Supplement of

State updating in the Xin'anjiang Model: Joint assimilating streamflow and multi-source soil moisture data via Asynchronous Ensemble Kalman Filter with enhanced Error Models

Gong Junfu et al.

Correspondence to: Cheng Yao (yaocheng@hhu.edu.cn)

S1. Error estimation

S1.1. Uncertainty in model forcing

In flood forecasting, the most critical model driving data is rainfall. We used log-normal multiplicative perturbation to characterize rainfall errors (McMillan et al., 2011; DeChant and Moradkhani, 2012; Gong et al., 2023):

$$\boldsymbol{P}_j^o(t_i) = \boldsymbol{\delta}^{\boldsymbol{P}}(t_i) \cdot \boldsymbol{P}(t_i) \tag{S1-1}$$

Where $\boldsymbol{P}(t_i) = [P_1(t_i), \dots, P_{N_p}(t_i)]^{\boldsymbol{T}} \in \mathcal{R}^{N_p}$ is the rainfall observation vector; N_p is the dimensionality of the rainfall observations; $\boldsymbol{\delta}^{\boldsymbol{P}}(t_i)$ is lognormal perturbation matrix. The errors in the precipitation measurement are assumed to be spatially independent, so that, $\boldsymbol{\delta}^{\boldsymbol{P}}(t_i)$ is also a diagonal matrix. The diagonal element is $\delta_n^P(t_i), (n = 1, \dots, N_p)$, and $ln\, \delta_n^P(t_i) \sim N(\mu_{lnp}, \sigma_{lnp})$ follows a lognormal distribution with the mean of 1.0 and standard deviation of σ_p. Additionally, a first-order autoregressive model is employed to represent the temporal correlation in precipitation measurement errors. At each time step, the perturbation is mathematically adjusted as follows:

$$ln\, \delta_n^P(t_i) = \mu_{lnp} + \alpha_{lnp}[ln\, \delta_n^P(t_{i-1}) - \mu_{lnp}] + \varphi\sigma_{lnp}(1 - {\alpha_{lnp}}^2)^{0.5} \tag{S1-2}$$

Where $\mu_{lnp} = -0.5{\sigma_{lnp}}^2$; α_{lnp} is autocorrelation coefficient for precipitation measurement errors.

S1.2. Uncertainty in observations

The observation error is generalized as functions of the corresponding observed values (Weerts & El Serafy, 2006; Clark et al., 2008; Alvarez-Garreton et al., 2015):

$$\boldsymbol{y}_j^{\boldsymbol{o}}(t_i) = [\boldsymbol{I} + \boldsymbol{\delta}^{\boldsymbol{y}}(t_i)] \cdot \boldsymbol{y}(t_i) \tag{S1-3}$$

Where $\boldsymbol{y}_j^{\boldsymbol{o}}(t_i) \in \mathcal{R}^{N_y}$ represents the perturbed observation vector for the jth ensemble. $\boldsymbol{I}$ is identity matrix; $\boldsymbol{\delta}^{\boldsymbol{y}}(t_i)$ is Gaussian perturbation matrix. Assuming that the observation errors are spatially independent, $\boldsymbol{\delta}^{\boldsymbol{y}}(t_i) \in \mathcal{R}^{N_y \times N_y}$ is a diagonal matrix with diagonal elements $\delta_n^y(t_i), \left(n = 1, \dots, N_y\right)$. When assimilating soil moisture observations, the diagonal elements follow a normal distribution $\delta_n^y(t_i) \sim N\left(0, \sigma_{ys}\right)$, and similarly, $\delta_n^y(t_i) \sim N\left(0, \sigma_{yd}\right)$ is used when assimilating discharge observations. Furthermore, we employ a first-order autoregressive model to account for the temporal correlation in observation errors. At time step t, the perturbation is adjusted using the formula:

$$\delta_n^y(t_i) = \mu_y + \alpha_y\left[\delta_n^y(t_{i-1}) - \mu_y\right] + \varphi\sigma_y(1 - {\alpha_y}^2)^{0.5} \tag{S1-4}$$

Where $\mu_y = 0$; φ is a standard Gaussian noise; σ_y is the standard deviation, which, as previously stated, takes the values σ_{ys} or σ_{yd}; α_y is the autocorrelation coefficient, with values of α_{ys} when assimilating soil moisture observations, or α_{yd} when assimilating discharge observations.

S1.3. Uncertainty in model state

In the assimilation of observed outlet discharge, the key model state variable updated is cumulative channel flow. This variable represents the outflow from each sub-basin on the routing calculation unit, denoted as QC. This state variables are perturbed using a Gaussian error function (Li et al., 2014), and the state variables of the jth ensemble are changed to:

$$\boldsymbol{X}_j^{f\prime}(t_i) = \boldsymbol{X}_j^f(t_i) + \boldsymbol{\delta}^{\boldsymbol{x}}(t_i) \cdot \boldsymbol{X}_j^f(t_i) \tag{S1-5}$$

Where $\boldsymbol{X}_j^f(t_i) = [QC_{j,1}(t_i), QC_{j,2}(t_i), .., QC_{j,N_x}(t_i)]^{\boldsymbol{T}} \in \mathcal{R}^{N_x}$, and when assimilating discharge observations, N_x is set as the number of sub-reaches; $\boldsymbol{\delta}^{\boldsymbol{x}}(t_i)$ is diagonal Gaussian perturbation matrix, and the diagonal element is $\delta_n^x(t_i)$, $(n = 1, \dots, N_x)$, and $\delta_n^x(t_i) \sim N(0, \sigma_d)$.

When assimilating observed soil moisture, the model state variables representing soil humidity need to be updated. Specifically, this refers to the tension water storage (including upper, and lower layer tension water) and the free water storage in the Xin'anjiang model. The perturbation of soil state variables follows the same form as in Eq. (S1-5), but $\boldsymbol{X}_j^f(t_i) = [\boldsymbol{W}_j(t_i), \boldsymbol{WU}_j(t_i), \boldsymbol{WL}_j(t_i), \boldsymbol{S}_j(t_i)]^T \in \mathcal{R}^{N_x}$, and N_x is set as $4 * N_{sub}$ in this case (N_{sub} is the number of sub-basins). Here, $\boldsymbol{W}_j(t_i) = [W_{j,1}(t_i), .., W_{j,N_{sub}}(t_i)]$, $\boldsymbol{WU}_j(t_i) = [WU_{j,1}(t_i), .., WU_{j,N_{sub}}(t_i)]$, $\boldsymbol{WL}_j(t_i) = [\mathrm{W}L(t_i), .., WL_{j,N_{sub}}(t_i)]$, and $\boldsymbol{S}_j(t_i) = [S_{j,1}(t_i), .., S_{j,N_{sub}}(t_i)]$. The diagonal elements of $\boldsymbol{\delta}^{\boldsymbol{x}}(t_i)$ become $\delta_n^x(t_i) \sim N(0, \sigma_s)$.

We introduce the Bias-corrected Gaussian Error Model (BGEM) (Ryu et al., 2009) to reduce biases arising from the need to adhere to physical constraints, which occur when Gaussian perturbations push variables beyond their limits, leading to truncation errors in hydrological model predictions. The BGEM method is accomplished by running an unperturbed model in parallel with the ensemble model simulations. The average bias, $\boldsymbol{\delta}^{\boldsymbol{b}}(t_i)$, of the perturbed state variables at time step t_i is calculated using the formula:

$$\boldsymbol{\delta}^{\boldsymbol{b}}(t_i) = \frac{1}{N_e} \sum_{n=1}^{N_e} \left[\boldsymbol{X}_j^{f\prime}(t_i) - \boldsymbol{X}^b(t_i)\right] \tag{S1-6}$$

Where $\boldsymbol{X}^b(t_i)$ represents the soil moisture derived from the undisturbed simulation. The bias-corrected set of state variables is then acquired by deducting the average bias $\boldsymbol{\delta}^{\boldsymbol{b}}(t_i)$ from the perturbed variables $\boldsymbol{X}_j^{f\prime}(t_i)$.

$$\widetilde{\boldsymbol{X}}_j^{f\prime}(t_i) = \boldsymbol{X}_j^{f\prime}(t_i) - \boldsymbol{\delta}^{\boldsymbol{b}}(t_i) \tag{S1-7}$$

S1.4. Maximum a posteriori estimation method

The hyperparameters required in the aforementioned error model include σ_{ys}, α_{ys}, σ_{yd}, α_{yd}, σ_{lnp}, α_{lnp}, σ_s, and σ_d. To identify the globally optimal values of these hyperparameters, the Maximum a posteriori estimation method (MAP) is applied. This method aims to maximize the probability density of the hyperparameters with given the observed historical flood events.

This section offers a concise overview of the MAP. For a comprehensive understanding of the implementation of the method, refer to our previous study by Gong et al. (2023).

Following Bayesian theory, the posterior probability density is expressed as a product of the prior probability density and the conditional probability density of historical observations:

$$p(\boldsymbol{\psi}|\boldsymbol{y}) \propto \chi(\boldsymbol{\psi}) = p(\boldsymbol{\psi}) \times p(\boldsymbol{y}|\boldsymbol{\psi}) \tag{S1-8}$$

Where $\boldsymbol{\psi} = (\sigma_{ys},\ \alpha_{ys},\ \sigma_{yd},\ \alpha_{yd}, \sigma_{lnp},\ \alpha_{lnp},\ \sigma_s,\ \sigma_d)$ is hyperparameter array; $p(\boldsymbol{y}|\boldsymbol{\psi})$ is the conditional probability density of the historical observations $\boldsymbol{y}$ given $\boldsymbol{\psi}$; $p(\boldsymbol{\psi})$ is a prior probability density, calculated as the product of individual probabilities:

$$p(\boldsymbol{\psi}) = p(\sigma_{ys})p(\alpha_{ys})p(\sigma_{yd})p(\alpha_{yd})p(\sigma_{lnp})p(\alpha_{lnp})p(\sigma_s)p(\sigma_d) \tag{S1-9}$$

The conditional probability density $p(\boldsymbol{y}|\boldsymbol{\psi})$ is determined as:

$$p(\boldsymbol{y}|\boldsymbol{\psi}) = \prod_{i=1}^{N_y}\prod_{t=1}^{N_t} p(y_{i,t}|\boldsymbol{\psi}) \tag{S1-10}$$

For global optimization, the Shuffled Complex Evolution (SCE-UA) method (Duan et al., 1992) is employed, setting the objective function in a negative logarithmic format:

$$f_{OBJ} = \frac{1}{N_f}\sum_{m=1}^{N_f} -\ln\,[\chi(\boldsymbol{\psi})] \tag{S1-11}$$

S2. Multi-source soil moisture data fusion

Table S2-1. List of WKNN model parameters

Sub-basin	*W*		*WU*		*WL*		*S*	
	K	*p*	*K*	*p*	*K*	*p*	*K*	*p*
1	2	1	19	1	18	1	4	1
2	15	1	19	1	15	1	6	1
3	19	1	13	1	12	1	4	1
4	14	1	10	1	6	1	4	1
5	2	1	18	1	16	1	4	1
6	19	1	14	1	14	1	4	1
7	18	1	19	1	19	1	2	1
8	7	1	16	1	18	1	4	1
9	11	1	19	1	10	1	5	1
10	16	1	14	1	7	1	5	1

S3. Evaluation metrics

S3.1. Normalized Nash-Sutcliffe efficiency coefficient (NNSE)

The range of Nash-Sutcliffe efficiency coefficient (NSE) is $(-\infty, 1]$. For a flood event, a higher NSE indicates that the simulated discharge process is closer to the observed discharge process. When the simulated hydrograph coincides with the observed hydrograph, the NSE value is equal to 1.0. The NSE is calculated as following:

$$NSE = 1 - \frac{\sum_t \left(\bar{y}_t^{sim} - y_t\right)^2}{\sum_t (y_t - y_{ave})^2} \tag{S3-1}$$

Where y_t is the observed discharge at time step t; y_{ave} is the temporal mean of the observed discharge in a flood event; $\bar{y}_t^{sim}$ is the simulated discharge (ensemble mean discharge for the ensemble run or the simulated discharge of the Xin'anjiang model for the deterministic run).

The potential issue arising from the lower limit of negative infinity in NSE can be addressed by employing a specific equation that normalizes the NSE. This approach rescales the NSE to fall within the (0,1] range, resulting in what is termed the Normalized Nash-Sutcliffe Efficiency (NNSE) (Nossent and Bauwens, 2012):

$$NNSE = \frac{1}{2 - NSE} \tag{S3-2}$$

It's important to note that $\mathrm{NSE} = 1$ equates to $\mathrm{NNSE} = 1$, $\mathrm{NSE} = 0$ translates to $\mathrm{NNSE} = 0.5$, and $\mathrm{NSE} = -\infty$ corresponds to $\mathrm{NNSE} = 0$. The mean value of NNSE for multiple flood events is denoted as MNNSE.

S3.2 Root mean square error (RMSE)

The range of root mean square error (RMSE) is $[0, \infty)$. A smaller value indicates higher accuracy of the simulated discharge.

$$RMSE = \sqrt{\frac{1}{N_t} \sum_{t=1}^{N_t} \left(\bar{y}_t^{sim} - y_t\right)^2} \tag{S3-3}$$

To compare the performance difference between the assimilation (Ens) and the open loop (OL), the ratio between them is calculated, denoted as R(.):

$$R_{RMSE} = \frac{RMSE_{Ens}}{RMSE_{OL}} \tag{S3-4}$$

The range of $\mathrm{R_{RMSE}}$ values extends from $[0, \infty)$. When $\mathrm{R_{RMSE}}$ is less than 1.0, it indicates that the accuracy of ensemble mean of assimilation run surpasses that of the OL. Further, the mean value of $\mathrm{R_{RMSE}}$ for multiple flood events is denoted as $\mathrm{MR_{RMSE}}$.

S3.3. Continuous ranked probability score (CRPS)

Overall performance about ensemble simulation is assessed using the Continuous Ranked Probability Score (CRPS) (Hersbach, 2000). The CRPS is regarded as the integral of the Brier score over all possible threshold values for the variable. For non-ensemble systems, CRPS that simplifies to mean absolute error (MAE).

$$CRPS = \int_{-\infty}^{\infty} [P(x) - H(x - x_a)]^2\, dx \tag{S3-5}$$

$$P(x) = \int_{-\infty}^{x} \rho(y) dy \tag{S3-6}$$

$$H(z) = \begin{cases} 0, z < 0 \\ 1, z \geq 0 \end{cases} \tag{S3-7}$$

where $P(x)$ is the cumulative distribution; x is the forecast value; and x_a is the observed value. For the ensemble system, with equal weights for each ensemble, CRPS can be calculated by following.

Let $y_{t,0}^{sim} = -\infty$ and $y_{t,N_e+1}^{sim} = \infty$, and arrange the ensemble members in ascending order to obtain the ordered simulated discharge ensemble $\boldsymbol{\acute{Y}_t^{sim}}$, where the ensemble members satisfy the following equation:

$$\acute{y}_{t,i}^{sim} < \acute{y}_{t,j}^{sim}, i < j \leq N_e \tag{S3-8}$$

The cumulative distribution function can be expressed as:

$$P\left(\acute{y}_t^{sim}\right) = p_i \equiv \frac{i}{N_e}\ , \acute{y}_{t,i}^{sim} < y_t^{sim} < \acute{y}_{t,i+1}^{sim} \tag{S3-9}$$

Based on Eq. (S3-5), the CRPS at time step t can be expressed as:

$$CRPS_t = \sum_{n=0}^{N_e} \alpha_n p_n^2 + \beta_n (1 - p_n)^2 \tag{S3-10}$$

where the values of α_n and β_n can be found in Table S3-1. Ultimately, the time-averaged CRPS for a flood event can be obtained through the following expression:

$$CRPS_{ave} = \sum_{n=0}^{N_e} \alpha_{ave,n} p_n^2 + \beta_{ave,n} (1 - p_n)^2 \tag{S3-11a}$$

$$\alpha_{ave,n} = \frac{1}{N_t} \sum_{t=1}^{N_t} \alpha_{t,n} \tag{S3-11b}$$

$$\beta_{ave,n} = \frac{1}{N_t} \sum_{t=1}^{N_t} \beta_{t,n} \tag{S3-11c}$$

Table S3-1. Calculation table of α_n and β_n

		α_n	β_n
$n = N_e$	$y_t > \acute{y}_{t,N_e}^{sim}$	$y_t - \acute{y}_{t,N_e}^{sim}$	0
$0 < n < N_e$	$y_t > \acute{y}_{t,n+1}^{sim}$	$\acute{y}_{t,n+1}^{sim} - \acute{y}_{t,n}^{sim}$	0
	$\acute{y}_{t,n+1}^{sim} > y_t > \acute{y}_{t,n}^{sim}$	$y_t - \acute{y}_{t,n}^{sim}$	$\acute{y}_{t,n+1}^{sim} - y_t$
	$y_t < \acute{y}_{t,n}^{sim}$	0	$\acute{y}_{t,n+1}^{sim} - \acute{y}_{t,n}^{sim}$
$n = 0$	$y_t < \acute{y}_{t,1}^{sim}$	0	$\acute{y}_{t,1}^{sim} - y_t$

The ratio of $\mathrm{CRPS_{ave}}$ between Ens and OL is denoted as $\mathrm{R_{CRPS}}$:

$$R_{CRPS} = \frac{CRPS_{ave,Ens}}{CRPS_{ave,OL}} \tag{S3-12}$$

Further, the mean value of $\mathrm{R_{RMSE}}$ for multiple flood events is denoted as $\mathrm{MR_{CRPS}}$.

S3.4. Reliability part of continuous ranked probability score (RELI)

The CRPS, as a composite indicator, can be decomposed into several distinct components. Among these, the reliability part, denoted as RELI, shares similarities with the performance of the Rank Histogram, effectively quantifying the reliability of the ensemble (Hersbach, 2000):

$$RELI_{ave} = \sum_{n=0}^{N_e} g_{ave,n}\,(o_{ave,n} - p_n)^2 \tag{S3-13}$$

When $0 < n < N_e$:

$$o_{ave,n} = \frac{\beta_{ave,n}}{\alpha_{ave,n} + \beta_{ave,n}} \tag{S3-14a}$$

$$g_{ave,n} = \alpha_{ave,n} + \beta_{ave,n} \tag{S3-14b}$$

When $n = 0$:

$$o_{ave,0} = \frac{1}{N_t}\sum_{t=1}^{N_t} FH(\acute{y}_{t,1}^{sim} - y_t) \tag{S3-15a}$$

$$g_{ave,0} = \frac{\beta_{ave,0}}{o_{ave,0}} \tag{S3-15b}$$

When $n = N_e$:

$$o_{ave,N_e} = \frac{1}{N_t}\sum_{t=1}^{N_t} FH(\acute{y}_{t,N_e}^{sim} - y_t) \tag{S3-16a}$$

$$g_{ave,N_e} = \frac{\alpha_{ave,N_e}}{1 - o_{ave,0}} \tag{S3-16b}$$

where FH(.) is the Heaviside function.

Denote the ratio of $\mathrm{RELI_{ave}}$ between the assimilated system and the OL as:

$$R_{RELI} = \frac{RELI_{ave,Ens}}{RELI_{ave,OL}} \tag{S3-17}$$

Further, the mean value of $\mathrm{R_{RELI}}$ for multiple flood events is denoted as $\mathrm{MR_{RELI}}$.

S4. Data Overview

This hydro-meteorological data utilized in the study spanning from 2014 to 2023, provided by the Hunan Provincial Hydrological Bureau, including evaporation, precipitation, and discharge data. Within the catchment, there are 17 rain gauges, one evaporation observation station, and four discharge observation stations. Evaporation data are derived from daily pan evaporation measurements using the E-601 pan, with hourly values calculated as 1/24th of the daily measurements. Notably, with only one evaporation observation station in the catchment, it is assumed that the observed evaporation is spatially uniform. When multiple rain gauges exist within a sub-catchment, the area-averaged rainfall is calculated as the arithmetic mean of all gauge observations. For discharge observation stations, Wuqiangxibashang (WQXBS) serves as the outlet observation station, while the remaining three stations Hexi (HX), Pushi (PS), and Gaochetou (GCT) measure inflow. Hourly observations of precipitation and discharge are intermittent, thus hourly data are only available during flood events, with daily data available at other times. Fifteen flood events from 2014 to 2018 were used for model calibration, and fourteen events from 2019 to 2023 for model validation. Considering soil moisture data availability, six flood events in 2023 were used for assimilation studies. For an overview of these flood events, refer to Table S3-1.

Table S4-1. List of flood events investigated in this study

	Serial number	Start date	End date	Observed Peak flow (m^3/s)
calibration	No.2014052300	2014/05/23 00:00	2014/05/27 20:00	17356
	No.2014070300	2014/07/03 00:00	2014/07/06 08:00	22705
	No.2014071400	2014/07/14 00:00	2014/07/19 00:00	35725
	No.2015060121	2015/06/01 21:00	2015/06/07 01:00	17762
	No.2015060718	2015/06/07 18:00	2015/06/10 18:00	12017
	No.2015062023	2015/06/20 23:00	2015/06/24 09:00	19196
	No.2016050703	2016/05/07 03:00	2016/05/11 06:00	13051
	No.2016062017	2016/06/20 17:00	2016/06/21 21:00	12472
	No.2016062720	2016/06/27 20:00	2016/06/30 03:00	14996
	No.2016070311	2016/07/03 11:00	2016/07/08 12:00	22278
	No.2017052208	2017/05/22 08:00	2017/05/25 19:00	8872
	No.2017062711	2017/06/27 11:00	2017/07/05 12:00	32147
	No.2017081121	2017/08/11 21:00	2017/08/16 00:00	13091
	No.2018053010	2018/05/30 10:00	2018/06/03 16:00	7348
	No.2018092518	2018/09/25 18:00	2018/09/27 05:00	8518
validation	No.2019051905	2019/05/19 05:00	2019/05/22 00:00	14024
	No.2019070700	2019/07/07 00:00	2019/07/16 12:00	14046
	No.2020070800	2020/07/08 00:00	2020/07/09 18:00	25963
	No.2020071823	2020/07/18 23:00	2020/07/20 16:00	18688
	No.2020091500	2020/09/15 00:00	2020/09/21 08:00	20829
	No.2021050300	2021/05/03 00:00	2021/05/05 00:00	8021
	No.2021051112	2021/05/11 12:00	2021/05/27 00:00	13347
	No.2021060300	2021/06/03 00:00	2021/06/07 00:00	8391
	No.2023040308	2023/04/03 08:00	2023/04/05 14:00	6192
	No.2023050416	2023/05/04 16:00	2023/05/06 17:00	4747
	No.2023052008	2023/05/20 08:00	2023/05/22 18:00	5660
	No.2023062100	2023/06/21 00:00	2023/06/25 19:00	6940
	No.2023063000	2023/06/30 00:00	2023/07/01 14:00	9317
	No.2023072516	2023/07/25 16:00	2023/07/27 18:00	8449

[a] The flood events utilized for assimilation research are indicated by bold text with an underline.

The Wuqiangxi Catchment houses 10 soil moisture monitoring sites, established between 2018 and 2023. Of these, eight sites have sensing depths of 20, 40, and 60 cm, with observations taken every 2 hours. The remaining two stations have depths of 10, 20, and 40 cm, with observations every 8 hours. For detailed information about these soil moisture monitoring sites, refer to Table S3-2. In the absence of a monitoring site in the No.6 sub-basin, we utilize data from the Daheping station, situated in the No.3 sub-basin, due to its close spatial proximity.

Table S4-2. List of soil moisture monitoring sites

Sub-basin	Soil moisture monitoring sites	Sensing depths (cm)	Start date of data availability	Observation interval (h)
1	Wuqiangxi	20/40/60	2022/10/09	2
1	Qijiaping	20/40/60	2022/10/09	2
2	Qinglang	20/40/60	2022/10/09	2
3	Daheping	20/40/60	2022/11/22	2
4	Madiyi	20/40/60	2022/10/28	2
5	Guanzhuang	20/40/60	2022/10/09	2
6	Daheping	20/40/60	2022/11/22	2
7	Gaoqitou	20/40/60	2023/01/12	2
8	Yuanling	20/40/60	2023/01/12	2
9	Maxipu	10/20/40	2019/01/01	8
10	Pushi	10/20/40	2018/01/01	8

The soil moisture reanalysis data, sourced from the China Meteorological Administration Land Data Assimilation System (CLDAS V2.0) near-real-time dataset (https://data.cma.cn/data/cdcdetail/dataCode/NAFP_CLDAS2.0_NRT.html), released by the China Meteorological Administration. It has a spatial resolution of 0.0625° and a temporal resolution of one day, providing profile soil moisture across four layers (0-10 cm, 10-40 cm, 40-100 cm, 100-200 cm) from 2017 onwards (Liu et al., 2019). This study utilizes data on the top three soil layers (0-10 cm, 10-40 cm, 40-100 cm) from 2018 to 2023.

The DEM data were downloaded from the Geospatial Data Cloud (https://www.gscloud.cn/sources/accessdata/305?pid=302), selecting the SRTM digital elevation model with a resolution of 90 m. The soil texture map was sourced from the Harmonized World Soil Database version 1.2 (HWSD V1.2) (https://www.fao.org/soils-portal/data-hub/soil-maps-and-databases/harmonized-world-soil-database-v12/en/), with a spatial resolution of approximately 1 km (30 arc-seconds). This study utilizes data on the percentage of sand and clay in the topsoil (0-30 cm) and subsoil (30-100 cm) layers, along with USDA soil texture classification data.

References

Alvarez-Garreton, C., Ryu, D., Western, A. W., Su, C. H., Crow, W. T., Robertson, D. E., and Leahy, C.: Improving operational flood ensemble prediction by the assimilation of satellite soil moisture: comparison between lumped and semi-distributed schemes, Hydrol. Earth Syst. Sci., 19, 1659-1676, https://doi.org/10.5194/hess-19-1659-2015, 2015.

Clark, M. P., Rupp, D. E., Woods, R. A., Zheng, X., Ibbitt, R. P., Slater, A. G., Schmidt, J., and Uddstrom, M. J.: Hydrological data assimilation with the ensemble Kalman filter: Use of streamflow observations to update states in a distributed hydrological model, Adv. Water Resour., 31, 1309-1324, https://doi.org/10.1016/j.advwatres.2008.06.005, 2008.

DeChant, C. M. and Moradkhani, H.: Examining the effectiveness and robustness of sequential data assimilation methods for quantification of uncertainty in hydrologic forecasting, Water Resour. Res., 48, W04518, https://doi.org/10.1029/2011WR011011, 2012.

Duan, Q. Y., Sorooshian, S., and Gupta, V.: Effective and efficient global optimization for conceptual rainfall - runoff models, Water Resour. Res., 28, 1015-1031, https://doi.org/10.1029/91WR02985, 1992.

Gong, J. F., Weerts, A. H., Yao, C., Li, Z. J., Huang, Y. C., Chen, Y. F., Chang, Y. F., and Huang, P. N.: State updating in a distributed hydrological model by ensemble Kalman filtering with error estimation, J. Hydrol., 620, 129450, https://doi.org/10.1016/j.jhydrol.2023.129450, 2023.

Hersbach, H.: Decomposition of the continuous ranked probability score for ensemble prediction systems, Weather Forecast., 15, 559-570, https://doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2, 2000.

Li, Y., Ryu, D., Western, A. W., Wang, Q. J., Robertson, D. E., and Crow, W. T.: An integrated error parameter estimation and lag-aware data assimilation scheme for real-time flood forecasting, J. Hydrol., 519, 2722-2736, https://doi.org/10.1016/j.jhydrol.2014.08.009, 2014.

Liu, J. G., Shi, C. X., Sun, S., Liang, J. J., and Yang, Z. L.: Improving Land Surface Hydrological Simulations in China Using CLDAS Meteorological Forcing Data, J. Meteorolog. Res., 33, 1194-1206, https://doi.org/10.1007/s13351-019-9067-0, 2019.

McMillan, H., Jackson, B., Clark, M., Kavetski, D., and Woods, R.: Rainfall uncertainty in hydrological modelling: An evaluation of multiplicative error models, J. Hydrol., 400, 83-94, https://doi.org/10.1016/j.jhydrol.2011.01.026, 2011.

Nossent, J. and Bauwens, W.: Application of a normalized Nash-Sutcliffe efficiency to improve the accuracy of the Sobol'sensitivity analysis of a hydrological model, EGU General Assembly 2012, Vienna, 22-27 April, 237, https://ui.adsabs.harvard.edu/abs/2012EGUGA..14..237N/abstract, 2012.

Ryu, D., Crow, W. T., Zhan, X. W., and Jackson, T. J.: Correcting Unintended Perturbation Biases in Hydrologic Data Assimilation, J. Hydrometeorol., 10, 734-750, https://doi.org/10.1175/2008JHM1038.1, 2009.

Weerts, A. H. and El Serafy, G. Y. H.: Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models, Water Resour. Res., 42, W09403, https://doi.org/10.1029/2005WR004093, 2006.