generation of synthetic retrievals, the least realistic is probably the assumption of a 10-cm vertical measurement depth. This assumption was made in order to make the observational support of remote sensing retrievals consistent with calibrated values of SAC model upper-zone layer depth obtained from the MOPEX experiment. However, a 10-cm measurement depth is larger than typical estimates for the vertical penetration depth of remotely-sensed surface soil moisture retrievals (usually between 1 and 5 cm). Consequently, the impact of smaller measurement depths must be considered. Figure 11c displays results for the systematic reduction of the upper-zone depth in the SAC model to values smaller than 10 cm. It reveals a general tendency for the difference between Case 3 and Case 5 results to increase upon a decrease in the upper-zone depth of the SAC model. There are several reasons for this tendency. First, utilizing a thin upper-zone in the SAC model prompts the model to produce higher amounts of direct sur-

face runoff relative to indirect, sub-surface runoff. Such a shift is critical because the basis of improved Case 5 results (relative to Case 3) is the presence of substantial amounts of direct surface runoff (Fig. 9). In addition, the use of a thinner upper-zone decreases correlation between observations of the upper-zone and the non-observed lower-zone. This, in turn, limits the ability of the EnKS to accurately constrain lower-zone soil moisture variables. Consequently, our choice of an unrealistically thick upper-zone likely reduces the relative positive impact of introducing our rainfall correction scheme into hydrologic forecasting.

## 8 Operational prospects

All results presented here are based on a synthetic twin experimental methodology in which remotely-sensing surface soil moisture retrievals are artificially generated and assimilated into a hydrologic model. Such experiments are required as an initial proof-of-concept for new data assimilation systems. Nevertheless, it is important to consider the likelihood of duplicating encouraging synthetic results when using actual remote sensing data.

For instance, a key result in this analysis is the demonstration that adaptation of our dual rainfall and soil moisture correction scheme (Case 5 in Fig. 3) can improve SAC model runoff results above and beyond levels obtainable using the best state correction technique (Case 3 in Fig. 3). Consequently, an important issue is the degree to which assumptions and design decisions imbedded in our synthetic experiment methodology affect the magnitude of this difference. On this point, it should be noted that – in our particular synthetic twin methodology – state correction-only cases (Cases 2 and 3) retain an artificial advantage in that synthetic surface soil moisture retrievals are generated by the same model (the SAC model) that they are subsequently assimilated into. In the terminology of synthetic data assimilation experiments this is referred to as an identical-twin experiment. In contrast, rainfall correction results are based on the cross-assimilation of synthetic surface soil moisture retrievals (generated by the SAC model) into an API model. This difference means that our rainfall correction strategy is tested using a more challenging fraternal twin synthetic experiment in which observations generated by one model are assimilated into a second model. However, this lack of consistency is actually beneficial to the analysis. By choosing an easier identical twin set-up for state correction, relatively to the more difficult fraternal twin experiment applied for rainfall correction, we minimize the probability that increased runoff skill associated with our rainfall correction scheme is actually an artifact of our particular experimental approach. In this way, our synthetic approach is designed to maximize the credibility of key manuscript results. Likewise, our decision to assume a relatively thick (10 cm) upper-zone depth for the SAC model

may reduce the relative benefit of our proposed approach relative to existing state-correction procedures (Fig. 11c).

Conversely, there are additional aspects of our particular approach which have the opposite effect and may artificially enhance the relative benefit of our new approach. Figure 11a and b appear to demonstrate a tendency for limitations in retrieval accuracy and frequency to disproportionately affect our dual correction case (relative to state-correction only cases). This tendency suggests that overly optimistic assumptions concerning the frequency and accuracy of remote sensing retrievals will aid rainfall correction more than state correction. In addition, the tuning of  $\lambda$  in (9) is based here on the assumption that high-quality MOPEX rain gauges are available for calibration purposes. If comparably accurate rain gauge data is not available in an operational setting it is possible to calibrate  $\lambda$  using only satellite-based rainfall data. However, such alternative calibration is associated with a slight reduction in the performance of the rainfall correction procedure (Crow et al., 2009).

Another key consideration is the spatial and temporal scales at which our rainfall correction procedure is effective. At best, it can correct rainfall at time/space scales consistent with the ground resolution (typically 10–40 km) and revisit times (1 to 3 days) of satellite-based soil moisture retrievals. Real data results using the AMSR-E sensor indicate difficulties in correcting rainfall accumulations at time scales finer than about 2 days (Crow et al., 2009). Obviously, restricting correction to such coarse scales will limit the effectiveness of our approach when applied to hydrologic prediction applications – such as flash flood forecasting – requiring rainfall accumulation information at much finer space-time scales. Consequently, the highest potential for an operational application will likely be the prediction and monitoring of large-scale flooding events associated with prolonged periods (days to weeks) of excessive rainfall and flooding over large geographic regions ( $>100^2 \text{ km}^2$ ). The relatively large spatial scales associated with such events make real-time runoff monitoring a critical component of forecasting downstream flood peak timing stage height.

A final concern is the degree to which the adaptation of a reanalysis smoothing (rather than a sequential filtering) formulation will degrade the real-time functioning of a hydrologic forecasting/prediction system. The adoption of a smoothing framework will necessarily increase the latency of SAC model runoff predictions since it requires the acquisition of a soil moisture observation following a given storm period prior to the calculation of soil moisture and runoff for the same period. However, such delays may be small since, even in the absence of any soil moisture data assimilation, an operational system stills needs to wait until the acquisition of rainfall accumulation observations (presumably from some real-time rainfall observing system) to forecast stream flow. Consequently, the added delay required to obtain and process a subsequent soil moisture observation may not add substantial prediction latency to the system.

## 9 Summary

To date, efforts to improve hydrologic model stream flow predictions have focused on the sequential assimilation of remotely-sensed surface soil moisture to constrain pre-storm antecedent soil moisture conditions (see e.g. Crow et al., 2005). However, such approaches have not generally been successful at demonstrating clear value for remotely-sensed soil moisture retrievals in hydrologic applications. Here we propose an alternative reanalysis system (in Case 5 in Fig. 4) that reformulates the runoff prediction problem into a smoothing framework which simultaneously corrects both hydrologic model internal soil moisture states and external rainfall input feed into the model. Preliminary testing of the approach using a synthetic twin methodology suggests that, for a wide range of climatic conditions (Fig. 1), the approach can enhance the value of remotely-sensed soil moisture retrievals for runoff and stream flow prediction applications (Figs. 6 and 8) – particularly for high flow events in which direct, surface runoff processes play a dominant role in generating stream flow (Fig. 9). Since the advantages of our dual approach emerge only at relatively high levels of rainfall error (Fig. 10), its primary utility will likely be for large-scale flood forecasting in areas of the world lacking sufficient ground-based resources for real-time rainfall monitoring.

All preliminary work presented here is based on synthetic twin data assimilation experiments and must be confirmed by follow-on work aiming at verifying the approach with real data. Nevertheless, it is worth noting that our overall synthetic methodology is fundamentally conservative in that state correction is attempted using a less challenging identical twin set-up relative to the more challenging fraternal twin structure of the rainfall correction approach (see Sect. 7). Combined with completed validation studies (Crow et al., 2009), this suggests that expressions of added skill associated with our rainfall correction approach (above and beyond that achieved using only EnKF or EnKS state correction techniques) are likely credible representations of results obtainable from real data.

*Acknowledgements.* This work was partially supported by the NASA EOS Aqua Science and the NASA Terrestrial Hydrology Programs through grant NNH04AC301.

Edited by: N. Verhoest

## References

- Aubert, D., Loumagne, C., and Oudin, L.: Sequential assimilation of soil moisture and streamflow data in a conceptual rainfall-runoff model, *J. Hydrol.*, 280, 145–161, 2003.
- Burnash, R. J. C., Ferral, R. L., and McGuire, R. A.: A generalized streamflow simulation system: Conceptual modeling for digital computers, US National Weather Service and California Department of Water Resources, Joint Federal-State River Forecast Center, Sacramento, California, USA, 204 pp., 1973.
- Crow, W. T., Bindlish, R., and Jackson, T. J.: The added value of spaceborne passive microwave soil moisture retrievals for forecasting rainfall-runoff ratio partitioning, *Geophys. Res. Lett.*, 32, L18401, doi:10.1029/2005GL023543, 2005.
- Crow, W. T. and Van Loon, E.: The impact of incorrect model error assumptions on the sequential assimilation of remotely sensed surface soil moisture, *J. Hydrometeorol.*, 8(3), 421–431, 2006.
- Crow, W. T. and Bolten, J. D.: Estimating precipitation errors using spaceborne surface soil moisture retrievals, *Geophys. Res. Lett.*, 34, L08403, doi:10.1029/2007GL029450, 2007.
- Crow, W. T., Huffman, G. H., Bindlish, R., and Jackson, T. J.: Improving satellite-based rainfall accumulation estimates using spaceborne surface soil moisture retrievals, *J. Hydrometeorol.*, doi:10.1175/2008JHM986.1, in press, 2009.
- Dunne, S. and Entekhabi, D.: An ensemble-based reanalysis approach to land data assimilation, *Wat. Resour. Res.*, 41, W02013, doi:10.1029/2004WR003449, 2005.
- Entekhabi, D., Njoku, E. G., and Houser, P., et al.: The Hydrosphere State (HYDROS) mission concept: An earth system pathfinder for global mapping of soil moisture and land freeze/thaw, *IEEE Trans. Geosci. Rem. Sens.*, 42(10), 218–2195, 2004.
- Francois, C., Quesney, A., and Otle, C.: Sequential assimilation of ERS-1 SAR data into a coupled land surface-hydrological model using an extended Kalman filter, *J. Hydrometeorol.*, 4(2), 473–487, 2003.
- Georgakakos, K. P.: Analytical results for operational flash flood guidance, *J. Hydrol.*, 317(1–2), 81–103, 2005.
- Goodrich, D. C., Schmugge, T. J., Jackson, T. J., Unkrich, C. L., Keefer, T. O., Parry, R., Bach, L. B., and Amer, S. A.: Runoff simulation sensitivity to remotely-sensed initial soil water content, *Water Resour. Res.*, 30(5), 1393–1405, 1994.
- Jacobs, J. M., Meyers, D. A., and Whitfield, B. M.: Improved rainfall/runoff estimates using remotely-sensed soil moisture, *J. Am. Water Resour. Assoc.*, 39(2), 313–324, 2003.
- Kantamneni, V., Simon, D., and Schwartz, S.: Optimal filtering techniques for analytical streamflow forecasting, *Proceedings of the 18th International Conference on Systems Engineering*, Las Vegas, Nevada, USA, 130–135, 16–18 August 2005.
- Kerr, Y. H., Waldteufel, P., Wigneron, J.-P., Martinuzzi, J.-M., Font, J., and Berger, M.: Soil moisture retrieval from space: the soil moisture and ocean salinity mission (SMOS), *IEEE Trans. Geosci. Rem. Sens.*, 39(8), 1729–1735, 2001.
- Lakshmi, V.: Use of satellite remote sensing in hydrological predictions in ungauged basins, *Hydrol. Processes*, 18(5), 1029–1034, 2004.
- National Research Council: *Earth Science and Applications from Space: National Imperatives for the Next Decade and Beyond*, National Academy Press, Washington DC, USA, 400 pp., 2007.
- Oki, T., Nishimura, T., and Dirmeyer, P.: Assessment of annual runoff from land surface models using Total Runoff Integrating Pathways (TRIP), *J. Meteor. Soc. Jpn.*, 78, 235–255, 1999.
- Parajka, J., V. Naemi, G. Blöschl, W. Wagner, R. Merz, and K. Scipal, Assimilating scatterometer soil moisture data into conceptual hydrologic models at coarse scales, *Hydrology and Earth System Science*, 10, 353–368, 2006.
- Pauwels, R. N., Hoeben, R., Verhoest, N. E. C., De Troch, F. P., and Troch, P. A.: Improvements of TOPLATS-based discharge predictions through assimilation of ERS-based remotely-sensed soil moisture values, *Hydrol. Proc.*, 16, 995–1013, 2002.
- Reichle, R. H. and Koster, R. D.: Global assimilation of satellite surface soil moisture retrievals into the NASA Catchment land surface model, *Geophys. Res. Lett.*, 32, L02404, doi:10.1029/2004GL021700, 2005.
- Ryu, D., Crow, W. T., Zhan, X., and Jackson, T. J.: Correcting unintended perturbation biases in hydrologic data assimilation using Ensemble Kalman filter, *J. Hydrometeorol.*, in press, 2009.
- Schaake, J., Duan, Q., Koren, V., and Hall, A.: Toward improved parameter estimation of land surface hydrology models through the Model Parameter Estimation Experiment (MOPEX), *IAHS Publ.*, 270, 91–97, 2001.
- Weissling, B. P., Xie, H., Murray, K. E.: A multitemporal remote sensing approach to parsimonious streamflow modeling in a southcentral Texas watershed, USA, *Hydrol. Earth Syst. Sci.*, 4, 1–33, 2007, <http://www.hydrol-earth-syst-sci.net/4/1/2007/>.
- Wood, E. F., Sivapalan, M., and Beven, K.: Similarity and scale in catchment storm response, *Rev. Geophys.*, 28(1), 1–18, 1990.