Research and application of parameter estimation method in hydrological model based on dual ensemble Kalman filter

Mengtian Lu, Sicheng Lu, Weihong Liao, Xiaohui Lei, Zhaokai Yin and Hao Wang

ABSTRACT

Although field measurements and using long hydrological datasets provide a reliable method for parameters’ calibration, changes in the underlying basin surface and lack of hydrometeorological data may affect parameter accuracy in streamflow simulation. The ensemble Kalman filter (EnKF) can be used as a real-time parameter correction method to solve this problem. In this study, five representative Xin’anjiang model parameters are selected to study the effects of the initial parameter ensemble distribution and the specific function form of the parameter on the EnKF parameter estimation process for both single and multiple parameters. Results indicate: (1) the method of parameter calibration to determine the initial distribution mean can improve the assimilation efficiency; (2) there is mutual interference among the parameters during multiple parameters’ estimation which invalidates some conclusions of single-parameter estimation. We applied and evaluated the EnKF method in Jinjiang River Basin, China. Compared to traditional approaches, our method showed a better performance in both basins with long hydro-meteorological dataset (an increase of KGE from 0.810 to 0.887 and a decrease of bias from −1.08% to −0.74%); and in basins with a lack of hydrometeorological data (an increase of KGE from 0.536 to 0.849 and a decrease of bias from −15.55% to −11.42%).

Key words: ensemble Kalman filter, parameter estimation, sensitivity analysis, underlying surface change, ungauged basin

INTRODUCTION

In hydrological models, the parameter is the quantity describing the physical characteristics of the underlying surface, which is usually time invariant. With the development of hydrological science, some hydrological models have become more accurate in describing physical processes, but also more complex and contain more parameters (Zhao 1992; Arnold & Fohrer 2005). Most of the parameters are not easy to obtain directly through the basin measurement, thus, to obtain the accurate parameters, we need a long series of hydrological data, which require optimization to calibrate the parameters. For example, the impervious area ratio Im of the Xin’anjiang model can now be obtained by analyzing the land use of the watershed through remote sensing satellite images, but civil remote sensing satellites will not monitor the target watershed in real time, so the obtained data generally lags behind, resulting in the reduction of the accuracy of this parameter. What is more, the deep evapotranspiration coefficient parameter Ci is related to the proportion of deep-rooted plants in the watershed area. This parameter can only be obtained through artificial ground measurement, so it is a heavy workload to obtain the latest data for large watersheds. Therefore, a long series of rainfall–runoff data are needed to calibrate the parameters by...
using the optimization algorithm, so as to obtain more accurate parameters. The underlying surface condition of a basin usually changes with time, and there are some deviations in the data used for parameter calibration. These problems will be ignored when using fixed model parameters (Pathiraja et al. 2016). In addition, parameter calibration cannot be performed in ungauged basins. Usually, an interpolation algorithm (Wagner et al. 2012) or the parameter of a similar basin (Micovic & Quick 1999; Reichle & Walker 2002) is used to solve this problem. The above conditions may make the parameters deviate from the optimal value to a certain extent and reduce simulation accuracy. Therefore, it is necessary to apply the real-time parameter estimation method to the hydrological model to correct the parameters.

The convincing performance of data assimilation (DA) in real-time correction of hydrological models has attracted the attention of hydrologists. DA is a development process used to optimally combine model simulations and observations with appropriate modeling. Many DA methods have been applied in hydrology. Among them, the most commonly used methods are variational method, particle filters (PFs) and the Kalman filters (KFs). Compared with other methods, KFs are easier to implement and can produce even better results with lower computational requirements. In addition, KFs are very flexible in coupling with hydrological models and have more derived variants than any other method.

Kalman proposed the Kalman filter in 1958 (Kalman 1960), which is similar to the least-squares method in that it aims to minimize the error and only applies to linear models. In reality, hydrological models are usually non-linear. When the Kalman filter is applied to the non-linear model, filter divergence is often caused by error accumulation (Mirzaei & Roumeliotis 2007). For the non-linear model, Kalman proposed an extended Kalman filter (EKF), which has been successfully applied to the biological field (Svrcek et al. 1974). However, the EKF may still diverge under the condition of strong non-linearity. When Evensen applied EKF to the multilayer quasi-geostrophic model, he found the reason for the divergence of the EKF, and proposed the ensemble Kalman filter (EnKF), which eliminated the divergence phenomenon caused by the infinite error growth in the EKF due to an overly simplified closure in the error covariance equation (Evensen 1994). Later studies also reported EnKF was better than the EKF in performance (Reichle & Walker 2002). Burgers et al. (1998) improved EnKF and proposed that one should add random perturbations with the correct statistics to the observations and generate an ensemble of observations that is then used to update the ensemble of model states. In the following ten years, there were still many researchers trying to improve EnKF (Anderson 2001; Bishop et al. 2001; Whitaker & Hamill 2002; Tippett et al. 2003; Vrugt et al. 2005; Clark et al. 2008). The earliest EnKF is applied to state estimation (Houtekamer & Mitchell 1998, 2001). This method can not only estimate state variables but also use observations to estimate the value of imprecisely known model parameters (Anderson 2001).

For the first time, Moradkhani et al. (2004) applied EnKF in the field of hydrology, combined the parameter ensemble and variable ensemble to form a dual-parameter-state ensemble, used observations to estimate model parameters, and proved its applicability and practicability in the prediction. EnKF has been widely used in the field of hydrology and water resources. There are three ways to use EnKF: (1) assimilate observations to estimate the state (Reichle & Walker 2002; Komma et al. 2008; Xie & Zhang 2008; Fu et al. 2014; Lu et al. 2016); (2) assimilate observations to estimate the parameters (Pathiraja et al. 2016; Lei et al. 2019); (3) assimilate observations to estimate the state and parameters (Vrugt et al. 2005; Lu et al. 2013; Xie & Zhang 2013; Liu et al. 2017). Previous studies considered the parameters as a whole to study their characteristics in the estimation process or to study the effect of parameters corrected by the EnKF on simulation results. There are few studies that focus on the effects of selecting different parameters for estimation and setting different initial ensemble distributions on the estimation process.

In this study, according to the sensitivity of parameters and the function of parameters in the hydrological model, five representative parameters of the Xin’anjiang model are selected as the research object. This study is divided into two parts. The first part is numerical simulation, which mainly studies the characteristics of EnKF when different parameters are selected for estimation from two aspects: the initial distribution of parameters’ ensemble and the function of parameters in the model. According to these rules, some suggestions are put forward to improve the efficiency of EnKF in the hydrological model. The second part is the application, in which EnKF is applied to a new basin: Jinjiang River Basin, China, and it is verified if the EnKF method can solve the problem of prediction accuracy decline caused by the change of underlying surface, the lack of data and other reasons, and the correctness of some suggestions put forward in the first part is further verified.

**METHODS**

As mentioned above, this study selects five representative Xin’anjiang model parameters to study the applicability and practicability of the ensemble Kalman filter (EnKF) in hydrological models.
The Xin’anjiang model, which was originally used in flood prediction of Xin’anjiang hydropower station in China, was developed by Zhao and was officially published in 1980 (Zhao 1992). After decades of improvement, it has become one of the most influential hydrological models in the world, and has been widely used in streamflow simulation, flood prediction, and other aspects worldwide (Jayawardena et al. 2006; Yang et al. 2011; Rahman & Lu 2015). It is a conceptually based model based on the conception of hillslope hydrology. In this paper, the most commonly used improved form of the Xin’anjiang model, the three-source Xin’anjiang model, is used for research. It divides the whole watershed into many sub-watershed units. The parameters of the model have clear physical meaning, and the parameters are basically independent of each other. Compared with the past lumped hydrological models such as the two-source Xin’anjiang model, hymod model and SCS model, the three-source Xin’anjiang model is more detailed and more in line with the real situation in terms of evaporation, soil water, runoff division, and slope confluence. The infiltration module of the model calculates the total runoff generated by rainfall according to the concept of excess storage runoff. Water in the soil is divided into tension water (water below the field capacity) and free water (water above the field capacity), and their regulation and storage effects are considered, respectively. In the evaporation calculation, the model divides the tension water into upper layer, middle layer, and lower layer to calculate the evapotranspiration, respectively. In this model, the total runoff is divided into saturated surface runoff, soil water runoff, and groundwater runoff by using a simplified free water storage reservoir. In terms of routing calculation, the unit hydrograph method is generally used for surface runoff, and the linear reservoir method is used for routing of soil flow and underground flow. The inputs of the model are precipitation and potential evaporation. The precipitation is directly obtained from the hydrological station, and the potential evaporation is calculated from the meteorological data.

As there are many parameters in the Xin’anjiang model, the dynamically dimensioned search (DDS) algorithm, which is suitable to deal with the problem of parameter optimization in large calculations, is selected to calibrate the model parameters in the application. The objective function is Nash–Sutcliffe efficiency coefficient (NSE). The DDS algorithm was proposed by Tolson & Shoemaker (2007). This algorithm does not need to adjust the algorithm parameters because it has fast convergence speed and easily avoids local optimization. When dealing with the parameter optimization problem in large calculations, the performance of this algorithm is better than the traditional global optimization algorithms such as SCE. The DDS algorithm and its improved algorithm are widely used in model parameter calibration with more undetermined parameters (Huang et al. 2014; Yen et al. 2016).

The global sensitivity of the model parameters was analyzed using the LH-OAT method. The accuracy of streamflow simulation is evaluated by NSE, Kling–Gupta efficiency (KGE) and bias. In the numerical simulation experiment, the correlation coefficient is selected as the correlation index between the parameter ensemble and streamflow ensemble simulations at a certain time step. Some algorithms and evaluation indexes used in the study will be described in detail below.

**LH-OAT global sensitivity analysis method**

The LH-OAT sensitivity analysis method is a global sensitivity analysis method combining the Latin hypercube sampling (LH) and the one-factor-at-a-time algorithm (OAT) proposed by Griensven et al. (2006).

The sensitivity analysis steps are as follows:

1. Using LH sampling, the space is divided into \( n \) layers, and each layer is randomly sampled once to generate \( n \) LH sampling sets (including \( m \) parameters).
2. According to OAT, each parameter is changed slightly, so that each sampling set can obtain \( m \) group parameters.
3. \( n \times (1+m) \) group parameters are substituted into the model for calculation, and an objective function is selected to evaluate the calculation results.
4. Evaluate the sensitivity of each parameter according to the following equation:

\[
RS_{ij} = \left| \frac{M[e_1, e_2, \ldots, e_i(1+f_j), \ldots, e_p] - M[e_1, \ldots, e_p]}{M[e_1, \ldots, e_i(1+f_j), \ldots, e_p] + M[e_1, \ldots, e_p]} \right| \tag{1}
\]

where \( M[f] \) is the objective function; \( f_j \) is the minor change in parameter, which is directly proportional to the range of parameter values; \( e_i \) is the \( i \)th parameter, \( i \in [1, p] \); \( p \) is the total number of parameters; \( RS_{ij} \) is the relative sensitivity of parameters considering the range of parameter value; and \( j \) is a LH sampling set, \( j \in [1, n] \).
Kalman filtering

Standard Kalman filter

The Kalman filter is an unbiased estimation based on minimizing the variance of the analysis error under the assumption that the noise is Gaussian white noise in a linear system. The estimation process is as follows:

State prediction equation. According to the following equation, the real state of the next time can be obtained through the state of the current time:

\[ x_t = Fx_{t-1} + w_{t-1} \]  

where \( F \) is the state transition matrix that infers the state vector of time \( t \) from the state vector of time \( t-1 \), \( x_t \) is the real state vector of time \( t \), \( x_{t-1} \) is the real state vector of time \( t-1 \), and \( w_{t-1} \) is the model error, which obeys the Gaussian distribution.

According to the following equation, the predictions of state vector of the next time can be obtained:

\[ \hat{x}_t = F\hat{x}_{t-1} \]  

where \( \hat{x}_{t-1} \) is the analysis value of the state vector at time \( t-1 \), and \( \hat{x}_t \) is the prediction of the state vector at the next time.

Measurement equation. According to the following equation, the observation is represented by the real state vector at the current time:

\[ z_t = Hx_t + v_t \]  

where \( z_t \) is the observation value at time \( t \), \( H \) is the transition matrix that transforms the state vector into the observation, and \( v_t \) is the observation error, which obeys the Gaussian distribution.

State update equation. According to the following equation, when both the predictions and the observations have been obtained, a weight can be given by the Kalman gain matrix to fuse the observations and predictions:

\[ x_t^- = \hat{x}_t + K_t(z_t - H\hat{x}_t) \]  

where \( \hat{x}_t \) is the prediction of the state vector at time \( t \), \( K_t \) is the Kalman gain matrix, \( x_t^- \) is the updated analysis value of the state vector at time \( t \), and \( (z_t - H\hat{x}_t) \) is innovation.

Kalman gain matrix. \( P_t^- = E[(x_t^- - \hat{x}_t)(x_t^- - \hat{x}_t)^T] \) is the error in covariance matrix between the analysis value of the state vector and prediction. We make the trace of the above matrix 0 (to minimize the matrix variance) and obtain the following equation:

\[ K_t = P_t^-H^T(HP_t^-H^T + R_t) \]  

where \( R_t \) is the covariance matrix of observation error.

Error covariance matrix. We place \( K_t \) into \( P_t^- = E[(x_t^- - \hat{x}_t)(x_t^- - \hat{x}_t)^T] \) and obtain the following equation:

\[ P_t^- = (I - K_tH)P_t \]  

where \( I \) is the unit matrix. \( P_t \) can be recursive by the following equation:

\[ P_t = FP_{t-1}F^T + Q_t \]  

where \( Q_t \) is the covariance matrix of model error.
Ensemble Kalman filter

Equation (1) is changed to

\[ x_t = Mx_{t-1} + w_{t-1} \tag{9} \]

where \( M \) is a nonlinear prediction model.

The most important difference between the ensemble Kalman filter and the standard Kalman filter is that the ensemble Kalman filter establishes an initial state ensemble, and each element in the ensemble is taken as the initial value to start the assimilation. \( P_t \) in Equation (4) is no longer calculated by the recurrence formula in step 5, but by the following equation:

\[ P_t \approx E[(x_t - \bar{x}_t)(x_t - \bar{x}_t)^T] \tag{10} \]

where \( \hat{x}_i \) is the state vector at time \( t \), and \( \bar{x}_t \) is the state vector ensemble mean at time \( t \).

Dual ensemble Kalman filter

Equation (5) changes into

\[ X_t^- = \hat{X}_t + K_t(z_t - H\hat{X}_t) \tag{11} \]

where \( \hat{X}_t \) is the dual-parameter-state ensemble at time \( t \), \( \hat{X}_t = (x_t, p_{1t}, p_{2t}, \ldots, p_{mt}, \ldots, p_{nt}) \). \( p_{mt} \) is the vector of the value of the \( m \)-th parameter at time \( t \), and \( n \) is the total number of parameters to be evaluated. \( X_t^- \) is the updated analysis value of dual-parameter-state ensemble at time \( t \), and \( X_t^- = (x_t^-, p_{1t}^-, p_{2t}^-, \ldots, p_{mt}^-, \ldots, p_{nt}^-) \). \( p_{mt}^- \) is the vector of the analysis value of the \( m \)-th parameter at time \( t \).

Equation (10) changes into

\[ P_t \approx E[(\hat{X}_t - \bar{X}_t)(\hat{X}_t - \bar{X}_t)^T] \tag{12} \]

where \( \bar{X}_t \) is the mean value of the dual-parameter-state ensemble at time \( t \).

Generating a Gaussian white noise sequence

The noise with a Gaussian distribution is called Gaussian noise, and its power spectral density is a uniform distribution, which is called Gaussian white noise. In the ensemble Kalman filter, the initial values of the parameter and state ensembles are Gaussian white noise sequences. In this study, the approximate Gaussian white noise sequence is used as the initial value of each parameter ensemble. The generation steps are as follows:

(1) Generate uniformly distributed random numbers between 0 and 1 according to the following equations:

\[ r_i = (2053 \cdot r_{i-1} + 13849), \quad i = 1, 2, 3 \ldots \tag{13} \]

and

\[ \text{rnd}_i = \frac{(r_i - \text{mod}(r_i/m) \cdot m)}{m} \tag{14} \]

where \( \text{rnd}_i \) is the random number between 0 and 1, \( \text{mod}() \) is the rounding function, \( m = 2^{16} \), \( r_i \) is the random seed, and \( r_0 \) can be any value less than \( m \).

(2) Generate the Gaussian white noise sequence according to the following equation:

\[ y = \mu + \sigma \frac{\left( \sum_{i=1}^{n} \text{rnd}_i \right) - n}{\sqrt{n/12}} \tag{15} \]
where \( n \) is sufficiently large, the approximation of \( n=12 \) is quite good; \( \mu \) is the mean of noise, \( \sigma \) is the variance of noise, and \( y \) is the element of noise sequence.

Index

Nash–Sutcliffe efficiency coefficient (NSE)

In this study, the NSE is used as the objective function of parameter global sensitivity analysis and parameter calibration. The perfect value for NSE is 1.

\[
NSE = 1 - \frac{\sum_{t=1}^{n} (Q_t^o - Q_t^s)^2}{\sum_{t=1}^{n} (Q_t^o - \bar{Q}_o)^2}
\]

(16)

where \( Q_t^o \) is the flow simulation value at time \( t \), \( Q_t^s \) is the observation value at time \( t \), \( \bar{Q}_o \) is the mean of observation streamflow sequence, NSE is the Nash–Sutcliffe efficiency coefficient, and \( n \) is the length of streamflow sequence.

Kling–Gupta efficiency (KGE)

The KGE is used as the statistic metric to measure the performance of the evaluation period which equally weights correlation coefficient, streamflow bias, and streamflow standard deviation bias. The KGE function can be written as follows:

\[
KGE = 1 - \left( w_r (r - 1))^2 + (w_\sigma (\alpha - 1))^2 + (w_\beta (\beta - 1))^2 \right)
\]

(17)

where \( \alpha = \sigma_m / \sigma_o \) and \( \beta = \mu_m / \mu_o \), with \( w \) indicating the weight applied to each component. In the present study \( w = 1 \). KGE achieves an optimum value of 1.0 (when \( \mu_m = \mu_o \), \( \sigma_m = \sigma_o \), and \( r = 1 \)), and lower values indicate less skill. Here, \( \mu_m \) and \( \mu_o \) are the modeled and observed means, respectively, while \( \sigma_m \) and \( \sigma_o \) are the modeled and observed standard deviations, respectively. The symbol \( r \) denotes the correlation coefficient.

Correlation coefficient (r)

In this study, the correlation coefficient is used as the correlation index between parameter ensemble and streamflow ensemble at a certain time step in the numerical simulation experiment. When \( r \) is positive, it means positive correlation, and when \( r \) is negative, it means negative correlation. The greater the absolute value of \( r \), the stronger the correlation.

\[
r(x, y) = \frac{Cov(x, y)}{\sqrt{Var|x|Var|y|}}
\]

(18)

Bias

In this study, bias is used as an index to evaluate the water balance. The perfect value for bias is 0.

\[
BIAS = \frac{\sum_{t=1}^{n} (Q_t^o - Q_t^s)}{\sum_{t=1}^{n} Q_t^o}
\]

(19)

STUDY AREA AND DATA

The Jinjiang River is a tributary of Ganjiang River, a first-order tributary of the Yangtze River in China. It flows into the left bank of Ganjiang River, with a length of 294 km, a drainage area of 7,650 km², and an average annual flow of 222 m³/s. The forest coverage of the Jinjiang River Basin is over 70%. The rainfall is rich, with an average annual precipitation of 1,400–1,700 mm. The rainstorms are mainly from May to July. In this paper, the main study area is the drainage basin of Gaoan hydrological station, which is located on Jinjiang River, with an area of 6,215 km² (Figure 1). The elevation of the catchment...
is higher in the northwest, where most tributaries of the Jinjiang River originate, ranging from 18 to 1,096 m (Yin et al. 2018).

There is no large reservoir in the basin, and it is less disturbed by human activities than other basins. It is a representative natural basin, and it can obtain better streamflow simulation results without a coupling reservoir operation model or other models; therefore, it is suitable to study the adaptability and practicability of EnKF as the hydrological model itself.

The Danjiang River is a tributary of Hanjiang River in the north of the middle reaches of the Yangtze River, with a drainage area of 16,812 km$^2$ and an average annual precipitation of 400–1,100 mm. There are dense vegetation and primitive forests in the basin. Another study area of this paper is the drainage basin of Jingziguan hydrological station, which is located on the Danjiang River, with an area of 7,060 km$^2$ (Figure 1). The elevation of the basin is 163–1,970 m, and the river flows from the northwest to the southeast.

In this study, the study area of the numerical simulation part is Jinjiang River Basin; the main study area of the application part is Jinjiang River Basin, and the Danjiang River Basin is used to provide the initial distribution mean of parameters ensemble in the case of lack of data in Jinjiang River Basin.

The daily precipitation observation data of 20 precipitation stations from 2001 to 2009 and the hydrological station daily streamflow observation data from 2001 to 2009 were used to simulate the streamflow of the Gaoan hydrological station. The daily precipitation observation data of 14 precipitation stations from 2001 to 2009 and the daily streamflow observation data of the Jingziguan hydrological station from 2001 to 2009 were used to simulate the streamflow of the Jingziguan hydrological station. The Thiessen polygon method (Yen et al. 2016) was used to interpolate the precipitation data into the sub-basin areas. Other data including temperature, wind speed, relative humidity, and solar radiation were obtained from the observation data of three weather stations in or around the basin from 2001 to 2009 provided by China Meteorological Bureau. The DEM (http://www.resdc.cn/data.aspx?DataID=123), vegetation index grid data (http://www.resdc.cn/data.aspx?DataID=257), and spatial distribution grid data of three soil textures (http://www.resdc.cn/data.aspx?DataID=260) were provided by the

Figure 1 | Overview of the study area, including main rivers, rain stations, hydrological stations, weather stations, digital elevation model (DEM), and location of basin.
resource and environmental science data center of the Chinese Academy of Sciences, with a resolution of 1 km × 1 km (Figure 2). Flow length grid data and slope grid data were obtained by hydrological analysis of the DEM.

RESULTS AND DISCUSSION

In this section, the drainage basin of Gaoan hydrological station is selected as the study area. In numerical simulation, the randomly generated parameters are used as the true values of the Xin'anjiang model parameters. These true values will be used in the model to generate an ideal observed runoff. In the numerical simulation test, the true values of some parameters are set as unknowns, and the ideal observed runoff is input into the ensemble Kalman filter algorithm to test whether the algorithm can accurately find the true values of these parameters and, at the same time, evaluate the performance of the algorithm when using different parameters as unknowns. The two-year daily streamflow data generated by the observation precipitation data of the Gaoan station drainage basin from 2008 to 2009 are selected for the observation streamflow for the numerical simulation experiment. Table 1 shows the true value and range of the model parameters.

Global sensitivity analysis of model parameters and selection of research objects

When there are many undetermined parameters in the model, no matter the optimization algorithm or data assimilation algorithm used to optimize the parameters of the model, to ensure the efficiency of the algorithm, all parameters are not input into the algorithm for calculation at the same time, but the parameters that have a greater impact on the simulation results are selected for calculation. The global sensitivity of the parameters is an important index to evaluate the effect of the parameters on the simulation results. In this study, the LH-OAT global sensitivity analysis method is used, and the range of each parameter is divided into 200 layers. Sensitivity analysis is performed for 13 parameters of the Xin'anjiang model, and the sensitivity analysis results are shown in Figure 3. According to the global sensitivity and the role of parameters in the model, this study selects five parameters $I_m$, $PETM$, $B$, $C_i$, and $K_i$ for numerical simulation. These parameters directly affect the main processes of the Xin'anjiang model. The parameter $I_m$ is a parameter in the process of saturation excess

---

**Figure 2** | Underlying surface of the study area: (a) DEM (m); (b) vegetation index; (c) flow length (m); (d) slope (%); (e) sand content (%); (f) silt content (%); (g) clay content (%).
runoff, which determines how much water will penetrate into the soil, thus affecting the surface runoff, soil runoff, and underground runoff respectively; parameter $B$ is also a parameter in the process of saturation excess runoff, which reflects the heterogeneity of soil water storage capacity in the basin. Runoff occurs first where the soil water capacity is low. Therefore, this parameter affects the total amount of the runoff and the time of runoff generation in the basin; parameter PETM is a parameter of evapotranspiration process, which directly affects the evaporation of soil water and the total amount of the runoff in the basin; parameter $K_i$ is the parameter of water distribution process, which determines the proportion of free water outflow in soil to soil flow in a certain period of time, and directly affects the total discharge of soil flow; the parameter $C_i$ is a parameter of the routing process, which determines the routing time of the flow in soil.

Analysis of factors affecting single-parameter estimation

This section analyzes three factors that affect the time required for the parameter ensemble to stabilize under the condition of single-parameter estimation. They are the mean of Gaussian white noise used to generate the initial parameter ensemble (hereinafter referred to as the mean initial distribution value), the variance in Gaussian white noise used to generate the initial parameter set (hereinafter referred to as the initial distribution variance), and the roles of parameters used in the model.

Table 1 | True value and range of model parameters

| Name   | Meaning                                                                 | True value | Lower limit | Upper limit |
|--------|-------------------------------------------------------------------------|------------|-------------|-------------|
| C      | Coefficient of the deep layer                                           | 0.252761   | 0.1         | 0.3         |
| $I_M$ (%) | Percentage of impervious and saturated areas in the catchment           | 0.0177124  | 0           | 0.7         |
| $U_M$ (mm) | Averaged soil water storage capacity of the upper layer                  | 30.36697   | 5           | 100         |
| $L_M$ (mm) | Averaged soil water storage capacity of the lower layer                  | 121.02594  | 50          | 300         |
| $D_M$ (mm) | Averaged soil water storage capacity of the deep layer                   | 12.275438  | 5           | 100         |
| $S_m$ (mm) | Areal mean free water capacity of the surface soil layer                | 45.35443   | 5           | 100         |
| $E_X$ | Exponent of the free water capacity curve                               | 0.77414    | 0.5         | 2           |
| $K_{CI}$ | Out coefficients of the free water storage to groundwater relationships | 0.159807   | 0.05        | 0.65        |
| $K_i$ | Out coefficients of the free water storage to interflow relationships   | 0.7405706  | 0.65        | 0.8         |
| $C_g$ | Recession constants of the groundwater storage                          | 0.960716   | 0.9         | 0.999       |
| $C_i$ | Recession constants of the interflow storage                           | 0.8681724  | 0.05        | 0.95        |
| PETM   | Correction coefficient of evaporation                                  | 0.805537   | 0.5         | 1           |

Figure 3 | Results of sensitivity analysis.
### Initial distribution mean

We change the initial distribution means of $I_M$, PETM, $B$, $C$, and $K$, and initial distribution variance is fixed as $(bl2-bl1)\times0.1$, where $bl1$ is the lower limit of parameter value and $bl2$ is the upper limit of parameter value. Each parameter is estimated separately, and other parameters are set as the true value.

The ensemble number is set to 100, the standard deviation of the model error is set to 0.1, and the standard deviation of the observation error is set to 0.1. Parameter updates are performed at every time step except for special instructions. It is set that when the elements in the ensemble at a certain time reach the following conditions, the ensemble is considered to reach stability:

- The difference between the maximum and the minimum in the ensemble is less than $(bl2-bl1)\times0.01$, and no longer greater than $(bl2-bl1)\times0.01$.
- The difference between the ensemble mean at a certain time and the ensemble mean at the end of the estimation process is less than $(bl2-bl1)\times0.01$ and no longer greater than $(bl2-bl1)\times0.01$.
- The difference between the parameter ensemble mean and the parameter true value is less than $(bl2-bl1)\times0.01$.

In the process of estimation, if a parameter in an ensemble exceeds the upper or lower limit, the following equation shall be used for correction:

$$p' = bl + Rand \times (bl2 - bl1)\times0.1$$

(20)

where $p'$ is the corrected value, $bl$ is the upper limit (lower limit) of the parameter, and $Rand$ is a random number between 0 and 1.

According to Figure 4(a)–4(e), when there is only one parameter to be estimated, the closer the initial distribution mean is to the true parameter value, the shorter the time step required for the parameter ensemble to stabilize. For the parameters of PETM, $C$, and $K$, the true values coincide with the initial distribution mean, which needs the shortest time step to stabilize, but there are some exceptions. First, it can be seen from Figure 4 that when the initial distribution mean is close to the upper and lower limits, the time to stabilize does not conform to the above conclusion. This is because there are elements exceeding the upper or lower limits specified by the model in the initial parameter ensemble; therefore, it is necessary to correct these elements to return to the range, and that affects the estimation process, which is also the reason that the time step required for $I_M$ to stabilize when the initial distribution mean is equal to the true parameter value is not the shortest. Second, it can be seen that the time step required for $B$ to stabilize when the initial distribution mean is equal to the true value is not the shortest. This is because parameter $B$ only affects the streamflow when the precipitation is not equal to 0. When the precipitation is 0, only the model error and observation error effect the value of $K_i(z_t - H\hat{x}_t)$ in Equation (4), which makes the parameter update in a random direction and changes the time step required for $B$ to stabilize. To further verify this conclusion, on the basis of the original solution, parameter $B$ is updated only when precipitation is not equal to 0. The calculation results are shown in Figure 4(f). It can be seen that the time required for the parameter ensemble to stabilize is basically the same as when the parameter is updated at all times, and except for the part where the initial distribution mean is close to the upper and lower limits, all the experimental results meet the conclusion that the closer the initial distribution mean is to the true parameter value, the shorter the time step required for the parameter ensemble to stabilize.

### Initial distribution variance

We change the initial distribution variance of $I_M$, PETM, $B$, $C$, and $K$, and the initial distribution mean is fixed as $(bl2+bl1)/2$. Each parameter is estimated separately, and other parameters are set as the true value. Other settings are the same as the previous section.

It can be seen from Figure 5 that when the initial distribution variance is less than the blue line, with the increase in initial distribution variance, the time step required for different parameter ensembles to stabilize generally demonstrates the three trends of increasing, decreasing, and increasing first and then decreasing.

According to Equations (6) and (12), the following equation can be obtained:

$$K_{tm} = E[(p_{tm} - p_{tn})(\hat{x}_t - \bar{x}_t)^T](E[(\hat{x}_t - \bar{x}_t)(\hat{x}_t - \bar{x}_t)^T] + R_t)$$

(21)

$K_{tm}$ is the matrix of Kalman gain of the $m$-th parameter at time $t$, $p_{tm}$ is the vector of the value of the $m$-th parameter at time $t$ in, and $p_{tn}$ is the mean of ensemble of the $m$-th parameter value at time $t$. 

Hydrology Research Vol 53 No 1, 74
When the initial distribution variance increases, the uncertainty of the initial ensemble increases, and both $E[(p_{tm} - p_{lm})(\hat{x}_t - \bar{x}_t)^T]$ and $E[(\hat{x}_t - \bar{x}_t)(\hat{x}_t - \bar{x}_t)^T]$ increase, and according to Equation (21), the Kalman gain will also increase. According to Equation (11), larger Kalman gain makes the algorithm believe less in the old predicted value, making it easier for the parameters to approach the true value.

In Figure 5(c), it can be seen that the time step required to stabilize parameter $B$ increases with the increase in initial distribution variance because the true value of $B$ is 0.249109, and the initial distribution mean of $B$ is set to 0.25, which is very close to the true value. Low uncertainty of the parameter ensemble is required for rapid stabilization; therefore, increasing the initial distribution variance makes it more difficult for the parameter ensemble to stabilize. In Figure 5(a), it can be seen that the time required for $I_M$ to stabilize decreases with the increase in initial distribution variance. The true value of $I_M$ is
0.0177124, which is quite different from the mean of initial ensemble. Therefore, to make the parameter ensemble stabilize in a short time, it is more important to shorten the time of the parameter ensemble to close to the true value than to reduce the uncertainty, thus increasing the variance of the initial distribution will reduce the stabilization time step. In Figure 5(b) and 5(d), the true values of \textit{PETM} and \textit{Ci} are neither around the mean of the initial distribution nor quite different. When the initial distribution variance is too small, because it is difficult for the parameter ensemble to approach the real value, the time step required to stabilize is long. When the initial distribution variance is too large, the uncertainty of the ensemble is still high, and the time step required to stabilize is also long. When the initial distribution variance is moderate, the parameter ensemble can not only approach the true value rapidly, but also reduce the uncertainty quickly, so \textit{PETM} and \textit{Ci} can stabilize after a short time step. To further verify the above conclusion, the parameter \textit{PETM} is estimated with the initial distribution variance of 0.002, 0.02, and 0.04. Figure 6 shows the estimation process of the first 200 time steps of \textit{PETM}. In these three cases, after 61, 29, and 22 time steps, the difference between the ensemble mean and the true value is less than $(bl_2-bl_1)/c_2(0.02)$. Based on the above time steps, after 69, 82, and 103 time steps, the parameter ensemble stabilizes, and the total time steps to stabilize are 130, 111, and 125, respectively.

In Figure 5(e), the position of the true value of \textit{Ki} is similar to that of \textit{PETM} and \textit{Ci}, but it generally decreases with the increase in initial distribution variance, because the condition $K_i + K_g < 1$ of the Xin'anjiang model affects the process of parameter estimation. Although it may not be reasonable to do so in the physical model, to further verify the above conclusion, the condition of $K_i + K_g < 1$ is ignored in the Xin'anjiang model, and then the same calculation is performed for
It is found that the time step required to stabilize decreases first and then increases with the increase in initial distribution variance (Figure 5(f)), which conforms to the above conclusion.

In Figure 5, when the initial distribution variance is greater than the blue line, the ensemble elements exceeding the upper and lower limits will be corrected back to the parameters allowed by the model, which means that the uncertainty can no longer be increased, so the estimation process is basically similar, and it will stabilize at a similar time step.

In conclusion, the most favorable initial distribution variance for the parameter ensemble stabilization depends on the difference between the ensemble mean and the true value of the parameter. When the initial distribution mean is close to the true value, the small variance in the initial distribution is conducive to making the ensemble stabilize. When the initial distribution mean is far from the true value, the large initial distribution variance is conducive to making the ensemble stabilize. When the initial distribution mean is not close to or far from the true value, the moderate variance in the initial distribution is conducive to making the ensemble stabilize. However, in practice, we cannot know the true value of the parameter, so it is difficult to improve assimilation efficiency by determining an initial distribution variance.

Functions of parameters in the model

To improve simulation accuracy, it is preferred to use high sensitivity parameters for assimilation, although they are not all suitable. It can be seen from Figures 4 and 5 that in all cases, the time step required for high sensitivity parameter $I_M$ to stabilize is less than 30 steps. $I_M$ is the parameter with the shortest stabilization time step. However, the time step required for the high sensitivity parameter $PETM$ to stabilize is similar to the low sensitivity parameter $K_r$, which is longer than that for medium sensitivity parameters $B$ and $C_r$. To analyze the causes of the above phenomena, the high sensitivity parameters $I_M$ and $PETM$ are selected, the initial distribution mean value is fixed as $(b_{l2}+b_{l1})/2$, the initial distribution variance is fixed as $(b_{l2}-b_{l1}) \times 0.01$, each parameter is estimated separately, and other parameters are set as the true value. Other settings are the same as the previous section. The results are shown in Figure 7.

According to Equation (21), EnKF updates the parameters according to the covariance matrix between the parameter and streamflow simulation ensembles at time $t$. Covariance is an index to evaluate the correlation between two ensembles. The higher the correlation between parameters and streamflow, the more accurate the result of parameter estimation is, and the more easily it stabilizes. It can be seen from Figure 7(d) that in the estimation process of $PETM$, the correlation coefficient is close to 1 at some time steps, indicating that the parameter is almost linearly related to the streamflow, but the range and

![Figure 6](http://iwaponline.com/hr/article-pdf/53/1/65/995143/nh0530065.pdf)
mean of the parameter ensemble are almost unchanged. The reason is that *PETM* at time *t* not only affects the streamflow at time *t*, but also affects the streamflow at time *t+1, t+2…*, which is very common in hydrological models. However, EnKF does not consider such a situation, resulting in the innovation that $K_t(z_t - H\hat{x}_t)$ in Equation (5) may be 0. For example, at some time steps, some elements in the ensemble are not real values; however, because the variables such as capacity of soil water are not reasonable due to the effect of parameters of non-true values at other time steps, the error of streamflow is 0, so the innovation is 0, and the parameter update is not performed. In addition, the effect of parameters of other time steps may reduce the correlation between the parameter and the simulation of streamflow, making it more difficult for the parameter ensemble to stabilize. On the contrary, $I_M$ only affects the streamflow at the current time, so it can estimate the parameter more accurately when the correlation coefficient is close to 1 and can stabilize rapidly. It should be noted that $I_M$ could only be updated correctly when the precipitation was not 0. Although the precipitation is 0 at the time steps of 1–9, the reason why the parameter ensemble changes at the time steps of 1–9 in Figure 7(a) is that there are model observation errors causing the parameter ensemble to update in a random direction.

In conclusion, it is necessary to analyze not only the sensitivity of parameters, but also the functions of parameters in the model, and then select the suitable parameters for estimation.

**Analysis of factors affecting multi-parameter estimation**

In this section, we analyze the time step required for the parameter ensemble to stabilize under different initial distribution means and different initial distribution variances in the case of two-parameter estimation, three-parameter estimation, and four-parameter estimation. In addition, we discuss the different points of multi-parameter estimation and single-parameter estimation and the interaction among parameters during multi-parameter estimation. We set the following solutions:

A(1): Select parameters $I_M$ and $C_l$ and change the mean of their initial distributions together. The variance of their initial distributions is fixed to $(bI2-bI1) \times 0.1$. The two parameters are estimated together, and the other parameters are set to the true value.
A(2): On the basis of A(1), parameter $B$ is added. The initial distribution mean of $B$ is fixed to $(b_l+b_l)/2$, and the initial distribution variance is fixed to $(b_l-b_l)\times0.1$. These three parameters are estimated together, and the other parameters are set to the true values.

A(3): On the basis of A(1), parameters $B$ and $K_i$ are added. The initial distribution mean of $B$ and $K_i$ is fixed to $(b_l+b_l)/2$, and the initial distribution variance is fixed to $(b_l-b_l)\times0.1$. These four parameters are estimated together, and the other parameters are set to the true value.

B(1): Select parameters $C_i$ and PETM and change their variance of initial distribution together. The initial distribution mean is fixed to $(b_l+b_l)/2$. The two parameters are estimated together, and the other parameters are set to the true value.

B(2): On the basis of B(1), parameter $B$ is added. The initial distribution mean of $B$ is fixed to $(b_l+b_l)/2$, and the initial distribution variance is fixed to $(b_l-b_l)\times0.1$. The three parameters are estimated together, and the other parameters are set as the true values.

B(3): On the basis of A(1), parameters $B$ and $IM$ are added. The initial distribution mean of $B$ and $IM$ is fixed to $(b_l+b_l)/2$, and the initial distribution variance is fixed to $(b_l-b_l)\times0.1$. The four parameters are estimated together, and the other parameters are set to the true value.

Other settings are the same as the previous section. In Figure 8(a) and 8(b), by adding the parameters to be estimated, the overall parameters still tend to take the shortest time to stabilize when the initial distribution mean value is closer to the true value, although this feature becomes less obvious. In Figure 8(c), only $C_i$ and PETM are estimated. As in single-parameter estimation, the ensemble of $C_i$ has the feature that the time required to stabilize first decreases and then increases with the increase in initial distribution variance. When the variance is approximately 0.02, the time step required to stabilize is the shortest, but by adding estimation parameters, this characteristic becomes less obvious. PETM in Figure 8(d) has a similar situation. Equation (21) is a calculation of the Kalman gain established in the cases of single-parameter and multi-parameter estimations, which indicates that the calculation of Kalman gain of each parameter will not consider whether other parameters are reasonable in the case of multi-parameter estimation, thus, it may cause mutual interference between parameters. Comparing Figures 4–8, it can be seen that the mutual influence of these parameters may be favorable or unfavorable to the parameter estimation, but it is unfavorable in most cases.

It can be seen from Figure 8(a) that adding other parameters together with $IM$ for multi-parameter estimation has little effect on the time required for parameter $IM$ to stabilize. As analyzed in the previous section, $IM$ only affects the streamflow of current time step and $IM$ is highly sensitive. This parameter has a high correlation with the streamflow when the precipitation is not zero, and the time step required to stabilize is very short, thus, adding other parameters will hardly affect the estimation process of $IM$. $C_i$ in Figure 8(b) is moderately sensitive, which directly affects the interflow and thus affects the streamflow of other time steps. After adding $B$, which is more sensitive than $C_i$ and stabilizes more easily in single-parameter estimation, the time step required for $C_i$ to stabilize increases slightly, while after adding $K_i$, which is less sensitive than $C_i$ and does not stabilize easily in single-parameter estimation, the time step required for $C_i$ to stabilize remains almost unchanged. In conclusion, the correlation between streamflow and the parameters that required shorter time steps to stabilize in the case of single-parameter estimation is not easily affected in the case of multi-parameter estimation. Therefore, this parameter ensemble may stabilize first in the case of multi-parameter estimation, while the time step required for other parameters to stabilize may increase. However, it may be easier to disrupt the correlation between other parameters and streamflow after adding the parameters which are highly sensitive and do not stabilize easily, thereby significantly reducing the time step required for other parameters to stabilize. However, this is not absolute, because updating EnKF occurs locally, and the high global sensitivity of parameters does not mean that the local parameter sensitivity is also high, although it can represent a general trend. In Figure 8(c), the time required for $C_i$ to stabilize after adding PETM and $B$ on the basis of parameter $C_i$ increases significantly, and in Figure 8(d), the time required for PETM to stabilize after adding $C_i$ and $B$ on the basis of parameter PETM conforms to the above conclusions. In addition, it can be seen from Figure 8(c) and 8(d) that the time required for $C_i$ and PETM to stabilize after adding $IM$ is not obvious, because although $IM$, as a highly sensitive parameter, more easily disrupts the correlation between other parameters and streamflow, the parameter itself can stabilize after a short time step and will not affect the estimation process of other parameters.

In conclusion, to improve the simulation accuracy of the hydrological model and ensure the performance of the ensemble Kalman filter, it is suggested that: (1) high sensitivity parameters that required shorter times to stabilize in the case of
Figure 8 | (a) and (b) Time step required for $I_m$ and $C_i$ to stabilize under different initial distribution means in different solutions. (c) and (d) Time step required for $C_i$ and PETM to stabilize under different initial distribution variance in different solutions.
single-parameter estimation should be selected as the estimation parameters in the case of multi-parameter estimation; (2) low sensitivity parameters have little effect on improving the simulation accuracy and are difficult to stabilize, so it is not necessary to select them for estimation; (3) parameters with high sensitivity that need a long time step to stabilize due to various reasons should not be estimated. Although such parameters can still stabilize in numerical simulation, they may not in application.

**APPLICATION**

In this section, the main study area is Jinjiang River Basin and two cases are designed to evaluate the applicability and practicability of the ensemble Kalman filter, and further verify some conclusions drawn in numerical simulation: (A) ‘there is a long series of hydrometeorological data in the basin, but due to the change of underlying surface or other reasons, the optimization algorithm may not get the optimal parameters of the forecast period’ and (B) ‘there is a lack of hydrometeorological data in the basin’. In (A), the result of parameter calibration by optimization algorithm is used as the assimilation basis, and in (B), the results of parameter calibration by optimization algorithm in a similar basin, Danjiang River Basin, are used as the initial distribution mean of parameter ensemble in assimilation. In this study, the similarity of the two basins is evaluated in terms of topography, vegetation, and soil. Each index is shown in Table 2. The mean of each pixel of grid data is taken for slope, vegetation index, and soil texture content. The two basins are located in a similar latitudinal zone, experience monsoon influence, have a humid climate and vegetation index of 0.8 indicating dense vegetation coverage (http://www.resdc.cn/data.aspx?DataID=257), and have similar area, max flow length, and slope. The soil texture is dominated by sand content in the two basins, and the silt content is basically the same; however, the soil viscosity in the Gaoan Basin is higher. The analysis shows that the underlying surface of the two basins is generally similar, and the parameters can be transferred.

The solution is as follows:

**A(1):** The daily streamflow data of the Gaoan hydrological station from 2001 to 2007 are used to calibrate the parameters by the DDS algorithm, and the parameters obtained from the calibration are used to simulate the station streamflow from 2008 to 2009. The calibration period is from January 2008 to June 2008, and the validation period is from July 2008 to December 2009. Only the simulation results of the validation period are evaluated.

**A(2):** The mean of the initial distribution of $I_M$, $PETM$, $B$, $C_i$, and $K_i$ is set as the result of the calibration, and the variance in initial distribution is fixed as $(bl2-bl1) \times 0.1$. Each parameter is estimated separately, and other parameters are set as the calibration result. The calibration period is from January 2008 to June 2008, and the validation period is from July 2008 to December 2009. Only the simulation results of the validation period are evaluated.

The ensemble number is set to 100, the standard deviation of the model error is set to one-tenth of the mean of the ensemble of simulation streamflow, and the standard deviation of the observation error is set to one-tenth of the observation. Parameter updates are performed at every time step. When the elements in the ensemble at a certain time reach the following conditions, the ensemble is considered to stabilize:

1. The difference between the maximum and the minimum in the ensemble is less than $(bl2-bl1) \times 0.05$, and no longer greater than $(bl2-bl1) \times 0.05$.
2. The difference between the ensemble mean at a certain time and the ensemble mean at the end of the estimation process is less than $(bl2-bl1) \times 0.05$, and no longer greater than $(bl2-bl1) \times 0.05$.

Note: since the true value of the parameter cannot be known in practice, the condition that the difference between the ensemble parameter mean and the true parameter value is less than $(bl2-bl1) \times 0.01$ is ignored.

**A(3):** $I_M$, $B$, $C_i$, and $K_i$ are estimated together. Other settings are the same as in A(2).

| Hydrological station | Latitude | Area (km²) | Max of flow length (km) | Slope (%) | vegetation index | Sand content (%) | Silt content (%) | Clay content (%) |
|----------------------|----------|------------|--------------------------|----------|-----------------|-----------------|-----------------|-----------------|
| Gaoan                | 28°00′N–28°30′N | 6,215      | 18.25                    | 3.108    | 0.804           | 36.437          | 29.489          | 34.073          |
| Jingziguan           | 33°00′N–34°30′N | 7,060      | 19.69                    | 6.385    | 0.814           | 47.301          | 31.599          | 21.100          |
B(1): The daily streamflow data of the Jingziguan hydrological station from 2001 to 2007 are used to calibrate the parameters by DDS algorithm. The Nash of the calibration period is 0.839. The parameters obtained from the calibration are used for the drainage basin of the Gaoan hydrological station to simulate the streamflow from 2008 to 2009. Other settings are the same as in A(1).

B(2): Set the mean of the initial distribution of $I_M$, $PETM$, $B$, $C_i$, and $K_i$ as the result of calibration of the Jingziguan hydrological station, and the variance in initial distribution is fixed as $(b l 2 - b l 1) \times 0.1$. Each parameter is estimated separately, and other parameters are set as a result of calibration of the Jingziguan hydrological station. Other settings are the same as A(2).

B(3): $I_M$, $B$, $C_i$, and $K_i$ are estimated together. Other settings are the same as B(2).

In Table 3, in A(2), except for $PETM$ proposed in a previous section, which is not suitable for assimilation and $K_i$, which is a low sensitivity parameter, the other three parameters can stabilize after assimilation and compared with A(1), the simulation is optimized (as shown in Figures S1–S5 in the Supplementary material, A(1) is the blue line in all the figures). In A(3), due to the interaction of parameters in multi-parameter estimation, some parameters are not reasonable when they stabilize, so although NSE and KGE increase (as shown in Figure S6 in the Supplementary material and bias in Table 3), the absolute value of bias also increases significantly. In solution A, the results of streamflow simulation after assimilation are generally better than that after parameter calibration by the DDS algorithm, which shows that the parameters obtained in the calibration period due to the underlying surface change or other reason are not fully suitable for the validation period, while EnKF can effectively adjust the parameters and improve the streamflow simulation accuracy.

In B(2), except for the solution of estimating $C_i$, the simulation of the solution of estimating other parameters is not much better than that of B(1), and some indexes become worse instead (as shown in Figures S7–S11 in the Supplementary material, B(1) is the blue line in all the figures). This is because solution A has performed parameter calibration, and the values of other parameters are reasonable except for those being estimated, thus, the estimation parameters more easily approach the true value, but solution B uses parameters of similar basins, some of which may be inconsistent with the underlying basin surface, so it is difficult to find the true value of estimation parameters. In B(3), four parameters are estimated together, which reduces the parameters that may not be consistent with the underlying basin surface, and the two indexes are significantly improved (as shown in Figure S12 in the Supplementary material).

In conclusion, in the basin where long data series can be used for parameter calibration, it is suggested to select a small number of parameters with high sensitivity that are suitable for assimilation as the estimation parameters to avoid interference among parameters in the assimilation. In the ungauged basin, it is suggested to use the parameters of a similar basin and select as many parameters suitable for assimilation as the number to be estimated to reduce the parameters inconsistent with the underlying basin surface.

Table 3 | Calculation results of evaluation indexes of each solution

| Solution       | NSE   | KGE   | Bias     | State      |
|----------------|-------|-------|----------|------------|
| A(1)           | 0.829 | 0.810 | −1.08%   | /          |
| A(2) $I_M$     | 0.865 | 0.865 | −1.02%   | Stability  |
| A(2) $B$       | 0.829 | 0.810 | −1.02%   | Stability  |
| A(2) $K_i$     | 0.837 | 0.825 | −0.91%   | Instability|
| A(2) $C_i$     | 0.854 | 0.893 | −0.74%   | Stability  |
| A(2) $PETM$    | 0.802 | 0.673 | −30.10%  | Instability|
| A(3) $I_M$, $B$, $K_i$, $C_i$ | 0.875 | 0.887 | −9.36%   | Stability: $I_M$, $B$, $C$; Instability: $K_i$ |
| B(1)           | 0.524 | 0.536 | −15.55%  | /          |
| B(2) $I_M$     | 0.460 | 0.476 | −7.04%   | Stability  |
| B(2) $B$       | 0.523 | 0.537 | −15.38%  | Stability  |
| B(2) $K_i$     | 0.547 | 0.561 | −15.96%  | Instability|
| B(2) $C_i$     | 0.781 | 0.822 | −14.41%  | Stability  |
| B(2) $PETM$    | 0.614 | 0.566 | −39.31%  | Instability|
| B(3) $I_M$, $B$, $K_i$, $C_i$ | 0.816 | 0.849 | −11.42%  | Stability: $I_M$, $B$, $C$; Instability: $K_i$ |
CONCLUSIONS

In this study, the five representative parameters of $I_M$, $PETM$, $B$, $C_i$, and $K_i$ of the Xin’anjiang model are selected as the research objects to analyze the characteristics of parameter estimation using the EnKF in numerical simulation and application.

After the numerical simulation analysis, the following conclusions are obtained: (1) the closer the mean of the initial parameter ensemble distribution is to the true value, the easier it is to stabilize. Generally, the parameter after calibration is closer to the true value than the default value. Therefore, for basins with long series of data, the method of parameter calibration to determine the initial distribution mean can improve the assimilation efficiency. (2) The most favorable value of the variance in the initial parameter ensemble distribution for the parameter ensemble to stabilize depends on the distance between the ensemble mean value and the true parameter value. In practice, it is almost impossible for us to know the true value of parameters, so it is difficult to improve assimilation efficiency by determining an initial distribution variance. (3) The parameters, which are highly sensitive like $PETM$, but are difficult to stabilize because the parameters at a certain time will affect the streamflow simulation at other times and greatly influence the correlation between other parameters and streamflow, are not suitable for assimilation. This conclusion is further verified in the application part. (4) Similar to $I_M$, it has high sensitivity and needs a short time step to stabilize in the case of single-parameter estimation, which is not easily affected by other parameters in the case of multi-parameter estimation. (5) Some conclusions obtained in the case of single-parameter estimation may not be applicable in the case of multi-parameters due to the interaction between parameters.

The application of EnKF in the Jinjiang River Basin shows that it can solve the problem of streamflow simulation accuracy degradation caused by the change in underlying surface, the lack of hydrometeorological data, and other reasons. In the basin where long series of data can be used for parameter calibration, it is suggested to select a small number of parameters with high sensitivity and assimilation suitability as the parameters to be estimated. In the ungauged basin, it is suggested to use the parameters of a similar basin and select as many parameters suitable for assimilation as the parameters to be estimated. The study area selected in this study is close to the natural state, and the underlying surface is minimally affected by human activities. The better application of EnKF in the basin does not show that the method is still applicable to the basin with large-scale reservoir and large-scale water diversion facilities. The applicability study of such a basin may be a future topic of investigation.

ACKNOWLEDGEMENTS

This paper was jointly supported by National Key R&D Program of China (2017YFC0406004). The first two authors contributed equally to the work.

DATA AVAILABILITY STATEMENT

All relevant data are available from an online repository or repositories.

REFERENCES

Anderson, J. L. 2001 An ensemble adjustment Kalman filter for data assimilation. *Monthly Weather Review* **129**, 2884–2903.

Arnold, J. G. & Fohrer, N. 2005 SWAT2000: current capabilities and research opportunities in applied watershed modelling. *Hydrological Processes* **19**(3), 563–572.

Bishop, C. H., Etherton, B. J. & Majumdar, S. J. 2001 Adaptive sampling with the ensemble transform Kalman filter. Part I: theoretical aspects. *Monthly Weather Review* **129**(3), 420–436.

Burgers, G., van Leeuwen, P. J. & Evensen, G. 1998 Analysis scheme in the ensemble Kalman filter. *Monthly Weather Review* **128**(6), 1719–1724.

Clark, M. P., Rupp, D. E., Woods, R. A., Zheng, X., Ibbitt, R. P., Slater, A. G., Schmidt, J. & Uddstrom, M. J. 2008 Hydrological data assimilation with the ensemble Kalman filter: use of streamflow observations to update states in a distributed hydrological model. *Advances in Water Resources* **31**(10), 1309–1324.

Evensen, G. 1994 Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics. *Journal of Geophysical Research* **99**(C5), 10143–10162.

Fu, X., Yu, Z., Luo, L., Lu, H., Yang, T., Gu, H., Yang, C., Chen, J. & Wang, T. 2014 Investigating soil moisture sensitivity to precipitation and evapotranspiration errors using sib2 model and ensemble Kalman filter. *Stochastic Environmental Research and Risk Assessment* **28**(3), 681–693.
Griensven, A., Meixner, T. & Grunwald, S. A. 2006 Global sensitivity analysis tool for the parameters of multi-variable catchment models. Journal Of Hydrology 324 (1–4), 10–23.

Houtekamer, P. L. & Mitchell, H. L. 1998 Data assimilation using an ensemble Kalman filter technique. Monthly Weather Review 126 (3), 796–811.

Houtekamer, P. L. & Mitchell, H. L. 2001 A sequential ensemble Kalman filter for atmospheric data assimilation. Monthly Weather Review 129 (1), 123–137.

Huang, X., Liao, W., Lei, X., Jia, Y., Wang, Y., Wang, X., Jiang, Y. & Wang, H. 2014 Parameter optimization of distributed hydrological model with a modified dynamically dimensioned search algorithm. Environmental Modelling & Software 52, 98–110.

Jayawardena, A. W., Muttil, N. & Lee, J. H. W. 2006 Comparative analysis of data-driven and GIS-based conceptual rainfall-runoff model. Journal Of Hydrologic Engineering 11 (1), 1–11.

Kalman, R. A. 1960 A new approach to linear filtering and prediction problems. Journal of Fluids Engineering 82, 35–45.

Komma, J., Bloschl, G. & Reszler, C. 2008 Soil moisture updating by Ensemble Kalman Filtering in real-time flood forecasting. Journal Of Hydrology 357 (3–4), 228–242.

Lei, X., Tian, Y. & Zhang, Z. 2019 Correction of pumping station parameters in a one-dimensional hydrodynamic model using the Ensemble Kalman filter. Journal Of Hydrology 568, 108–118.

Liu, Y., Wang, W. & Hu, Y. 2017 Investigating the impact of surface soil moisture assimilation on state and parameter estimation in SWAT model based on the ensemble Kalman filter in upper Huai River basin. Journal of Hydrology and Hydromechanics 65 (2), 123–133.

Lu, H., Hou, T., Horton, R. & Zhu, Y. 2013 The streamflow estimation using the Xinanjiang rainfall runoff model and dual state-parameter estimation method. Journal Of Hydrology 480, 102–114.

Lu, H., Crow, W. T., Zhu, Y., Fen, O. & Su, J. 2016 Improving streamflow prediction using remotely-sensed soil moisture and snow depth. Remote Sensing 8 (6), 503.

Micovic, Z. & Quick, M. C. 1999 A rainfall and snowmelt runoff modelling approach to flow estimation at ungauged sites in British Columbia. Journal Of Hydrology 226 (1–2), 101–120.

Mirzaei, F. M. & Roumeliotis, S. I. A. 2007 Kalman filter-based algorithm for IMU-camera calibration. C// Intelligent Robots and Systems, 2007. IROS 2007. In IEEE/RSJ International Conference on. IEEE.

Moradkhani, H., Soroosh, S., Gupta, H. V. & Houser, P. R. 2004 Dual state-parameter estimation of hydrological models using ensemble Kalman filter. Advances in Water Resources 28 (2), 135–147.

Pathiraja, S., Marshall, L., Sharma, A. & Moradkhani, H. 2016 Hydrologic modeling in dynamic catchments: a data assimilation approach. Water Resources Research 52 (5), 3350–3372.

Rahman, M. M. & Lu, M. 2015 Model spin-up behavior for wet and dry basins: a case study using the Xinanjiang model. Water 7 (8), 4256–4275.

Reichle, R. H. & Walker, J. P. 2002 Extended versus ensemble Kalman filtering for land data assimilation. Journal of Hydrometeorology 3 (5), 728–740.

Svrcek, W. Y., Elliott, R. F. & Zajic, J. E. 1974 The extended Kalman filter applied to a continuous culture model. Biotechnology and Bioengineering 16 (6), 827–846.

Tippett, M. K., Anderson, J. L., Bishop, C. H., Hamill, T. M. & Whitaker, J. S. 2003 Ensemble square root filters. Monthly Weather Review 131, 1485–1499.

Tolson, B. A. & Shoemaker, C. A. 2007 Dynamically dimensioned search algorithm for computationally efficient water-shed model calibration. Water Resources Research 43 (01), W01413.1–W01413.16.

Vrugt, J. A., Diks, C. G. H., Gupta, H. V., Bouten, W. & Verstraten, J. M. 2005 Improved treatment of uncertainty in hydrologic modeling: combining the strengths of global optimization and data assimilation. Water Resources 41 (1), W01017.

Wagner, P. D., Fiener, P., Wilken, F., Kumar, S. & Schneider, K. 2012 Comparison and evaluation of spatial interpolation schemes for daily rainfall in data scarce regions. Journal of Hydrology 464, 388–400.

Whitaker, J. S. & Hamill, T. M. 2002 Ensemble data assimilation without perturbed observations. Monthly Weather Review 130, 1913–1924.

Xie, X. & Zhang, D. 2008 Data assimilation for distributed hydrological catchment modeling via ensemble Kalman filter. Advances in Water Resources 31 (10), 1309–1324.

Xie, X. & Zhang, D. 2015 A partitioned update scheme for state-parameter estimation of distributed hydrologic models based on the ensemble Kalman filter. Water Resources Research 49 (11), 7350–7365.

Yang, S., Dong, G. & Zheng, D. 2011 Coupling Xinanjiang model and SWAT to simulate agricultural non-point source pollution in Songtiao watershed of Hainan, China. Ecological Modelling 222 (20–22), 3701–3717.

Yen, H., Jeong, J. & Smith, D. R. 2016 Evaluation of dynamically dimensioned search algorithm for optimizing SWAT by altering sampling distributions and searching range. Journal of the American Water Resources Association 52 (2), 443–455.

Yin, Z. K., Liao, W. H., Lei, X. H., Wang, H. & Wang, R. 2018 Comparing the hydrological responses of conceptual and process-based models with varying rain gauge density and distribution. Sustainability 10 (9), 3209.

Zhao, R. 1992 The Xinanjiang model applied in China. Journal of Hydrology 135, 371–381.

First received 8 July 2021; accepted in revised form 8 October 2021. Available online 15 November 2021.