

Research Article

Improving the Distributed Hydrological Model Performance in Upper Huai River Basin: Using Streamflow Observations to Update the Basin States via the Ensemble Kalman Filter

Yongwei Liu,^{1,2} Wen Wang,^{1,2} Yiming Hu,¹ and Wei Cui¹

¹College of Hydrology and Water Resources, Hohai University, Nanjing 210098, China

²State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing 210098, China

Correspondence should be addressed to Wen Wang; w.wang@126.com

Received 31 March 2016; Accepted 22 May 2016

Academic Editor: Taesam Lee

Copyright © 2016 Yongwei Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This study investigates the capability of improving the distributed hydrological model performance by assimilating the streamflow observations. Incorrectly estimated model states will lead to discrepancies between the observed and estimated streamflow. Consequently, streamflow observations can be used to update the model states, and the improved model states will eventually benefit the streamflow predictions. This study tests this concept in upper Huai River basin. We assimilate the streamflow observations sequentially into the Soil and Water Assessment Tool (SWAT) using the ensemble Kalman filter (EnKF) to update the model states. Both synthetic experiments and real data application are used to demonstrate the benefit of this data assimilation scheme. The experiment shows that assimilating the streamflow observations at interior sites significantly improves the streamflow predictions for the whole basin. Assimilating the catchment outlet streamflow improves the streamflow predictions near the catchment outlet. In real data case, the estimated streamflow at the catchment outlet is significantly improved by assimilating the in situ streamflow measurements at interior gauges. Assimilating the in situ catchment outlet streamflow also improves the streamflow prediction of one interior location on the main reach. This may demonstrate that updating model states using streamflow observations can constrain the flux estimates in distributed hydrological modeling.

1. Introduction

Catchment hydrological modeling is critical for water resources management and planning (e.g., the flood control and drought monitoring). However, the accuracy and reliability of the hydrological modeling are generally subjected to various uncertainties associated with model forcing inputs, model parameters, and model structures [1, 2]. Data assimilation (DA) techniques, owing to the capability to merge modeling and observations by jointly considering the uncertainties in them, have been commonly used in hydrology to improve the rainfall-runoff simulation and prediction [3–6]. Previous researches have demonstrated that assimilating the remote sensing data, for example, the satellite soil moisture products [4, 7, 8], into hydrological modeling could improve the hydrological process.

Assimilating the discharge observations has more advantages in rainfall-runoff improvement, since the discharge observations are more reliable and more directly related to the streamflow modeling. This concept is intensively tested in lumped hydrological modeling [9–12]. In distributed modeling, the studies on streamflow assimilation keep growing in recent years. Clark et al. [13] employed streamflow observations to update the states of a distributed model (TopNet) to improve the streamflow prediction and demonstrated that transforming streamflow into log space before computing the covariance would improve the filter performance of the ensemble Kalman filter (EnKF) [14]. Xie and Zhang [5] proved the capability of the EnKF in the improvement of parameter estimation and thus improved the hydrological modeling in SWAT by assimilating the streamflow observations in synthetic experiments. Lee et al. [15] validated the

potential of updating soil moisture states of a distributed model by assimilating streamflow and in situ soil moisture data using a variational assimilation approach for the prediction of streamflow and soil moisture. Streamflow assimilation has the potential to improve the distributed hydrological simulations and predictions via updating the model states in the model propagating process, but to date, this potential is not fully exploited; still few studies evaluate this forecast improvements of distributed hydrological modeling that are possible from the improved estimates of basin states via assimilating streamflow observations. However, in theory, the distributed models are expected to outperform the lumped ones in rainfall-runoff modeling as it could account for the spatial variability of meteorological inputs and the underlying land surface conditions of a catchment [16]. Unlike the lump models that can only produce streamflow prediction at the basin outlet, the distributed models have the capability to provide the streamflow prediction at interior locations. This makes the streamflow assimilation in distributed models not only benefit the streamflow simulations and predictions at the upstream of the observation sites but also can be attempted to improve that of the neighboring or downstream areas. This property is especially important for the streamflow predictions of ungauged basins. Considering the significance of the streamflow assimilation in the improvement of the distributed hydrological simulations and predictions and still few studies on this aspect at present, the study on streamflow assimilation in distributed models needs to be strengthened and their potential in improving distributed hydrological modeling needs to be further explored.

Besides, the state variables of the distributed models are characterized with spatial variability, which are usually controlled by the underlying land surface characteristics of the catchment (e.g., the topography, land cover, and soil types). This makes the state update probably be influenced by the underlying land surface conditions; thus the performance of streamflow assimilation in improving rainfall-runoff modeling is also affected. Having a good understanding of these factors would be important in improving the streamflow assimilation performance. However, to our knowledge, the influencing factors which related the land surface conditions on streamflow assimilation in distributed modeling are rarely reported by now.

In this study, we investigate the capability of streamflow assimilation in improving the distributed hydrological modeling of upper Huai River basin via updating the basin states. This investigation is based on the Soil and Water Assessment Tool (SWAT) and the ensemble Kalman filter (EnKF) method. Both the synthetic experiment and real application are considered; the synthetic experiment assesses the upper bound of the streamflow assimilation performance under idealized conditions of unbiased model parameters and well-known modeling and observation errors, while the real application provides a more realistic streamflow assimilation performance. Besides, we also analyze the influence of the underlying land surface conditions (e.g., slopes) on the performance of streamflow assimilation.

2. Methodology

2.1. Soil and Water Assessment Tool (SWAT). SWAT is a physically based basin scale distributed model developed by USDA (United States Department of Agriculture) Agricultural Research Service. It has been widely used in predicting the impact of land management practices on water, sediment, and agricultural chemical yields in large, complex catchments with varying soil, land use, and management conditions over long periods of time [17]. In SWAT, the catchment is geographically partitioned into different subbasins; then each subbasin is further delineated into several Hydrological Response Units (HRUs) based on soil, land cover, and slopes. HRUs are the basic calculation units of the model, on which the surface runoff, soil water, lateral flow, and ground water are produced. The generated water from HRUs is aggregated at subbasins and discharges to the outlet of the catchment through river routing process.

The surface runoff on a given day (Q_{surf}) is simulated by the modified SCS curve number method:

$$Q_{\text{surf}} = \frac{(P - 0.2S)^2}{(P + 0.8S)}, \quad (1)$$

where P is the precipitation on a given day (mm) and S is the retention parameter (mm) that can be described by

$$S = \frac{25400}{CN} - 254, \quad (2)$$

where CN is the SCS curve number for a given day; it is determined by the initial curve number CN_2 [17] and the profile soil water content on the same day.

The reduced rainfall by surface runoff recharging to soil profile is redistributed using the storage routing technique at each soil layer in the root zone as

$$SW_{ly,2} = SW_{ly,1} + \omega_{\text{perc},ly} - Q_{\text{lat},ly} - E_{a,ly}, \quad (3)$$

where $SW_{ly,1}$ and $SW_{ly,2}$ are the soil water content at the beginning and end of the day (mm) in layer ly , respectively, $\omega_{\text{perc},ly}$ is the percolation from the overlaying layer (mm) to layer ly , $E_{a,ly}$ is the actual evapotranspiration drawn from layer ly during the day, and $Q_{\text{lat},ly}$ is the lateral flow generated in layer ly calculated by

$$Q_{\text{lat}} = \sum_{ly=1}^n \frac{0.048 \cdot (SW_{ly} - FC_{ly}) \cdot K_{\text{sat},ly} \cdot \text{slp}}{(\phi_{\text{soil},ly} - \phi_{FC,ly}) \cdot L_{\text{hill}}}, \quad (4)$$

where n is the number of layers on soil profile, FC_{ly} is the soil water content of layer ly at field capacity (mm), $K_{\text{sat},ly}$ is the saturated hydraulic conductivity (mm/h) of layer ly , slp is the slopes (m/m), $\phi_{\text{soil},ly}$ is the total porosity of the soil layer ly (mm/mm), $\phi_{FC,ly}$ is the porosity of layer ly when it is at field capacity, and L_{hill} is the hill slope length (m).

For groundwater modeling, SWAT incorporates two aquifers in each subbasin: the shallow (unconfined) and deep (confined) aquifer. Only the water in the shallow aquifer contributes to flow in the main channel of the subbasin, where the water balance is described by

$$aq_{sh,2} = aq_{sh,1} + \omega_{\text{rchrg},sh} - Q_{gw} - W_{\text{revap}} - W_{\text{pump},sh}, \quad (5)$$

where $aq_{sh,1}$ and $aq_{sh,2}$ are the water stored in the shallow aquifer at the beginning and end of the day (mm), $\omega_{rchrge,sh}$ is the recharge entering the shallow aquifer during the day (mm), Q_{gw} is the ground water flow to the main channel during the day (mm), W_{revap} is the water moving into the soil zone in response to the water deficiency for the day (mm), and $W_{pump,sh}$ is the water removing from shallow aquifer by pumping during the day (mm).

The surface runoff (Q'_{surf}) and lateral flow (Q'_{lat}) to the main channel on a given day are calculated by (6) and (7). One has

$$Q'_{surf} = (Q_{surf} + Q_{stor,t-1}) \cdot \left(1 - \exp \left[\frac{-\text{surlag}}{t_{conc}} \right] \right), \quad (6)$$

where $Q_{stor,t-1}$ is the surface runoff lagged from the previous day (mm), surlag is the surface runoff lag coefficient, and t_{conc} is the time of concentration for the surface runoff. One has

$$Q'_{lat} = (Q_{lat} + Q_{latstor,t-1}) \cdot \left(1 - \exp \left(\frac{-1}{TT_{lag}} \right) \right), \quad (7)$$

where $Q_{latstor,t-1}$ is the lateral flow lagged from the previous day (mm) and TT_{lag} is the lateral flow travel time (days).

The channel water balance in the reach can be generally expressed as

$$V_{stored,2} = V_{stored,1} + V_{in} - V_{out}, \quad (8)$$

where $V_{stored,1}$ and $V_{stored,2}$ are the volume of water in the reach at the beginning and end of the day respectively (m^3), V_{in} and V_{out} are the volume of water flow into and out of the reach during the day (m^3), and V_{out} is calculated by (9) in variable storage routing method (see [17]):

$$V_{out} = SC \cdot (V_{in} + V_{stored,1}), \quad (9)$$

where SC is the storage coefficient with an upper limit of 1.0 calculated by

$$SC = \frac{2\Delta t}{2TT + \Delta t}, \quad (10)$$

where Δt is the length of time step (s) and TT is the reach travel time (s), which is a result of the main channel length divided by the discharge rate.

2.2. The Ensemble Kalman Filter (EnKF). Based on the theory of the linear Kalman filter [18], the ensemble Kalman filter (EnKF) incorporates a Monte Carlo method to generate the state ensemble with a certain distribution to represent its probability. The generated state ensemble is propagated forward in time using the model F , with forcing inputs $u(t)$, model parameters α , and system uncertainty $w(t)$ as

$$X(t) = F[X(t-1), u(t), \alpha, w(t)], \quad (11)$$

where $X(t)$ is the predicted model state ensemble at time t and the state variables to be included will be detailed in Section 4.1. F represents SWAT model in this study, and

the model uncertainty $w(t)$ is assumed to be a Gaussian distribution multiplied to the system.

When the observations are available, the predicted model state ensemble $X(t)$ is related to the observation ensemble $Z(t)$ using an operator H as

$$Z(t) = H[X(t)] + v(t), \quad (12)$$

where $v(t)$ represents the observation error being assumed to be a Gaussian distribution with zero mean and covariance of R . The state update is obtained by

$$X^a(t) = X(t) + K(t)(Z(t) - H[X(t)]), \quad (13)$$

where $X^a(t)$ is the updated state ensemble and $K(t)$ is the Kalman gain, which determines the weight of modeling and observation in state update and is calculated by the forecast error covariance and the observation error covariance as

$$K(t) = P_{FH}(P_{HH} + R)^{-1}, \quad (14)$$

where P_{FH} is the cross error covariance between the predicted states $X(t)$ and the measurement predictions $H[X(t)]$ and P_{HH} is the forecast error covariance of the measurement predictions [19].

2.3. Evaluation Method. In order to assess the performance of streamflow assimilation, the root mean square error, Nash-Sutcliffe coefficient of efficiency, Pearson's correlation coefficient and normalized error reduction index are used.

The root mean square error (RMSE) can be expressed by [20]

$$\text{RMSE} = \left(\frac{1}{n} \sum_{i=1}^n (Q_i^{\text{sim}} - Q_i^{\text{obs}})^2 \right)^{1/2}, \quad (15)$$

where n is the total time step and Q_i^{obs} and Q_i^{sim} are the measured and simulated streamflow at time i .

The Nash-Sutcliffe coefficient (NSE) is expressed by [21]

$$\text{NSE} = 1 - \frac{\sum_{i=1}^n (Q_i^{\text{sim}} - Q_i^{\text{obs}})^2}{\sum_{i=1}^n (Q_i^{\text{obs}} - \bar{Q}^{\text{obs}})^2}, \quad (16)$$

where \bar{Q}^{obs} indicates the mean value of the measured streamflow for the whole period.

Pearson's correlation coefficient is obtained by [22]

$$\text{CC} = \frac{\sum_{i=1}^n (S_i^{\text{sim}} - \bar{S}^{\text{sim}})(Q_i^{\text{sim}} - \bar{Q}_i^{\text{sim}})}{\sqrt{\sum_{i=1}^n (S_i^{\text{sim}} - \bar{S}^{\text{sim}})^2 \cdot \sum_{i=1}^n (Q_i^{\text{sim}} - \bar{Q}_i^{\text{sim}})^2}}, \quad (17)$$

where CC represents the correlation coefficient, S_i^{sim} is the simulated states at time i , and \bar{S}^{sim} indicates the mean value of simulated states for the whole period.

The normalized error reduction index (NER) is expressed by [23]

$$\text{NER}(\%) = 100 \cdot \left(1.0 - \frac{\text{RMSE}_{\text{EnKF}}}{\text{RMSE}_{\text{EnOL}}} \right), \quad (18)$$

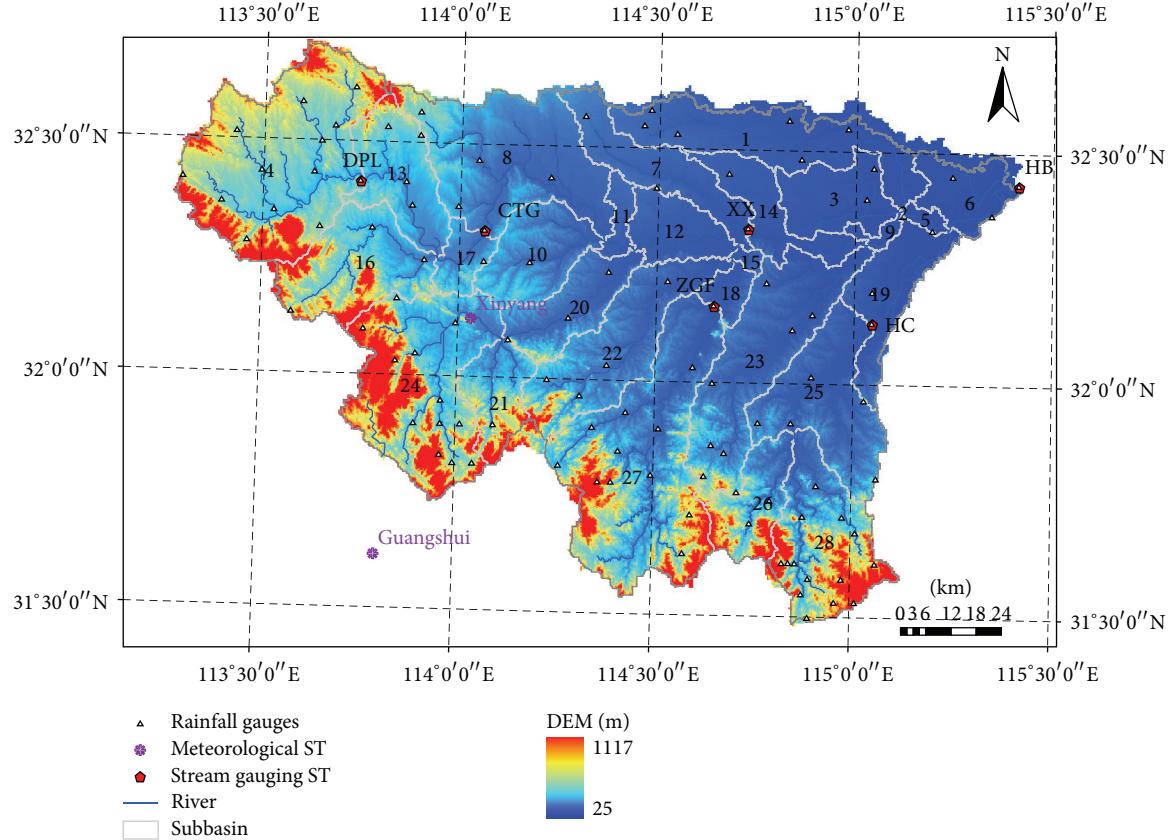


FIGURE 1: Upper Huai River basin (note: ST represents station; DPL, CTG, ZGF, XX, HC, and HB represent Dapoling, Changtaiguan, Zhuganfu, Xixian, Huangchuan, and Huaibin station, resp.; 1 to 28 represent the subbasin number).

where $\text{RMSE}_{\text{EnKF}}$ and $\text{RMSE}_{\text{EnOL}}$ are the root mean square error of the variable in EnKF and EnOL (detailed in Section 4.3), respectively. The larger value of NER means EnKF outperforms the EnOL. $\text{NER} = 1.0$ indicates that the EnKF results overlap the truth or observations.

3. Study Area and Data Used

The study catchment is located between $113^{\circ}15'E \sim 116^{\circ}00'E$ and $31^{\circ}30'N \sim 33^{\circ}00'N$ in the upper Huai River basin in China, with an area of 16005 km^2 and an elevation from 25 m to 1117 m (Figure 1). It is dominated by flat area except for the west and south west mainly covered by mountains and hills. The basin locates in the transition zone between the northern subtropical and the warm temperate zone; the annual average precipitation is about 900 mm and the annual average temperature is around 15°C . The rainfall is considerably influenced by monsoons during flood season from June to September, so the precipitation with 50% to 80% falls in this period. The major land cover of this catchment is agriculture, forest, and brush. Agriculture is the dominant land use type, the majority of which is rice (34.98%) and wheat (32.49%).

The input data required by SWAT model mainly includes the meteorological forcing data and the underlying land

surface data. The meteorological input data includes precipitation, maximum/minimum temperature, solar radiation, wind speed, and relative humidity. The precipitation is provided by 106 local rainfall gauges in the basin, while for each subbasin, the precipitation input is the interpolated value of the 106 rainfall stations using the Thiessen polygon method. The 5 types of subsequent meteorological data above are collected from 2 meteorological stations (Xinyang and Guangshui station) in or near the catchment (Figure 1). The land surface data contains the topography data (digital elevation data, DEM), soil category, and land cover data. The DEM data is downloaded from the Shuttle Radar Topography Mission (SRTM) with a spatial resolution of 90 m (<http://srtm.csi.cgiar.org/SELECTION/inputCoord.asp>). The soil data is resampled from a soil map at a scale of 1: 100000 collected from Soil Handbook of Henan province. According to the Soil Handbook, there are 7 types of soil in this basin; the area proportions, soil texture, and the corresponding USDA (United States Department of Agriculture) classification are present in Table 1. Besides, the land use data is resampled from a year-1995 land use map at a scale of 1: 210000 provided by the government of Xinyang city.

There are 6 streamflow gauges in the catchment (Figure 1), which provides daily streamflow discharges for model calibration, validation, and data assimilation. The locations of all the 6 streamflow stations are set as the subbasin outlet. And

TABLE 1: Soil classification and its area proportions in the upper Huai River basin.

Soil code	Source soil	Clay (%)	Silt (%)	Sand (%)	Rock (%)	USDA soil texture	Area proportion (%)
S-1	Huanghetu	23.43	65.1	11.5	0	Silt loam	21.52
S-2	Shuidatu	16.46	71	12.5	0	Silt loam	34.62
S-3	Huangzongrang	17.03	39.4	43.6	0	Loam	13.7
S-4	Cugutu	7.05	34	35.9	23.1	Sandy loam	7.35
S-5	Shizhitu	9.16	44.1	46.7	0	Loam	7.8
S-6	Huichaotu	12.86	51.8	35.3	0	Silt loam	4.72
S-7	Shajiangheitu	20.32	65.5	14.2	0	Silt loam	9.81

TABLE 2: State variables in SWAT model chosen for correlation analysis.

Order	State variable	Description	Units	Level
1	Q_{stor}	Amount of surface runoff lagged or stored on a given day	mm	HRU
2	$Q_{latstor}$	Amount of lateral flow lagged or stored on a given day	mm	HRU
3	aq_{sh}	Amount of water stored in shallow aquifer on a given day	mm	HRU
4	SW_{ly}	Amount of water stored in the soil layer (ly) at the end of day	mm	HRU * N_{lay}
5	V_{stored}	Water stored in reach at the end of day	m^3	Subbasin

Note: N_{lay} is the number of soil layers on profile in a hydrologic response unit (HRU).

the catchment is partitioned into 28 subbasins (Figure 1) in total based on the DEM data. Then, it is further delineated into 82 HRUs according to the slope, soil, and land use information. Besides, the soil on profile is divided into 4 layers to maximum.

4. Implementation of Streamflow Assimilation in SWAT

4.1. Selecting State Variables to Be Updated in SWAT. Considering the complicated model physics and structures and the large number of state variables in SWAT, there is a possibility that the spurious correlations exist in a high dimensional state vector and, thus, bring large degrees of freedom to the state update in streamflow assimilation. In order to reduce the freedom in state update, only the state variables that are strongly dependent on the streamflow measurement predictions (i.e., the state variables sensitive to the streamflow modeling) will be updated. Based on the physical significance of the state variables in SWAT, the five variables are preliminarily chosen (Table 2). Then, the objective is to analyze the correlations of the preliminary selected five variables with the simulated streamflow at the six runoff gauges (Figure 1) to finally determined variables to be updated in the assimilation.

The analysis results are present in Figure 2, taking the fifteenth calculation unit (HRU) and the sixth subbasin as an example. It can be found that compared to the surface runoff storage (Q_{stor}) and reach streamflow storage (V_{stored}), the correlation coefficients (CC) of the lateral flow storage ($Q_{latstor}$), shallow aquifer storage (aq_{sh}), and soil water storage (SW_{ly}) of 4 layers are comparatively low. For other 81 calculation units and 27 subbasins, the analysis results are similar to the above conclusions. Therefore, in this study, only

Q_{stor} and V_{stored} are selected to be updated. It is noteworthy that the predicted streamflow at measurement sites is also included in the state vector, but they are not updated as they are diagnostic variable rather than state variable [13]. As the predicted streamflow at measurement locations is included in the state vector, there is an exact match between the observations and their model equivalent. Therefore, the state transition matrix H from modeling to observations in EnKF (12) can be constructed to have a value of 1 for elements where there is a model prediction of the observation and 0 where there is no equivalent of observation [24].

4.2. Modeling and Observation Error. In this study, the uncertainty of hydrological modeling comes from the errors of meteorological forcing inputs (e.g., precipitation), model parameters, and model structures. Each ensemble member in EnKF is perturbed using the assumptions summarized in Table 3. The error of precipitation is assumed to be uncorrelated both on temporal scale between continuous time steps and on spatial scale with different stations. This design of the precipitation error considers the previous studies on the system measurement error of rainfall at precipitation gauges in China [25]. The small error design for the predicted model states (i.e., the surface runoff storage and the reach flow storage) is to avoid rapid changes between continuous time steps in the modeling process. The assumed error on the measurement prediction of streamflow flux is to account for the errors of model structure. Besides, the ranging standard deviation (SD) for the sensitive parameters (Table 4) is to ensure that the parameters still stay in their physical threshold after perturbation. The assumed error for streamflow observations considers the characteristics of the uncertainty of streamflow measurement [12] and the error assumption adopted in previous studies [13].

TABLE 3: Error parameters in the synthetic experiment.

Variables	Error distribution	Mean	Standard deviation	Bound
Precipitation	Gaussian, multiplicative	1	0.2	(0, —)
Predicted states	Gaussian, multiplicative	1	0.01	(0, —)
Predicted streamflow	Gaussian, multiplicative	1	0.2	(0, —)
Parameters	Gaussian distribution	1	Given in Table 4	
Observed streamflow	Gaussian, multiplicative	1	0.1	(0, —)

TABLE 4: The sensitive parameters in SWAT for rainfall-runoff modeling and their standard deviation (SD) of Gaussian multiplicative error in streamflow assimilation.

Parameters	Units	Description	Level	SD of GME
CN ₂	—	Initial SCS runoff curve number for moisture condition II	HRU	0.05
ESCO	—	Soil evaporation compensation factor	HRU	0.2
ALPHA_BF	Days	Baseflow recession constant	HRU	0.2
GW_REVAP	—	Ground “revap” coefficient	HRU	0.2
GWQMN	mm	Threshold depth of water in the shallow aquifer required for return flow to occur	HRU	0.2
REVAPMN	mm	Threshold depth of water in the shallow aquifer for “revap” or percolation to the deep aquifer to occur	HRU	0.2
OV_N	—	Manning’s “n” value for overland flow	HRU	0.1
SOL_AWC	mm/mm	Available water capacity of the soil layer	HRU * N _{lay}	0.2
CH_N(1)	—	Manning’s “n” values for the tributary channels	HRU	0.2
CH_N(2)	—	Manning’s “n” values for the main channel	Subbasin	0.2
SURLAG	—	Surface runoff lag coefficient	Basin	0.2

Note: N_{lay} is the number of soil layers on profile in a hydrologic response unit (HRU); SD represents standard deviation; GME represents Gaussian multiplicative error.

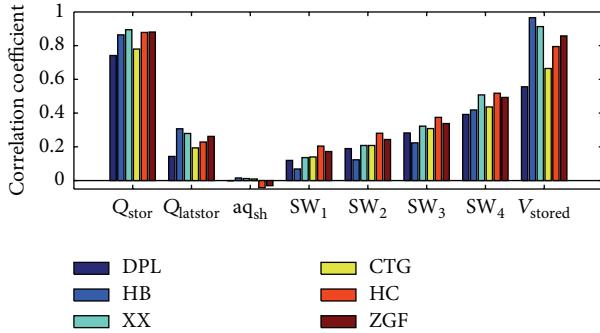


FIGURE 2: Correlation between the state variables of the fifteenth calculation units and the sixth subbasin on the horizontal axis and the simulated streamflow at six streamflow gauges (DPL, HB, XX, CTG, HC, and ZGF in Figure 1).

4.3. Synthetic and Real Data Assimilation. In the synthetic experiment, one reference field is randomly picked out from a Gaussian distribution with given model inputs. Then, the model is propagated forward to obtain a reference modeling process, which is regarded as the synthetic truth. The synthetic streamflow observations are generated by randomly perturbing the synthetic true streamflow drawn from the

reference simulation using a Gaussian multiplicative error with the same standard deviation as that of the streamflow observation error. An ensemble integration with 200 members of SWAT model is performed with known errors (Table 3) of model inputs, model parameters, and states, which is regarded as the Open loop (EnOL) run. In the EnKF run, the assimilation integration is performed by introducing the synthetic streamflow observations to the stochastic modeling process with the same errors as that of the EnOL run. Here, two cases are designed: (1) assimilating the synthetic streamflow observations at 5 interior sites (i.e., Dapolding, Changtaiguan, Xixian, Zhuganfu, and Huangchuan); (2) assimilating the synthetic streamflow observations at the catchment outlet (Huaibin). The performance of streamflow assimilation is illustrated by comparing the predicted streamflow obtained by EnKF and EnOL with the synthetic truth as a reference on the whole basin.

In real data assimilation, the error set of model and observation for EnOL and EnKF is identical to that of the synthetic experiment. Similar to the synthetic experiment, we consider the following two cases: (1) the streamflow measurements at the five interior discharge gauges are assimilated and the in situ streamflow measurements at the catchment outlet are used for validation; (2) the streamflow measurements

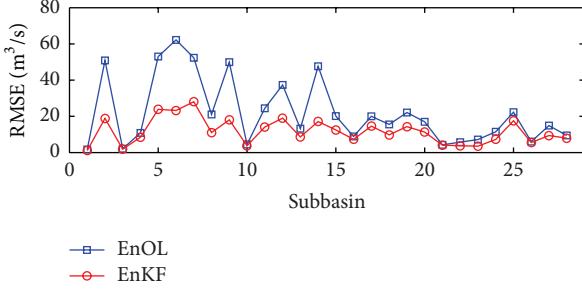


FIGURE 3: The RMSE (15) of the estimated streamflow by EnOL and EnKF at all subbasin outlets in the assimilation of the streamflow observations at the five interior sites.

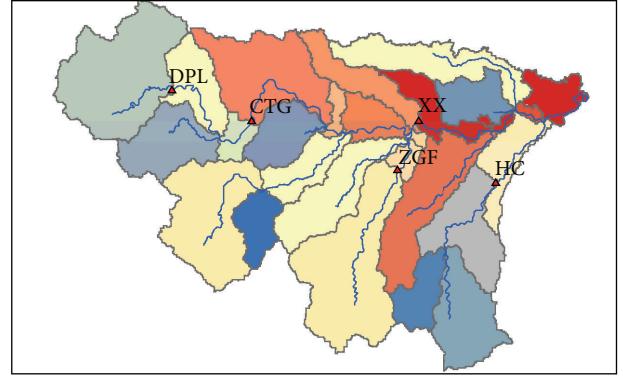
at the catchment outlet are assimilated and the measured streamflow at the five interior sites is adopted for validation.

5. Results

5.1. Synthetic Experiment

5.1.1. Assimilation of the Streamflow Observations at Interior Sites. Figure 3 compares the root mean square error (RMSE) of the estimated streamflow at all subbasin outlets (located at different sites on the reach) under the assimilation of the streamflow observations at the five interior sites to that without streamflow assimilation for the whole assimilation period from Jan. 1, 1996 to Dec. 31, 1997. It can be found that the RMSE is reduced by EnKF for all subbasin outlets. However, for different subbasins, the streamflow assimilation performance shows large difference. This can be better illustrated by the normalized reduction of RMSE (i.e., NER), the distribution of which on the whole basin is present in Figure 4. To make the NER distribution on the basin clear, the NER value at the subbasin outlet is put on the subbasin where it is located. We can see that the estimated streamflow at the subbasin outlets obtains different levels of improvement by EnKF as the NER ranges from 5.43% to 64.05%. It seems that this improvement shows an increasing trend from upstream to downstream of the main reach where the discharges gauges Dapoling, Changtaiguan, and Xixian (DPL, CTG, and XX) are located. The predicted streamflow at the catchment outlet obtains significant improvement compared with the fact that no data assimilation is performed.

Figure 5 presents the time series of the estimated streamflow obtained by EnKF and EnOL at the catchment outlet. It can be seen in Figure 5(a) that the outlet streamflow is significantly improved by EnKF for the whole assimilation period. In terms of the ensemble mean of the simulated streamflow, EnKF obtains significant improvement on the accuracy as it is close to the synthetic truth with the RMSE reduced from $62.18 \text{ m}^3/\text{s}$ to $23.24 \text{ m}^3/\text{s}$ (NER = 63%); the NSE and the correlation coefficient (CC) increased from 0.97 and 0.98 to approximately 1.0, respectively. The uncertainty of the simulated streamflow in EnOL is significantly reduced by EnKF, as the ensemble spread is largely decreased in EnKF. Besides, it is clear in Figures 5(b) and 5(c) that this improvement on high flow is significant, the simulated



▲ Discharge
— River
□ Subbasin
NER (%) value
64.05
5.43

FIGURE 4: The NER (18) of the estimated streamflow for all subbasin outlets obtained by assimilating the streamflow observations at five interior gauges (DPL, CTG, ZGF, XX, and HC). The NER value at the subbasin outlet is displayed on the subbasin where it is located.

streamflow over $1000 \text{ m}^3/\text{s}$, of which the RMSE is reduced from $189.7 \text{ m}^3/\text{s}$ to $70 \text{ m}^3/\text{s}$ with the NER approaching 63.1%. Although the improvement on low flow is not clearly shown in Figure 5(a) due to the large gaps between high flow and low flow, it is still significant as the simulated streamflow less than $1000 \text{ m}^3/\text{s}$ of which the RMSE is decreased from $49.4 \text{ m}^3/\text{s}$ to $18.6 \text{ m}^3/\text{s}$ with the NER reaching 62.3%.

The improvement of streamflow prediction by EnKF should be benefited from the real-time updating of the surface runoff storage (Q_{stor}) and the reach flow storage (V_{stored}) on the basis of the observed streamflow, because according to design, the sole difference between EnKF and EnOL is that both Q_{stor} and V_{stored} are updated in EnKF but not in EnOL. Q_{stor} and V_{stored} in EnKF are compared with those in EnOL with their synthetic truth as a reference. Figure 6 displays the RMSE distribution of Q_{stor} for 82 HRUs in EnKF and EnOL. It can be found that the RMSE of Q_{stor} is significantly reduced by EnKF as both the mean and the spread of RMSE are considerably decreased. Figure 7 presents the RMSE of V_{stored} in EnOL and EnKF and the normalized reduction of RMSE (NER) for V_{stored} on 28 subbasins. By comparing Figures 7(a) and 7(b), we can see that the RMSE of V_{stored} on most subbasins are substantially reduced by EnKF, especially on subbasins 6, 8, 12, and 14. This could be better illustrated by Figure 7(c), where the NER on all subbasins are above zero and those on subbasins 6, 8, 12, and 14 are over 70%. Overall, Q_{stor} and V_{stored} updated by EnKF significantly approach their synthetic truth.

For different subbasins and HRUs, the impact of streamflow assimilation on the update of V_{stored} (Figure 7) and Q_{stor} (Figure 6) is significantly different, as the net reduction calculated by $\text{RMSE}_{\text{EnOL}} - \text{RMSE}_{\text{EnKF}}$ and the normalized reduction of RMSE (18) for V_{stored} on different subbasins show large difference and the RMSE distribution for Q_{stor} changes

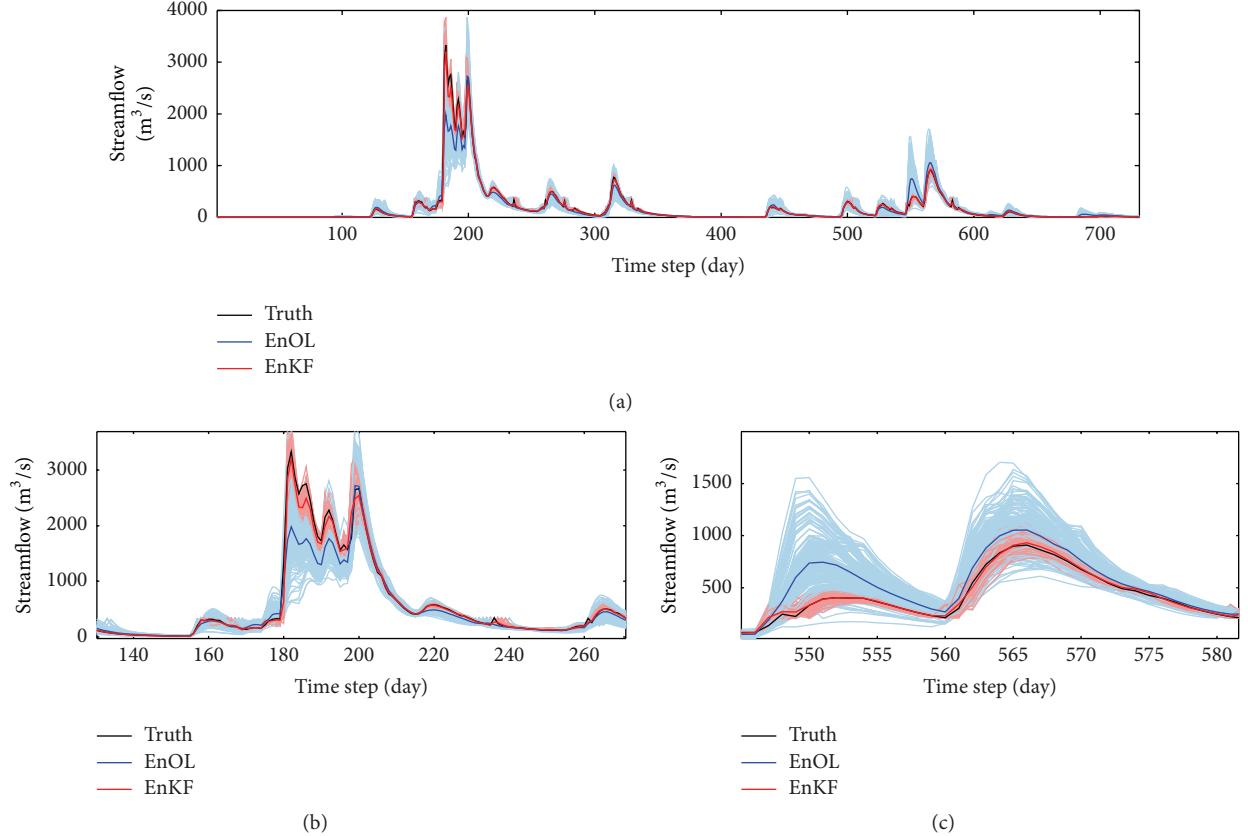


FIGURE 5: Comparison of the estimated streamflow by EnKF and EnOL at the basin outlet (Huaibin station) in synthetic experiment: (a) the whole period from Jan. 1, 1996 to Dec. 31, 1997; (b) and (c) details of the period with high flows. The light blue and red lines indicate the ensemble members of the EnOL and EnKF, respectively.

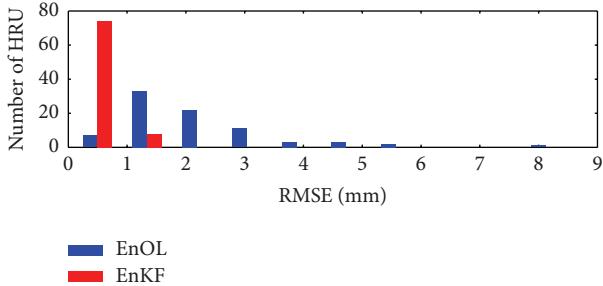


FIGURE 6: The RMSE (15) distribution of the estimated surface runoff storage (Q_{stor}) for 82 HRUs using EnKF and EnOL.

significantly after state update. This difference can be partly related to the underlying land surface conditions.

Figure 8 shows the impact of the slopes, soil category, and land cover on the update of the surface runoff storage (Q_{stor}) as the three factors determine the HRU delineation. Figure 8(a) shows that the Q_{stor} on the HRUs with low slopes ($<5^\circ$) obtain large net reduction of RMSE ($\text{RMSE}_{\text{EnOL}} - \text{RMSE}_{\text{EnKF}}$), which contributes significant improvement of rainfall-runoff modeling. Figures 8(b) and 8(c) illustrate that the large reduction of RMSE occurs on the HRUs

with soil category of Huanghetu (S-1), Shuidaitu (S-2), and Shajiangheitu (S-7) (all the three soils belong to silt loam in USDA soil classification) and with land cover of rice (RICE) and mixed agriculture mainly covered by wheat and corn (AGRC).

Figure 9 presents the impact of the main channel length of the subbasins on the state update of the reach flow storage (V_{stored}) in terms of the normalized reduction of RMSE (NER) for V_{stored} . It shows that the less significant state update of V_{stored} ($\text{NER} < 25\%$) mainly occurs on subbasins with short main channel length (<18 km). This can be interpreted from the channel routing equations (8), (9), and (10) in SWAT. The main channel with short length reduces the channel flow travel time (TT) to a relatively small value, which makes the storage coefficient (SC) approaches its upper limit of 1.0 (10). Thus, the V_{stored} is decreased to zero value of (8) and (9) frequently in the model running process, which leads to the update of this variable with no significance. Therefore, the subbasin with short main channel tends to have less significant update of V_{stored} in the assimilation. Besides, it seems that the NER displays an increasing trend with the increase of the main channel length, which indicates the strong dependency of the state update of the reach flow storage on the length of the reach itself.

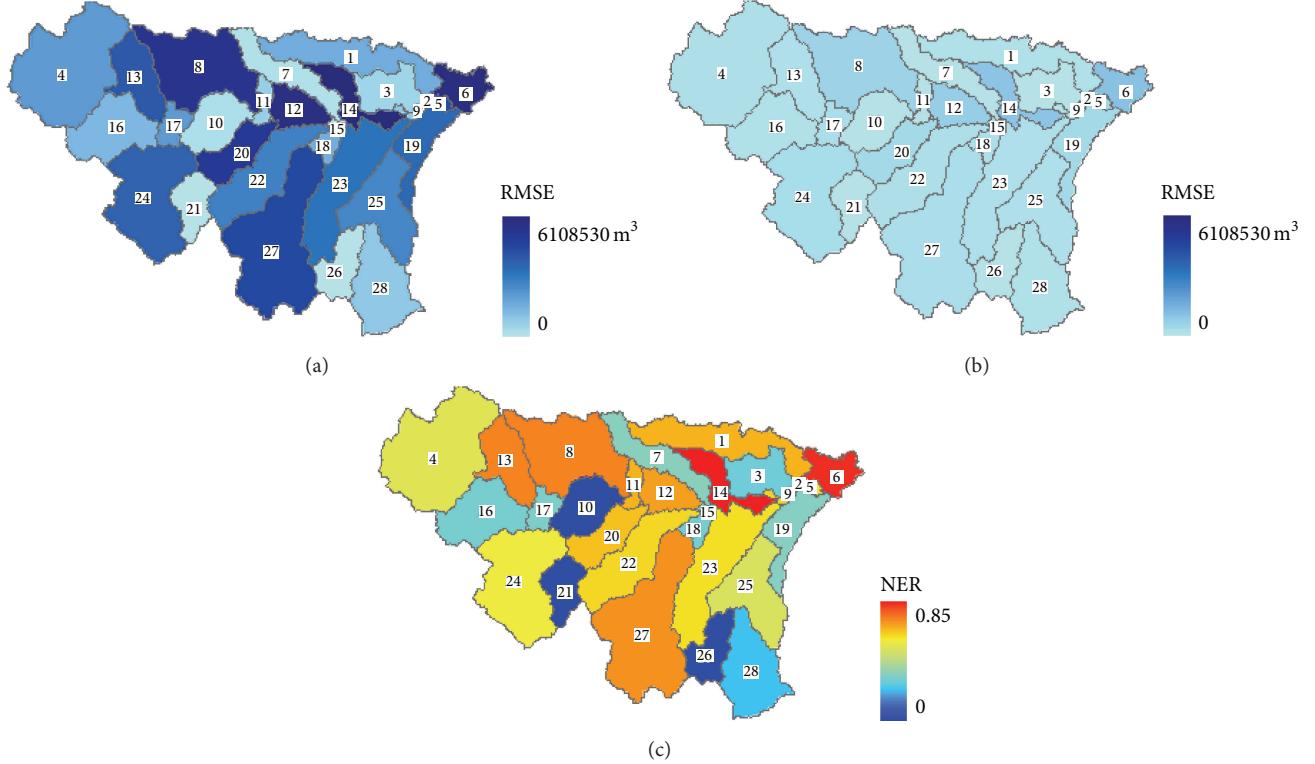


FIGURE 7: RMSE of the estimated river flow storage (V_{stored}) on 28 subbasins by (a) EnOL and (b) EnKF for the whole assimilation period. (c) NER (18) of the estimated V_{stored} . 1 to 28 are the subbasin number.

5.1.2. Assimilation of the Streamflow Observations at Catchment Outlet. Figure 10 compares the root mean square error (RMSE) of the estimated streamflow at all subbasin outlets (located at different sites of the reach) in the case of the assimilation of the streamflow observations at the catchment outlet (HB, i.e., Huaibin station) to that without streamflow assimilation for the whole assimilation period. It can be seen that the estimated streamflow at the outlet of several subbasins (e.g., subbasins 1, 5, 6, 9, and 14) is improved by assimilating the streamflow at the catchment outlet as their RMSE are reduced. The outlet streamflow of the whole basin (the outlet of subbasin 6) obtains the most significant improvement in terms of the degree of the RMSE reduction. This can be better illustrated by the normalized reduction of RMSE (i.e., NER), the distribution of which for all subbasin outlets is present in Figure 11. To make the NER distribution on the basin clear, the NER values at the subbasin outlet are put on the whole subbasin where it is located. It seems that the improvement of the streamflow modeling shows a decreasing trend from downstream to upstream by assimilating the catchment outlet streamflow observations, especially for the sites located at the main reach. This trend is consistent with that of the correlation coefficients between the simulated streamflow at the catchment outlet and that at the subbasin outlets (Figure 12), as they also present a decreasing trend from downstream to upstream on the whole basin. Hence, it can be inferred that the assimilation of the streamflow observations tends to be effective in improving the streamflow estimation for the catchments where the

streamflow modeling is closely related to that of the measurement predicted streamflow.

5.2. Real Data Application

5.2.1. Assimilation of the Streamflow Measurements at the Five Interior Gauges. Figure 13 compares the estimated streamflow at the catchment outlet obtained by EnOL and EnKF under the assimilation of the streamflow measurements at the five interior gauges. It can be seen that the simulated streamflow in EnOL is considerably improved (e.g., the day around 260, 320, and 450) by the EnKF based state update for the entire data assimilation period, which can be fully illustrated by the reduction of RMSE from $271.58 \text{ m}^3/\text{s}$ to $204.7 \text{ m}^3/\text{s}$ with a reduction of 25% (NER = 25%) and the increase of NSE from 0.62 to 0.78 and the correlation coefficient from 0.79 to 0.9. However, it can also be found from Figures 13(b) and 13(c) that this assimilation shows some limited capability in the error correction of streamflow prediction when there are large biases between the simulated and the observed streamflow. This limitation of the streamflow assimilation performance might be related to the unsatisfied (or possibly biased) model parameters, which can hardly be avoided in parameter calibration, especially for the distributed models with complicated model physics and structures and large numbers of parameters. Besides, the biased parameters will deteriorate the state update of hydrological modeling in the assimilation process [10].

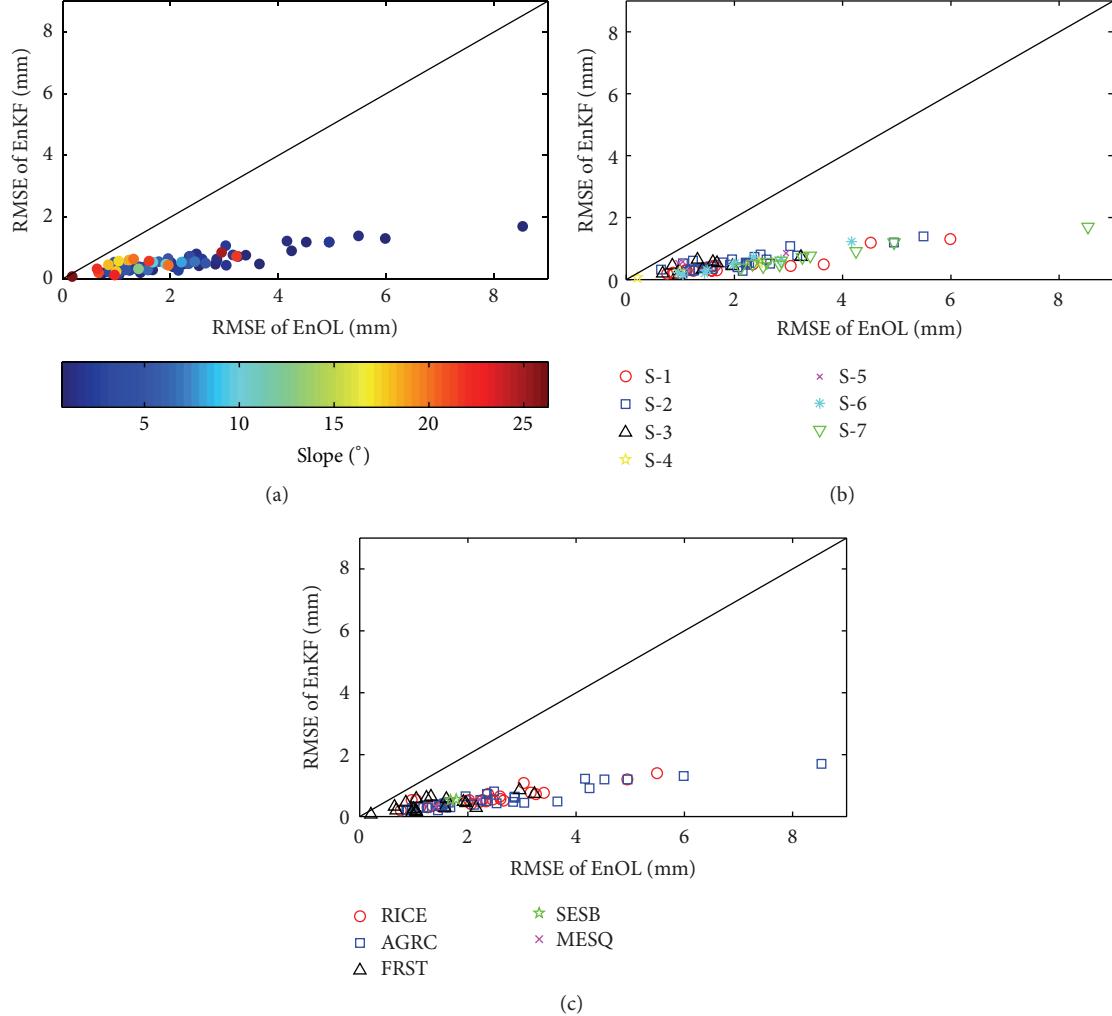


FIGURE 8: Comparison of the RMSE (15) for the estimated surface runoff storage (Q_{stor}) on 82 HRUs by EnKF and EnOL with varying (a) slope steepness, (b) soil category, and (c) land cover. S-1, S-2, S-3, S-4, S-5, S-6, and S-7 are the seven types of soil (Table 1). RICE, AGRC, FRST, SESB, and MESQ represent the land cover of rice, mixed agriculture, mixed forest, the *Sesbania*, and honey mesquite, respectively. The solid black line is the 1:1 line.

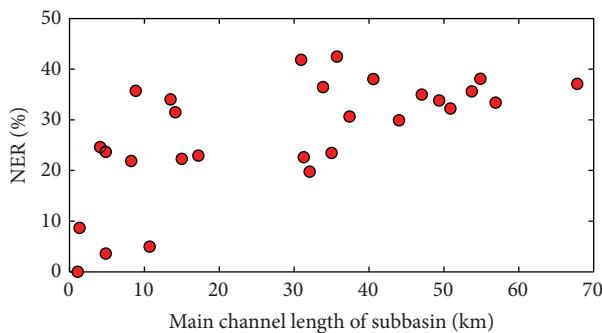


FIGURE 9: The NER (18) of the estimated reach flow storage on 28 subbasins with varying main channel length of subbasins.

5.2.2. Assimilation of the Streamflow Measurements at the Catchment Outlet

Table 5 summarizes the statistics of the

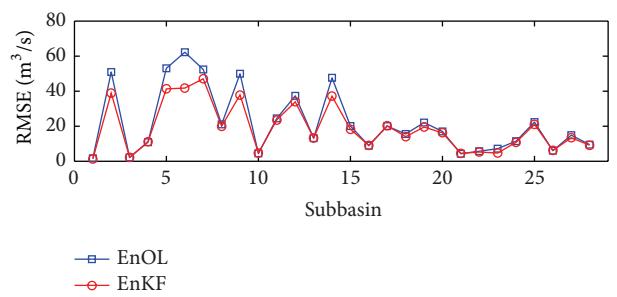


FIGURE 10: The RMSE (15) of the estimated streamflow obtained by EnOL and EnKF at the subbasin outlet in the assimilation of the streamflow observations at the catchment outlet.

estimated streamflow by assimilating the streamflow measurements at the catchment outlet (Huabbin station) comparing with the fact that no data assimilation is performed. It

TABLE 5: The statistics of the estimated streamflow obtained by EnOL and EnKF at the five interior discharge gauges.

Discharge gauges	EnOL			EnKF			NER (%)
	RMSE (m^3/s)	NSE	CC	RMSE (m^3/s)	NSE	CC	
Dapoling	39.82	0.46	0.71	39.01	0.49	0.71	2.03
Changtaiguan	72.99	0.41	0.66	72.96	0.41	0.66	0.04
Zhuganfu	56.51	0.64	0.81	60.12	0.6	0.78	-6.4
Huangchuan	67.36	0.59	0.77	68.07	0.58	0.76	-1.05
Xixian	199.98	0.55	0.74	187.38	0.61	0.78	6.3

Note: CC represents the correlation coefficient.

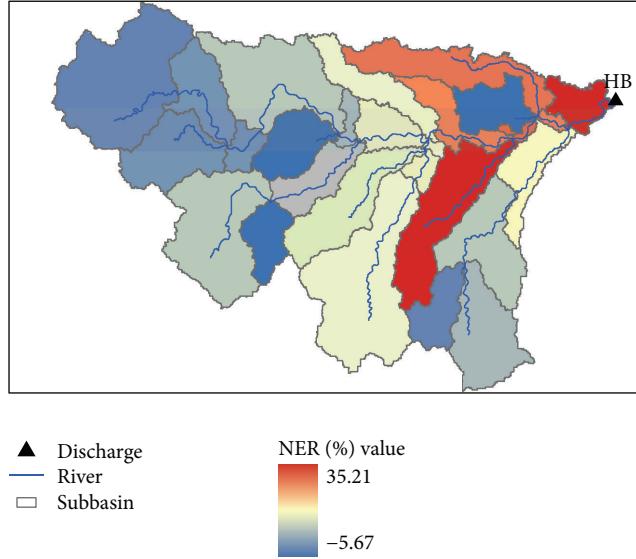


FIGURE 11: The NER (18) of the estimated streamflow for all subbasin outlets obtained by assimilating the streamflow observations at the catchment outlet (HB station). The NER value at the subbasin outlet is displayed on the subbasin where it is located.

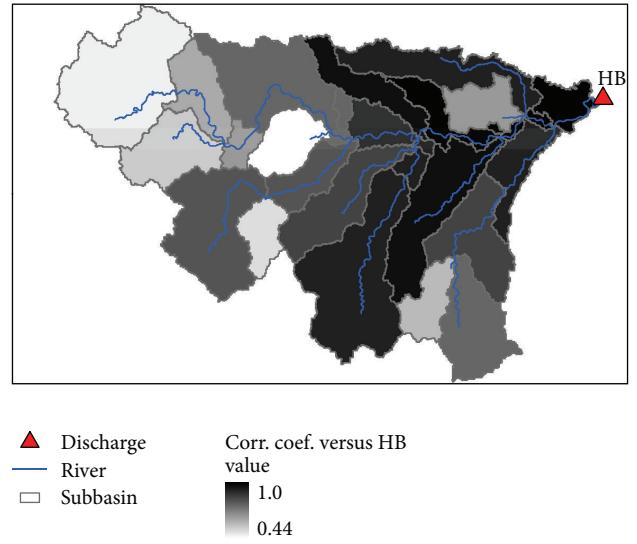


FIGURE 12: The correlation coefficients (CC) between the simulated streamflow at all subbasin outlets and that at the catchment outlet (HB station). The value of CC for one subbasin outlet is displayed on the subbasin where it is located. “Corr. Coef. versus HB” indicates the correlation coefficients between the streamflow at the subbasin outlet and that at HB station.

can be found that the streamflow at Xixian station obtains a considerable improvement by the catchment outlet streamflow assimilation, as the RMSE is decreased and both the NSE and CC are increased by EnKF. However, the simulated streamflow at Zhuganfu and Huangshan is worse than that without streamflow assimilation. It might be caused by the over updates of the model state variables as the correlation coefficients between the streamflow at the measurement site (Huainan station) and that at the interior gauges for model simulation are comparatively higher than that for observation (Figure 14). This can be related to the inadequacy in modeling the spatial variability of the hydrological process due to the insufficient spatial variability of model inputs and parameters [13]. Besides, the assimilation of the catchment outlet streamflow has nonsignificant impacts on the streamflow modeling of Dapoling and Changtaiguan, which can be partly explained by the relatively low correlation coefficients between their simulated streamflow and that at the measurement sites (Huainan station).

6. Conclusions

This study uses both the synthetic experiment and real data application to investigate the capability of streamflow assimilation in improving the streamflow predictions of upper Huai River basin via updating the basin states in distributed hydrological modeling process. The two sensitive model state variables (i.e., the surface runoff storage and the reach flow storage) in SWAT are updated using the streamflow observations (1) at the five interior sites and (2) at the catchment outlet.

The synthetic experiment shows that the predicted streamflow at all subbasin outlets is improved by assimilating the synthetic streamflow observations at the five interior sites. This improvement shows an increasing trend from upstream to downstream of the main reach, and the predicted streamflow at the catchment outlet obtains the most significant improvement. This improvement of the catchment streamflow modeling is benefited from the updating of

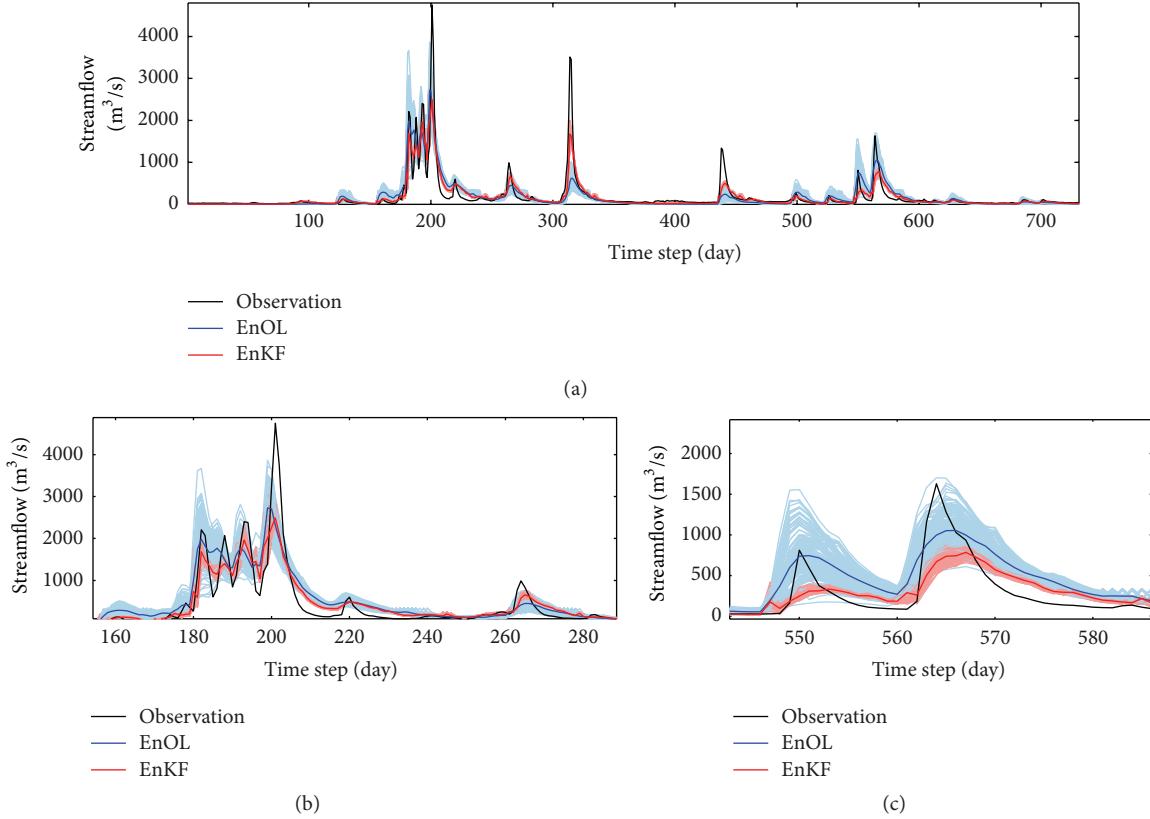


FIGURE 13: Comparison of the simulated streamflow by EnKF and EnOL at the basin outlet (HB station) in real data application: (a) the whole period from Jan. 1, 1996 to Dec. 31, 1997; (b) and (c) details of the period with high flow. The light blue and red lines represent the ensemble members of EnOL and EnKF, respectively.

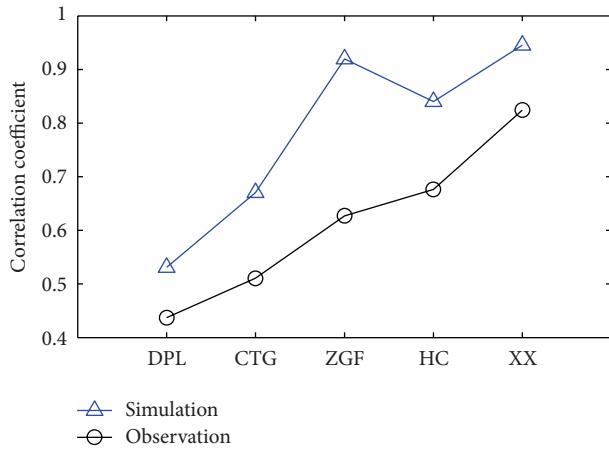


FIGURE 14: The correlation coefficient between the streamflow at the catchment outlet (HB station) and that at the five interior gauges (DPL, CTG, ZGF, HC, and XX station) for both observation and model simulations.

the surface runoff storage and reach flow storage of the whole basin as the two state variables significantly approach their synthetic truth after assimilation. It is found that the significant improvement of the surface runoff storage generally occurs on the calculation units with the slopes

less than 5°. We also find that the main channel length of the subbasins has a considerable impact on the updating of reach flow storage (V_{stored}), as V_{stored} with short main channel tend to get less significant improvement compared to that with long main channel. Besides, the assimilation of the streamflow observations at the catchment outlet improves the estimated streamflow at the outlet of several subbasins near the catchment outlet. The improvement of the streamflow modeling shows a decreasing trend from downstream to upstream, especially for the sites located at the main reach, which is consistent with that of the correlation coefficients between the simulated streamflow at the catchment outlet and that at the subbasin outlets.

The real data application shows that a large improvement of the estimated streamflow at the catchment outlet is obtained by updating the basin states using the in situ streamflow measurements at the five interior gauges in rainfall-runoff modeling. The assimilation of the in situ catchment outlet streamflow measurements improves the streamflow modeling at Xixian, produces a slight worse impact on the streamflow estimation at Zhuganfu and Huangchuan, and has nonsignificant influence on the streamflow modeling of the upstream of Dapoling and Changtaiguan. In this study, the method for model and observation error parameter estimation is subjective and empirical to a certain extent as it generally relies on experience and conclusions from

other previous studies. More rigorous and theoretical basis methods for quantifying the model and observation error parameters, for example, the adaptive assimilation approach [2, 26] and the maximum a posterior approach [27], should be attempted in future studies. Besides, the streamflow assimilation performance in distributed hydrological modeling relies on the capability of the distributed model in characterizing the spatial variability of the real hydrological process, which is usually controlled by the spatial variability of the model inputs and parameters. More works should be paid to improving the spatial variability of model forcing data, and new and optimal methods for parameter estimations and calibrations are needed to produce more accurate and reliable model parameters on spatial scale [28].

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

This study was supported by the National Natural Science Fund Project of China, the Response Mechanism of the Watershed Hydrological Processes under Meteorological Drought (41371050), the Specialized Research Fund for the Doctoral Program of Higher Education of China (20130094110007), the Fundamental Research Funds for the Central Universities (2015B05514), and the Graduate Research and Innovation Program for Ordinary University of Jiangsu Province, China (CXZZ13_0248). The authors thank Susan-steel Dunne from Delft University of Technology for her precious suggestions.

References

- [1] N. K. Ajami, Q. Duan, and S. Sorooshian, “An integrated hydrologic Bayesian multimodel combination framework: confronting input, parameter, and model structural uncertainty in hydrologic prediction,” *Water Resources Research*, vol. 43, no. 1, Article ID W01403, 2007.
- [2] R. H. Reichle, W. T. Crow, and C. L. Keppenne, “An adaptive ensemble Kalman filter for soil moisture data assimilation,” *Water Resources Research*, vol. 44, no. 3, Article ID W03423, 2008.
- [3] J. A. Vrugt, H. V. Gupta, B. Ó. Nualláin, and W. Bouten, “Real-time data assimilation for operational ensemble streamflow forecasting,” *Journal of Hydrometeorology*, vol. 7, no. 3, pp. 548–565, 2006.
- [4] W. T. Crow and D. Ryu, “A new data assimilation approach for improving runoff prediction using remotely-sensed soil moisture retrievals,” *Hydrology and Earth System Sciences*, vol. 13, no. 1, pp. 1–16, 2009.
- [5] X. Xie and D. Zhang, “Data assimilation for distributed hydrological catchment modeling via ensemble Kalman filter,” *Advances in Water Resources*, vol. 33, no. 6, pp. 678–690, 2010.
- [6] Y. Li, D. Ryu, A. W. Western, Q. J. Wang, D. E. Robertson, and W. T. Crow, “An integrated error parameter estimation and lag-aware data assimilation scheme for real-time flood forecasting,” *Journal of Hydrology*, vol. 519, pp. 2722–2736, 2014.
- [7] C. Alvarez-Garreton, D. Ryu, A. W. Western, W. T. Crow, and D. E. Robertson, “The impacts of assimilating satellite soil moisture into a rainfall–runoff model in a semi-arid catchment,” *Journal of Hydrology*, vol. 519, pp. 2763–2774, 2014.
- [8] C. Alvarez-Garreton, D. Ryu, A. W. Western et al., “Improving operational flood ensemble prediction by the assimilation of satellite soil moisture: comparison between lumped and semi-distributed schemes,” *Hydrology and Earth System Sciences*, vol. 19, no. 4, pp. 1659–1676, 2015.
- [9] D. Aubert, C. Loumagne, and L. Oudin, “Sequential assimilation of soil moisture and streamflow data in a conceptual rainfall–runoff model,” *Journal of Hydrology*, vol. 280, no. 1–4, pp. 145–161, 2003.
- [10] H. Moradkhani, S. Sorooshian, H. V. Gupta, and P. R. Houser, “Dual state-parameter estimation of hydrological models using ensemble Kalman filter,” *Advances in Water Resources*, vol. 28, no. 2, pp. 135–147, 2005.
- [11] A. H. Weerts and G. Y. H. El Serafy, “Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models,” *Water Resources Research*, vol. 42, no. 9, Article ID W09403, 2006.
- [12] Y. Li, D. Ryu, A. W. Western, and Q. J. Wang, “Assimilation of stream discharge for flood forecasting: the benefits of accounting for routing time lags,” *Water Resources Research*, vol. 49, no. 4, pp. 1887–1900, 2013.
- [13] M. P. Clark, D. E. Rupp, R. A. Woods et al., “Hydrological data assimilation with the ensemble Kalman filter: use of streamflow observations to update states in a distributed hydrological model,” *Advances in Water Resources*, vol. 31, no. 10, pp. 1309–1324, 2008.
- [14] G. Evensen, “Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics,” *Journal of Geophysical Research*, vol. 99, no. 5, pp. 10143–10162, 1994.
- [15] H. Lee, D.-J. Seo, and V. Koren, “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,” *Advances in Water Resources*, vol. 34, no. 12, pp. 1597–1615, 2011.
- [16] M. B. Smith, D.-J. Seo, V. I. Koren et al., “The distributed model intercomparison project (DMIP): motivation and experiment design,” *Journal of Hydrology*, vol. 298, no. 1–4, pp. 4–26, 2004.
- [17] S. L. Neitsch, J. G. Arnold, J. R. Kiniry, and J. R. Williams, *Soil and Water Assessment Tool Theoretical Documentation Version 2009*, Texas Water Resources Institute, College Station, Tex, USA, 2011.
- [18] R. E. Kalman, “A new approach to linear filtering and prediction problems,” *Journal of Basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.
- [19] S. Dunne and D. Entekhabi, “An ensemble-based reanalysis approach to land data assimilation,” *Water Resources Research*, vol. 41, no. 2, Article ID W02013, 2005.
- [20] Z. Yu, X. Fu, L. Luo et al., “One-dimensional soil temperature simulation with Common Land Model by assimilating in situ observations and MODIS LST with the ensemble particle filter,” *Water Resources Research*, vol. 50, no. 8, pp. 6950–6965, 2014.
- [21] J. E. Nash and J. V. Sutcliffe, “River flow forecasting through conceptual models part I—a discussion of principles,” *Journal of Hydrology*, vol. 10, no. 3, pp. 282–290, 1970.

- [22] H. Lü, T. Hou, R. Horton et al., "The streamflow estimation using the Xinanjiang rainfall runoff model and dual state-parameter estimation method," *Journal of Hydrology*, vol. 480, pp. 102–114, 2013.
- [23] H. Chen, D. Yang, Y. Hong, J. J. Gourley, and Y. Zhang, "Hydrological data assimilation with the Ensemble Square-Root-Filter: use of streamflow observations to update model states for real-time flash flood forecasting," *Advances in Water Resources*, vol. 59, pp. 209–220, 2013.
- [24] G. Evensen, "The Ensemble Kalman filter: theoretical formulation and practical implementation," *Ocean Dynamics*, vol. 53, no. 4, pp. 343–367, 2003.
- [25] Z. Ren, G. Wang, and F. Zou, "The research of precipitation measurement errors in China," *Acta Mechanica Sinica*, vol. 61, no. 5, pp. 621–627, 2003.
- [26] W. T. Crow and R. H. Reichle, "Comparison of adaptive filtering techniques for land surface data assimilation," *Water Resources Research*, vol. 44, no. 8, Article ID W08423, 2008.
- [27] Q. J. Wang, D. E. Robertson, and F. H. S. Chiew, "A Bayesian joint probability modeling approach for seasonal forecasting of streamflows at multiple sites," *Water Resources Research*, vol. 45, no. 5, Article ID W05407, 2009.
- [28] T. Xu and A. J. Valocchi, "A Bayesian approach to improved calibration and prediction of groundwater models with structural error," *Water Resources Research*, vol. 51, no. 11, pp. 9290–9311, 2015.



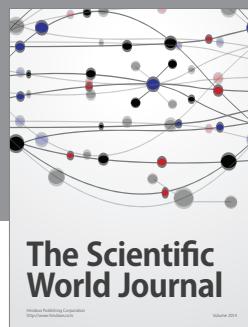
Journal of
Geochemistry



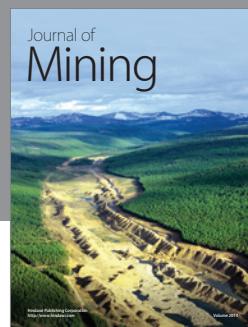
Scientifica



International Journal of
Ecology



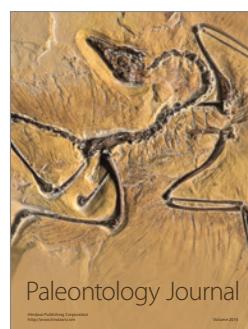
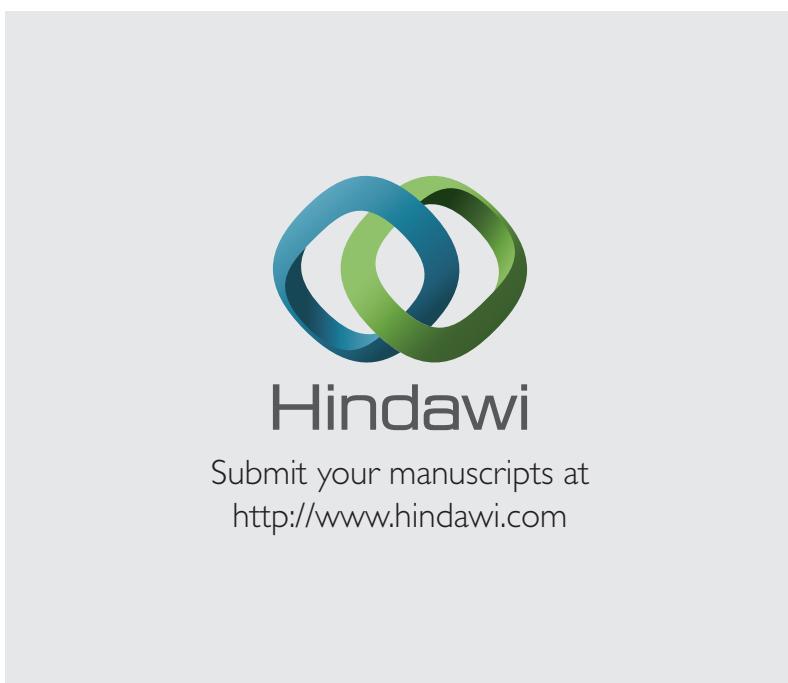
The Scientific
World Journal



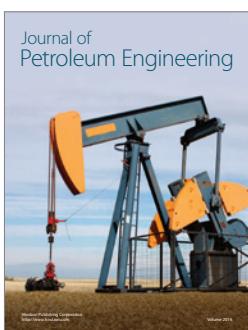
Journal of
Mining



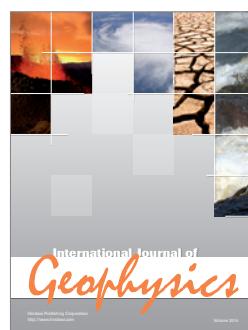
Journal of
Earthquakes



Paleontology Journal



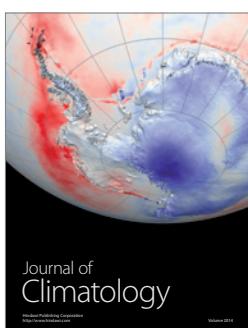
Journal of
Petroleum Engineering



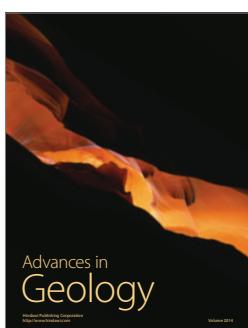
International Journal of
Geophysics



Advances in
Meteorology



Journal of
Climatology



Advances in
Geology



Advances in
Oceanography



International Journal of
Oceanography



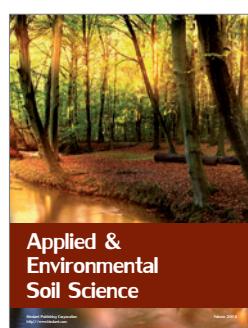
International Journal of
Mineralogy



Journal of
Geological Research



International Journal of
Atmospheric Sciences



Applied &
Environmental
Soil Science



Journal of
Computational
Environmental Sciences