Combined Assimilation of Streamflow and Satellite Soil Moisture with the Particle Filter and Geostatistical Modeling

Hongxiang Yan and Hamid Moradkhani

Remote Sensing and Water Resources Lab
Department of Civil and Environmental Engineering, Portland State University

Abstract

Assimilation of satellite soil moisture and streamflow data into a distributed hydrologic model has received increasing attention over the past few years. This study provides a detailed analysis of the joint and separate assimilation of streamflow and Advanced Scatterometer (ASCAT) surface soil moisture into a distributed Sacramento Soil Moisture Accounting (SAC-SMA) model, with the use of recently developed particle filter-Markov chain Monte Carlo (PF-MCMC) method. Performance is assessed over the Salt River Watershed in Arizona, which is one of the watersheds without anthropogenic effects in Model Parameter Estimation Experiment (MOPEX). A total of five data assimilation (DA) scenarios are designed and the effects of the locations of streamflow gauges and the ASCAT soil moisture on the predictions of soil moisture and streamflow are assessed. In addition, a geostatistical model is introduced to overcome the significantly biased satellite soil moisture and also discontinuity issue. The results indicate that: (1) solely assimilating outlet streamflow can lead to biased soil moisture estimation; (2) when the study area can only be partially covered by the satellite data, the geostatistical approach can estimate the soil moisture for those uncovered grid cells; (3) joint assimilation of streamflow and soil moisture from geostatistical modeling can further improve the surface soil moisture prediction. This study recommends that the geostatistical model is a helpful tool to aid the remote sensing technique and the hydrologic DA study.

Keywords: Data assimilation; Particle filter; Markov chain Monte Carlo; Soil moisture; Geostatistical modeling; ASCAT

1 Introduction

Soil moisture is a key variable of the earth system, with consequent impacts on the water, energy, and biogeochemical cycles [1]. An accurate representation of soil moisture is crucial for climate prediction, such as floods and droughts, and for better guidance in agriculture and water resources planning [2,3]. Soil moisture have also been analyzed to understand the changing water, energy, and carbon fluxes in the context of climate and land use change [4–9]. However, these soil moisture datasets suffer from poor resolution, spatial heterogeneity, and bias issues [10–12]. One possible approach to reduce soil moisture predictive uncertainty is to integrate information from multiple sources (models, *in-situ*, and remotely sensed) through data assimilation (DA). DA has been emphasized as one of the key elements to improve hydrologic prediction in last decade [13–20].

Applications of DA in the hydrologic community can be classified according to the study scale (single-scale or multi-scale) and the assimilated data (univariate or multivariate) [4,21]. Most DA studies belong to the univariate and single-scale scenario [22–28], where the univariate and single-scale indicate the assimilation of a single data type (e.g. streamflow) and the observation data scale coincides with the model scale (e.g. outlet of a watershed), respectively. There are also DA applications that assimilate univariate but multi-scale datasets [29–31]. Multi-scale means that the spatial resolution of the assimilated observations is different from the model output resolution, and upscaling/downscaling techniques are usually required [10,24]. Currently, there is an increasing trend in the multivariate scenario (i.e., assimilating more than one data type). The multiple data can be from different satellite sensors [11,17,32,33], a combination of satellite and *in-situ*/data-driven data [34,35], and different *in-situ* data [36,37].

With the development of remote sensing techniques, unprecedented spatial and temporal resolution of soil moisture data are available across a range of scales [38–41]. As a result, assimilation of remotely sensed soil moisture into hydrologic models has been receiving increasing attention [24,42–44]. These studies either focus on the soil moisture prediction [43,45–47] or streamflow prediction [48–50]. For the latter case, there is still no consensus in the community about the improvement in streamflow forecasting skill from the assimilation of satellite soil moisture [48,51–55]. Currently, assimilation of satellite soil moisture is still an active research area, as some key aspects of the assimilation framework have not been fully understood to date [49,50,56]. These aspects include: (1) characterizing the errors in model simulations; (2) the observation data error; (3) observation data discontinuity; (4) scale issues; (5) the optimal rescaling technique; and (6) the most suitable DA method [24,45,57–59]. In the following two paragraphs, we discuss two aspects of these challenges and propose possible solutions.

DA Algorithm. Several DA algorithms have been used in soil moisture assimilation, such as the extended Kalman filter (EKF), the ensemble Kalman filter (EnKF), the variational (VAR) algorithm, and the particle filter (PF) [60]. For instances, Aubert et al. [61] used the EKF to assimilate the *in-situ* soil moisture and streamflow into the lumped GR4J model; Reichle et al. [62] assimilated the microwave soil moisture into a hydrologic model using the four-dimensional (4-D) VAR technique. But the DA approach used in the majority of satellite soil moisture assimilation studies is state updating of the EnKF [13,33,53]. However, two limitations exist in these studies: (1) The Gaussian error assumption within EnKF is unsuitable for hydrological cases and the final performance is often suboptimal [27,63,64]; (2) Under the climate and land use change, the stationary parameter assumption is challenged [9,19,20]. Alternatively, the PF algorithm with parameter updating technique is suggested as a more robust DA technique for hydrological studies in a changing world [14,18,22,24,27]. Compared with EnKF, the PF can relax the Gaussian error assumption, maintain water balance, and provide a more complete representation of state/parameter posterior distribution [23,46,47,65]. The PF technique can be further improved by combining with Markov chain Monte Carlo (MCMC). The PF-MCMC was first proposed in statistical literature by Andrieu et al. [66]. Moradkhani et al. [27] re-designed the PF-MCMC and introduced it to the hydrologic community by integrating the variable variance multiplier [63] for appropriate perturbation of observation and also including the parameter updating to the whole DA scheme. Vrugt et al. [26] also used the PF-MCMC for state-parameter updating using a parameter optimization and assimilation approach.

Data Discontinuity. Due to the temporal and spatial limitations of many satellite instruments, it is common that not all the watershed grid cells can be measured at the same time. For instance, the overpass of Soil Moisture and Ocean Salinity (SMOS) is at minimum every three days [39]. In addition, the C/X bands have higher attenuation in the presence of vegetation, and these measurements are significantly biased for dense vegetated areas [41]. Han et al. [67] and Yin et al. [68] found that the quality of satellite soil moisture data impacted their assimilation and that assimilation with biased soil properties can worsen surface fluxes characterization. In order to overcome sensor limitations and improve the accuracy of soil moisture estimates at uncovered/biased grid cells, a geostatistical method–general Gaussian approach [69] is used in this study. The advantages of this method are: (1) the general Gaussian approach can predict the soil moisture data at the uncover/biased grid cells instead of relying on the localization concept [45]; (2) the general Gaussian approach is a more robust model than the traditional geostatistical variogram model [69].

Based on the above discussion, there are two objectives of this study: (1) investigate the effects of PF-MCMC on the assimilation of satellite soil moisture for soil moisture and streamflow prediction; (2) introduce the geostatistical model to overcome the data discontinuity issue, and study the effects of assimilation of soil moisture estimated from geostatistical model. Here we used the synthetic Advanced Scatterometer (ASCAT) soil moisture [40] and a fully distributed hydrological model–Sacramento Soil Moisture Accounting (SAC-SMA) with Muskingum-Cunge routing method. We aim to answer the following five science questions:

(1) What are the effects of jointly/separately assimilating streamflow and soil moisture data on streamflow prediction by PF-MCMC?

(2) What are the effects of assimilating streamflow from internal/outlet gauges on the streamflow prediction at a watershed scale?

(3) Given the limitation of remote sensing instruments (for instance, the vulnerability to radio frequency interference and uncovered regions), can the general Gaussian approach accurately fill in the soil moisture for these regions?

(4) What are the effects of assimilating soil moisture derived from the geostatistical model on soil moisture and streamflow prediction?

This paper is organized as follows: Section 2 describes the theory of PF-MCMC, the hydrological model, and the geostatistical model. Section 3 illustrates our experiment design, including the study area, data error, synthetic data, DA scenarios, and performance verifications. Finally, the results and discussions are provided in Section 4.

2 Methodology

2.1 Sequential Bayesian Theory

Following Moradkhani [4], the state-space model that describes the generic non-linear earth system are as follows:

$$y_t = h(x_t) + v_t \tag{1}$$

$$x_t = f(x_{t-1}, u_t, \theta) + w_t \tag{2}$$

where $x_t \in \mathbb{R}^n$ is a vector of the uncertain state variables at current time step, $y_t \in \mathbb{R}^m$ is a vector of observation data, u_t is the uncertain forcing data, $\theta \in \mathbb{R}^d$ is the model parameters, $h(\cdot)$ is the non-linear function relates the states x_t to the observations y_t, w_t represents the model error, and v_t indicates the observation error. The errors w_t and v_t are assumed to be white noise with mean zero and covariance Q_t and R_t, respectively.

Following Moradkhani et al. [14], the posterior distribution of the state variables x_t given a realization of the observations $y_{1:t}$ is written as follows:

$$\begin{aligned} p(x_t|y_{1:t}) = p(x_t|y_{1:t-1}, y_t\,) &= \frac{p(y_t|x_t)p(x_t|y_{1:t-1})}{p(y_t|y_{1:t-1})} \\ &= \frac{p(y_t|x_t)p(x_t|y_{1:t-1})}{\int p(y_t|x_t)p(x_t|y_{1:t-1})dx_t} \end{aligned} \tag{3}$$

$$p(x_t|y_{1:t-1}) = \int p(x_t, x_{t-1}|y_{1:t-1})dx_{t-1} = \int p(x_t|x_{t-1})p(x_{t-1}|y_{1:t-1})dx_{t-1} \tag{4}$$

where $p(y_t|x_t)$ is the likelihood, $p(x_t|y_{1:t-1})$ is the prior distribution, and $p(y_t|y_{1:t-1})$ is the normalization factor.

In practice, the Equation (3) does not have an analytic solution except for few special cases (e.g., the linear system with Gaussian assumption). Instead, the posterior distribution $p(x_t|y_{1:t})$ is approximated using a set of MC random samples.

2.2 Particle Filter-Markov Chain Monte Carlo (PF-MCMC)

The PF-MCMC [27] is an extension of PF-SIR [14]. The application of the MCMC to PF leads to a more complete characterization of the parameter posteriors and reducing risk of sample impoverishment. The PF-MCMC consists of two steps: (1) generating the random replicates of model states forecasts and parameters with equal weights; and (2) updating forecasted states, parameters, and weights when new observations become available. This leads to the posterior density $p(x_t|y_{1:t})$, which is approximated as:

$$p(x_t|y_{1:t}) \approx \sum_{i=1}^{N} w_t^{i+} \delta\left(x_t - x_t^i\right) \quad (5)$$

where w_t^{i+} is the posterior weight of the ith particle, δ is the Dirac delta function, and N is the ensemble size. Following Moradkhani et al. [27], the normalized weights are calculated as follows:

$$w_t^{i+} = w_t^{i-} \frac{p\left(y_t|x_t^i, \theta_t^i\right)}{\sum_{i=1}^{N} w_t^{i-} p\left(y_t|x_t^i, \theta_t^i\right)} \quad (6)$$

where w_t^{i-} is the prior particle weights, and $p\left(y_t|x_t^i, \theta_t^i\right)$ can be computed from the likelihood $L\left(y_t|x_t^i, \theta_t^i\right)$. Generally, a Gaussian distribution is used to estimate $L\left(y_t|x_t^i, \theta_t^i\right)$:

$$L\left(y_t|x_t^i, \theta_t^i\right) = \frac{1}{\sqrt{(2\pi)^m |R_t|}} exp\left[-\frac{1}{2}\left(v_t^i\right)^T {R_t}^{-1}\left(v_t^i\right)\right] \quad (7)$$

where $v_t^i = y_t - h\left(x_t^i\right)$ is the residual.

To obtain approximate samples from $p(x_t|y_{1:t})$, a resampling operation is required. The sampling importance resampling (SIR) algorithm [14] is suggested to resample the particles with a probability greater than the uniform probability. After resampling, all the particle weights are set equal to $1/N$. To avoid the sample impoverishment, a perturbation of the resampled parameters is recommended. Then, a proposal distribution is formed to generate proposed parameters $\theta_t^{i,p}$:

$$\theta_t^{i,p} = \theta_t^{i+} + \epsilon_t^i, \qquad \epsilon_t^i \sim N\left[0, s_t Var\left(\theta_t^{i-}\right)\right] \tag{8}$$

where θ_t^{i+} is the parameters after SIR, $Var\left(\theta_t^{i-}\right)$ is the variance of the prior parameters at the current time step, and s_t is a small tuning time-variant parameter. To reject the parameter samples $\theta_t^{i,p}$ that move outside the filtering posterior distribution, a metropolis acceptance ratio α is used to determine whether to accept the proposed parameters:

$$\alpha = min\left[1, \frac{p\left(x_t^{i,p}, \theta_t^{i,p} | y_{1:t}\right)}{p\left(x_t^{i+}, \theta_t^{i+} | y_{1:t}\right)}\right] \tag{9}$$

where $p\left(x_t^{i,p}, \theta_t^{i,p} | y_{1:t}\right)$ is the proposed joint probability distribution:

$$p\left(x_t^{i,p}, \theta_t^{i,p} | y_{1:t}\right) \propto p\left(y_t | x_t^{i,p}, \theta_t^{i,p}\right) \cdot p\left(x_t^{i,p} | \theta_t^{i,p}, y_{1:t-1}\right) \cdot p\left(\theta_t^{i,p} | y_{1:t-1}\right) \tag{10}$$

$$x_t^{i,p} = f\left(x_{t-1}^{i+}, u_t^{i+}, \theta_t^{i,p}\right) \tag{11}$$

where $x_t^{i,p}$ is a sample from the proposal state distribution and u_t^{i+} is the resampled forcing data. The advantage of the MCMC move is that it does not adjust the state variables therefore still retains the water balance.

Moradkhani et al. [27] modified the variable variance multiplier (VVM) method [63] which automatically finds the most fitting tuning factor s_t in Equation (9). The tuning factor is self-adaptive based on the spread of the previous states. The procedure of VVM is as follows:

$$\widehat{\omega}_t = |E(y_t') - y_t| \tag{12}$$

$$ub_t = \begin{cases} y_t'^{75} - E(y_t'), & E(y_t') < y_t \\ E(y_t') - y_t'^{25}, & E(y_t') > y_t \end{cases} \tag{13}$$

$$er_t = \tau\left[median\left(\frac{\widehat{\omega}_{(t-lag):t}}{ub_{(t-lag):t}}\right) - 1\right] + 1 \tag{14}$$

$$s_t = er_t \cdot E\left[s_{(t-lag):t}\right] \quad (15)$$

where $E(y_t')$ is the forecast expected value, y_t is the observation, $y_t'^{25}$ and $y_t'^{75}$ are the 25th and 75th forecast quantiles, respectively; τ is the smoothing value and is set to 0.5; the lag time is set to 100 as suggested by Moradkhani et al. [27].

2.3 General Gaussian Model

The general Gaussian model is used in this study to predict soil moisture for uncovered/biased grid cells. We used the physical covariates which directly ties to soil moisture in the model. As opposed to traditional geostatistical modeling (such as kriging), which considers the covariates as fixed, the general Gaussian model treats the covariates as random variables [69]:

$$Y_t(n) = \mu_t + \eta[\chi(n)] \quad (16)$$

where $Y_t(n)\ (m \times 1)$ is the vector of observations at a finite number of locations $n = (n_1, n_2, \dots, n_m)$ in the study region; μ_t is a fixed mean parameters for m locations at time step t; $\chi(n) = \left[\chi_1(n), \chi_2(n), \dots, \chi_p(n)\right]$ is a vector of p covariates associated with the locations n**;** and η is the zero-mean stationary Gaussian process, which is characterized by a covariance matrix:

$$Cov\{\eta[\chi(n_j)], \eta[\chi(n_k)]\} = C_t[\chi(n_j), \chi(n_k)] \quad (17)$$

where $k = 1,2,\dots,m$, $j = 1,2,\dots,m$, and $C_t[\cdot]$ is an isotropic exponential covariance function, where

$$C_t[\chi(n_j), \chi(n_k)] = \alpha_t^2 + \beta_t^2 exp\left[-\|n_j - n_k\|/\rho_t\right] \quad (18)$$

where α_t^2 is the nugget, β_t^2 is the partial sill, and ρ_t is the range parameters.

The goal of the general Gaussian model is to predict the soil moisture value $Y_t(n_{m+1})$ for uncovered grid cells n_{m+1}. After fitting the Equation (18), the ordinary kriging method is used here to estimate the uncovered grid cell $Y_t(n_{m+1})$ [70,71]. In this study, three covariates directly tied to soil moisture are considered: elevation, slope, and aspect. More details are provided in Section 4.5. The general assumption behind the geostatistical model is that model is a second-order stationary and isotropic, where for sites n_j and n_k, the isotropic exponential covariance $C_t[\chi(n_j), \chi(n_k)]$ depends only on the distance $\|n_j - n_k\|$.

2.4 Sacramento Soil Moisture Accounting Model

The SAC-SMA model is a nonlinear conceptual rainfall-runoff model with spatially lumped parameters. The model was first developed by Burnash et al. [72] and it is used operationally by the National Weather Service River Forecast Centers for streamflow forecasting. The model includes two soil moisture zones, an upper and a lower zone. The upper zone is responsible for surface runoff and interflow, while the lower zone controls baseflow. Short-term storage of water in the soil is accounted in the upper zone, while the long-term storage of groundwater is in the lower zone. When the upper zone water storage is satisfied, the upper zone water can move vertically into the lower zone and horizontally as interflow. Excess runoff is routed to the watershed outlet using a Nash cascade of three linear reservoirs. A total of six interdependent soil water states are estimated in SAC-SMA model: upper zone tension water content (UZTWC), upper zone free water content (UZFWC), lower zone tension water content (LZTWC), lower zone free primary water content (LZFPC), lower zone free secondary water content (LZFSC), and basin saturated fraction (ADIMP). Precipitation and potential evapotranspiration (PET) are the required forcing data for SAC-SMA model.

The 17 SAC-SMA model parameters are summarized in Table 1. The synthetic truth parameters used to generate the synthetic streamflow and soil moisture are also shown in Table 1. Figure 1 presents the flowchart for the combined DA and geostatistical modeling.

Please place Table 1 here

Please place Figure 1 here

----

3 Experiment Design

In order to assimilate the ASCAT soil moisture data, the SAC-SMA is implemented in a distributed manner, where the runoff in each grid cell is routed along the stream segments to the watershed outlet using the Muskingum-Cunge routing method [73]. Synthetic observations of streamflow and soil moisture are generated using the SAC-SMA model with a predefined parameter value (Table 1). The daily synthetic soil moisture (degree of saturation) is used to represent the ASCAT soil moisture product. The benefit of a synthetic study is to allow a direct comparison between model simulations and the "truth" such that the systematic biases between remotely sensed and model based soil moisture estimates can be avoided; otherwise rescaling method would be necessary [24,74].

3.1 Study Area

The study area is a sub-watershed of Salt River basin (HUC 150601), located in the west of Arizona, east of the city of Phoenix (Figure 2). The target watershed crosses four counties (Gila, Navajo, Apache, and Graham) of Arizona with an area of 7,379 km^2. The soil texture within the watershed is clay and the major land use is forest (75%) and shrub (22%). A total of five USGS streamflow gauges are operated within the watershed. Based on these gauges, five sub-watersheds are delineated (Figure 2). The studied watershed is one of the selected watersheds in the Model Parameter Estimation Experiment (MOPEX) [75], therefore the effects of water management can be ignored.

\---

Please place Figure 2 here

\---

3.2 Synthetic ASCAT Soil Moisture

The ASCAT sensor is a C-band (5.255 GHz) active microwave scatterometer on board of a series of three polar orbiting Meteorological Operational (METOP) satellites. The first satellite (METOP-A) was launched in October 2006, the second (METOP-B) in September 2012, and the third (METOP-C) is scheduled for launch in 2018 [44]. Two nominal spatial resolutions of the ASCAT backscatter measurements are available: 50 km and 25 km. Complying with the Nyquist–Shannon sampling theorem, the grid spacing of the 50 km product is 25 km, and the 25 km product is 12.5 km [40]. Unlike other remotely sensed products, ASCAT soil moisture is expressed in terms of degree of saturation (range between 0 and 1), not the volumetric soil moisture (m^3/m^3) itself.

To represent daily ASCAT 12.5×12.5 km^2 surface soil moisture product, the study watershed was distributed to 12.5×12.5 km^2 grid cells and a total of 46 grid cells were delineated

(Figure 2). To generate the synthetic ASCAT surface soil moisture product, it is assumed that the SAC-SMA model upper zone soil moisture states (UZTWC and UZFWC) are directly observed by remotely sensed surface soil moisture retrievals [55]. Hence, the synthetic ASCAT soil moisture in terms of degree of saturation S can be estimated as:

$$S = \frac{UZTWC + UZFWC}{UZTWM + UZFWM} \tag{19}$$

where UZTWM is the upper zone tension-water maximum and UZFWM is the upper zone free-water maximum (Table 1).

3.3 Forcing Data

Daily precipitation data from Oct. 1 2005–Sep. 30 2007 were acquired from Oregon State University Parameter-elevation Regressions on Independent Slopes Model (PRISM) 4-km^2 grid cell data (http://www.prism.oregonstate.edu/). For each 12.5×12.5 km^2 grid cell in the study area, the mean area precipitation data were upscaled from the finer PRISM grid cell data. The PET data in the same period were obtained from the Moderate-resolution Imaging Spectroradiometer (MODIS) global terrestrial evapotranspiration dataset [76]. The PET product MOD16 is a 1-km^2 resolution data at 8-day intervals. After disaggregating MOD16 to daily data, the 1-km^2 daily PET data were upsclaed to obtain the 12.5×12.5 km^2 mean area PET data.

The spatial pattern of the mean daily precipitation and PET calculated from the whole time period are presented in Figure 3. It is noted that elevation is the main determinant of precipitation patterns, where precipitation is significantly enhanced (more than double) on higher-elevation grid cells compared to lowlands. For PET data, much of the variation is likely caused by local topographic effects and land cover. Figure 3 shows how PET slightly decreases with increasing elevation and transformation from shrub to forest.

Please place Figure 3 here

3.4 Data Assimilation Scenarios

Our preliminary results show that by solely assimilating the synthetic ASCAT surface soil moisture, the PF-MCMC (with parameter updating) can lead to a significant difference in the parameter estimation of the model. As a result, the skill of streamflow prediction is significantly degraded. The univariate assimilation of surface soil moisture is not sufficient to constrain the hydrologic model parameters. Similar result was also reported by Plaza et al. [46]. In addition, Lee et al. [36] and Wanders et al. [35] suggested that the benefits of satellite soil moisture are largest when they are assimilated simultaneously with streamflow observations.

The two main goals of this study are to investigate the performance of PF-MCMC on soil moisture and streamflow predictions, and to introduce the geostatistical model to overcome the satellite data discontinuity issue and assimilation of the soil moisture estimated from the geostatistical model (Figure 1). Several other factors affect the investigation, such as the gauge location and the satellite soil moisture. Therefore, a total of five scenarios were designed to explore difference approaches in DA performance. The details of these five scenarios are summarized in Table 2.

The effects of assimilating outlet or internal gauges are compared in scenarios 1-3. The effects of soil moisture are investigated by jointly assimilating with streamflow in scenario 4.

Finally, scenario 5 jointly assimilates the streamflow and soil moisture data estimated from the geostatistical model.

Please place Table 2 here

3.5 Performance Metrics

Three metrics were used to assess the performance of the DA: the Nash-Sutcliffe efficiency (NSE), the root mean square error (RMSE), and the 95% exceedance ratio (ER95).

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{y}_t)^2} \tag{25}$$

$$NSE = 1 - \frac{\sum_{t=1}^{T}(y_t - \hat{y}_t)^2}{\sum_{t=1}^{T}(y_t - \bar{y})^2} \tag{26}$$

$$ER95 = \frac{1}{T}\sum_{t=1}^{T}\left(\hat{y}_{97.5\%,t} < y_t \; or \; \hat{y}_{2.5\%,t} > y_t\right) \times 100\% \tag{27}$$

where $\hat{y}_t$ is the ensemble mean prediction at time step t, $\bar{y}$ is the observation mean over the time steps T, $\hat{y}_{97.5\%,t}$ and $\hat{y}_{2.5\%,t}$ indicate the 97.5% and 2.5% percentiles for the ensemble predictions.

Both NSE and RMSE are the measures of the accuracy of expected value; whereas the ER95 is a probabilistic verification method where the observations fall outside the ensemble range just 5% of the time [77]. ER95 greater than five suggest the distribution is too narrow (over-confident) and ER95 less than five suggests the distribution is too wide (under-confident)

[27]. The assumption behind ER95 is that the ensemble predictions and observations are independent and identically distributed (i.i.d.) and the probability integral transform (PIT).

For streamflow DA verification, the NSE and ER95 metrics are used. The NSE can be considered as a normalized RMSE and compared directly in different gauges. The i.i.d. assumption behind ER95 is more likely to be met by streamflow data [64,78–80]. But for soil moisture, they are persistent on seasonal-to-interannual time scales and the independent assumption of ER95 is seriously challenged [24]. Therefore, for soil moisture DA verification, only the RMSE is used. In addition, the RMSE is a standard metric used in SMOS and Soil Moisture Active Passive (SMAP) satellite missions [81].

4 Results and Discussions

For all the five DA scenarios, following DeChant and Moradkhani [82] and Moradkhani et al. [27], the precipitation was perturbed with a lognormal distribution with a coefficient of variation of 0.25, the PET and streamflow were assumed to follow normal distribution with a coefficient of variation of 0.25 and 0.15, respectively. The white noise (standard deviation) for synthetic ASCAT soil moisture is set to 0.04 according to Wagner et al. [44].

The prior distributions of SAC-SMA parameters were uniformly distributed according to their default ranges presented in Table 1. The synthetic truth was also shown in Table 1. The initial parameters were sampled using the Latin hypercube sampling (LHS) method. The LHS is used due to its strength to properly sample the parameters by dividing the parameter space into regions of equal probability [14]. Parameter values are assumed to be uncorrelated in space.

The streamflow for the five gauges were generated using the same predefined SAC-SMA parameters shown in Table 1. For all scenarios, the SAC-SMA is applied in a distributed fashion. With the predefined model parameters, the synthetic soil moisture observations were generated for each grid cell based on Equation (19) and the runoff was routed to the streamflow segments. The streamflow of the five segments which are closest to the locations of the five streamflow gauges were used to represent the synthetic streamflow for the five gauges.

4.1 The Effects of Streamflow Gauge Location on PF-MCMC

In the distributed SAC-SMA model schematic, runoff generated in each grid cell was routed along the segments to the watershed outlet. Since the streamflow is structured by the hydrologic network, it is possible that by only assimilating the outlet streamflow, the prediction of internal gauge streamflow would be improved correspondingly. Similarly, by only assimilating the internal gauge streamflow, the outlet streamflow prediction is expected to be improved. The former schematic is generally described as "inverse routing" [83], and the latter schematic is related to "forward routing". These effects were examined in scenarios 1-2. Scenario 1 only assimilated the outlet streamflow; while scenario 2 jointly assimilated the streamflow of the four internal gauges. Table 3 summarized the PF-MCMC performance for these two scenarios.

Please place Table 3 here

For scenario 1, the outlet show an NSE value of 0.99, indicating a high accuracy in prediction. Meanwhile, the four internal gauges exhibited the same NSE value (0.99), except for gauge 3 (0.97). Gauge 3 had a less NSE value than the other gauges because of the high elevation. The contributing grid cells of gauge 3 had the highest elevations than other cells and received the largest precipitation in the watershed (Figure 3). It is therefore shown a high variance of perturbed precipitation and rainfall-runoff process. Overall, the average NSE value of the total five gauges was 0.99. This result indicated that by only assimilating the unbiased outlet streamflow based on PF-MCMC, the streamflow in each segment of the watershed can be correctly tracked through inverse routing.

To further examine the performance of this inverse routing method, the mean runoff data for each grid cell before routing was compared with the synthetic “true” runoff value. Figure 4 presents the mean daily synthetic, inverse routing runoff, and their mean RMSE values (mm/d) for each grid cell on all time steps. It is noted that the inverted runoff field (estimated for each grid cell and obtained by assimilating the outlet streamflow) shows a similar spatial pattern to the synthetic runoff. The largest daily mean RMSE value existed in the highest elevation cell with the value of only about 0.1 mm/d. This result indicated that by only assimilating the unbiased outlet streamflow data with PF-MCMC method, the fine scale runoff field inside the total watershed can be successfully inferred.

Please place Figure 4 here

For scenario 2, four internal gauge streamflow were jointly assimilated. The NSE values for all the five gauges were 0.99, which indicated a very well prediction in the whole watershed. This result suggested that by only assimilating the unbiased internal gauge streamflow, the outlet streamflow can be correctly predicted. As seen in Table 3, the average predictive uncertainty range (in terms of ER95) decreased with the increasing number of assimilated gauges (from significantly under-confident to a little over-confident). Scenario 3 assimilated the streamflow from the five gauges and provided the best results with NSE=1 for outlet gauge. Figure 5 show the streamflow hydrographs for the five gauges based on scenario 3. It is noted that the mean simulated streamflow and synthetic observation hydrographs were overlapped with each other, because this is a synthetic study with large ensemble size 400.

Please place Figure 5 here

4.2 The Effects of ASCAT Soil Moisture on PF-MCMC

The ASCAT surface soil moisture and outlet streamflow were jointly assimilated, and the results were presented in scenario 4 of Table 4. Results of scenario 1 were also presented in Table 4 to compare the effects of assimilating surface soil moisture on streamflow prediction. Besides the joint assimilation scheme, surface soil moisture had also been solely assimilated into the distributed hydrologic model. However, we found that by only assimilating the unbiased ASCAT surface soil moisture, the streamflow prediction skill would significantly decrease (NSE < 0) due to incorrect parameter estimation in the lower zone of the SAC-SMA model. Therefore only the joint assimilation results were presented here.

Please place Table 4 here

The surface soil moisture RMSE values in Table 4 represented the watershed average during the simulation period. Compared with scenario 1, scenario 4 show a significant decrease of RMSE value by over 65% in predicting the surface soil moisture field. However, the average NSE value show a slight decrease of about 2%. This means that although assimilating the surface soil moisture resulted in no improvement in streamflow prediction skill with PF-MCMC, the surface soil moisture field can be predicted significantly better.

Currently, there is no consensus in the community about the improvement in streamflow prediction skill from the assimilation of satellite soil moisture, since many factors affect the DA performance such as the DA algorithm, the particular model structure, the choice of bias correction technique, the appropriate quantification of observation and model forecast errors, the spatial mismatching, and the watershed topography and climatology [24,34]. Massari et al. [56] described this issue as a "complex recipe". Since this is a synthetic study, the observation error and model structure error can be explicitly quantified. The rescaling method is also unnecessary. The key difference in this study is the application of PF-MCMC technique combined with geostatistical modeling, which is implemented in dual state-parameter updating scenario.

In literature, the study of assimilation of satellite soil moisture on streamflow prediction can be classified into synthetic study/observing system simulation experiment (OSSE) and real case study [33,49,53,55]. However, neither scenarios consider the parameter updating concurrently with the state updating. For synthetic study, typical procedures include: (1) an open-loop (OL) simulation of the hydrologic model with high-quality forcing inputs; (2) generating synthetic "true" satellite soil moisture from the open-loop simulation and incorporating realistic errors; (3) an OL simulation of the hydrologic model with the same pre-defined parameters with lower quality forcing inputs; (4) assimilating the synthetic "truth" in the OL configuration from step (3). Finally, the DA results will be compared with the OL in step (1) to evaluate the impact of assimilating satellite soil moisture [42]. For the real case study, first, the hydrologic model is calibrated with the streamflow data (OL simulation). Next, the satellite soil moisture is assimilated into the calibrated model. Last, the DA results are compared with the OL simulations to determine if additional gains can be achieved beyond the optimized model [33,53,84].

For our synthetic study, we did not have two input forcing but updated the state-parameter together using the PF-MCMC assimilation method. The degrading outlet streamflow predictions in the joint assimilation are compared with the outlet streamflow univariate assimilation only, i.e., no comparison with OL result is made. Unlike the state-augmentation technique used in the EnKF [13], the PF-MCMC resamples the particle weights based on model forecasts and observations. In the univariate outlet streamflow assimilation, particle weights are determined based on streamflow. The model forecasts receive the highest weight when they are closest to the observations. However, in the joint assimilation scenario, the particle weights depend on the joint distribution of soil moisture and streamflow. For instance, even if the streamflow forecast is closest to the observations, this particle may receive low weight if soil moisture forecast is far away from the observation. For the real soil moisture assimilation study with PF-MCMC, please refers to Yan et al. [24].

4.3 The Effects of Soil Moisture Inferred from Geostatistical Model on PF-MCMC

Figure 6 presented the synthetic ASCAT soil moisture for both dry and wet days. The spatial pattern show a trend of increased soil moisture from southwest to northeast of the watershed. Figure 7 show the three standardized covariates: elevation, slope, and the cosine of aspect, which are the effective covariates to predict soil moisture [69]. These three covariate values were calculated based on the USGS National Elevation Dataset (NED) 30-m DEM data (http://nationalmap.gov/elevation.html). The cosine value of the aspect represents the northern-facing amount. From Figures 6-7, it is noted that the higher elevation consists higher soil moisture. The north-facing slope is easier to dry out.

In order to better meet the isotropic assumption of the general Gaussian model, an orthogonal transformation is used for the three standardized covariates value, as suggested by Leung and Cooley [69]. The orthogonal standardized covariate $\chi'(n)$ are calculated as: $\chi'(n) = \Omega^{-1/2} \cdot \chi(n)$, where Ω is the sample covariance matrix of covariates $\chi(n)$ in Equation (18).

Please place Figures 6-7 here

Because soil moisture varies at each time step, the regression coefficients in the model need to be optimized for each time step. As a result, a total of 730 general Gaussian models were built in this study and the maximum likelihood estimator (MLE) method was used to optimize each coefficient. After fitting the general Gaussian model, the predictions for uncovered cells were estimated using the ordinary kriging method [70,71]. Here, parts of the grid cells were treated as pseudo uncovered cells and the remaining grid cells were used to train the model. Finally, the RMSE is used to analyze the prediction skill based on these pseudo cells.

The questions that remain are: (1) what proportion of the watershed area should be used to train the model, (2) if the satellite data only cover 30% of the study watershed, can one accurately predict the soil moisture for the other uncovered areas? No specific criterion has been suggested in literature to date. Considering that the goal of the satellite missions (e.g., SMOS and SMAP) is to reduce the RMSE of the remotely sensed soil moisture to less than 0.04 m^3/m^3 as compared with the *in-situ* soil moisture data, we use this threshold RMSE to answer the above question.

In this study, different proportions of the watershed grid cells were randomly selected to train the model. The proportion started from 80% and decreased to 10%. For each proportion, the training cells were randomly selected. To avoid the random selection error, the RMSE for each proportion set was based on an average of 30 iterations. Figure 8 presented the RMSE values for prediction cells (pseudo uncovered cells), differing on the number of training cells. When the training data area decreased from 80% to 10%, the RMSE increased from 0.025 to 0.045. If only 10% of the watershed area was used for training the model, the RMSE for the 90% uncovered area was larger than the threshold 0.04. However, when using more than 20% of the watershed area to build the model, the RMSE value for the uncovered areas was less than 0.036, which is smaller than the threshold value of 0.04. This finding indicates that if the satellite data only covers 20% of the watershed or only the confident soil moisture retrievals for the 20% of the watershed is available, the general Gaussian model can predict the soil moisture for the remaining watershed under the same accuracy criteria as the satellite retrievals (RMSE<0.04). In other words, the general Gaussian model is a helpful tool to aid the remote sensing technique.

Please place Figure 8 here

Since using only 20% of the watershed area in the training model can give an acceptable soil moisture prediction (compared with the remotely sensed soil moisture retrieval error), the effects of assimilating the predicted soil moisture is examined in scenario 5. Here, we randomly selected the 20% grid cells (9 cells) to train the model, and predict the soil moisture for the remaining 80% cells (37 cells). The RMSE values based on the average of the whole time steps were presented in top left panel of Figure 9. The nine zero RMSE value cells in top-left panel of Figure 9 indicated the location of the training cells. It is noted that the majority of grid cells had the RMSE less than 0.04, except for some cells located on the high elevation. This also can be explained by the large precipitation value on these cells (Figure 3), resulting in high variability of the soil moisture than other cells. We assimilated the soil moisture from geostatistical model and outlet streamflow into SAC-SMA, and the PF-MCMC performance was presented in Table 4.

Please place Figure 9 here

Compared with scenario 1 (only assimilating the outlet streamflow), the average NSE decreased from 0.99 to 0.91 (8% decrease). Especially for gauge 3, the NSE decreased from 0.97 to 0.72 (26% decrease). The significant decrease of NSE value for gauge 3 is due to the assimilation of soil moisture for the high elevation cells. These cells had high bias soil moisture estimations from geostatistical model. Compared with scenario 4, the average NSE value decreased from 0.97 to 0.91 (6% decrease). However, for the surface soil moisture field, compared with scenario 1, the RMSE decreased from 0.54 to 0.25 (54% decrease).

This result indicates that if we only calibrated the hydrologic model with outlet streamflow, other states might be significant biased. Using these biased states datasets might lead to unreliable assessment for other purposes, i.e., floods and droughts estimation [33,42,85–87]. If the satellite data cannot cover the whole study area, general Gaussian model can be used to retrieve the soil moisture for these uncovered cells. By assimilating the soil moisture from general Gaussian model, the surface soil moisture field can be predicted more accurately as compared with the scenario without soil moisture assimilation (Figure 9).

The time evolution of two upper soil zone parameters (UZTWM and UZFWM) is shown in Figure 10. Two scenarios (scenario 1 and 5) were examined. For scenario 1 (only assimilating outlet streamflow), the posterior distributions of the two parameters did not converge to the "true" values and the uncertainty did not decrease over time, although the mean value of

UZFWM was close to the “true” value. For scenario 5, even the assimilated soil moisture were estimated from geostatistical model, the convergence of both parameters can be clearly seen.

Please place Figure 10 here

----

5 Conclusion

The PF-MCMC method is used to assimilate the synthetic ASCAT soil moisture retrievals (12.5×12.5 km^2) and the synthetic streamflow from five gauges (four internal and one outlet gauges) into the fully distributed SAC-SMA model at the same spatial scale. A total of five data assimilation scenarios are designed in this study. These scenarios explore the effects of the locations of streamflow gauges and the ASCAT soil moisture on the predictions of soil moisture and streamflow. The results suggest that:

(1) by only assimilating the unbiased outlet streamflow, the PF-MCMC can successfully track the runoff filed (inverse-routing) and the internal segment streamflow;

(2) it is possible to improve outlet streamflow prediction by assimilating the streamflow of the four internal gauges;

(3) higher accuracy can be achieved along the streamflow segment network by assimilating more streamflow gauge data.

Another main goal of this study was to introduce a geostatistical model to combine with PF-MCMC to overcome the satellite discontinuity issues where satellite data does not cover the whole study region or is significantly biased, and the dominant land cover is dense vegetation. Our results further suggest that:

(4) when satellite data cannot cover the whole study area or the land surface is dominant with dense vegetation, the general Gaussian model can be used to complement the soil moisture;

(5) with only 20% of the watershed covered with the satellite footprint, the soil moisture in the remaining 80% of the uncovered areas can be estimated using general Gaussian model within the expected satellite data quality threshold (RMSE<0.04);

(6) by jointly assimilating the soil moisture inferred from the general Gaussian model and outlet streamflow, the RMSE of the surface soil moisture prediction is significantly reduced when compared with the assimilation of outlet streamflow only.

Overall, these findings can further aid the application of satellite soil moisture data for even drought monitoring and forecasting where the soil moisture deficit is the main variable that characterizes the agricultural drought.

Acknowledgement

Partial financial support for this project was provided by the National Oceanic and Atmospheric Administration (NOAA) Modeling, Analysis, Predictions, and Projections (MAPP) (Grant No. NA140AR4310234) and National Science Foundation (NSF) Cyber-Innovation for Sustainability Science and Engineering (CyberSEES) (Grant No. CCF-1539605). We thank the three anonymous reviewers whose comments helped improve and clarify this manuscript.

Reference

[1] Seneviratne SI, Corti T, Davin EL, Hirschi M, Jaeger EB, Lehner I, et al. Investigating soil moisture-climate interactions in a changing climate: A review. Earth-Science Rev 2010;99:125–61. doi:10.1016/j.earscirev.2010.02.004.

[2] Rosolem R, Hoar T, Arellano A, Anderson JL, Shuttleworth WJ, Zeng X, et al. Translating aboveground cosmic-ray neutron intensity to high-frequency soil moisture profiles at sub-kilometer scale. Hydrol Earth Syst Sci 2014;18:4363–79. doi:10.5194/hess-18-4363-2014.

[3] Chen W, Huang C, Shen H, Li X. Comparison of ensemble-based state and parameter estimation methods for soil moisture data assimilation. Adv Water Resour 2015. doi:10.1016/j.advwatres.2015.08.003.

[4] Moradkhani H. Hydrologic remote sensing and land surface data assimilation. Sensors 2008;8:2986–3004. doi:10.3390/s8052986.

[5] Shuttleworth J, Rosolem R, Zreda M, Franz T. The COsmic-ray Soil Moisture Interaction Code (COSMIC) for use in data assimilation. Hydrol Earth Syst Sci 2013;17:3205–17. doi:10.5194/hess-17-3205-2013.

[6] Reichle RH, De Lannoy GJM, Forman BA, Draper CS, Liu Q. Connecting Satellite Observations with Water Cycle Variables Through Land Data Assimilation: Examples Using the NASA GEOS-5 LDAS. Surv Geophys 2014;35:577–606. doi:10.1007/s10712-013-9220-8.

[7] Milly PCD, Betancourt J, Falkenmark M, Hirsch RM, Kundzewicz ZW, Lettenmaier DP, et al. Climate change. Stationarity is dead: whither water management? Science 2008;319:573–4. doi:10.1126/science.1151915.

[8] Najafi MR, Moradkhani H. Multi-model ensemble analysis of runoff extremes for climate change impact assessments. J Hydrol 2015;525:352–61. doi:10.1016/j.jhydrol.2015.03.045.

[9] Yan H, Edwards FG. Effects of land use change on hydrologic response at a watershed scale, Arkansas. J Hydrol Eng 2013;18:1779–85. doi:10.1061/(ASCE)HE.1943-5584.0000743.

[10] Reichle RH, Koster RD. Bias reduction in short records of satellite soil moisture. Geophys Res Lett 2004;31. doi:10.1029/2004GL020938.

[11] Draper CS, Reichle RH, De Lannoy GJM, Liu Q. Assimilation of passive and active microwave soil moisture retrievals. Geophys Res Lett 2012;39. doi:10.1029/2011GL050655.

[12] Yuan X, Ma Z, Pan M, Shi C. Microwave remote sensing of short-term droughts during crop growing seasons. Geophys Res Lett 2015;42:4394–401. doi:10.1002/2015GL064125.

[13] Reichle R, McLaughlin DB, Entekhabi D. Hydrologic data assimilation with the ensemble Kalman filter. Mon Weather Rev 2002;130:103–14. doi:10.1175/1520-0493(2002)130<0103:HDAWTE>2.0.CO;2.

[14] Moradkhani H, Hsu K-L, Gupta H, Sorooshian S. Uncertainty assessment of hydrologic

model states and parameters: Sequential data assimilation using the particle filter. Water Resour Res 2005;41:W05012. doi:10.1029/2004WR003604.

[15] Moradkhani H, Sorooshian S, Gupta H V., Houser PR. Dual state-parameter estimation of hydrological models using ensemble Kalman filter. Adv Water Resour 2005;28:135–47. doi:10.1016/j.advwatres.2004.09.002.

[16] Xia Y, Mitchell K, Ek M, Sheffield J, Cosgrove B, Wood E, et al. Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. J Geophys Res 2012;117. doi:10.1029/2011JD016048.

[17] De Lannoy GJM, Reichle RH, Arsenault KR, Houser PR, Kumar S, Verhoest NEC, et al. Multiscale assimilation of Advanced Microwave Scanning Radiometer-EOS snow water equivalent and Moderate Resolution Imaging Spectroradiometer snow cover fraction observations in northern Colorado. Water Resour Res 2012;48. doi:10.1029/2011WR010588.

[18] Liu Y, Weerts a. H, Clark M, Hendricks Franssen HJ, Kumar S, Moradkhani H, et al. Advancing data assimilation in operational hydrologic forecasting: Progresses, challenges, and emerging opportunities. Hydrol Earth Syst Sci 2012;16:3863–87. doi:10.5194/hess-16-3863-2012.

[19] Pathiraja S, Marshall L, Sharma A, Moradkhani H. Hydrologic modeling in dynamic catchments: A data assimilation approach. Water Resour Res 2016. doi:10.1002/2015WR017192.

[20] Pathiraja S, Marshall L, Sharma A, Moradkhani H. Detecting non-stationary hydrologic model parameters in a paired catchment system using data assimilation. Adv Water Resour 2016;94:103–19. doi:10.1016/j.advwatres.2016.04.021.

[21] Montzka C, Pauwels VRN, Franssen HJH, Han X, Vereecken H. Multivariate and multiscale data assimilation in terrestrial systems: A review. Sensors (Switzerland) 2012;12:16291–333. doi:10.3390/s121216291.

[22] Parrish MA, Moradkhani H, Dechant CM. Toward reduction of model uncertainty: Integration of Bayesian model averaging and data assimilation. Water Resour Res 2012;48. doi:10.1029/2011WR011116.

[23] Montzka C, Moradkhani H, Weihermüller L, Franssen H-JH, Canty M, Vereecken H. Hydraulic parameter estimation by remotely-sensed top soil moisture observations with the particle filter. J Hydrol 2011;399:410–21. doi:10.1016/j.jhydrol.2011.01.020.

[24] Yan H, DeChant CM, Moradkhani H. Improving soil moisture profile prediction with the particle filter-Markov chain Monte Carlo method. IEEE Trans Geosci Remote Sens 2015;53:6134–47. doi:10.1109/TGRS.2015.2432067.

[25] Dechant CM, Moradkhani H. Improving the characterization of initial condition for ensemble streamflow prediction using data assimilation. Hydrol Earth Syst Sci 2011;15:3399–410. doi:10.5194/hess-15-3399-2011.

[26] Vrugt J a, ter Braak CJF, Diks CGH, Schoups G. Hydrologic data assimilation using particle Markov chain Monte Carlo simulation: Theory, concepts and applications. Adv Water Resour 2013;51:457–78. doi:10.1016/j.advwatres.2012.04.002.

[27] Moradkhani H, Dechant CM, Sorooshian S. Evolution of ensemble data assimilation for uncertainty quantification using the particle filter-Markov chain Monte Carlo method. Water Resour Res 2012;48:W12520. doi:10.1029/2012WR012144.

[28] Abaza M, Anctil F, Fortin V, Turcotte R. Sequential streamflow assimilation for short-term hydrological ensemble forecasting. J Hydrol 2014;519:2692–706. doi:10.1016/j.jhydrol.2014.08.038.

[29] Parada LM, Liang X. Optimal multiscale Kalman filter for assimilation of near-surface soil moisture into land surface models. J Geophys Res D Atmos 2004;109:1–21. doi:10.1029/2004JD004745.

[30] Pan M, Wood EF, McLaughlin DB, Entekhabi D, Luo L. A Multiscale Ensemble Filtering System for Hydrologic Data Assimilation. Part I: Implementation and Synthetic Experiment. J Hydrometeorol 2009;10:794–806. doi:10.1175/2009JHM1088.1.

[31] Sahoo AK, De Lannoy GJM, Reichle RH, Houser PR. Assimilation and downscaling of satellite observed soil moisture over the Little River Experimental Watershed in Georgia, USA. Adv Water Resour 2013;52:19–33. doi:10.1016/j.advwatres.2012.08.007.

[32] Barbu AL, Calvet J-C, Mahfouf J-F, Lafont S. Integrating ASCAT surface soil moisture and GEOV1 leaf area index into the SURFEX modelling platform: a land data assimilation application over France. Hydrol Earth Syst Sci 2014;18:173–92.

[33] Kumar S V., Peters-Lidard CD, Mocko D, Reichle R, Liu Y, Arsenault KR, et al. Assimilation of remotely sensed soil moisture and snow depth retrievals for drought estimation. J Hydrometeorol 2014;15:2446–69. doi:10.1175/JHM-D-13-0132.1.

[34] Brocca L, Moramarco T, Melone F, Wagner W, Hasenauer S, Hahn S. Assimilation of surface- and root-zone ASCAT soil moisture products into rainfall-runoff modeling. IEEE Trans Geosci Remote Sens 2012;50:2542–55. doi:10.1109/TGRS.2011.2177468.

[35] Wanders N, Karssenberg D, De Roo A, De Jong SM, Bierkens MFP. The suitability of remotely sensed soil moisture for improving operational flood forecasting. Hydrol Earth Syst Sci 2014;18:2343–57. doi:10.5194/hess-18-2343-2014.

[36] Lee H, Seo DJ, Koren V. 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 2011;34:1597–615. doi:10.1016/j.advwatres.2011.08.012.

[37] Samuel J, Coulibaly P, Dumedah G, Moradkhani H. Assessing Model State and Forecasts Variation in Hydrologic Data Assimilation. J Hydrol 2014;513:127–41. doi:10.1016/j.jhydrol.2014.03.048.

[38] Njoku EG, Jackson TJ, Lakshmi V, Chan TK, Nghiem S V. Soil moisture retrieval from AMSR-E. IEEE Trans Geosci Remote Sens 2003;41:215–28. doi:10.1109/TGRS.2002.808243.

[39] Kerr YH, Waldteufel P, Wigneron J-P, Delwart S, Cabot F, Boutin J, et al. The SMOS Mission: New Tool for Monitoring Key Elements ofthe Global Water Cycle. Proc IEEE 2010;98. doi:10.1109/JPROC.2010.2043032.

[40] Bartalis Z, Wagner W, Naeimi V, Hasenauer S, Scipal K, Bonekamp H, et al. Initial soil moisture retrievals from the METOP-A Advanced Scatterometer (ASCAT). Geophys Res

Lett 2007;34. doi:10.1029/2007GL031088.

[41] Entekhabi D, Njoku EG, O'Neill PE, Kellogg KH, Crow WT, Edelstein WN, et al. The soil moisture active passive (SMAP) mission. Proc IEEE 2010;98:704–16. doi:10.1109/JPROC.2010.2043918.

[42] Kumar S V., Harrison KW, Peters-Lidard CD, Santanello J a., Kirschbaum D. Assessing the impact of L-band observations on drought and flood risk estimation: A decision theoretic approach in an OSSE environment. J Hydrometeorol 2014. doi:10.1175/JHM-D-13-0204.1.

[43] Kumar S V., Peters-Lidard CD, Santanello J a., Reichle RH, Draper CS, Koster RD, et al. Evaluating the utility of satellite soil moisture retrievals over irrigated areas and the ability of land data assimilation methods to correct for unmodeled processes. Hydrol Earth Syst Sci Discuss 2015;12:5967–6009. doi:10.5194/hessd-12-5967-2015.

[44] Wagner W, Hahn S, Kidd R, Melzer T, Bartalis Z, Hasenauer S, et al. The ASCAT soil moisture product: A review of its specifications, validation results, and emerging applications. Meteorol Zeitschrift 2013;22:5–33. doi:10.1127/0941-2948/2013/0399.

[45] Han X, Li X, Rigon R, Jin R, Endrizzi S. Soil Moisture Estimation by Assimilating L-Band Microwave Brightness Temperature with Geostatistics and Observation Localization. PLoS One 2015;10:e0116435. doi:10.1371/journal.pone.0116435.

[46] Plaza DA, De Keyser R, De Lannoy GJM, Giustarini L, Matgen P, Pauwels VRN. The importance of parameter resampling for soil moisture data assimilation into hydrologic models using the particle filter. Hydrol Earth Syst Sci 2012;16:375–90. doi:10.5194/hess-16-375-2012s.

[47] Montzka C, Grant JP, Moradkhani H, Franssen H-JH, Weihermüller L, Drusch M, et al. Estimation of radiative transfer parameters from L-band passive microwave brightness temperatures using advanced data assimilation. Vadose Zo J 2013;12:1–17. doi:10.2136/vzj2012.0040.

[48] Chen F, Crow WT, Starks PJ, Moriasi DN. Improving hydrologic predictions of a catchment model via assimilation of surface soil moisture. Adv Water Resour 2011;34:526–36. doi:10.1016/j.advwatres.2011.01.011.

[49] Alvarez-Garreton C, Ryu D, Western AW, Crow WT, Robertson DE. The impacts of assimilating satellite soil moisture into a rainfall–runoff model in a semi-arid catchment. J Hydrol 2014;519:2763–74. doi:10.1016/j.jhydrol.2014.07.041.

[50] Brocca L, Melone F, Moramarco T, Wagner W, Naeimi V, Bartalis Z, et al. Improving runoff prediction through the assimilation of the ASCAT soil moisture product. Hydrol Earth Syst Sci 2010;14:1881–93. doi:10.5194/hess-14-1881-2010.

[51] Scipal K, Scheffler C, Wagner W. Soil moisture-runoff relation at the catchment scale as observed with coarse resolution microwave remote sensing. Hydrol Earth Syst Sci Discuss 2005;2:417–48. doi:10.5194/hessd-2-417-2005.

[52] Parajka J, Naeimi V, Blöschl G, Wagner W, Merz R, Scipal K. Assimilating scatterometer soil moisture data into conceptual hydrologic models at the regional scale. Hydrol Earth Syst Sci Discuss 2006;10:353–68.

[53] Ridler M-E, Madsen H, Stisen S, Bircher S, Fensholt R. Assimilation of SMOS-derived

soil moisture in a fully integrated hydrological and soil-vegetation-atmosphere transfer model in Western Denmark. Water Resour Res 2014;50:8962–81.

[54] Alvarez-Garreton C, Ryu D, Western AW, Su C-H, Crow WT, Robertson DE, et al. Improving operational flood ensemble prediction by the assimilation of satellite soil moisture: comparison between lumped and semi-distributed schemes. Hydrol Earth Syst Sci Discuss 2014;11:10635–81. doi:10.5194/hessd-11-10635-2014.

[55] Crow WT, Ryu D. A new data assimilation approach for improving runoff prediction using remotely-sensed soil moisture retrievals. Hydrol Earth Syst Sci 2009;13:1–16. doi:10.5194/hess-13-1-2009.

[56] Massari C, Brocca L, Tarpanelli A, Moramarco T. Data Assimilation of Satellite Soil Moisture into Rainfall-Runoff Modelling: A Complex Recipe? Remote Sens 2015;7:11403–33. doi:10.3390/rs70911403.

[57] Yilmaz MT, Crow WT. The Optimality of Potential Rescaling Approaches in Land Data Assimilation. J Hydrometeorol 2013:650–60. doi:10.1175/JHM-D-12-052.1.

[58] Zhuo L, Han D. Misrepresentation and amendment of soil moisture in conceptual hydrological modelling. J Hydrol 2016;535:637–51. doi:10.1016/j.jhydrol.2016.02.033.

[59] Seo D-J, Koren V, Cajina N. Real-Time Variational Assimilation of Hydrologic and Hydrometeorological Data into Operational Hydrologic Forecasting. J Hydrometeorol 2003;4:627–41. doi:10.1175/1525-7541(2003)004<0627:RVAOHA>2.0.CO;2.

[60] Reichle RH. Data assimilation methods in the Earth sciences. Adv Water Resour 2008;31:1411–8. doi:10.1016/j.advwatres.2008.01.001.

[61] Aubert D, Loumagne C, Oudin L. Sequential assimilation of soil moisture and streamflow data in a conceptual rainfall–runoff model. J Hydrol 2003;280:145–61.

[62] Reichle RH, McLaughlin DB, Entekhabi D. Variational data assimilation of microwave radiobrightness observations for land surface hydrology applications. IEEE Trans Geosci Remote Sens 2001;39:1708–18. doi:10.1109/36.942549.

[63] Leisenring M, Moradkhani H. Analyzing the uncertainty of suspended sediment load prediction using sequential data assimilation. J Hydrol 2012;468-469:268–82. doi:10.1016/j.jhydrol.2012.08.049.

[64] Dechant CM, Moradkhani H. Examining the effectiveness and robustness of sequential data assimilation methods for quantification of uncertainty in hydrologic forecasting. Water Resour Res 2012;48:1–15. doi:10.1029/2011WR011011.

[65] Dong J, Steele-Dunne SC, Judge J, van de Giesen N. A particle batch smoother for soil moisture estimation using soil temperature observations. Adv Water Resour 2015;83:111–22. doi:10.1016/j.advwatres.2015.05.017.

[66] Andrieu C, Doucet A, Holenstein R. Particle Markov chain Monte Carlo methods. J R Stat Soc Ser B (Statistical Methodol 2010;72:269–342. doi:10.1111/j.1467-9868.2009.00736.x.

[67] Han X, Franssen HJH, Montzka C, Vereecken H. Soil moisture and soil properties estimation in the Community Land Model with synthetic brightness temperature observations. Water Resour Res 2014;50:6081–105. doi:10.1002/2013WR014586.

[68] Yin J, Zhan X, Zheng Y, Liu J, Hain CR, Fang L. Impact of Quality Control of Satellite Soil Moisture Data on Their Assimilation into Land Surface Model. Geophys Res Lett 2014;41:7159–66. doi:10.1002/2014GL060659.

[69] Leung S, Cooley D. A comparison of a traditional geostatistical regression approach and a general Gaussian process approach for spatial prediction. Stat 2014;3:228–39. doi:10.1002/sta4.57.

[70] Yan H, Moradkhani H. A regional Bayesian hierarchical model for flood frequency analysis. Stoch Environ Res Risk Assess 2015;29:1019–36. doi:10.1007/s00477-014-0975-3.

[71] Yan H, Moradkhani H. Toward more robust extreme flood prediction by Bayesian hierarchical and multimodeling. Nat Hazards 2016;81:203–25. doi:10.1007/s11069-015-2070-6.

[72] Burnash RJC, Ferral RL, McGuire RA. A generalized streamflow simulation system, conceptual modeling for digital computers 1973.

[73] Barry DA, Bajracharya K. On the Muskingum-Cunge flood routing method. Environ Int 1995;21:485–90.

[74] Su CH, Ryu D, Crow WT, Western AW. Beyond triple collocation: Applications to soil moisture monitoring. J Geophys Res Atmos 2014;119:6419–39. doi:10.1002/2013JD021043.

[75] Duan Q, Schaake J, Andréassian V, Franks S, Goteti G, Gupta H V., et al. Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops. J. Hydrol., vol. 320, 2006, p. 3–17. doi:10.1016/j.jhydrol.2005.07.031.

[76] Mu Q, Zhao M, Running SW. Improvements to a MODIS global terrestrial evapotranspiration algorithm. Remote Sens Environ 2011;115:1781–800. doi:10.1016/j.rse.2011.02.019.

[77] Hamill TM. Interpretation of Rank Histograms for Verifying Ensemble Forecasts. Mon Weather Rev 2001;129:550–60. doi:10.1175/1520-0493(2001)129<0550:IORHFV>2.0.CO;2.

[78] Schaefer GL, Cosh MH, Jackson TJ. The USDA Natural Resources Conservation Service Soil Climate Analysis Network (SCAN). J Atmos Ocean Technol 2007;24:2073–7. doi:10.1175/2007JTECHA930.1.

[79] Yan H. Magnitude and frequency of floods for rural, unregulated streams of Tennessee by L-Moments method. University of Arkansas; 2012.

[80] Yan H, Moradkhani H. Bayesian model averaging for flood frequency analysis. World Environ. Water Resour. Congr. 2014, vol. 2014, 2014, p. 1886–95. doi:10.1061/9780784413548.189.

[81] Jackson TJ, Cosh MH, Bindlish R, Starks PJ, Bosch DD, Seyfried M, et al. Validation of advanced microwave scanning radiometer soil moisture products. IEEE Trans Geosci Remote Sens 2010;48:4256–72. doi:10.1109/TGRS.2010.2051035.

[82] Dechant C, Moradkhani H. Radiance data assimilation for operational snow and

streamflow forecasting. Adv Water Resour 2011;34:351–64. doi:10.1016/j.advwatres.2010.12.009.

[83] Pan M, Wood EF. Inverse streamflow routing. Hydrol Earth Syst Sci 2013;17:4577–88. doi:10.5194/hess-17-4577-2013.

[84] Liu Q, Reichle RH, Bindlish R, Cosh MH, Crow WT, de Jeu R, et al. The Contributions of Precipitation and Soil Moisture Observations to the Skill of Soil Moisture Estimates in a Land Data Assimilation System. J Hydrometeorol 2011;12:750–65. doi:10.1175/JHM-D-10-05000.1.

[85] Wang A, Lettenmaier DP, Sheffield J. Soil moisture drought in China, 1950-2006. J Clim 2011;24:3257–71. doi:10.1175/2011JCLI3733.1.

[86] Samaniego L, Kumar R, Zink M. Implications of parameter uncertainty on soil moisture drought analysis in Germany. J Hydrometeorol 2013;14:47–68. doi:10.1175/JHM-D-12-075.1.

[87] Wood EF, Schubert SD, Wood AW, Peters-Lidard CD, Mo KC, Mariotti A, et al. Prospects for Advancing Drought Understanding, Monitoring, and Prediction. J Hydrometeorol 2015;16:1636–57. doi:10.1175/JHM-D-14-0164.1.

Table 1. The summary of parameters in the SAC-SMA model. The values in the parenthesis are the synthetic truths used to generate the synthetic streamflow and soil moisture.

Parameters	Description	Range and Synthetic Truth
Storage		
UZTWM	upper zone tension-water maximum, mm	10-300 (75)
UZFWM	upper zone free-water maximum, mm	5-150 (70)
LZTWM	lower zone tension-water maximum, mm	10-500 (335)
LZFPM	lower zone free primary maximum, mm	10-1000 (215)
LZFSM	lower zone free secondary maximum, mm	5-400 (68)
ADIMP	additional impervious area	0-0.4 (0.2)
Recession		
UZK	upper zone recession coefficient, day^{-1}	0.1-0.75 (0.25)
LZPK	lower zone primary recession coefficient, day^{-1}	0.0001-0.05 (0.01)
LZSK	lower zone secondary recession coefficient, day^{-1}	0.01-0.35 (0.05)
Percolation and other		
ZPERC	maximum percolation rate	5-350 (85)
REXP	percolation equation exponent	1-5 (2.6)
PCTIM	impervious fraction of watershed	0-0.1 (0.035)
PFREE	free water percolation from upper to lower zone	0-0.1 (0.05)
Routing		
Kq	Nash-cascade routing coefficient, day^{-1}	0.1-0.5 (0.23)
Not estimated		
RIVA	riparian vegetated area	0.01
SIDE	deep recharge to channel baseflow	0
RSERV	lower zone free water not transferable to tension water	0.3

Table 2. The summary of five data assimilation scenarios.

Scenario	Assimilated Data	Description
1	Streamflow	Outlet
2	Streamflow	Internal gauges 1-4
3	Streamflow	5 gauges
4	Streamflow + Soil moisture	Outlet + Synthetic truth soil moisture
5	Streamflow + Soil moisture	Outlet + Soil moisture from geostatistical model

Table 3. The summary of DA performance for scenarios 1-3.

Metric	Gauge	Assimilated Data and Scenario		
		Outlet gauge (Scenario 1)	Internal 4 gauges (Scenario 2)	Total 5 gauges (Scenario 3)
NSE	Outlet	0.99	0.99	1.00
	1	0.99	0.99	0.99
	2	0.99	0.99	1.00
	3	0.97	0.99	0.99
	4	0.99	0.99	0.99
	Avg.	0.99	0.99	0.99
ER95(%)	Outlet	1.0	7.0	3.2
	1	0.7	7.5	4.4
	2	0.6	6.2	1.2
	3	0.0	7.5	2.1
	4	1.2	8.2	6.7
	Avg.	0.7	7.3	3.5

Table 4. The summary of DA performance for scenarios 4-5. Scenario 1 is also presented to compare the surface moisture filed. The metrics NSE and ER95 are used for streamflow while RMSE is used for surface soil moisture (in terms of degree of saturation).

Metric	Gauge	Assimilated Data and Scenario		
		Outlet gauge (Scenario 1)	Outlet gauge + Synthetic truth soil moisture (Scenario 4)	Outlet gauge + Soil moisture from geostatistical model (Scenario 5)
NSE	Outlet	0.99	0.98	0.97
	1	0.99	0.97	0.95
	2	0.99	0.98	0.94
	3	0.97	0.92	0.72
	4	0.99	0.98	0.97
	Avg.	0.99	0.97	0.91
ER95(%)	Outlet	1.0	0.0	1.6
	1	0.7	1.5	1.8
	2	0.6	0.0	1.9
	3	0.0	1.5	3.8
	4	1.2	0.3	1.2
	Avg.	0.7	0.7	2.1
RMSE	Basin Avg.	0.054	0.018	0.025

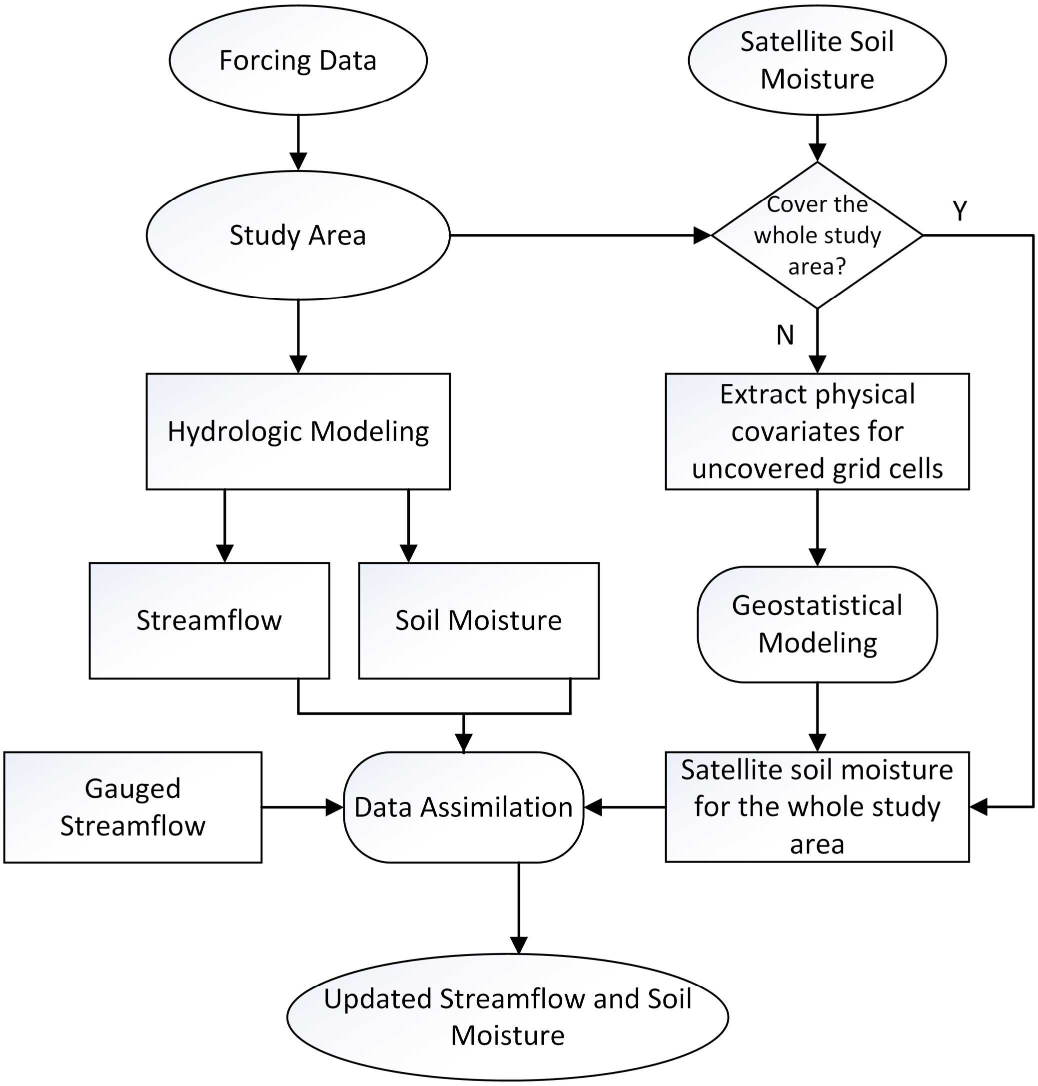

Figure 1. The flowchart of the combined data assimilation with geostatistical modeling.

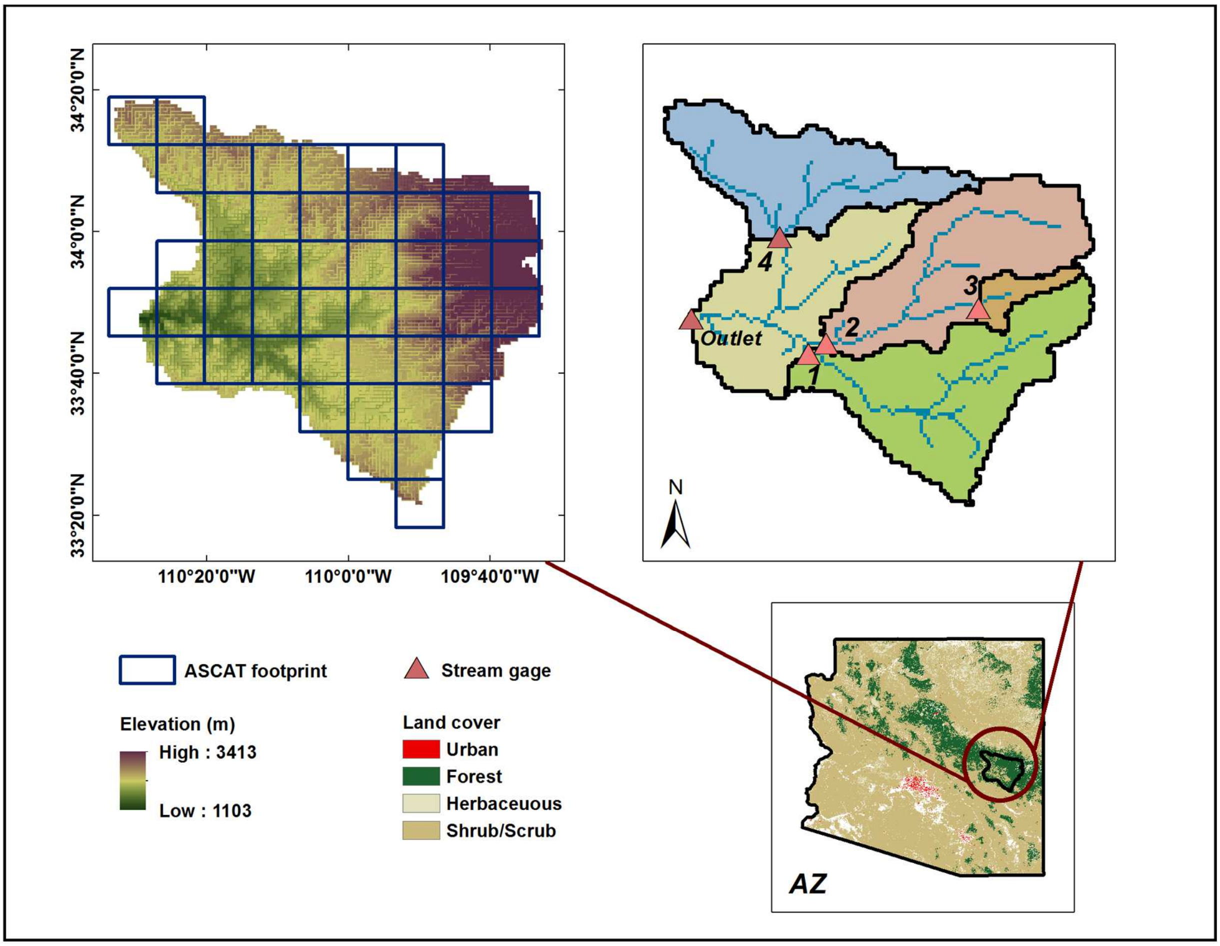

Figure 2. The location of the study watershed, the delineation of the five sub-watersheds, and the footprint (12.5×12.5 km^2) of synthetic ASCAT data over the watershed.

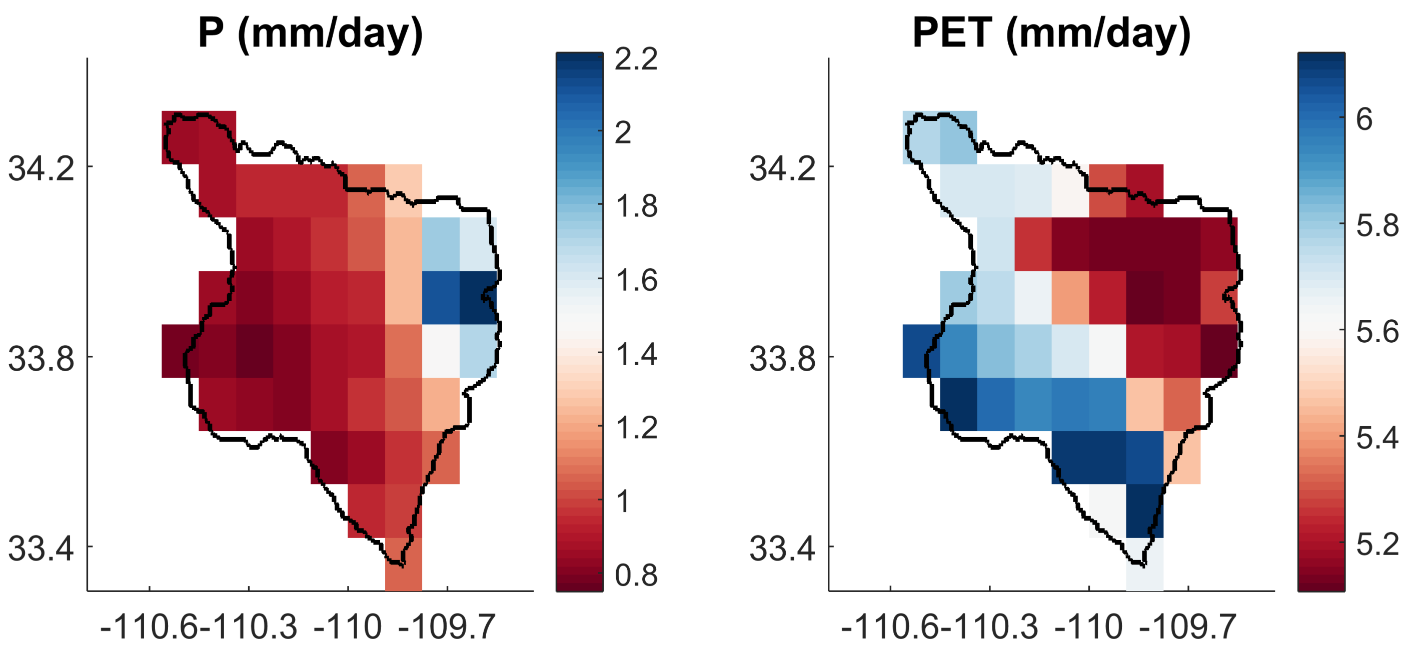

Figure 3. The spatial pattern of mean daily forcing data, precipitation (P) and potential evapotranspiration (PET), from Oct. 1 2005–Sep. 30 2007.

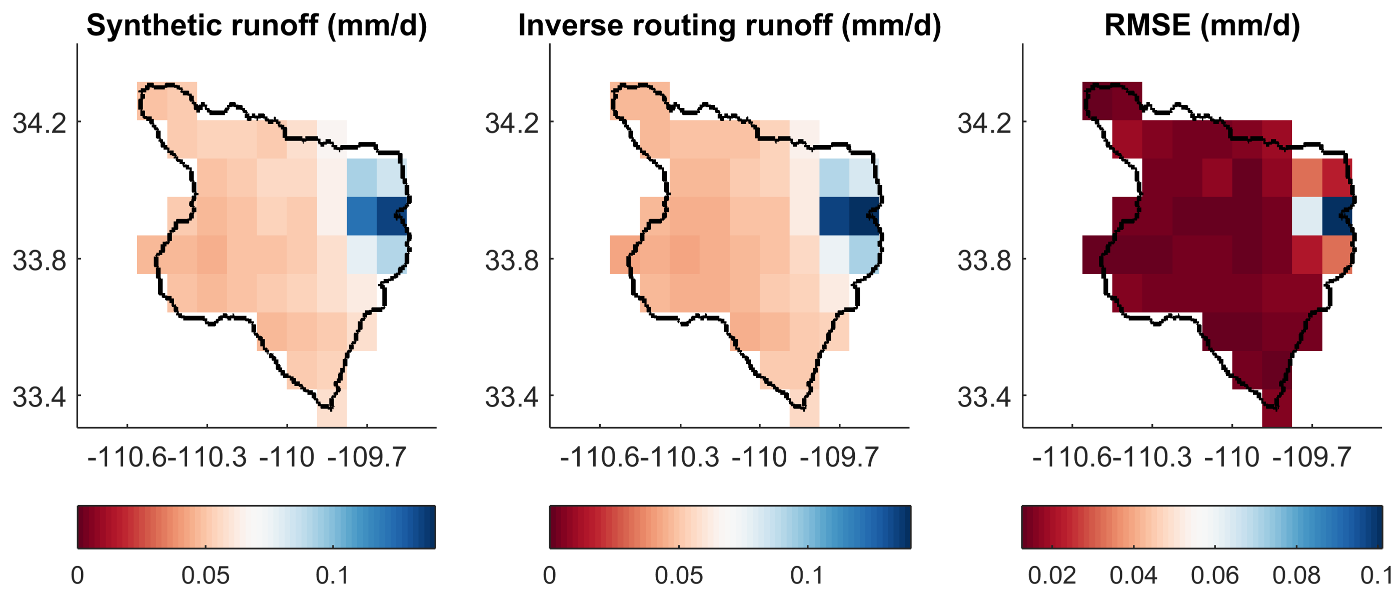

Figure 4. The mean daily synthetic runoff, the inverse routing runoff, and their RMSE values, based on the simulated period Oct. 1 2005–Sep. 30 2007.

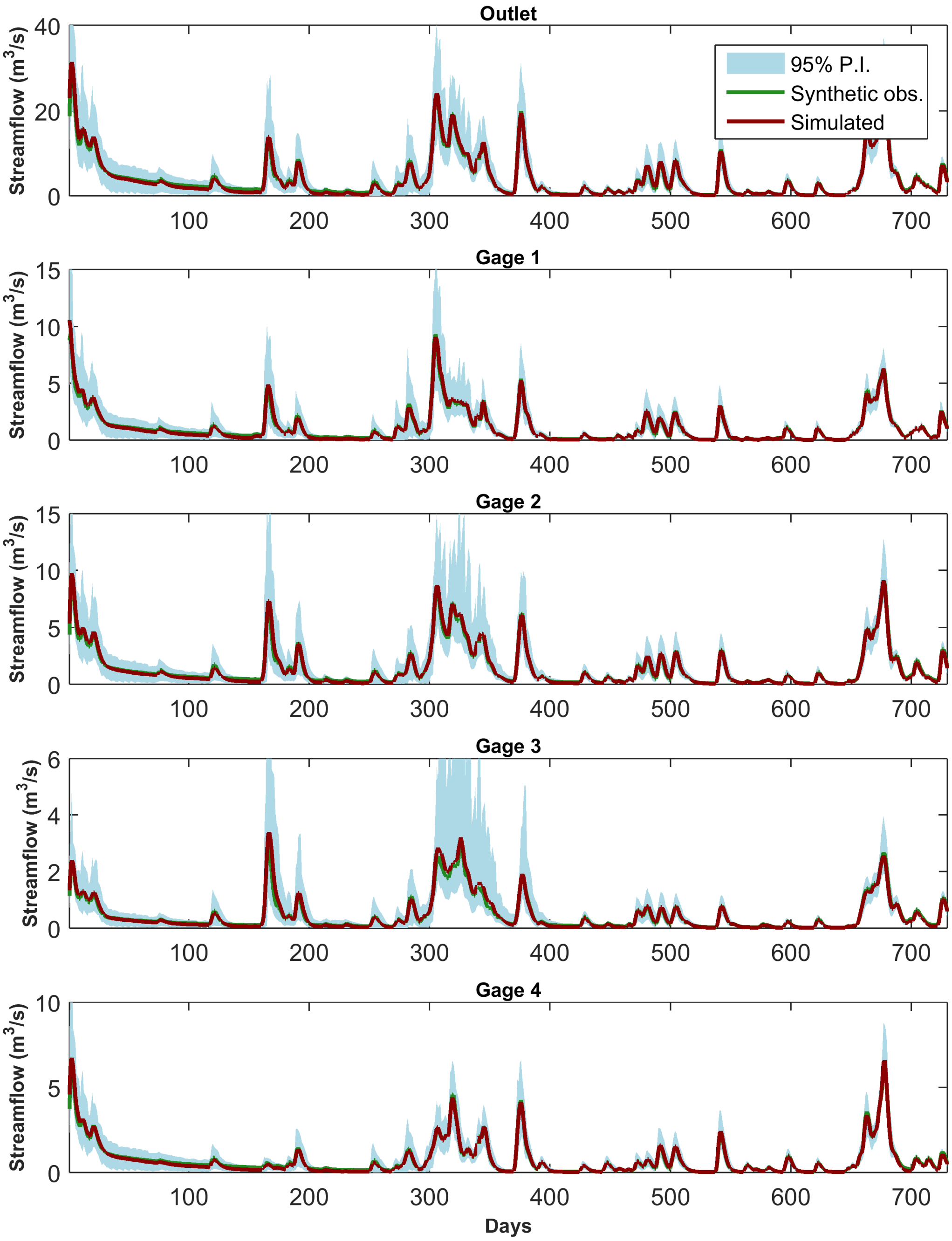

Figure 5. The streamflow hydrographs for the five gauges from scenario 3 (assimilating five gauge streamflow).

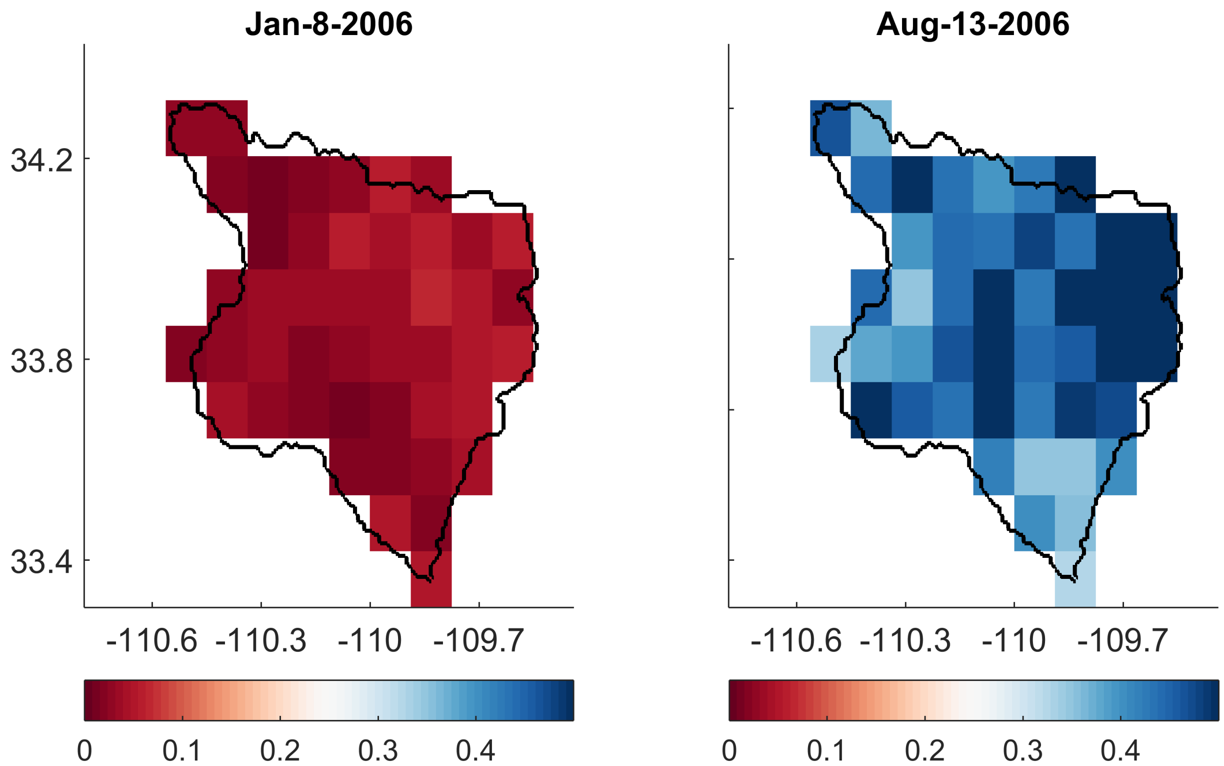

Figure 6. Synthetic ASCAT soil moisture (degree of saturation) for the dry (left) and wet (right) days.

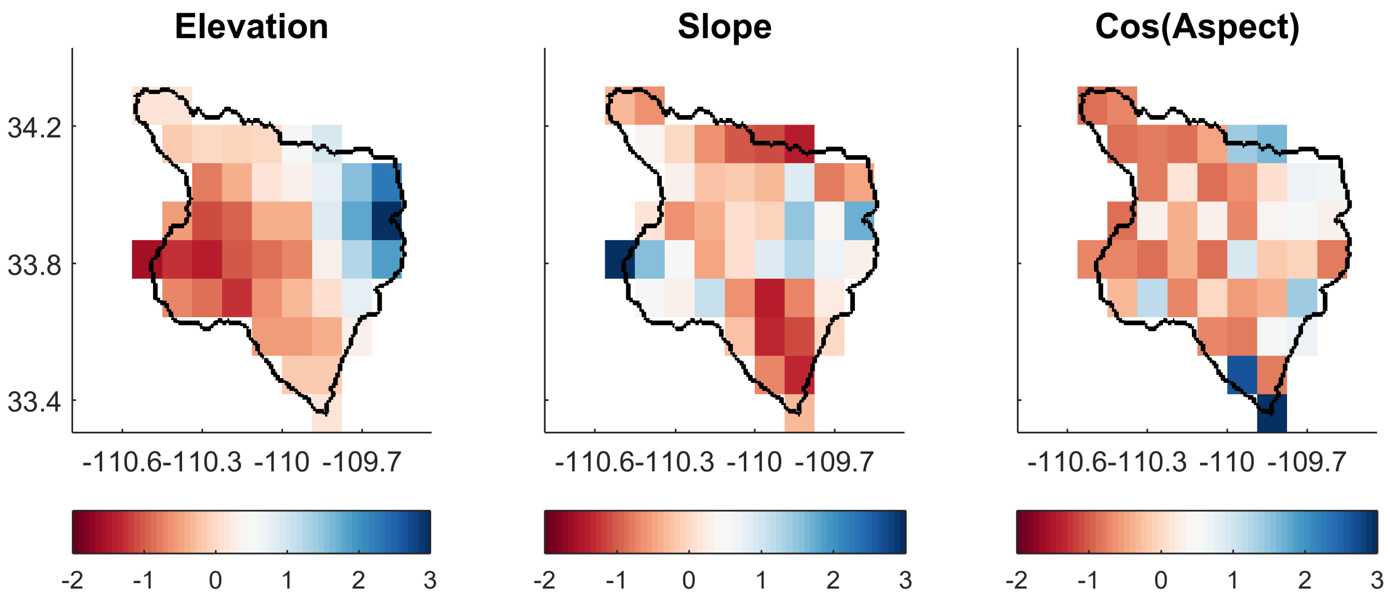

Figure 7. Standardized covariates: elevation, slope, and cos (aspect) for each grid cell.

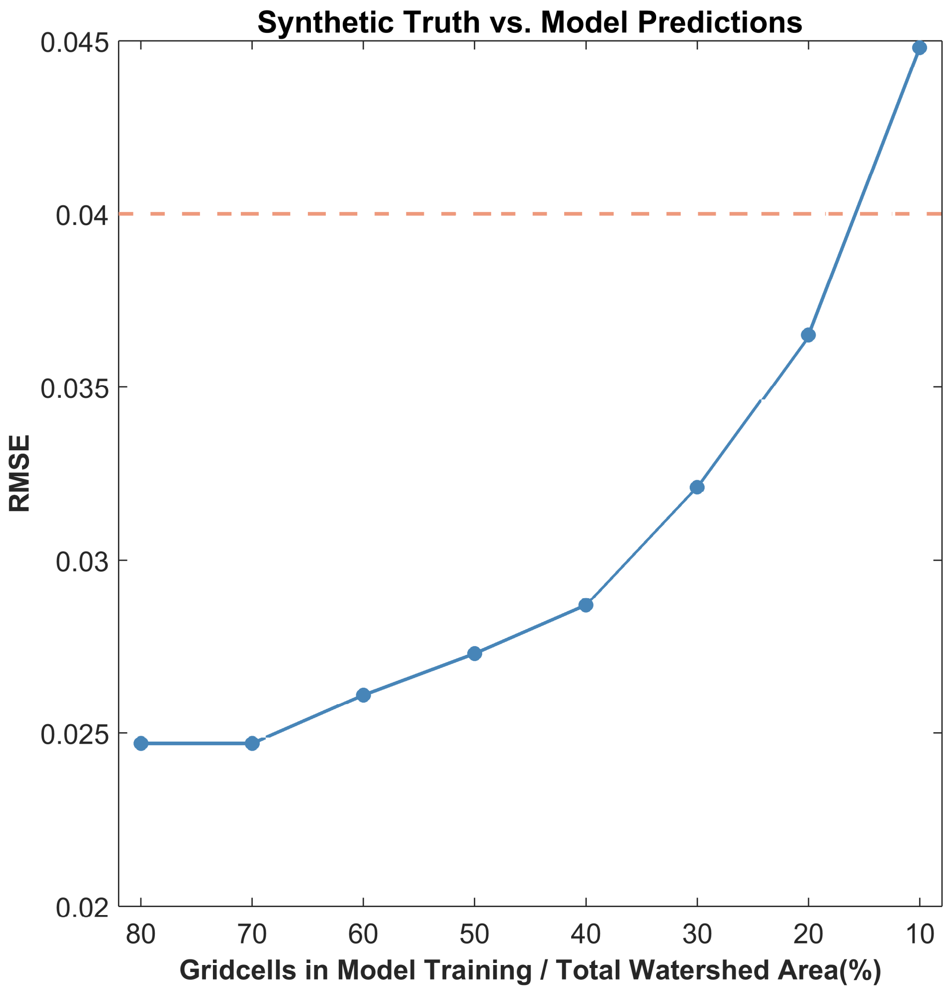

Figure 8. The RMSE for the prediction cells against the training cells for general Gaussian model with respect to the percent of the total watershed area. The *x-axis* indicates the number of total watershed grid cells, which are used to train the model (the remaining grid cells are used to validate the model). The *dash line* indicates the recommended ASCAT RMSE threshold for the study area.

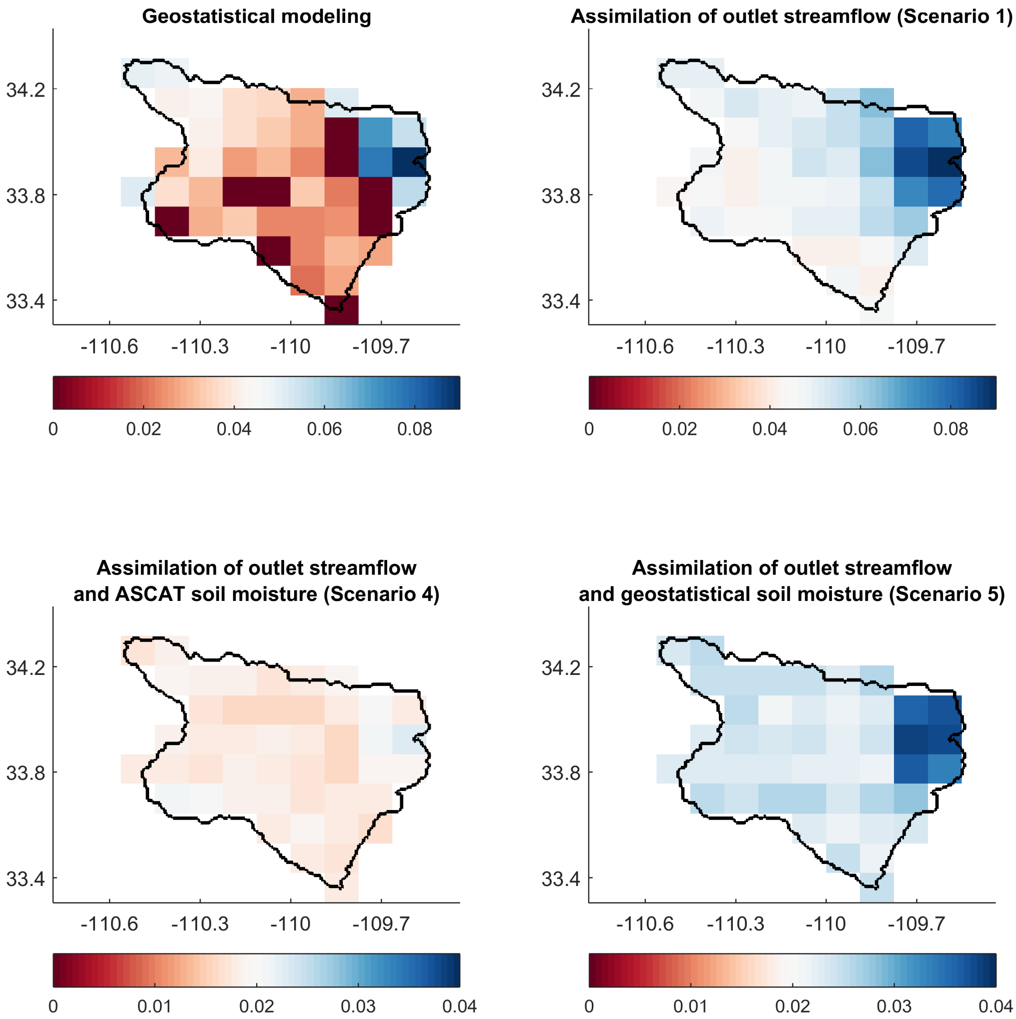

Figure 9. The RMSE values (in terms of degree of saturation) for the surface soil moisture (synthetic truth vs. geostatistical model/DA predictions) from different scenarios. The top-left panel shows the performance of geostatistical modelling (the nine zero RMSE grid cells indicate the cells used in model training). The top-right panel shows the high uncertainty of soil moisture field if only outlet streamflow is assimilated. The bottom-left panel indicates that the joint assimilation of outlet streamflow and satellite soil moisture can significantly improve the soil

moisture field. The bottom-right panel suggests that if the study area cannot be covered by satellite data, assimilating the soil moisture from geostatistical model can significantly improve the soil moisture predictions.

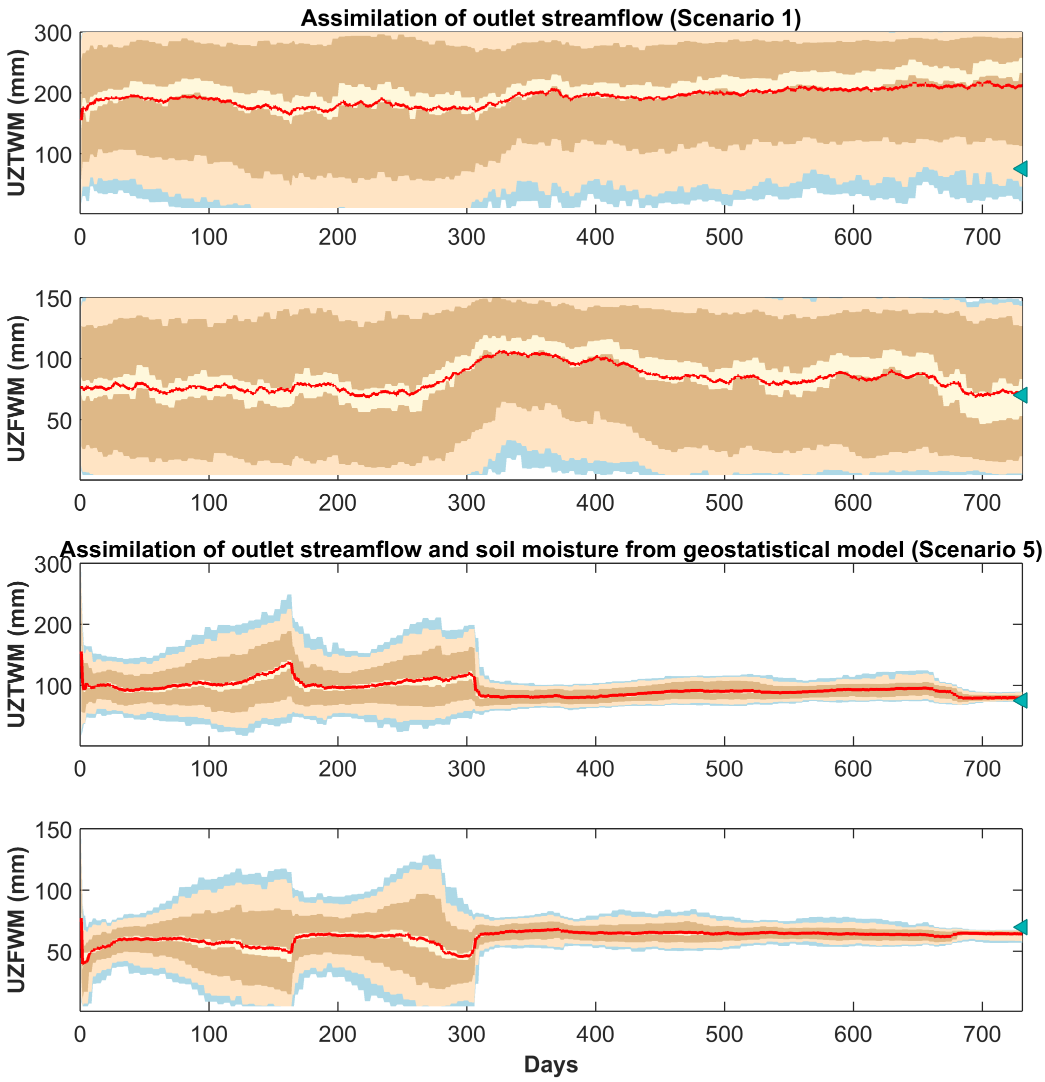

Figure 10. Evolution of two upper soil zone model parameters (UZTWM and UZFWM) for two scenarios. The shaded areas correspond to 95, 68, and 10 percentiles of prediction intervals. The line is the mean value and the symbol on the right *y-axis* is the predefined parameter value. The convergence of parameters can be seen with the assimilation of soil moisture from the geostatistical model.