Advances in Water Resources 67 (2014) 32–45

Contents lists available at ScienceDirect

Advances in Water Resources

journal homepage: www.elsevier.com/locate/advwatres

Improving the estimation of hydrological states in the SWAT model via the ensemble Kalman smoother: Synthetic experiments for the Heihe River Basin in northwest China

Fangni Lei [a,b], Chunlin Huang [b,*], Huanfeng Shen [a,*], Xin Li [b]

[a] School of Resource and Environmental Sciences, Wuhan University, Wuhan, Hubei 430079, China
[b] Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences, Lanzhou, Gansu 730000, China

ARTICLE INFO

Article history:
Received 29 January 2013
Received in revised form 29 January 2014
Accepted 23 February 2014
Available online 3 March 2014

Keywords:
Soil moisture
SWAT
Data assimilation
Heterogeneity
EnKS

ABSTRACT

Data assimilation as a method to predict variables, reduce uncertainties and explicitly handle various sources of uncertainties has recently received widespread attention and has been utilized to combine in situ and remotely sensed measurements with hydrological models. However, factors that significantly influence the capability of data assimilation still need testing and verifying. In this paper, synthetic surface soil moisture data are assimilated into the Soil and Water Assessment Tool (SWAT) model to evaluate their impact on other hydrological variables via the ensemble Kalman smoother (EnKS), using data from the Heihe River Basin, northwest China. The results show that the assimilation of surface soil moisture can moderately improve estimates of deep layer soil moisture, surface runoff and lateral flow, which reduces the negative influences of erroneous forcing and inaccurate parameters. The effects of the spatially heterogeneous input data (land cover and soil type) on the performance of the data assimilation technique are noteworthy. Moreover, the approaches including inflation and localization are specifically diagnosed to further extend the capability of the EnKS.

1. Introduction

Soil moisture plays a fundamental role in the study of hydrology, meteorology, and agriculture. From a hydrologic viewpoint, the significance of soil moisture is in its role in partitioning precipitation into runoff and infiltration and controlling storage and drainage [14,25,28]. In recent years, remote sensing has provided a crucial solution for assimilating surface soil moisture into hydrologic models, thereby improving the prediction of substantial hydrologic variables, such as root-zone soil moisture, evapotranspiration, and surface runoff [49,52,56].

Since the 1990s, remote sensing has demonstrated its utility in the spatial and temporal characterization of surface soil moisture and has overcome the limitations of the sparsely distributed point measurements of traditional in situ networks [25,47]. With the launch of the SMOS (Soil Moisture and Ocean Salinity) satellite, and the SMAP (Soil Moisture Active and Passive) mission on the agenda, satellite-based global observations of surface soil moisture (0–5 cm) will soon furnish complementary information to enhance the estimations of other hydrologic states, especially for flood forecasting [15,29]. The information on the moisture condition in the subsurface and root-zone layers is crucial for improving flood forecasting, water resource management, and the comprehension of hydrologic processes. A number of studies have focused on how to integrate the surface soil moisture into land surface and hydrologic models from the local to the global scale [4,8,24,27,33,37,39,51,59,60]. However, there is still a lack of knowledge on how to efficiently employ remotely sensed soil moisture data in catchment-scale hydrologic models, especially in distributed and semi-distributed models [12,23,57,62].

Data assimilation techniques have been utilized in meteorology and oceanography for decades, and the commonly applied methods can be roughly divided into two categories: variational techniques and filter techniques [13,17]. Both techniques have recently been adopted in hydrologic data assimilation frameworks. Reichle et al. [50] used a four-dimensional variational method to estimate large-scale soil moisture profiles, and Lee et al. [31] carried out a series of real-world experiments with streamflow data assimilated into a distributed hydrologic model via a variational method. To examine the behavior of the ensemble Kalman filter (EnKF) in a hydrologic data assimilation framework, Xie and Zhang [62] diagnosed the augmented state-parameter assimilation in

* Corresponding authors. Tel.: +86 0931 4967975 (C. Huang). Tel.: +86 027 68777080 (H. Shen).
E-mail addresses: huangcl@lzb.ac.cn (C. Huang), shenhf@whu.edu.cn (H. Shen).

http://dx.doi.org/10.1016/j.advwatres.2014.02.008
0309-1708/

catchment modeling, and Chen et al. [6], Han et al. [23] both assimilated synthetic soil moisture and in situ measurements into a semi-distributed hydrologic model to improve the estimation of important variables. The results were positive in these synthetic experiments. However, in practical applications, where more uncertainties arise, more difficulties will be encountered. The recent advancements in the Particle filter (PF) have suggested that the promising Markov Chain Monte Carlo (MCMC) reduces the parameter sample impoverishment with a manageable ensemble size. Both the synthetic and real experiments achieved more accurate streamflow predictions with PF-MCMC in comparison to the traditional PF approach [43]. Meanwhile, the smoother techniques, such as the ensemble Kalman smoother (EnKS), have been proved to perform well in land surface assimilation and other domains [13,18,21,46,58], but until now have been inadequately investigated for use in hydrological data assimilation. Therefore, the effectiveness and capabilities of this technique need to be tested and verified in a hydrological data assimilation framework. The potential influencing factors (especially the spatially heterogeneous geographic information) and the remedies for the limited ensemble size and the sizeable model forecast error are also explored in detail.

In this work, the study region is located in the upper reaches of the Heihe River Basin in northwest China. Synthetic experiments are designed to assess the potential of EnKS in improving the prediction of hydrologic variables, including root-zone soil water content (SWC), evapotranspiration, surface runoff, and lateral flow, by assimilation of the surface SWC under different conditions. Additionally, various other factors are diagnosed in detail. The remainder of this paper is organized as follows. Section 2 presents the basic principles of the methodologies and the hydrologic processes of the SWAT model, especially emphasizing the soil water dynamics. Section 3 describes the study area and basic information, as well as the design and implementation of the experiments. The diagnostics of the capability and sensitivity of EnKS are discussed in Section 4. Finally, Section 5 provides a brief summary and describes the directions of future work.

2. Method and model

2.1. The ensemble Kalman filter

Because the chosen method is an important variant of the EnKF, we explain its mechanism starting from the EnKF. The EnKF was first proposed by Evensen [16], based upon the standard Kalman filter formulations and Monte Carlo method to estimate states and parameters in large high-dimensional nonlinear dynamics. The error covariance of the model forecast state is computed based on the ensemble generation produced by adding Gaussian white noise to the model state and/or forcing data. Let $y^f(t)$ be the ensemble of the model states at time t, which is propagated forward using the full nonlinear hydrological model, $A[\bullet]$. Thus, the model forecasting can be expressed as:

$$y^f(t) = A[y^a(t-1), \alpha, u(t), w(t)] \tag{1}$$

where $y^a(t-1)$ is the analysis results at time $t-1$, which combines the model forecast and observations whenever available, and α, $u(t)$, and $w(t)$ are the time-invariant parameters, forcing data, and forecasting error, respectively.

The observation $z(t)$ is related to states through the measurement operator $H[\bullet]$:

$$z(t) = H[y^f(t), \beta] + \varepsilon(t) \tag{2}$$

where $\varepsilon(t)$ represents independently and identically Gaussian distributed observation errors with the covariance matrix R, and β is the measurement parameters related to the observational operator. Whenever observations are available, the model forecast is updated using the Kalman gain, which weights the relative uncertainty of the simulated estimation and observation:

$$y^a(t) = y^f(t) + K\{z(t) + \varepsilon(t) - H[y^f(t)]\} \tag{3}$$

where the Kalman gain K is calculated from the ensemble statistics:

$$K = PH^T(HPH^T + R)^{-1} \tag{4}$$

Here, P is the model forecast error covariance matrix. In practice, PH^T is the cross covariance between the predicted model state $y^f(t)$ and its transformed value $H[y^f(t)]$ in the observation space, and HPH^T is the error covariance of the transformed forecasting states $H[y^f(t)]$. Matrices PH^T and HPH^T are computed using the ensemble realizations of the model forecast and its transformation in the observation space. Each ensemble member is updated individually, and the averaged ensemble diagnostic variables are treated as "true" or analysis values.

2.2. The ensemble Kalman smoother

The EnKS [18] is an extension of the EnKF in which the states and observations are distributed in time and space, and the augmented state vector is updated via the EnKF. The EnKF solution is the "first guess" of the EnKS, and a smoother solution can be found by including the impact of measurements backward in time [17].

The EnKS requires only forward model runs and no backward integrations of the model equations [13,18]. The forward model runs through to the end of the smoother window to achieve the forecasting states. Thus, the augmented state vector Y containing states y at the time steps of interest is as follows:

$$Y = [y_1\ y_2 \cdots y_n]^T \tag{5}$$

In addition, the augmented measurement vector Z contains all the observations in the smoother window that are temporally and spatially independent:

$$Z = [z_1\ z_2 \cdots z_m]^T \tag{6}$$

The above-mentioned EnKF formulas are then utilized to yield a solution. The implementation of the EnKS requires that the ensemble during the prior times must be stored and able to be updated whenever new observations become available [13].

2.2.1. Inflation

The successful application of the EnKF and EnKS is highly dependent on the accurate estimation of the model forecast and the observational error covariance matrices. Many researchers have been making efforts to improve the estimation of error covariance matrices and forecast bias to take into account limited ensemble size and ambiguous model intrinsic bias. Recently, Wu et al. [61] have introduced the second-order least squares statistic (SLS) and an iterative feedback framework to generate a more accurate background forecast error covariance matrix. The iteration is conducted within one time step, and thus the notation of the time step is eliminated in the following formulae. Without the consideration of adjusting observational error covariance matrix R, the inflation factor γ_i for the background error matrix can be achieved by minimizing the objective function $L[\bullet]$:

$$L(\gamma_i) = Tr\left[\left(dd^T - \gamma_i HP_iH^T - R\right)\left(dd^T - \gamma_i HP_iH^T - R\right)^T\right] \tag{7}$$

where d is the filter innovation between the actual and predicted observations, $d = z - H(y^f)$ and P_i is the forecast error covariance

matrix of i th iteration with $P_1 = P$. Then, an improved analysis state y_i^a can be updated by including this inflation factor.

$$y_i^a = y^f + \gamma_i P_i H^T \left(H\gamma_i P_i H^T + R\right)^{-1} d \tag{8}$$

Thus, the estimation of the forecast error covariance matrix P_{i+1} could be improved with the feedback information from the updated states y_i^a:

$$P_{i+1} = \frac{1}{N-1}\sum_{j=1}^{N}\left(y_j^f - y_i^a\right)\left(y_j^f - y_i^a\right)^T \tag{9}$$

where N is the ensemble size. In real applications, where the true states are uncertain, the analysis state, which combines the model forecast and observations together, would be a more accurate estimate of the actual situation than the model forecast. Therefore, the forecast error covariance matrix is adjusted to $\gamma_{i+1}P_{i+1}$, where

$$\gamma_{i+1} = \frac{Tr[HP_{i+1}H^T(dd^T - R)]}{Tr[HP_{i+1}H^T HP_{i+1}H^T]} \tag{10}$$

is estimated by minimizing the objective function $L(\gamma_{i+1})$. The iteration for computing the inflation factor and forecast error covariance matrix would not stop until the criterion is satisfied; that is, the difference of the objective function value between two successive iterations is less than a predefined threshold.

2.2.2. Localization

The inflation of the forecast error covariance matrix can thwart the filter divergence to some extent. However, small background correlation errors between distinctive states and remote observations still exist. Localization is a possible solution to filter these errors and to then obtain a compactly supported background error covariance matrix [26]. The kernel component of localization is the construction of the correlation function. In this study, a fifth-order piecewise rational function of Gaspari and Stephen [20] is adopted to generate the important correlation function ρ. The distance $D_{i,j}$ is the Euclidean distance between the two state points y_i and y_j. We define an influential length scale L and let $F = \sqrt{10/3} * L$. The ratio between the actual distance and the influential length is $\delta = D_{i,j}/F$. Thus, the correlation function is as follows:

$$\rho(i,j) = \begin{cases} -\frac{\delta^5}{4} + \frac{\delta^4}{2} + \frac{5\delta^3}{8} - \frac{5\delta^2}{3} + 1, & 0 \leqslant \delta \leqslant 1 \\ \frac{\delta^5}{12} - \frac{\delta^4}{2} + \frac{5\delta^3}{8} + \frac{5\delta^2}{3} - 5\delta + 4 - \frac{2}{3\delta}, & 1 < \delta \leqslant 2 \\ 0, & \delta > 2 \end{cases} \tag{11}$$

Apparently, this function is isotropic and decreases monotonically with the ratio δ, which mimics the real-world situation between different states [22]. The Kalman gain K is subsequently refreshed:

$$K = \left[(\rho \circ P)H^T\right]\left[H(\rho \circ P)H^T + R\right]^{-1} \tag{12}$$

where the notation $\circ$ denotes the element-by-element matrix multiplication (also called the Schur product). With the local support from a fifth-order correlation function, redundant or invalid correlation errors associated with remote observations and states are filtered out [26]. In this study, the influential length scale is referred to as the temporal scale, where the correlation between two states at different time steps decreases with the time difference increasing. The temporal compactness or localization conforms to the reality and is quite reasonable with respect to hydrologic phenomena, such as rainfall-runoff processes.

2.3. The Soil and Water Assessment Tool

The Soil and Water Assessment Tool (SWAT) is a watershed-scale, time-continuous, semi-distributed and physically based hydrologic model. It was developed to assess the impact of land management practices and climate variability on the water availability and quality response of complex watersheds with heterogeneous land use and soil conditions [1]. The SWAT model operates on a daily step and is widely utilized for its effectiveness and capability in water resource management.

In a SWAT simulation, a watershed is first subdivided into subbasins, according to the DEM, then each subbasin is further divided into hydrologic response units (HRUs), based on the spatially heterogeneous land cover, soil characteristics and slope. The hydrology is separated into two major partitions: (a) the land phase and (b) the routing phase. Here, we exclusively focus attention upon the fundamental physical processes of flow generation and soil water dynamics. The 2009 version of SWAT is used in this study, and detailed information on the hydrological processes is given by Neitsch et al. [53].

2.3.1. Flow generation

Hydrological processes, including surface runoff, are based on a water balance equation, with a daily step at the HRU scale:

$$SW_t = SW_0 + \sum_{i=1}^{t}\left(P_i - Q_{surf,i} - ET_i - W_{perc,i} - Q_{gw,i}\right) \tag{13}$$

where SW_0 is the initial SWC at the beginning of the simulation (mm H_2O), and SW_t is the SWC at the end of day t (mm H_2O). P_i is the total precipitation on day i, and $Q_{surf,i}$, ET_i, $W_{perc,i}$, and $Q_{gw,i}$ are the daily amounts of surface runoff, evapotranspiration, percolation, and return flow on day i (mm H_2O), respectively.

In the SWAT model, the total streamflow Q_i consists of surface runoff $Q_{surf,i}$, subsurface lateral flow $Q_{lat,i}$ and groundwater flow $Q_{gw,i}$ on day i:

$$Q_i = Q_{surf,i} + Q_{lat,i} + Q_{gw,i} \tag{14}$$

Surface runoff is the main component of streamflow and determines the soil moisture in profiles. It can be estimated using either the modified SCS curve number (CN) procedure or the Green–Ampt infiltration method, depending on the availability of daily or hourly precipitation data. Here, in this paper, the modified SCS CN method is chosen according to the available daily precipitation data.

The processes are carried out at the HRU level and, subsequently, water from these generations is aggregated at the subbasin level. Through the routing phase of the hydrologic cycle, the total runoff is generated in the channel network, producing the streamflow at the outlet.

2.3.2. Soil water dynamics

Soil moisture plays an important role in determining other processes, including surface runoff, lateral flow, and evapotranspiration. According to the water balance equation, the dynamics for soil water in each layer at the HRU scale on a daily basis can be expressed as follows:

$$SW'_{ly,i} = SW_{ly,i} + W_{perc,ly,i} - Q_{lat,ly,i} - ET_{a,ly,i} \tag{15}$$

where $SW'_{ly,i}$ and $SW_{ly,i}$ represent the SWC at the end and beginning of day i, respectively; $W_{perc,ly,i}$ is the net percolation that enters the layer ly; and $Q_{lat,ly,i}$ and $ET_{a,ly,i}$ are the lateral flow and evapotranspiration drawn from the layer ly, respectively. For the surface soil layer, the amount of infiltration is the difference between the precipitation and the surface runoff.

To diagnose the performance of data assimilation on the hydrologic cycle of the SWAT model, we exclusively select four primary

states, as shown in Table 1. The specifications of these states are shown by Neitsch et al. [45].

3. Experiment

3.1. Study area and model data

In this paper, the Babaohe watershed located in the upper reaches of the Heihe River Basin in semi-arid northwest China is selected (Fig. 1). This watershed varies greatly from 2669 to 4974 m in elevation and covers 2350.44 km^2. Grassland and brush are the two dominant land-cover types, and the main soil types are alpine meadow soil, alpine frost desert soil and chestnut soil [34]. The area receives approximately 526 mm of annual precipitation, of which nearly 70% falls between June and September. The daily average temperatures range from −14 °C to 20 °C.

The topographic information, land use, soil type, and daily meteorological record datasets were required for incorporation into ArcSWAT, an ArcGIS interface for the SWAT model. The spatial delineation of the subbasins and HRUs is based on the DEM, land cover and soil data, whose horizontal resolution is an important factor when deciding the number of subbasins and HRUs. In this work, the stream network and subbasins are delineated from the 300 m horizontal resolution DEM by resampling the DEM data at 30×30 m via bilinear interpolation, which is enough for the streamflow simulation and can somewhat reduce the computational cost [5,3]. Based on the topographic information, the watershed is divided into 27 subbasins, and the further division of subbasins into HRUs is based on the land cover and soil information. The land cover information is provided from the Chinese National Land Cover Data (NLCD) Set 2000, generated by visual interpretation of the Landsat TM and ETM+ images acquired primarily in the year 2000 [36]. The Chinese NLCD 2000 data for the Babaohe watershed is reclassified into six different land cover types (Table 2). The soil obtained from the Harmonized World Soil Database [19] is reclassified into seven types (Table 3). Thus, we obtained 27 subbasins after watershed delineation and 292 HRUs after the overlay of land cover, soil data and slope.

An additional complication is the lack of uniform soil layer depths between HRUs, resulting from the original soil data, which makes it difficult to obtain the forecast cross covariance term when performing the data assimilation [6]. Therefore, the soil profiles across the watershed are resampled into five vertically consistent layers, with the bottom depths of 5, 15, 30, 60 and 100 cm. The surface soil layer has a depth of 5 cm to be consistent with the surface soil moisture data retrieved from remote sensing.

The daily meteorological data, including precipitation, maximum/minimum temperature, wind speed, relative humidity and solar radiation, were obtained from three meteorological stations: Qilian (38.19 N, 100.24 E), Minle (38.44 N, 100.83 E) and Yongchang (38.18 N, 101.58 E), located in the vicinity of this watershed. Based on the mechanism in the ArcSWAT interface, the meteorological data assigned to a subbasin comes from the nearest station [38]. Three subbasins situated in the southeast of the Babaohe watershed are assigned records from the Yongchang station with extremely inaccurate measured precipitation. The influence of inaccurate precipitation quantities will be shown in the next section. The daily streamflow data have been measured at the outlet of the Babaohe watershed, the Qilian hydrologic station, since 1978. These measurements from the stations are utilized to calibrate the model and generate "synthetic true" states and observations in the Observing System Simulation Experiment (OSSE) framework [40]. However, the absence of surface soil moisture observations at the HRU spatial level necessitates an additional meteorological forcing dataset from another distinct source. Datasets simulated by the Weather Research and Forecasting model (WRF) with resolutions of 5 km/1 h were validated over the Heihe River Basin [48] and therefore aggregated into the daily data. Special forcing datasets assigned to the three stations were obtained from the corresponding grid points.

3.2. Experiment design

3.2.1. SWAT calibration

To construct the data assimilation framework, the SWAT model is calibrated for three years using the recorded daily streamflow data measured at the outlet from January 1, 2005 to December 31, 2007. Before calibration, the model is spun-up for three years from January 1, 2002 to December 31, 2004. A sensitivity analysis is also performed for the sake of the parameters being selectively optimized. According to the sensitivity analysis, a total of 16 parameters are chosen, including ESCO, CN2, CANMX, SOL_K, HRU_SLP, GWQMN, SOL_AWC, ALPHA_BF, CH_K2, EPCO, SURLAG, SOL_ALB, GW_DELAY, CH_N2, GW_REVAP, and SLSUBBSN, sorted by the degree of sensitivity (Table 4). The validation covers the whole year 2008. To assess the performance of the SWAT model, two evaluation metrics are chosen: the Nash–Sutcliffe efficiency and the coefficient of determination [44]. During the calibration period, they are 0.55 and 0.68, respectively. During the validation period, the corresponding values are 0.52 and 0.64, respectively. Both metrics show the effectiveness of the parameters and their applicability to the data assimilation framework [23,44].

Among these parameters, ESCO and EPCO are important in determining the vertical soil water coupling strengths between two adjacent soil layers [30,35]. As ESCO decreases and EPCO increases, more evaporative and plant uptake demands will be drawn from the deeper layers when water demand cannot be satisfied by the upper layers, thus strengthening the vertical coupling within the soil profile [6]. After calibration, they are 0.84 and 0.20, respectively, representing a moderately loose coupling strength.

3.2.2. Assimilation experiments

In this study the performance of the EnKS and improvements derived from the inflation factor and localization, associated with a semi-distributed hydrologic model, are the main objectives. Therefore, a typical data assimilation framework is constructed following the specification of the OSSE [40]. Four key components are important. (1) A "true" or "control" hydrologic simulation is generated with station-based meteorological forcing records and calibrated parameterization scheme. (2) Based on the "true" scenario, synthetic observations (surface SWC) are obtained by taking into account the observational error. (3) The ensemble of

Table 1
List of selected dynamic states.

Order	State variable	Definition	Level
1	SW_SOL	Amount of water stored in soil layer for each HRU (mm H_2O)	HRU×Nlay
2	SUR_Q	Amount of surface runoff contribution to stream flow in the main channel from the HRU (mm H_2O)	HRU
3	LAT_Q	Amount of lateral flow contribution to stream flow from the HRU (mm H_2O)	HRU
4	ET_A	Actual evapotranspiration (soil evaporation and plant transpiration) from the HRU (mm H_2O)	HRU

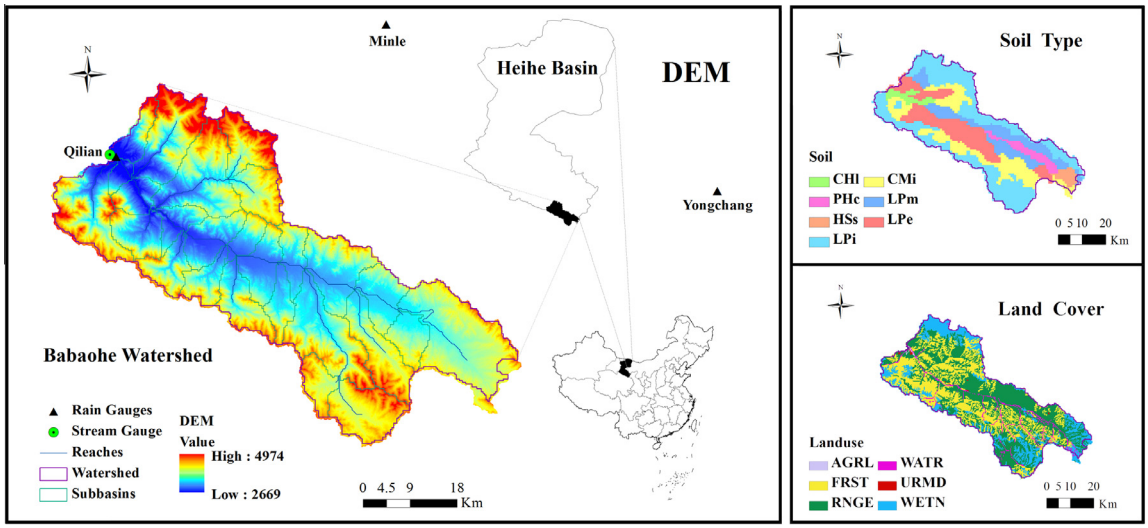

Fig. 1. Study area: the Babaohe watershed, northwest China.

Table 2
Land cover information of the Babaohe watershed.

Land cover classification		Area	
		(km^2)	Proportion (%)
RNGE	Range, grasses	1032.77	43.94
FRST	Mixed forest	798.16	33.96
WETN	Wetland, non-forested	426.72	18.15
WATR	Water	85.27	3.63
URMD	Residential, medium density	4.17	0.18
AGRL	Agricultural land	3.35	0.14

Table 3
Soil types of the Babaohe watershed.

Soil types		Area		Hydrologic soil group
		(km^2)	Proportion (%)	
LPi	Gelic leptosols	900.10	38.30	A
CMi	Gelic cambisols	443.85	18.88	C
LPe	Eutric leptosols	442.40	18.82	B
LPm	Mollic leptosols	349.42	14.87	B
PHc	Calcaric phaeozems	94.51	4.02	C
CHl	Luvic chernozems	68.69	2.92	D
HSs	Terric histosols	51.47	2.19	B

the meteorological forcing dataset is generated based on the WRF model simulation and then utilized to trigger the SWAT model. Within this "ensemble open loop" (EnOL) scenario, the hydrologic parameters are calibrated in a previous and well-warmed initial condition, which is achieved through a three-year spin-up period as the true scenario. (4) The EnKS, the key technique of combining the model forecast and the observations, is further improved with novel methods, such as the novel SLS-based inflation factor [61]. These approaches are capable of generating a more accurate background error covariance matrix and thus improving the estimates. Particularly, the SLS-based inflation factor and localization are diagnosed in this hydrologic data assimilation framework. The flowchart of the OSSE experiment is depicted in Fig. 2.

Specifically, the twin synthetic experiments are conducted to inspect how the assimilation of surface SWC improves the root-zone SWC and other hydrologic states with different implementations of the EnKS from June 1 to September 30, 2008. The original standard implementation of the EnKS (EnKS-O scenario) is first applied to examine its performance in hydrologic data assimilation. Moreover, the assimilation with improved or enhanced implementation of the EnKS via the SLS-based inflation factor and localization (EnKS-E) is also conducted. The comparison

Table 4
Description of the calibrated model parameters.

Order	Variable name	Definition	Level
1	ESCO	Soil evaporation compensation factor	HRU
2	CN2	Initial SCS runoff curve number for moisture condition II	HRU
3	CANMX	Maximum canopy storage	HRU
4	SOL_K	Saturated hydraulic conductivity	HRU
5	HRU_SLP	Average slope steepness	HRU
6	GWQWN	Threshold depth of water in the shallow aquifer required for return flow to occur	HRU
7	SOL_AWC	Available water capacity of the soil layer	HRU×Nlay
8	ALPHA_BF	Baseflow alpha factor	HRU
9	CH_K2	Effective hydraulic conductivity in main alluvium	SUBBASIN
10	EPCO	Plant uptake compensation factor	HRU
11	SURLAG	Surface runoff lag coefficient	BASIN
12	SOL_ALB	Moist soil albedo	HRU×Nlay
13	GW_DELAY	Groundwater delay time	HRU
14	CH_N2	Manning's "*n*" value for the main channel	SUBBASIN
15	GW_REVAP	Groundwater "revap" coefficient	HRU
16	SLSUBBSN	Average slope length	HRU

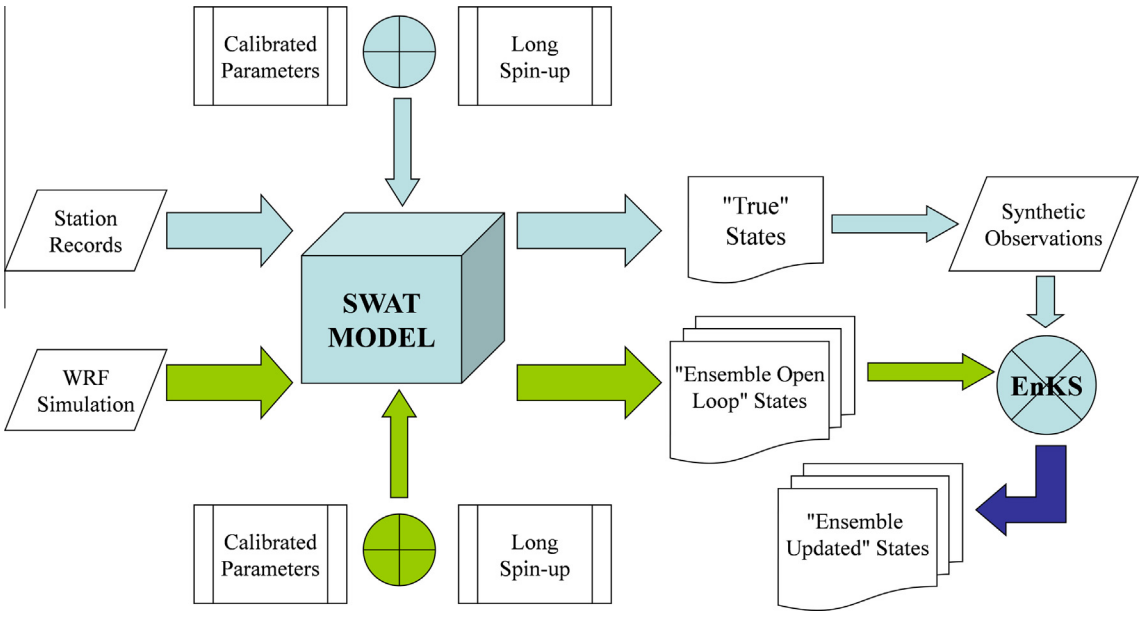

Fig. 2. Flowchart of the OSSE experiment.

between the twin experiments may be sufficient to investigate the capability and efficiency of the EnKS.

Within the iteration of the calculation of the inflation factor and the background error covariance matrix, the localization and boundary truncation are repeated to reduce the analysis error. This adjusts the analysis value to the minimum (10^{-6}) or maximum (saturated soil water content minus 10^{-6}) value whenever the updated soil moisture estimate exceeds the boundary limitation. This boundary truncation confines the final output to a reasonable range and eliminates the violation caused by numerical computation. In practical experiments, the objective function converges after 4–6 iterations; thus, the shift of mean values by one boundary truncation could be removed.

3.2.3. Evaluation metrics

To evaluate the performance of the assimilation, the time series of SWC and other representative hydrological states obtained from the true, ensemble open loop, and assimilation scenarios are compared to each other. Two metrics are chosen, root-mean-square error (RMSE) and mean bias error (MBE):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(S_i - T_i)^2}{N}} \tag{16}$$

$$MBE = \frac{\sum_{i=1}^{N}(S_i - T_i)}{N} \tag{17}$$

where T_i and S_i represent the ith value of true and simulated (ensemble open loop or assimilation scenario) hydrologic variables, respectively, and N is the total number.

To measure the magnitude of the improvement, a normalized error reduction (NER) index, which has been used in previous research [6], is introduced:

$$NER = 1 - \frac{RMSE_a}{RMSE_o} \tag{18}$$

where $RMSE_o$ and $RMSE_a$ represent the root-mean-square errors of the ensemble open loop and assimilation scenarios, respectively. NER varies from negative infinity to 1.0, with 1.0 representing the analysis results identical to the true value after assimilation. A negative NER means deterioration in the assimilation after the open loop.

Moreover, a typical measure, called the normalized root-mean-square-error ratio (NRR) [41], is used to investigate the quality of the generated ensemble in this ensemble assimilation and analysis framework.

$$NRR = \sqrt{\frac{2N}{N+1}} * \frac{RMSE^T}{RMSE^N} \tag{19}$$

where $RMSE^T$ and $RMSE^N$ represent the time-averaged RMSE of the ensemble mean and the mean RMSE of the ensemble members, respectively, T is the total time step, and N is the ensemble size. The identical value of NRR is 1, $NRR > 1$ indicates too narrow a spread, and $NRR < 1$ indicates an ensemble that is too wide.

3.3. Implementation of EnKS into SWAT

3.3.1. Specifications of the forcing data

To a large degree, the accuracy of the model output is dependent upon the quality of the input datasets, including their spatial and temporal resolution [2,3]. Among the datasets, precipitation is one of the most important because of its role in determining hydrologic processes between the surface and vertical soil layers [54,55]. The rainfall data are often obtained from rain gauge stations, which generally capture only a fraction of the true precipitation. In this paper, the precipitation is assumed to have a lognormal distributed error following the criteria of Moradkhani and Meskele [41], which assumes a heteroscedastic error variance:

$$\mu_{\ln P} = \ln\left(\frac{P_t^2}{\sqrt{P_t^2 + (\alpha_p \cdot P_t)^2}}\right) \tag{20}$$

$$\sigma_{\ln P} = \sqrt{\ln\left(\frac{(\alpha_p \cdot P_t)^2}{P_t^2} + 1\right)} \tag{21}$$

$$P_t^i = \exp(\mu_{\ln P} + \varphi_{P,i} \cdot \sigma_{\ln P}) \tag{22}$$

where P_t and P_t^i are the observed and perturbed precipitation at time t, respectively. The log transformation of P_t^i is a Gaussian distribution with mean, $\mu_{\ln P}$ and standard deviation, $\sigma_{\ln P}$. α_P is the variance scaling factor of precipitation with a set value of 0.5, and $\varphi_{P,i}$ is a standard normally distributed random number, with i varying from 1 to 50 because the ensemble size is set to 50.

The ensemble of temperature (maximum and minimum temperatures) is generated as follows:

$$T_t^i = T_t + \alpha_T \cdot \varphi_{T,i} \tag{23}$$

where T_t and T_t^i are the observed and perturbed temperature at time t, respectively, $\varphi_{T,i}$ is a standard normally distributed random number with a homoscedastic error assumption, and α_T is the variance scaling factor of temperature with a value of 2.0.

3.3.2. Specifications of the observations

The aggregated daily meteorological forcing datasets by the WRF simulation are employed to generate the "ensemble open loop" scenario. Meanwhile, the simulation with forcing based on station measurements is treated as the "true" or "control" scenario. The synthetic observations (surface SWC) are generated from the "true" scenario by adding random perturbation:

$$z_o = z \cdot (1 + \alpha_z \cdot \varphi_z) \tag{24}$$

where z_o represents the synthetic surface SWC observations, z is the "true state" obtained from the true scenario, and φ_z is a normally distributed random number with the observational variance scaling factor, α_z, and 0.1 is assigned to this scaling factor.

4. Results and discussion

To illustrate the capability of the EnKS to reduce the influences of inaccurate meteorological data and poor initial conditions, two assimilation experiments are conducted to analyze the impact of assimilating surface soil moisture on root-zone soil moisture and other primary hydrologic states under different configurations during the period from June 1 to September 30, 2008. In the following parts, the simulation results forced with station-based meteorological datasets are referred to as "TRUE". The ensemble mean results of the open loop simulation with WRF simulation as input are referred to as "EnOL". The ensemble mean values of the data assimilation via the original standard EnKS are referred to as "EnKS-O". Moreover, the improved EnKS with the inflation factor and localization is also applied to further improve estimation and is referred to as "EnKS-E". Unless otherwise noted, the length scale of the localization is set to 10.

4.1. Soil water content and predictor hydrologic variables estimation

The watershed area-averaged SWC of the five soil layers within TRUE, EnOL, EnKS-O, and EnKS-E are demonstrated in Fig. 3. In general, soil water content is slightly overestimated in the EnOL scenario, despite a relatively low amount of total precipitation. Compared to the EnOL scenario, both the original standard EnKS and the improved EnKS with the inflation factor and localization have improved the SWC estimation in the upper three layers (0–30 cm; Fig. 3a–c). Particularly in the EnKS-E scenario, the soil moisture estimation maintains good agreement with the true value (Fig. 3a). The error statistics are listed in Table 5. For the surface SWC estimation, the RMSE (MBE) is 1.19 (0.27) for EnOL, 0.90 (−0.43) for EnKS-O, and 0.29 (−0.01) for EnKS-E. Moreover, the

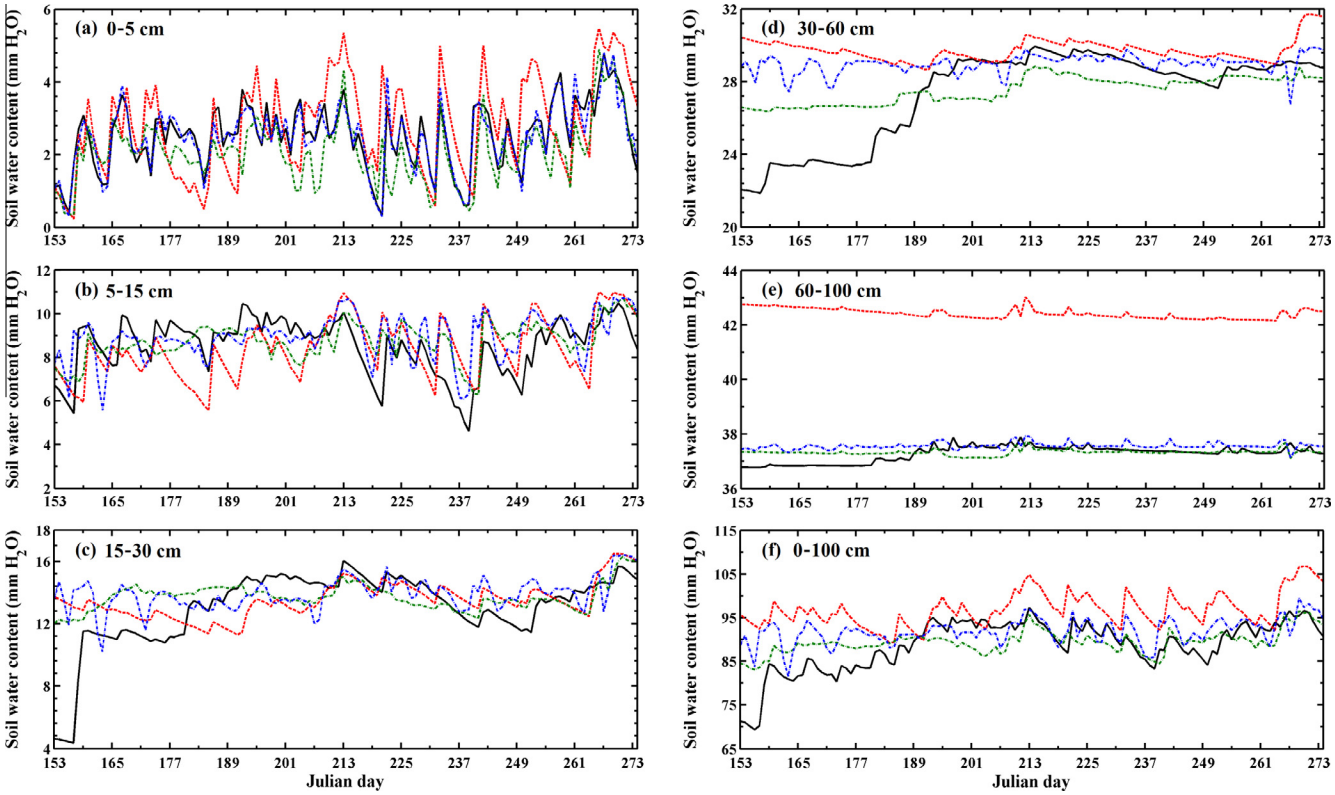

Fig. 3. Comparison of the area-averaged SWC between the TRUE (black solid line), the EnOL (red dashed line), the EnKS-O (green dash-dot line), and the EnKS-E (blue dash-dot line) scenarios from June 1st to September 30th, 2008: (a) 0–5 cm, (b) 5–15 cm, (c) 15–30 cm, (d) 30–60 cm, (e) 60–100 cm, and (f) 0–100 cm (unit: mm H_2O). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 5
Error statistics of the hydrologic variables between the results with ensemble simulation/assimilation and TRUE during June 1st to September 30th, 2008. Here, EnOL, EnKS-O, and EnKS-E represent the ensemble mean results of the open loop simulation, the assimilation with the standard EnKS, and the assimilation with the enhanced EnKS, respectively.

Metric		Variables								
		SW(1)	SW(2)	SW(3)	SW(4)	SW(5)	SW	ET	SURQ	LATQ
NRR	EnOL	1.38	1.27	1.21	1.35	1.40	1.40	1.37	1.39	1.34
	EnKS-O	1.34	1.24	1.20	1.33	1.36	1.34	1.40	1.15	1.32
	EnKS-E	1.04	1.23	1.23	1.30	1.33	1.27	1.40	1.13	1.32
RMSE	EnOL	1.19	1.55	2.28	3.50	5.15	9.77	1.16	1.07	0.78
	EnKS-O	0.90	1.21	2.12	1.95	0.30	4.68	1.22	0.48	0.71
	EnKS-E	0.29	1.24	2.41	2.89	0.39	5.81	1.32	0.53	0.71
RMSE-A	EnOL	1.40	2.08	3.50	4.41	5.40	12.20	1.29	1.11	0.83
	EnKS-O	1.50	2.46	4.59	3.81	0.67	8.85	1.76	0.56	0.86
	EnKS-E	0.34	1.85	3.65	3.60	0.58	7.60	1.47	0.55	0.75
MBE	EnOL	0.27	−0.10	0.43	2.42	5.14	8.15	−0.85	0.09	−0.11
	EnKS-O	−0.43	0.25	0.64	0.22	0.05	0.73	−0.78	−0.01	−0.40
	EnKS-E	−0.01	0.41	0.85	1.65	0.28	3.19	−0.94	−0.01	−0.16

overestimation in the fifth layer is significantly reduced, and thus promising profile SWC estimations are achieved in both scenarios (Fig. 3e and f). Compared to EnOL, the RMSE is reduced to 4.68 for EnKS-O and 5.81 for EnKS-E. The reduced RMSE values are observed in all soil layers and demonstrate the positive effects of both approaches.

The comparisons between the watershed-averaged SWC estimations achieved through different implementations of the EnKS seem to exhibit that the enhanced approach has a non-dominated advantage over the original standard version. Except for the surface soil layer, the RMSE value of the EnKS-E is larger than the corresponding value of the EnKS-O. However, the area-weighted RMSE (RMSE-A) of all the HRUs in the watershed is smaller in the EnKS-E scenario, indicating a steady and smooth performance with improved estimation over the study basin. For example, the area-weighted RMSE of the surface (profile) SWC for EnOL, EnKS-O, and EnKS-E scenarios is 1.40 (12.20), 1.50 (8.85), and 0.34 (7.60), respectively. Within the implementation of the standard EnKS, overestimation and underestimation in different HRUs mutually compensates each other, leading to a deceitfully better estimation. Actually, the ensemble of the forecast states is too narrow with a NRR larger than 1 for all variables. The background error covariance matrix is underestimated, leading to an inferior performance of the EnKS. Nevertheless, the SLS-based inflation factor provides a moderate multiplicative factor to account for the limited ensemble size and the inaccurate model error estimation. Further improvements are obtained by filtering out the redundant and improper remote observations within the implementation of localization.

Because the SWC plays a crucial role in determining the land phase of the hydrologic cycle, the changes in the SWC also subsequently influence other hydrologic components, such as evapotranspiration, surface runoff, and lateral flow. The estimates of important hydrological variables within the different scenarios are presented in Fig. 4. The EnOL scenario underestimates these hydrologic variables, especially the evapotranspiration and the lateral flow with negative MBE. Notwithstanding the promising results observed in the SWC estimation, only trivial improvements exist in surface runoff and lateral flow. Moreover, the increased RMSE values of the evapotranspiration for both approaches reveal a destructive effect of the data assimilation. The underestimation of evapotranspiration still persists even after the assimilation of surface SWC. The RMSE (MBE) is 1.16 (−0.85) for the EnOL, 1.22 (−0.78) for the EnKS-O, and 1.32 (−0.94) for the EnKS-E. As for the surface runoff, a remarkable overestimation of peak values emerges within the EnOL. Despite the correction of the absolute peak values, the lag time between different rainfall events has barely improved. Meanwhile, the lateral flow is slightly improved with reductions in the RMSE from 0.78 for EnOL to 0.71 for the two approaches. For the predictor hydrologic variables, the assimilation of surface soil moisture has few improvements over the open loop simulation. The information contained in the surface soil moisture is insufficient to impact these hydrologic variables. Further enhancements can be achieved by including other data sources, such as in situ streamflow records [7].

From the ensemble quality point of view, the EnOL scenario has a quite narrow spread, indicating an overconfident estimation of the forcing datasets and the parameter uncertainties with NERs larger than 1. The NRRs in Table 5 demonstrate the quality of the ensemble for different scenarios and variables. With the implementation of standard and enhanced EnKS models, the NRR is reduced. In particular, the enhanced EnKS with the inflation factor and localization, which adjust the background error covariance matrix, is better at characterizing the ensemble spread. Therefore, the NRR of the surface SWC is reduced from 1.38 to 1.34 and 1.04 for the EnKS-O and EnKS-E scenarios, respectively. For evapotranspiration and lateral flow, the improvements are trivial.

4.2. The spatial variation of the hydrologic variables

The temporal evolution of the SWC is improved by the assimilation of the surface SWC into the SWAT model, especially with the implementation of the enhanced EnKS with the inflation factor and localization. This section illustrates the impacts of the spatially heterogeneous input information, including meteorological data, soil type and land cover, on the performance of the standard and enhanced EnKSs. Usually, an HRU map is created by overlaying a land cover map on a soil map before the SWAT simulation and assimilation, to demonstrate the spatial variation of improvement. The time-averaged RMSE maps of the surface and profile SWCs, the evapotranspiration and the lateral flow are shown in Fig. 5. The spatially distributed RMSE map effectively characterizes the spatial variety of improvement and exposes the factors that influence the performance.

The results demonstrate, through smaller RMSE values, that the enhanced EnKS in general outperforms the original standard implementation of EnKS. Compared to the surface SWC in EnOL (Fig. 5a), the EnKS-O (Fig. 5b) reveals a slight change in the southeast. For other parts of the watershed, no obvious improvements or even deteriorations emerge. Due to the mechanism of the SWAT model, the rain gauge closest to the subbasin is assigned to the trigger model. However, due to the limited availability of the original data, only three rain stations are utilized in this work, which may be inadequate [54]. Nonetheless, the uncertainty introduced

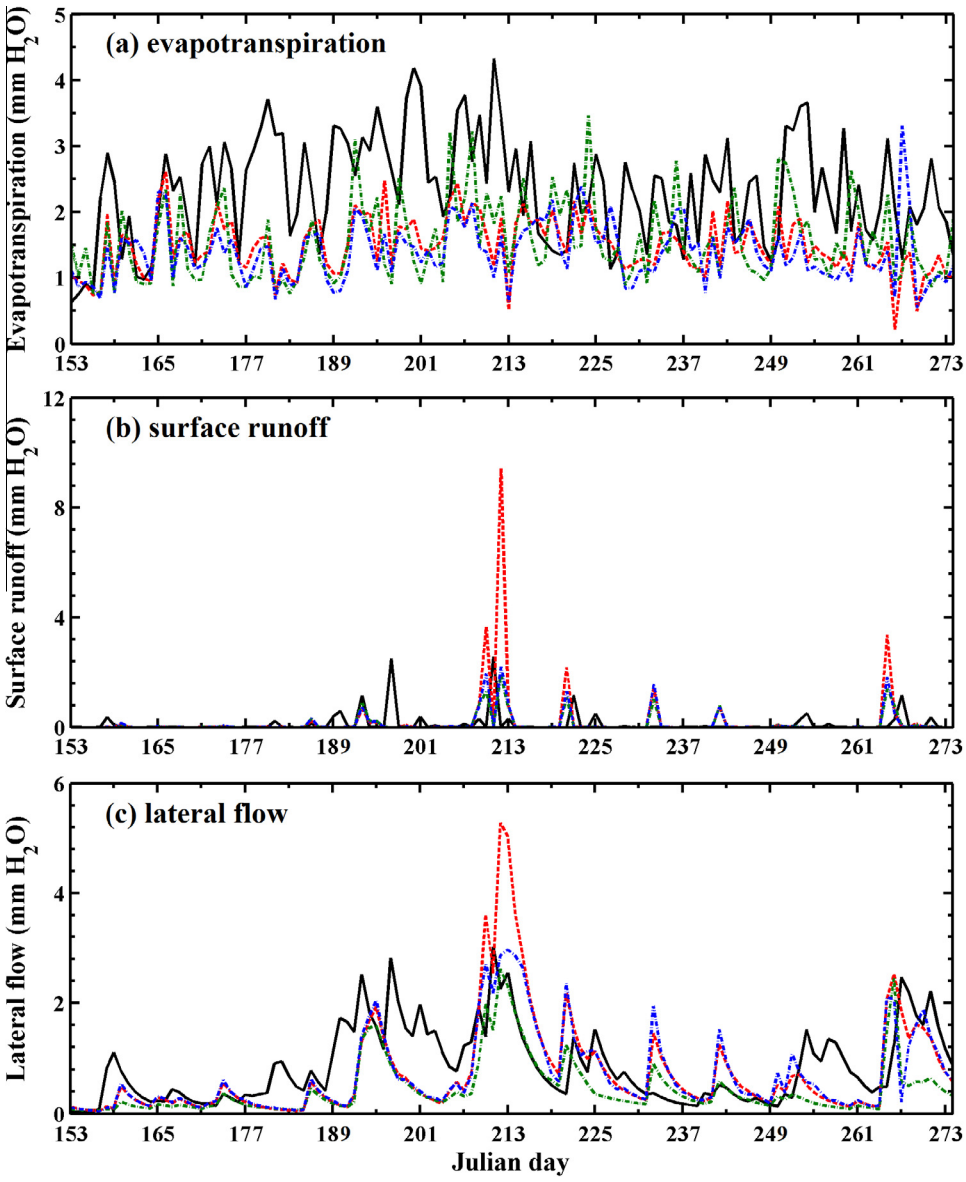

Fig. 4. Comparison of area-averaged predictor hydrologic variables between the TRUE (black solid line), the EnOL (red dashed line), the EnKS-O (green dash-dot line), and the EnKS-E (blue dash-dot line) scenarios from June 1st to September 30th, 2008: (a) evapotranspiration, (b) surface runoff, (c) lateral flow (unit: mm H_20). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

by meteorological forcing is greatly eliminated by the enhanced EnKS, as shown in Fig. 5c. As for the profile soil water content, both approaches have reduced the estimation error. Obviously, much more promising and steady improvements of the profile SWC are achieved in the EnKS-E scenario (Fig. 5f). The background error covariance matrix is inflated and better characterized through localization. However, the RMSE of evapotranspiration over the entire watershed have increased following the assimilation of the surface SWC in both implementations.

An interesting phenomenon emerges with the estimation of the lateral flow: the margins of the watershed have relatively larger errors. The distribution of the RMSE correlates with the soil type map (Fig. 1), signifying a relationship between the estimation of lateral flow and the physical soil characteristics. The significant impacts from the inaccurate forcing data observed in the surface and profile soil layers and the evapotranspiration (Fig. 5a, d, and g) have disappeared. Therefore, we can conclude that the lateral flow is significantly influenced by the original soil physics instead of by the extrinsic forcing datasets, such as precipitation.

4.3. The influence of land cover and soil type

The land cover and soil type determine the relevant parameters and physical properties that govern the movement of water and air throughout the soil profile [45], and thus affect the estimation of important hydrologic variables. The different soil hydrologic groups imply distinct infiltration characteristics. Group A, which is classified by LPi, indicates a high infiltration rate and water transmission, even when thoroughly wetted. Correspondingly, groups B, C, and D have moderate, low and very low infiltration rates, respectively. A high infiltration rate indicates a strong vertical correlation, and vice versa. The type of land cover, which plays a prominent role in determining the value of the CANMX parameter, significantly influences the infiltration, surface runoff and evapotranspiration. Technically, both land cover and soil type affect the behavior of the EnKS [11]. The investigations below are based on the EnKS-E scenario. Because the area proportion varies between different land cover and soil types, the diagnosis of the magnitudes is not easy. Fig. 6 shows the box and whisker plots

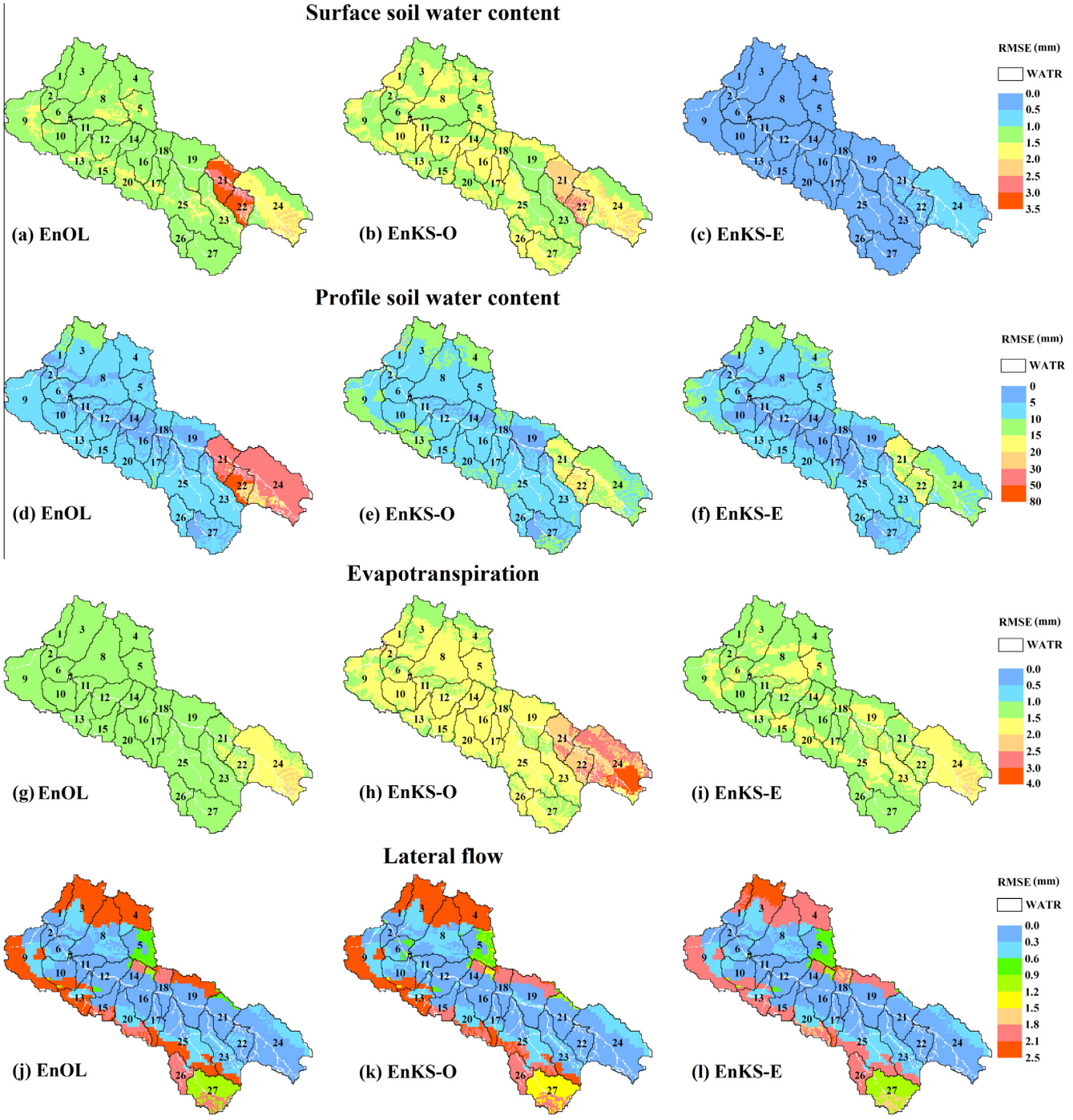

Fig. 5. Time-averaged RMSE distribution maps of the surface SWC, the profile SWC, the evapotranspiration, and the lateral flow from June 1st to September 30th, 2008. Numbers on each map indicate the subbasin number.

of the NER (normalized error reduction) under various conditions. In terms of the land cover type (Table 2), WETN (Fig. 6a), the third largest area of land cover, demonstrates a broad distribution of NERs, indicating large differences in the improvements. Due to the many factors that influence the hydrologic processes of wetlands, the assimilation method behaves arbitrarily across the watershed. RNGE (Fig. 6a, c, and e), comprising the majority of the land cover, exhibits a relatively smaller distribution for the surface and profile SWCs. Simultaneously, the statistics reveal a greater improvement in RNGE (higher median value) and a limited correction of FRST (lower median value). Not only is the median of FRST small in comparison to RNGE but a wide-ranging distribution also exists, implying unstable improvements. The spatially heterogeneous land cover affects the physics of the hydrologic processes and thus influences the ensemble spread of the EnOL/EnKS-E scenario.

As for the soil type (Table 3), the improvements of the surface and profile SWCs under different soil types are varied, as shown in Fig. 6b and d. The PHc demonstrates a wide spread with small area proportion for surface SWC. However, the NERs of most soil types for the profile soil layer range from zero to 0.8, indicating that factors influencing the final estimation are variable. The soil

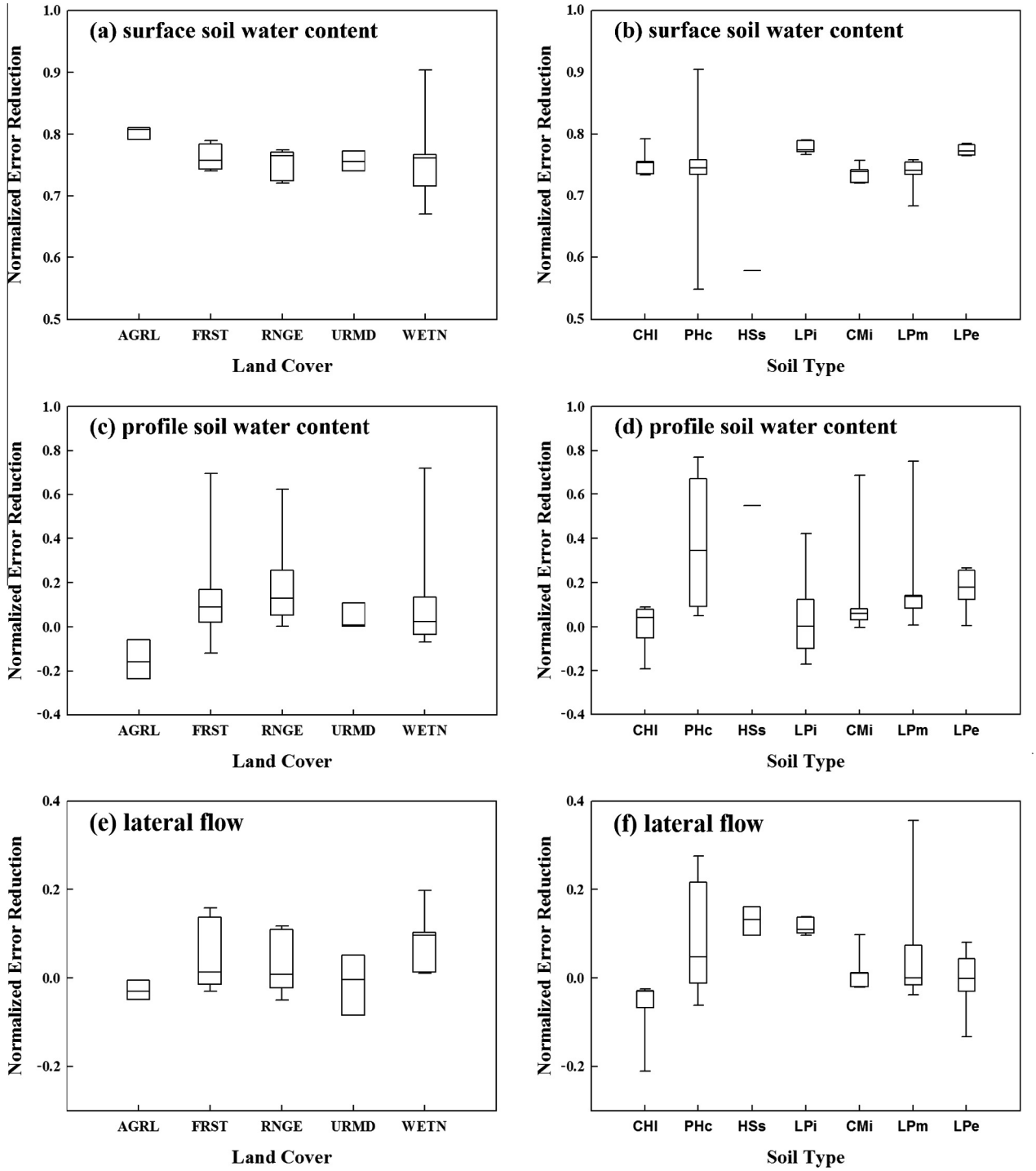

Fig. 6. Box and whisker plots of NER of the surface SWC (first row), the profile SWC (second row), and lateral flow (third row) in the EnKS-E under differing conditions: land cover (left) and soil types (right). Higher NERs indicate a greater improvement upon assimilation. The upper and lower boundaries of the box indicate the 25th percentile and 75th percentile, respectively, the line within the box marks the median value, and the whiskers above and below the box indicate the 90th and 10th percentiles.

type influences the subsurface hydrology, including infiltration and lateral flow. The statistics show an obvious correlation between different soil types and the degrees of improvement of lateral flow. The lateral flow in the profile soil layer (Fig. 6f) is significantly influenced by the infiltration rate, with a greater improvement in LPi and a moderate improvement in CHl. The vertical coupling is strengthened in soil types with a high infiltration rate; thus, the lateral flow is moderately improved through the assimilation of the surface SWC.

4.4. Influence of adjusting the forecast error covariance matrix

Despite the explanations given above, there are other factors that influence the behavior of the EnKS in the hydrologic data assimilation, such as the physics of model [10], the ensemble size and the observational error estimation [42]. Therefore, some specific investigations are needed to better characterize the performance of this approach. In this section, we focus on the sole impact with the SLS-based inflation factor, or localization. Further,

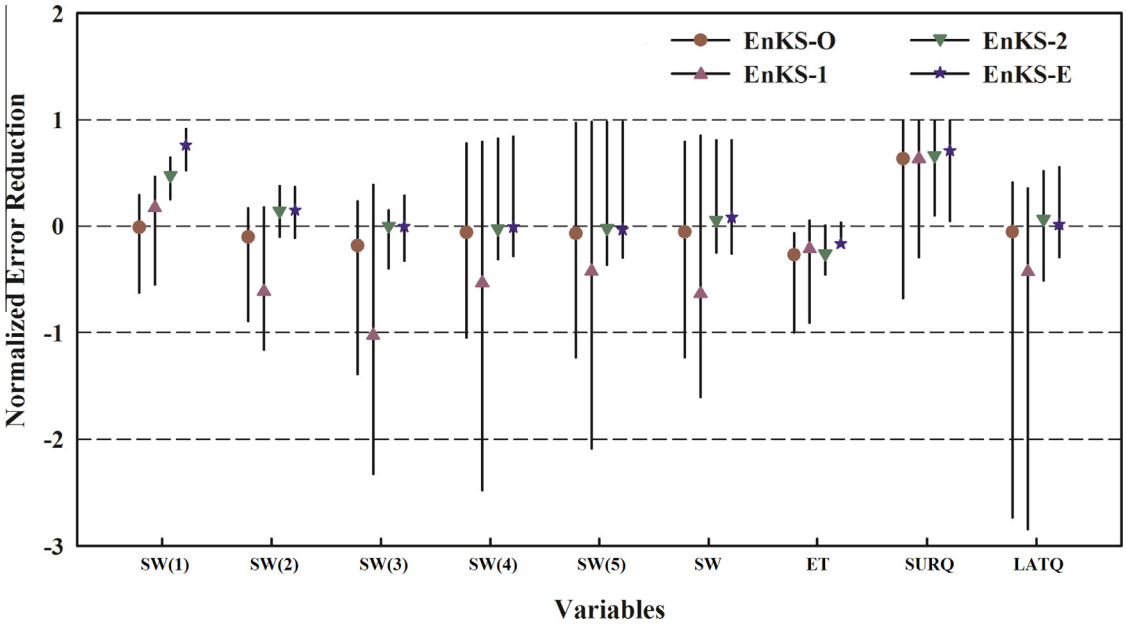

Fig. 7. Effects of the adjusting methods on the estimation of different variables.

the length scale of localization is also diagnosed to examine its effect on various states.

Fig. 7 shows the range and median values of NERs for all HRUs in four scenarios, i.e., the original standard EnKS (EnKS-O), the improved EnKS with inflation factor only (EnKS-1), the enhanced EnKS with localization only (EnKS-2), and an implementation of the EnKS with both approaches (EnKS-E). In the four scenarios, the surface SWCs are improved, with median values larger than zero. Additionally, a larger median value for the surface SWC is observed in the EnKS-E scenario. Significant improvements are achieved through the inflation factor, evidenced by increased median values (surface SWC), and stable performances are achieved via the localization, with a small range for all variables. Despite the fact that the inflation factor approach degraded the estimate of the SWCs in deep layers with decreased NER median values, the estimates are improved by the benefits of localization. Specifically, a narrow distribution of NERs in the EnKS-2 and EnKS-E reveal that the EnKS performs steadily when localization is applied.

Overall, the best performance of the EnKS emerges when both the SLS-based inflation factor and the localization are applied. The SLS-based inflation factor and the localization are both necessary in this hydrologic data assimilation framework to improve the final results. In our synthetic experiments, the bias correction proposed by De Lannoy et al. [9] was also utilized. However, the performance of the EnKS was poorer, and thus this method is not demonstrated here.

The length scale, the only parameter in the localization technique, plays an important role in determining the correlation function and thus the covariance matrix. Therefore, we examine the performance of the EnKS and the improvement of the state estimates with variable length scales, in which the length scale changes from 1 to 30 while the other conditions are held constant. Fig. 8 shows the varying normalized RMSE values of the different states with the increasing length scale. The curves change moderately with varying influential length scales. However, the trends in the estimation errors for the upper three layers and the deeper layers remain different with different length scales. The estimation error of the surface SWC increases with the increasing of length scale, which indicates that the surface SWC is mainly affected by the current observation. In particular, the RMSEs of the second and third soil layers decrease with the increasing length scale initially. After remaining constant for a period, the estimation values increase again. However, the fourth and root-zone soil layers have improved results with the increasing length scale. The response to the influential length scale varies with different soil layers. For specific application, the length scale should be chosen carefully. A short length scale should be adopted when the accuracy of the surface soil moisture is under specific consideration. When a larger length scale is utilized, the root-zone soil moisture is improved and the subsurface hydrology is better characterized. To counterbalance the improvements of the adjacent soil layers, a medium length scale must be adopted. In our study, the performance of the surface SWC is the main focus, and thus a length scale of 10 is chosen.

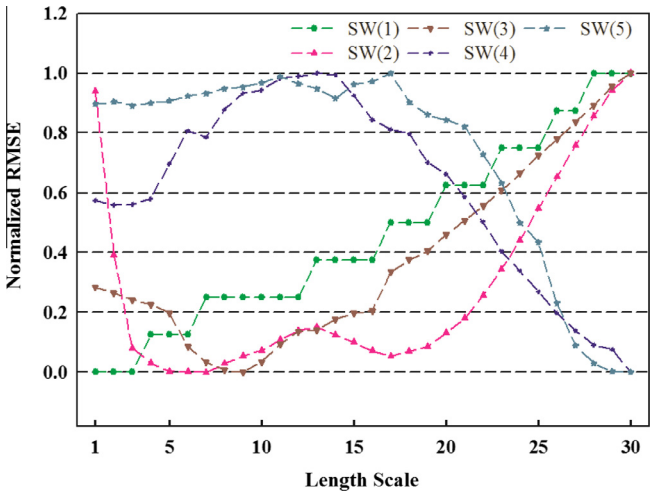

Fig. 8. Normalized RMSEs for different soil layers as a function of the length scale.

5. Conclusions

In this paper, synthetic experiments were conducted to investigate the influence of assimilating the surface SWC via the EnKS on other primary hydrologic variables, implemented in a

semi-distributed hydrologic model (SWAT). Three scenarios were conducted for each experiment: the TRUE scenario, the EnOL scenario, and the EnKS scenario. Specifically, the original and enhanced implementations of EnKS were utilized. The experiments were implemented in different situations to investigate the various factors that might influence the performance of the EnKS, including spatially heterogeneous input data and different methods of adjusting the forecast error covariance matrix. Our synthetic experiments demonstrated that the EnKS could effectively update the profile SWC and other predictor variables by assimilating surface SWC with calibrated static hydrologic parameters. The understanding of the model structure was very important because it greatly influences the outcome of the data assimilation. The EnKS behaved promisingly with other adjusting approaches, such as the SLS-based inflation factor and localization. Specifically, the localization improved the final results with a more stable performance. However, the length scale of the localization also influenced the estimation of different soil layers. In practical applications, a counterbalance between the surface and root-zone SWC estimates should be taken into account. Approaches that improve the estimation of the background error matrix are vital to account for the limited ensemble size and inaccurate model error specifications, and these approaches are crucial in the hydrologic data assimilation framework. Moreover, the performance of the EnKS varied across the watershed due to the spatial variation of the soil and land cover. The soil types with higher infiltration had a more significant improvement, especially in terms of the profile soil layer and the lateral flow, and the EnKS performed reasonably well under the RNGE land cover type. The results also showed the influence of inaccurate meteorological forcing dataset on the outcome.

Our analysis was subject to several factors that could be addressed in the future. First, we need to examine the performance of the EnKS and these adjustment approaches under real circumstances. In operational hydrologic forecasting, the performance of the data assimilation may be reduced due to the unknown uncertainties of the observation and model error specifications, the measurement availability, the time-varying hydrologic parameters and inaccurate forcing datasets [37]. Lately, the utilization of the variable variance multiplier as a flexible adjustment of the ensemble spread has gained popularity in handling fluctuating uncertainties [32,43] and could be further extended to future endeavors. Meanwhile, the effect of spatially heterogeneous input data should be diagnosed carefully with different data sources and circumstances. Finally, the exploration of the improved smoother methods, such as the ensemble moving batch smoother [13] is necessary, as well as the introduction of the MCMC approach into the smoothing framework [43]. Our ongoing research will focus mainly on these issues.

Acknowledgments

This work is supported by the National Science Foundation of China (NSFC) under Grant 91325106, the Major State Basic Research Development Program under Grant 2011CB707103, the NSFC under Grant 41271358, and the One Hundred Person Project of the Chinese Academy of Sciences under Grant 29Y127D01. The authors thank the anonymous reviewers for their constructive comments to improve the quality and clarity of this manuscript.

References

[1] Arnold JG, Srinivasan R, Muttiah RS, Williams JR. Large area hydrologic modeling and assessment part I: model development 1. J Am Water Resour Assoc 1998;34(1):73–89. <http://dx.doi.org/10.1111/j.1752-1688.1998.tb05961.x>.

[2] Bárdossy A, Das T. Influence of rainfall observation network on model calibration and application. Hydrol Earth Syst Sci 2008;12(1):77–89. <http://dx.doi.org/10.5194/hess-12-77-2008>.

[3] Bormann H. Impact of spatial data resolution on simulated catchment water balances and model performance of the multi-scale TOPLATS model. Hydrol Earth Syst Sci 2006;10(2):165–79. <http://dx.doi.org/10.5194/hess-10-165-2006>.

[4] 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 Remot Sens 2012;50(7):2542–55. <http://dx.doi.org/10.1109/TGRS.2011.2177468>.

[5] Chaubey I, Cotter A, Costello T, Soerens T. Effect of DEM data resolution on SWAT output uncertainty. Hydrol Process 2005;19(3):621–8. <http://dx.doi.org/10.1002/hyp.5607>.

[6] 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(4):526–36. <http://dx.doi.org/10.1016/j.advwatres.2011.01.011>.

[7] Clark MP, Rupp DE, Woods RA, Zheng X, Ibbitt RP, Slater AG, et al. Hydrological data assimilation with the ensemble Kalman filter: use of streamflow observations to update states in a distributed hydrological model. Adv Water Resour 2008;31(10):1309–24. <http://dx.doi.org/10.1016/j.advwatres.2008.06.005>.

[8] Das N, Mohanty B, Cosh M, Jackson T. Modeling and assimilation of root zone soil moisture using remote sensing observations in Walnut Gulch watershed during SMEX04. Remote Sens Environ 2008;112(2):415–29. <http://dx.doi.org/10.1016/j.rse.2006.10.027>.

[9] De Lannoy GJM, Reichle RH, Houser PR, Pauwels VRN, Verhoest NEC. Correcting for forecast bias in soil moisture assimilation with the ensemble Kalman filter. Water Resour Res 2007;43(9):W09410. <http://dx.doi.org/10.1029/2006WR005449>.

[10] 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(4):W04518. <http://dx.doi.org/10.1029/2011WR011011>.

[11] Draper C, Reichle R, Lannoy GD, Liu Q. Assimilation of passive and active microwave soil moisture retrievals. Geophys Res Lett 2012;39(4):L04401. <http://dx.doi.org/10.1029/2011GL050655>.

[12] Dumedah G, Coulibaly P. Evolutionary assimilation of streamflow in distributed hydrologic modeling using in-situ soil moisture data. Adv Water Resour 2013;53:231–41. <http://dx.doi.org/10.1016/j.advwatres.2012.07.012>.

[13] Dunne S, Entekhabi D. An ensemble-based reanalysis approach to land data assimilation. Water Resour Res 2005;41(2):W02013. <http://dx.doi.org/10.1029/2004WR003449>.

[14] Entekhabi D, Nakamura H, Njoku EG. Solving the inverse problem for soil moisture and temperature profiles by sequential assimilation of multifrequency remotely sensed observations. IEEE Trans Geosci Remot Sen 1994;32(2):438–48. <http://dx.doi.org/10.1109/36.295058>.

[15] Entekhabi D, Njoku EG, O'Neill PE, Kellogg KH, Crow WT, Edelstein WN, et al. The soil moisture active passive (SMAP) mission. PIEEE 2010;98(5):704–16. <http://dx.doi.org/10.1109/JPROC.2010.2043918>.

[16] Evensen G. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J Geophys Res 1994;99(C5):10143–62. <http://dx.doi.org/10.1029/94JC00572>.

[17] Evensen G. The ensemble Kalman filter: theoretical formulation and practical implementation. Ocean Dyn 2003;53(4):343–67. <http://dx.doi.org/10.1007/s10236-003-0036-9>.

[18] Evensen G, Van Leeuwen PJ. An ensemble Kalman smoother for nonlinear dynamics. Mon Weather Rev 2000;128(6):1852–67. <http://dx.doi.org/10.1175/1520-0493(2000)128<1852:AEKSFN>2.0.CO;2>.

[19] Nachtergaele F, Velthuizen H, Verelst L, Batjes N, Dijkshoorn J, Engelen VWP, et al. Harmonized world soil database (version 1.1). Rome, Italy: FAO; IIASA: Laxenburg, Austria; 2009.

[20] Gaspari G, Stephen EC. Construction of correlation functions in two and three dimensions. Quart J Roy Meteorol Soc 1999;125(554):723–57. <http://dx.doi.org/10.1002/qj.49712555417>.

[21] Grønnevik R, Evensen G. Application of ensemble-based techniques in fish stock assessment. Sarsia 2001;86(6):517–26. <http://dx.doi.org/10.1080/00364827.2001.10420490>.

[22] Hamill TM, Jeffrey SW, Chris S. Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. Mon Weather Rev 2001;129(11):2776–90. <http://dx.doi.org/10.1175/1520-0493(2001)129<2776:DDFOBE>2.0.CO;2>.

[23] Han E, Merwade V, Heathman GC. Implementation of surface soil moisture data assimilation with watershed scale distributed hydrological model. J Hydrol 2012;416(417):98–117. <http://dx.doi.org/10.1016/j.jhydrol.2011.11.039>.

[24] Han XJ, Li X, Franssen HH, Vereecken H, Montzka C. Spatial horizontal correlation characteristics in the land data assimilation of soil moisture. Hydrol Earth Syst Sci 2012;16(5):1349–63. <http://dx.doi.org/10.5194/hess-16-1349-2012>.

[25] Heathman GC, Cosh MH, Han E, Jackson TJ, McKee L, McAfee S. Field scale spatiotemporal analysis of surface soil moisture for evaluating point-scale in situ networks. Geoderma 2012;170:195–205. <http://dx.doi.org/10.1016/j.geoderma.2011.11.004>.

[26] Houtekamer PL, Mitchell HL. A sequential ensemble Kalman filter for atmospheric data assimilation. Mon Weather Rev 2001;129(1):123–37.

[27] Huang CL, Li X, Lu L, Gu J. Experiments of one-dimensional soil moisture assimilation system based on ensemble Kalman filter. Remote Sens Environ 2008;112(3):888–900. <http://dx.doi.org/10.1016/j.rse.2007.06.026>.
[28] Jackson T, Schmugge J, Engman E. Remote sensing applications to hydrology: soil moisture. Hydrolog Sci J 1996;41(4):517–30. <http://dx.doi.org/10.1080/02626669609491523>.
[29] Kerr YH, Waldteufel P, Wigneron JP, Delwart S, Cabot F, Boutin J, et al. The SMOS mission: new tool for monitoring key elements ofthe global water cycle. PIEEE 2010;98(5):666–87. <http://dx.doi.org/10.1109/JPROC.2010.2043032>.
[30] Kumar SV, Reichle RH, Koster RD, Crow WT, Peters-Lidard CD. Role of subsurface physics in the assimilation of surface soil moisture observations. J Hydrometeorol 2009;10(6):1534–47. <http://dx.doi.org/10.1175/2009JHM1134.1>.
[31] Lee H, Seo D, Liu Y, Koren V, McKee P, Corby R. Variational assimilation of streamflow into operational distributed hydrologic models: effect of spatiotemporal adjustment scale. Hydrol Earth Syst Sci Discuss 2012;9:93–138. <http://dx.doi.org/10.5194/hessd-9-93-2012>.
[32] Leisenring M, Moradkhani H. Analyzing the uncertainty of suspended sediment load prediction using sequential data assimilation. J Hydrol 2012;468–469:268–82. <http://dx.doi.org/10.1016/j.jhydrol.2012.08.049>.
[33] Li X, Huang CL, Che T, Jin R, Wang SG, Wang JM, et al. Development of a Chinese land data assimilation system: its progress and prospects. Prog Nat Sci 2007;17(8):881–92. <http://dx.doi.org/ 10.1080/10002007088537487>.
[34] Li Z, Xu Z, Shao Q, Yang J. Parameter estimation and uncertainty analysis of SWAT model in upper reaches of the Heihe River Basin. Hydrol Process 2009;23(19):2744–53. <http://dx.doi.org/10.1002/hyp.7371>.
[35] Li F, Crow WT, Kustas WP. Towards the estimation root-zone soil moisture via the simultaneous assimilation of thermal and microwave soil moisture retrievals. Adv Water Resour 2010;33(2):201–14. <http://dx.doi.org/10.1016/j.advwatres.2009.11.007>.
[36] Liu J, Liu M, Tian H, Zhuang D, Zhang Z, Zhang W, et al. Spatial and temporal patterns of China's cropland during 1990–2000: an analysis based on Landsat TM data. Remote Sens Environ 2005;98(4):442–56. <http://dx.doi.org/10.1016/j.rse.2005.08.012>.
[37] Liu Y, Weerts A, 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 Discuss 2012;9:3415–72. <http://dx.doi.org/10.5194/hess-16-3863-2012>.
[38] Winchell M, Srinivasan R, Luzio MD, Arnold J. ArcSWAT interface for SWAT 2009. Users' guide. Texas Agricultural Experiment Station, Temple, TX: Grassland, Soil and Water Research Laboratory, Agricultural Research Service, and Blackland Research Center; 2010.
[39] Margulis SA, McLaughlin D, Entekhabi D, Dunne S. Land data assimilation and estimation of soil moisture using measurements from the Southern Great Plains 1997 field experiment. Water Resour Res 2002;38(12):1299. <http://dx.doi.org/10.1029/2001WR001114>.
[40] Moradkahani H. Hydrologic remote sensing and land surface data assimilation. Sensors 2008;8:2986–3004. <http://dx.doi.org/10.3390/s8052986>.
[41] Moradkhani H, Meskele TT. Probabilistic assessment of the satellite rainfall retrieval error translation to hydrologic response. Springer, Water Science and Technology Library: Satellite Rainfall Applications for Surface Hydrology 2010:229–42. <http://dx.doi.org/10.1029/2004WR003604>.
[42] Moradkhani H, Hsu KL, Gupta HV, Sorooshian S. Uncertainty assessment of hydrologic model states and parameters: sequential data assimilation using the particle filter. Water Resour Res 2005;41(5):W05012. <http://dx.doi.org/10.1029/2004WR003604>.
[43] 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(12):W12520. <http://dx.doi.org/10.1029/2012WR012144>.
[44] Moriasi D, Arnold J, Van Liew M, Bingner R, Harmel R, Veith T. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Trans ASABE 2007;50(3):885–900. <http://dx.doi.org/10.13031/2013.23153>.
[45] Neitsch S, Arnold J, Kiniry J, Srinivasan R, Williams J. Soil and water assessment tool input/output file documentation, version 2005. Temple, TX: Blackland Research Center, USDA Agricultural Research Service; 2004.
[46] Ngodock HE, Jacobs GA, Chen M. The representer method, the ensemble Kalman filter and the ensemble Kalman smoother: a comparison study using a nonlinear reduced gravity ocean model. Ocean model 2006;12(3):378–400. <http://dx.doi.org/10.1016/j.ocemod.2005.08.001>.
[47] Njoku EG, Wilson WJ, Yueh SH, Dinardo SJ, Li FK, Jackson TJ, et al. Observations of soil moisture using a passive and active low-frequency microwave airborne sensor during SGP99. IEEE Trans Geosci Remot Sen 2002;40(12):2659–73. <http://dx.doi.org/10.1109/TGRS.2002.807008>.
[48] Pan X, Li X, Shi X, Han X, Luo L, Wang L. Dynamic downscaling of near-surface air temperature at the basin scale using WRF: a case study in the Heihe River Basin, China. Frontiers of Earth Science in China 2012;6(3):314–23. <http://dx.doi.org/10.1007/s11707-012-0306-2>.
[49] Reichle RH. Data assimilation methods in the earth sciences. Adv Water Resour 2008;31(11):1411–8. <http://dx.doi.org/10.1016/j.advwatres.2008.01.001>.
[50] Reichle RH, Entekhabi D, McLaughlin DB. Downscaling of radio brightness measurements for soil moisture estimation: a four-dimensional variational data assimilation approach. Water Resour Res 2001;37(9):2353–64. <http://dx.doi.org/10.1029/2001WR000475>.
[51] Reichle RH, McLaughlin DB, Entekhabi D. Hydrologic data assimilation with the ensemble Kalman filter. Mon Weather Rev 2002;130(1):103–14. <http://dx.doi.org/10.1175/1520-0493(2002)130<0103:HDAWTE>2.0.CO;2>.
[52] Ryu D, Crow WT, Zhan X, Jackson TJ. Correcting unintended perturbation biases in hydrologic data assimilation. J Hydrometeorol 2009;10(3):734–50. <http://dx.doi.org/10.1175/2008JHM1038.1>.
[53] Neitsch S, Arnold J, Kiniry J, Williams J. Soil and water assessment tool theoretical documentation, version 2009. Texas water resources institute technical rep. 406; College Station, Temple, TX: Texas A&M Univ. System; 2011.
[54] Starks P, Moriasi D. Spatial resolution effect of precipitation data on SWAT calibration and performance: implications for CEAP. Trans ASABE 2009;52(4):1171–80. <http://dx.doi.org/10.13031/2013.27792>.
[55] Strauch M, Bernhofer C, Koide S, Volk M, Lorz C, Makeschin F. Using precipitation data ensemble for uncertainty analysis in SWAT streamflow simulation. J Hydrol 2010;414(415):413–24. <http://dx.doi.org/10.1016/j.jhydrol.2011.11.014>.
[56] Tossavainen OP, Percelay J, Stacey M, Kaipio JP, Bayen A. State estimation and modeling error approach for 2-D shallow water equations and Lagrangian measurements. Water Resour Res 2011;47(10):W10510. <http://dx.doi.org/10.1029/2010WR009401>.
[57] Troch PA, Paniconi C, McLaughlin D. Catchment-scale hydrological modeling and data assimilation. Adv Water Resour 2003;26(2):131–6. <http://dx.doi.org/10.1016/S0309-1708(02)00087-8>.
[58] Van Leeuwen PJ. An ensemble smoother with error estimates. Mon Weather Rev 2001;129(4):709–28. <http://dx.doi.org/10.1175/1520-0493(2001)129<0709:AESWEE>2.0.CO;2>.
[59] Vrugt JA, Gupta HV, Nuallain BO, Bouten W. Real-time data assimilation for operational ensemble streamflow forecasting. J Hydrol 2005;7(3):548–65. <http://dx.doi.org/10.1175/JHM504.1>.
[60] Walker JP, Houser PR. A methodology for initializing soil moisture in a global climate model: assimilation of near-surface soil moisture observations. J Geophys Res 2001;106(D11):11761–74. <http://dx.doi.org/10.1029/2001JD900149>.
[61] Wu G, Zheng X, Wang L, Zhang S, Liang X, Li Y. A new structure for error covariance matrices and their adaptive estimation in EnKF assimilation. Quart J Roy Meteorol Soc 2013;139(672):795–804. <http://dx.doi.org/10.1002/qj.2002>.
[62] Xie X, Zhang D. Data assimilation for distributed hydrological catchment modeling via ensemble Kalman filter. Adv Water Resour 2010;33(6):678–90. <http://dx.doi.org/10.1016/j.advwatres.2010.03.012>.