



Improved Streamflow Simulation by Assimilating In Situ Soil Moisture in Lumped and Distributed Approaches of a Hydrological Model in a Headwater Catchment

Hongxia Li¹ · Yuanyuan Huang¹ · Yongliang Qi¹ · Yanjia Jiang¹ · Xuan Tang¹ · Elizabeth W. Boyer² · Carlos R. Mello³ · Ping Lan¹ · Li Guo¹

Received: 10 November 2023 / Accepted: 17 May 2024
© The Author(s), under exclusive licence to Springer Nature B.V. 2024

Abstract

Soil moisture data assimilation (SM-DA) is a valuable approach for enhancing streamflow prediction in rainfall-runoff models. However, most studies have focused on incorporating remotely sensed SM, and their results strongly depend on the quality of satellite products. Compared with remote sensing products, in situ observed SM data provide greater accuracy and more effectively capture temporal fluctuations in soil moisture levels. Therefore, the effectiveness of SM-DA in improving streamflow prediction remains site-specific and requires further validation. Here, we employed the Ensemble Kalman filter (EnKF) to integrate daily SM into lumped and distributed approaches of the Xinanjiang (XAJ) hydrological model to assess the importance of SM-DA in streamflow prediction. We observed a general improvement in streamflow prediction after conducting SM-DA. Specifically, the Nash-Sutcliffe efficiency increased from 0.61 to 0.65 for the lumped and from 0.62 to 0.70 for the distributed approaches. Moreover, the efficiency of SM-DA exhibits seasonal variation, with in situ SM proving particularly valuable for streamflow prediction during the wet-cold season compared to the dry-warm season. Notably, daily SM data from deep layers exhibit a stronger capability to improve streamflow prediction compared to surface SM. This indicates the significance of deep SM information for streamflow prediction in mountain areas. Overall, this study effectively demonstrates the efficacy of assimilating SM data to improve hydrological models in streamflow prediction. These findings contribute to our understanding of the connection between SM, streamflow, and hydrological connectivity in headwater catchments.

Keywords Data assimilation · Headwater catchment · Rainfall-runoff models · Soil moisture · Streamflow prediction

Hongxia Li and Yuanyuan Huang contributed equally to this work.

Extended author information available on the last page of the article

1 Introduction

Streamflow is a critical component of regional and global hydrological cycles. Accurate streamflow prediction plays an essential role in water resources management, including assessing the impact of climate and land use changes and aiding in reservoir operation (Mello et al. 2021). In practice, both lumped and distributed hydrological models are used for streamflow prediction across various environments (Li et al. 2009).

However, streamflow prediction by hydrological models is inevitably accompanied by uncertainties related to model inputs, parameters, and structure (Li et al. 2015; Bournas and Baltas 2021). These uncertainties arise from a lack of knowledge or inaccuracies when measuring or calculating the values of variables and parameters in hydrological models, resulting in discrepancies between modeled and actual streamflow values (Beven 2016). To tackle this issue, model calibration is commonly employed to minimize such discrepancies (Duan et al. 1992; Gotzinger and Bardossy 2008). However, calibration often heavily relies on observed data, typically utilizing only streamflow data in previous research, reducing the physical aspects of the models and constraining their application.

One effective approach to enhance streamflow prediction accuracy is to incorporate additional available information into forecasting (Gan et al. 2022; Ireson et al. 2022). With the advancements in monitoring technology, hydrological modeling can leverage various data sources such as soil moisture (SM), evapotranspiration (ET), groundwater, stream stage, temperature, and snow cover (Avellaneda et al. 2020; Koochi et al. 2022; Tekeli and Fouli 2017). Utilizing these data can improve model performance and reduce prediction uncertainty, increasing our understanding of the hydrological processes in the basin. The state of SM significantly impacts various hydrological processes within catchments, including rainfall-runoff transformations, aquifer recharge and actual evapotranspiration (Venegas-Cordero et al. 2023; Yin et al. 2022). Thus, accurately representing SM in hydrological models holds great importance for improving streamflow predictions. One way to integrate SM into hydrological models is through data assimilation, specifically SM data assimilation (SM-DA) (Alvarez-Garreton et al. 2015; Samuel et al. 2014). Data assimilation involves the integration of observational data from various sources with a predictive model, accompanied by real-time adjustments to the model's parameters or state variables. This process aims to minimize errors associated with uncertainties in both the data and the model itself. It is widely used for effectively integrating diverse observations and has increasingly been applied to streamflow forecasting. Due to the Ensemble Kalman filter (EnKF) method's flexibility, ease of use, robustness, and effective assimilation capabilities, it is one of the most popular approaches for SM-DA in hydrological modeling (Shi et al. 2014). The EnKF can simultaneously estimate and adjust parameters as state variables, allowing for the real-time adaptation of parameters over time. Additionally, it updates and corrects runoff simulation results in real-time by integrating new data. This ability to more accurately represent the actual situation enhances the prediction accuracy of the model.

In recent years, obtaining SM information has become easier through in situ and remote sensing observations (Caldwell et al. 2019; Li et al. 2021). Numerous studies have explored the assimilation of SM data into hydrological models to enhance streamflow prediction (Loizu et al. 2018; Nayak et al. 2021). However, conducting SM-DA has not yielded consistent results (Alvarez-Garreton et al. 2015; Patil and Ramsankaran 2017). Some studies have highlighted the improved representation of SM and superior streamflow prediction

with SM-DA (Brocca et al. 2012; Massari et al. 2015). For example, Ziliani et al. (2019) significantly enhanced flood prediction by assimilating SM data. Conversely, other studies have produced contrasting findings (Matgen et al. 2012; Liu et al. 2024). Nayak et al. (2021) used three different hydrological models with SM-DA, observing performance improvements in one model while deterioration in the other two models. Another example by Patil and Ramsankaran (2017) investigated the potential of assimilating remotely sensed soil moisture observations to improve streamflow predictions. The findings revealed moderate improvements in streamflow predictions. Therefore, although several studies have employed SM-DA for rainfall-runoff modeling and demonstrated promising results in enhancing streamflow predictions, the degree of improvement often varies depending on specific site characteristics. Additionally, the outcomes of SM-DA are influenced by factors such as model structure, catchment characteristics, and the spatiotemporal resolution of forcing data (Massari et al. 2015; Nayak et al. 2021). Despite this, it remains unclear which factors have the most significant impact on the performance of SM-DA (Alvarez-Garretón et al. 2015; Brocca et al. 2012; Massari et al. 2015).

Moreover, the incorporation of remote sensed SM data from sources such as Global Land Data Assimilation System (GLDAS), Soil Moisture and Ocean Salinity (SMOS), and Soil Moisture Active Passive (SMAP) in most studies is heavily influenced by the quality and scale of satellite observations (Khaki et al. 2019; Laiolo et al. 2016; Wu et al. 2022). On one hand, SM-DA using remote sensing data primarily focuses on surface SM (0–5 cm depth), while hydrological models often divide soils into multiple layers, making it challenging to accurately represent the state of deep soil moisture. Several studies assimilating in situ SM data typically rely on a limited number of sites predominantly situated in relatively flat terrains (Samuel et al. 2014), thereby leaving the performance of SM-DA in catchments with complex topography unclear. Therefore, it is crucial to assess the performance of SM-DA in streamflow prediction by utilizing an adequate number of in situ SM observations, encompassing both surface and subsurface SM, particularly in complex environments. This evaluation is necessary to identify the key factors that exert the most significant impact on model performance.

The objective of this paper is to evaluate the impact of assimilating in situ SM observations into lumped and distributed approaches of a rainfall-runoff model for streamflow simulation in a headwater catchment equipped with a network of 33 field monitoring sites. We investigate the effectiveness of SM assimilation using the Ensemble Kalman filter (EnKF) method to enhance streamflow predictions, as well as the disparity in streamflow predictions between lumped and distributed hydrological models after SM data assimilation in catchments characterized by complex terrains.

2 Study Area and Datasets

2.1 Study Catchment

The study was conducted in the Shale Hills catchment in central Pennsylvania, U.S.A. (Fig. 1a), a forested, V-shaped catchment covering an area of 0.08 km². The catchment is characterized by relatively steep slopes (ranging from 25 to 45%). A first-order stream forms within the catchment is a tributary of the Shaver's Creek. The catchment experiences

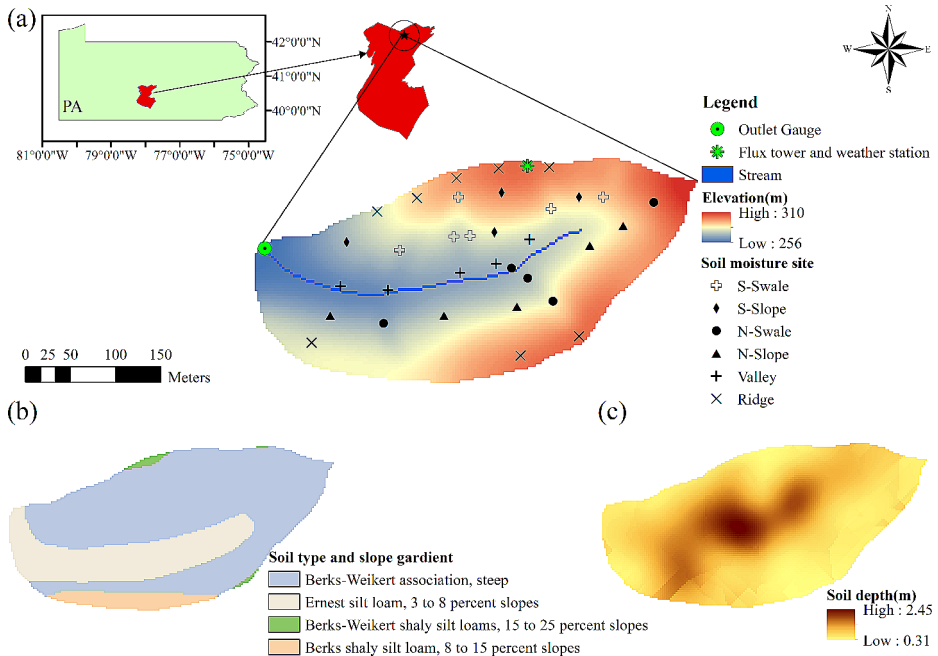


Fig. 1 (a) Geographical location, soil moisture monitoring sites, geomorphological features, and elevation of the Shale Hills catchment. (b) Distribution of soil types and slope degree. (c) Distribution of soil depth

a typical continental climate with a mean annual air temperature of 10 °C and mean annual precipitation of 1070 mm, distributed uniformly throughout the year. While most precipitation falls as snow during winter and early spring, rain is predominant from late spring to autumn (Liu and Lin 2015). Deciduous forests and a few evergreen coniferous forests and shrubs dominate the catchment. The majority of soils in the study catchment are silt loam and silty clay loam, with four soil types identified (Fig. 1b) (Lin 2006; Lin and Zhou 2008), and their depths range from 0.31 to 2.45 m within the Shale Hills catchment (Fig. 1c). The catchment's elevation ranges from 256 m at the outlet to 310 m at the highest ridge.

2.2 Datasets

The Shale Hills catchment was equipped with a real-time hydrologic (RTH) monitoring network, including a flux tower and a weather station (Fig. 1a). The RTH network recorded precipitation, wind speed, relative humidity, air temperature and net radiation at a temporal resolution of 10 min. Hourly streamflow data were obtained using water level readings from a V-shaped weir located at the catchment outlet (Fig. 1a) and an associated rating curve.

The Shale Hills catchment was established with 33 SM monitoring sites. The 33 sites were grouped into one of six soil-terrain units: Ridge, Valley, N-slope, N-swale, S-slope and S-swale (Fig. 1a). The grouping of these sites was based on a combination of soil types and slope characteristics to achieve adequate representation and coverage (Fan et al. 2020). These sensors were installed across multiple soil layers at each site to measure SM (m^3/m^3), with a temporal resolution of 10 min. Each site has three different sensor depths: the shal-

lowest (horizon A) is between 5 and 20 cm, the intermediate (horizon B), between 12 and 90 cm, and the deepest (horizon C), between 20 and 155 cm. Daily precipitation and streamflow data covering the period from January 01, 2008 to May 31, 2013 and SM data from May 19, 2011 to May 31, 2013 (downloaded from <http://www.czo.psu.edu>) were employed in this study (Table S1).

3 Methods

3.1 The Xinanjiang Hydrological Model (XAJ Model)

The XAJ model is a lumped rainfall-runoff model (Zhao 1992). It is primarily used in humid and semi-humid regions and has gained wide recognition for its application in streamflow prediction worldwide. The XAJ model comprises four sub-models: an evapotranspiration sub-model, a surface runoff generation sub-model, a runoff separation sub-model, and a runoff routing sub-model (Fig. S1). The catchment is divided into permeable and impermeable areas in the XAJ model. In impermeable areas, all effective rainfall is directly converted into surface runoff. In permeable areas, the model recognizes the heterogeneity of vertical soil distribution by subdividing the soil into upper, lower, and deep layers. The XAJ model calculates the evaporation of the three layers in the evapotranspiration module, followed by constructing a tension water capacity curve to estimate the runoff while considering the uneven distribution of the underlying surface. The free water reservoir structure separates the total runoff into three components: surface, interflow, and subsurface. The runoff is then transported to the outlet of each sub-basin through linear reservoirs to generate the outflow of the sub-basins. Finally, the runoff is calculated to the catchment's outlet using either the lag-and-route method or the Muskingum method.

Input data for the model include precipitation and potential evapotranspiration (PET), while the output is represented by runoff (or streamflow). In this study, PET was calculated using the Food and Agriculture Organization of the United Nations (FAO) Penman-Monteith equation (Cai et al. 2007) incorporating data on net radiation, wind speed, relative humidity, and air temperature obtained from the RTH network in the Shale Hills catchment. The XAJ model incorporates a total of 15 state variables and 18 parameters, including the Muskingum parameters (Table S2). For more detailed information on the XAJ model, please refer to Zhao (1992).

3.2 The Grid-XAJ Model

The Grid-XAJ model is a distributed hydrological model that builds upon the principles of the XAJ model and utilizes DEM data (Yao et al. 2012). The watershed is partitioned into rectangular grids based on the DEM information, with each grid serving as a computational unit within the Grid-XAJ framework. These units consist of two main components: a water balance component and a cell-to-cell flow routing component. The calculations for ET, runoff generation, and runoff separation within each grid follow the same principles as the XAJ model. The catchment calculation order for each grid is determined based on the grid catchment area matrix. The Muskingum method is applied to route the flow from each unit to the catchment outlet, considering the computed order among the DEM grid cells.

The Grid-XAJ model comprises 13 spatial uniform parameters, including K, C, B, IM, EX, KG, KI, CG, CI, CS, L, KE, and XE. These parameters in the Grid-XAJ model have the same physical meanings as in the XAJ model. The Grid-XAJ model assumes a uniform distribution of rainfall patterns and underlying surface conditions within each grid. The parameters W_m and S_m in each grid are related to the characteristics of the underlying surface and can be calculated based on soil texture properties. The specific calculation formula is as follows:

$$W_m = (\theta_f - \theta_w) \times L_a \quad (1)$$

$$S_m = (\theta_s - \theta_f) \times L_h \quad (2)$$

$$L_{al} = 0.167 \times L_a \quad (3)$$

$$L_{au} = 0.5 \times L_a \quad (4)$$

where θ_f is field capacity; θ_w is the wilting point; L_a is the thickness of aeration zone (mm); θ_s is saturated moisture content; L_h is the thickness of humus soil (mm); L_{al} is the thickness of the upper aeration zone (mm); L_{au} is the thickness of the lower aeration zone (mm). For more detailed information about these formulas, please refer to Yao et al. (2012).

3.3 Data Assimilation

3.3.1 Ensemble Kalman Filter

The Ensemble Kalman filter (EnKF) is a sequential data assimilation algorithm that employs a Monte Carlo-based approach (Evensen 1994). The EnKF consists of a forecast period and an update period. The dual EnKF approach simultaneously estimates state variables and model parameters using noisy observations. The method can be subdivided into two periods, i.e., parameter optimization and state update. At each assimilation time step, model parameters are updated first, followed by the update of model state variables. The detailed algorithmic process of the dual EnKF is as follows.

(1) Generating the initial state sets. Based on the Monte Carlo method, a set of random variables is generated by initializing the set of state variables $\{x_t^{i+}\}$ and parameters $\{\phi_t^{i+}\}$ of the model at time $t=1$, $i=1, \dots, N$.

In order to prevent over-updating in the process of parameter assimilation, kernel smoothing is performed on the parameter set:

$$\phi_{t+1}^{i-} \sim N(\alpha \phi_t^{i+} + (1 - \alpha) \bar{\phi}_t^+, h^2 V_t^+) \quad (5)$$

$$V_t^+ = \text{var}(\phi_t^+) \quad (6)$$

where ϕ is the parameter of assimilation, the superscripts “ $i-$ ” and “ $i+$ ” denote the forecast and updated values of the i^{th} member of the parameter set, respectively. $\alpha = \frac{3\delta-1}{2\delta}$, α values range from 0 to 1, and $h = \sqrt{1 - \alpha^2}$.

(2) Transferring state. Using the forecasted parameters (ϕ_{t+1}^{i-}), the set of forecast values of the model state variables at time $t+1$ is calculated:

$$x_{t+1}^{i-} = f(x_t^{i+}, \phi_{t+1}^{i-}, u_t, t) \quad (7)$$

where x_{t+1}^{i-} is the forecast value of the i^{th} state ensemble member at time $t+1$ and x_t^{i+} is the update value of the i^{th} state ensemble member at time t . u_t is the model inputs at time t , such as precipitation.

(3) Updating parameters. When observations are available at time t , the ensembles of parameters are updated according to the Kalman equation:

$$\phi_{t+1}^{i+} = \phi_{t+1}^{i-} + K_{t+1}^{\phi}((Y_{t+1}^i + v_{t+1}) - h(x_{t+1}^{i-}, \phi_{t+1}^{i-})), v_{t+1} \sim N(0, R_{t+1}) \quad (8)$$

$$K_{t+1}^{\phi} = P_{t+1}^{\phi} H^T (H P_{t+1}^{\phi} H^T + R_{t+1})^{-1} \quad (9)$$

$$P_{t+1}^{\phi} = \frac{1}{N-1} \sum_{i=1}^N (\phi_{t+1}^{i-} - \overline{\phi_{t+1}^{i-}})(\phi_{t+1}^{i-} - \overline{\phi_{t+1}^{i-}})^T \quad (10)$$

$$\overline{\phi_{t+1}^{i-}} = \frac{1}{N} \sum_{i=1}^N \phi_{t+1}^{i-} \quad (11)$$

where Y_{t+1}^i is the members of the observation at time $t+1$. The measured soil moisture data from the surface and subsurface layers were used in this study. v_{t+1} and R_{t+1} are the noise and covariance of the observation error, respectively. $h(g)$ is the observation operator characterizing the mapping process from the state variables x_{t+1}^{i-} and parameters ϕ_{t+1}^{i-} of the model to the observation Y_{t+1}^i ; K_{t+1}^{ϕ} is the Kalman gain of the modified parameter trajectory at time $t+1$; P_{t+1}^{ϕ} is the covariance of model forecasting error at time $t+1$; $\overline{\phi_{t+1}^{i-}}$ is the mean of the forecasted parameter members.

(4) Forecasting the state variable. Bringing the parameters obtained from Eq. (8) back to time t , and recalculating the state variables with the evolution of time.

$$x_{t+1}^{i-} = f(x_t^{i+}, \phi_{t+1}^{i+}, u_t, t) + \omega_t, \omega_t \sim N(0, Q_t) \quad (12)$$

where u_t and Q_t are the noise and covariance of the model structure.

(5) Updating the status variable. Using the observation data and Eq. (13) to update the state variables when the observation data are available.

$$x_{t+1}^{i+} = x_{t+1}^{i-} + K_{t+1}^X((Y_{t+1}^i + v_{t+1}) - h(x_{t+1}^{i-}, \phi_{t+1}^{i+})), v_{t+1} \sim N(0, R_{t+1}) \quad (13)$$

$$K_{t+1}^X = P_{t+1}^X H^T (H P_{t+1}^X H^T + R_{t+1})^{-1} \quad (14)$$

$$P_{t+1}^X = \frac{1}{N-1} \sum_{i=1}^N (x_{t+1}^{i-} - \overline{x_{t+1}^{i-}})(x_{t+1}^{i-} - \overline{x_{t+1}^{i-}})^T \quad (15)$$

$$\overline{x_{t+1}^-} = \frac{1}{N} \sum_{i=1}^N x_{t+1}^{i-} \quad (16)$$

where $h(g)$ is the observation operator characterizing the mapping process from the state variables x_{t+1}^{i-} and parameters ϕ_{t+1}^{i+} of the model to the observation Y_{t+1}^i ; K_{t+1}^X is the Kalman gain of the modified status variable trajectory at time $t+1$; P_{t+1}^X is the covariance of model forecasting error at time $t+1$; $\overline{x_{t+1}^-}$ is the mean of the forecasted state variables members.

(6) Examining the stop criterion. Return to step (2) if t is less than the total assimilation time T ; else, stop the loop.

3.3.2 SM-DA set up

The implementation framework of the EnKF for assimilating SM is depicted in Fig. 2. The dual EnKF scheme was employed to capture the associated covariances between model state variables and model parameters. Stochastic noises were assumed to follow a Gaussian distribution with predetermined variances, and they were added to model parameters to account for modeling errors (Loizu et al. 2018). Taking into account the computational cost and assimilation effect, as well as comparative analyses with existing research, this study has opted to set the ensemble size at 50 members.

During the data assimilation process, two parameters, namely the exponent of the distribution of tension water capacity (B) and the areal mean free soil water storage capacity (S_m), along with two state variables, the areal mean tension water storage of the upper soil layer (WU) and the mean tension water storage of the lower soil layer (WL), were adjusted and updated.

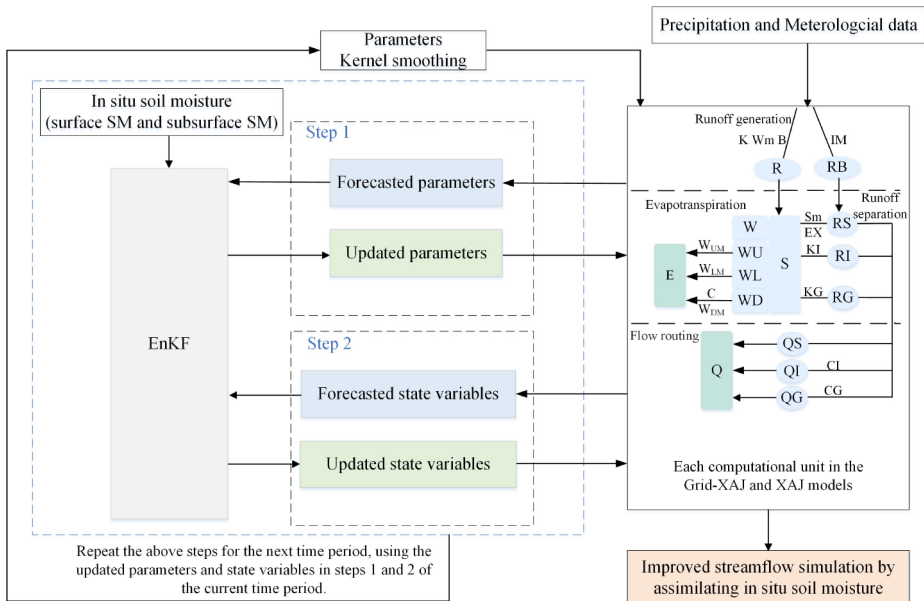


Fig. 2 Flow chart of the SM data assimilation set up in this study

3.4 Data Preprocessing

The soil moisture data from horizon A was used as the surface SM value. The mean value of soil moisture from horizons B and C was considered subsurface SM, as these horizons represent significantly deeper layers in this catchment (Primka et al. 2021).

The XAJ and Grid-XAJ model based on the Shale Hills catchment incorporates soil properties to represent soil water storage using three different soil layers: the upper soil layer (5–33 cm), lower soil layer (29–84 cm), and deep soil layer (34–101 cm). Initial experiments conducted during this study highlighted that the model's sensitivity to the deep soil layer was limited unless there were changes in the vertical distribution of soil properties. As a result, for data assimilation, the surface and subsurface soil water in the XAJ model were represented and updated using surface and subsurface SM observations.

The XAJ and Grid-XAJ models simulate soil moisture content for each layer. When assimilation observed SM data, the soil profile should be processed to divide the upper and lower “virtual” soil layers so that the soil moisture monitoring point are in the corresponding soil layers, thus the soil moisture content is converted into the corresponding moisture content data. The positions of the monitoring points for both surface and subsurface layers at the 33 stations are all located within the depth range of the upper and lower vadose zones in the corresponding grids of the Grid-XAJ model.

The conversion formulas for surface and subsurface SM to the upper and lower soil moisture content in the Grid-XAJ model are as follows:

$$WU_{ob,k} = (\theta_{sur,k}^{ob} - \theta_{w,k}) \times L_{au,k} \quad (17)$$

$$WL_{ob,k} = (\theta_{subsur,k}^{ob} - \theta_{w,k}) \times L_{al,k} \quad (18)$$

where $WU_{ob,k}$ is the upper soil moisture content corresponding to the observation site k ($k=1, \dots, 33$), mm; $\theta_{sur,k}^{ob}$ is surface SM, m^3/m^3 ; $\theta_{w,k}$ is the wilting point; $L_{au,k}$ is the thickness of upper aeration zone; $WL_{ob,k}$ is the lower soil moisture content corresponding to the observation site, mm; $\theta_{subsur,k}^{ob}$ is subsurface SM, m^3/m^3 ; $L_{al,k}$ is the thickness of lower aeration zone.

$$\overline{WU_{ob}} = \frac{1}{33} \sum_{k=1}^{33} WU_{ob,k} \quad (19)$$

$$\overline{WL_{ob}} = \frac{1}{33} \sum_{k=1}^{33} WL_{ob,k} \quad (20)$$

where $\overline{WU_{ob}}$, $\overline{WL_{ob}}$ is the mean of the upper and lower soil moisture content corresponding to the 33 observation sites, mm.

The $\overline{WU_{ob}}$ and $\overline{WL_{ob}}$ represent the measured soil moisture content of the corresponding upper and lower layers within the Shale Hills catchment in the XAJ model, respectively.

3.5 Evaluation of Streamflow Prediction Results

The optimization of parameters for the two approaches of the XAJ hydrological model was accomplished using the Genetic Algorithm (GA) method (Wang 1991). The calibration period encompassed January 01, 2008, to December 31, 2010, with the first three months allocated as the model warm-up period. Subsequently, the validation period extended from January 01, 2011, to May 31, 2013. To assess the impact on data assimilation output performance, the models were evaluated through two distinct runs: an Open Loop run, which involved no data assimilation, and an SM-DA run, which incorporated the assimilation of SM data. The simulation period considered for the SM-DA run spanned from May 19, 2011, to May 31, 2013.

The evaluation of the model's ability to simulate daily streamflow performance was conducted using the following metrics: the Nash-Sutcliffe efficiency coefficient (NSE, Eq. (21)), relative bias (BIAS, Eq. (22)), correlation coefficient (R, Eq. (23)) and assimilation efficiency (EFF, Eq. (24)). The calculation of these metrics is as follows:

$$\text{NSE} = 1 - \frac{\sum_{t=1}^T (Q_{\text{obs}}(t) - Q_{\text{sim}}(t))^2}{\sum_{t=1}^T (Q_{\text{obs}}(t) - \overline{Q_{\text{obs}}})^2} \quad (21)$$

Where T is the length of the streamflow time series (daily), $Q_{\text{obs}}(t)$ and $Q_{\text{sim}}(t)$ are the observed and simulated streamflow at time t , m^3/d , respectively, and $\overline{Q_{\text{obs}}}$ is the mean of the observed streamflow, m^3/d . In Eqs. (21), (22) and (23) the subscript sim refers to any of the different time series performance evaluated in this study, i.e., calibration (Q_{cal}), validation (Q_{val}), Open Loop (Q_{OL}) or data assimilation (Q_{DA}).

$$\text{BIAS} = \frac{\sum_{t=1}^T (Q_{\text{obs}}(t) - Q_{\text{sim}}(t))}{\sum_{t=1}^T Q_{\text{obs}}(t)} \times 100\% \quad (22)$$

$$R = \frac{\sum_{t=1}^T (Q_{\text{obs}}(t) - \overline{Q_{\text{obs}}})(Q_{\text{sim}}(t) - \overline{Q_{\text{sim}}})}{\sqrt{\sum_{t=1}^T (Q_{\text{obs}}(t) - \overline{Q_{\text{obs}}})^2} \sqrt{\sum_{t=1}^T (Q_{\text{sim}}(t) - \overline{Q_{\text{sim}}})^2}} \quad (23)$$

$$\text{EFF} = \left(1 - \frac{\sum_{t=1}^T (Q_{\text{DA}}(t) - Q_{\text{obs}}(t))^2}{\sum_{t=1}^T (Q_{\text{OL}}(t) - Q_{\text{obs}}(t))^2} \right) \times 100\% \quad (24)$$

The EFF is more than zero indicating that the assimilation run is better than the Open Loop run for streamflow simulation and vice versa.

4 Results

4.1 Dynamics of SM and Their Relation to Precipitation and Streamflow

The comparison of spatial patterns for surface and subsurface SM measurements is presented in Fig. 3a and b. Notably, the surface SM demonstrates higher levels of moisture in comparison to the subsurface SM, exhibiting low spatial variability and lacking any apparent spatial pattern (Fig. 3a). Conversely, the observed subsurface SM exhibits distinct spatial patterns (Fig. 3b). Wet areas predominantly align along the valley, while the ridgetop areas tend to be drier when compared to other soil-terrain units. These patterns may be attributed

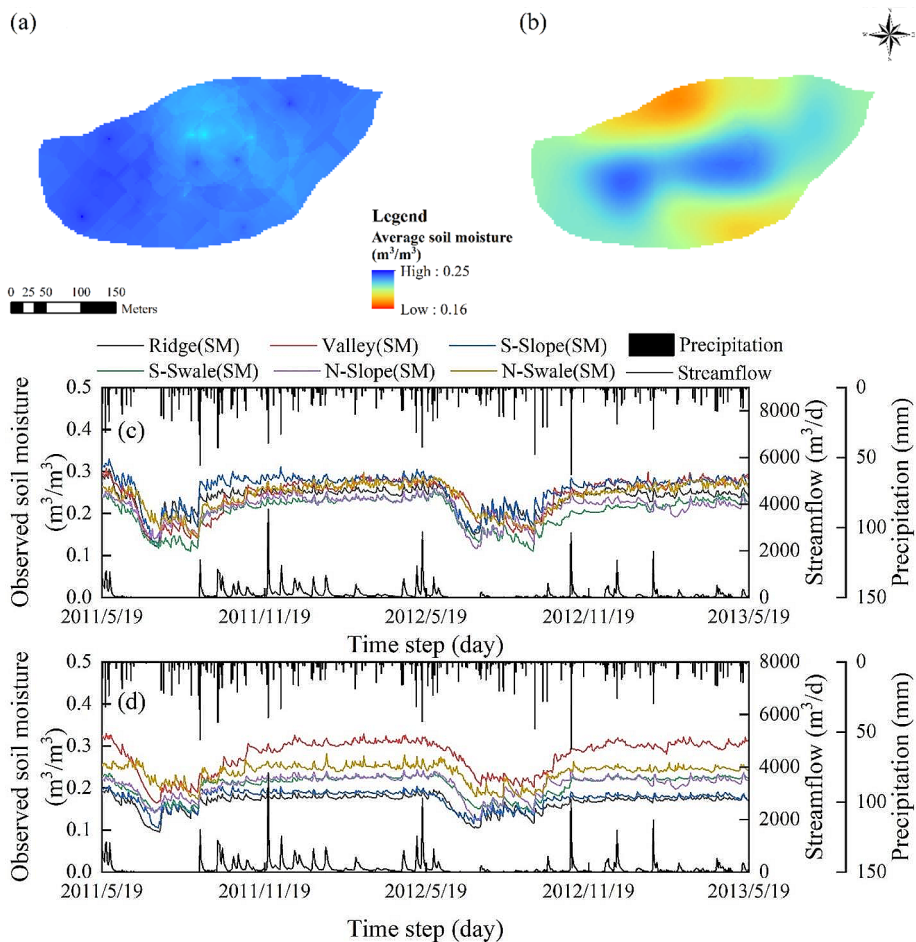


Fig. 3 Spatial distribution of the mean (a) surface SM and (b) subsurface SM measured from the sensor network in the Shale Hills catchment from 2011 to 2013 (interpolated for the entire catchment using the ordinary kriging method) and the time series of the daily unit-average of (c) surface SM and (d) subsurface SM observations in the six soil-terrain units, as well as the daily precipitation and streamflow during the same monitoring period

to the combined influence of topography and soil characteristics, manifesting in catchment wetting and drying cycles alongside precipitation and evapotranspiration processes (Fig. 1).

The temporal dynamics of surface and subsurface SM across six distinct soil-terrain units are presented in Fig. 3c and d, accompanied by daily precipitation and streamflow data during the monitoring period. Figure 3c and d suggests that the surface SM patterns remain consistent across all soil-terrain units (Fig. 3c), whereas subsurface SM exhibits greater variability influenced primarily by topography and soil depth within the catchment area. As depicted in Fig. 3d, the valley unit consistently displays the highest moisture content, reaching values of approximately $0.32 \text{ m}^3/\text{m}^3$ or above for most monitoring days, predominantly stored in the deeper subsurface layers. In contrast, the ridge unit exhibits lower moisture content, typically below $0.18 \text{ m}^3/\text{m}^3$, attributed to variations in precipitation and evapotranspiration rates, terrain characteristics, and soil properties. SM correlates well with streamflow, where variations in streamflow are partly dependent on the relative dryness/wetness of the soil, influencing the occurrence of overland flows on the hillslope.

4.2 Calibration and Validation of rainfall-runoff

We employed a 6-year time series spanning from 2008 to 2013 for calibration and validation. The predictive capabilities of both approaches for daily streamflow were compared. Figure 4 presents the results obtained by both models during the calibration and validation periods. Overall, both the XAJ and Grid-XAJ models demonstrated effectiveness in capturing the hydrologic characteristics of the catchment. However, in terms of accuracy, the distributed model showcased superior performance, indicated by consistently lower residual values.

The performance of the XAJ and Grid-XAJ models during both the calibration and validation periods was compared. The NSE for the XAJ model is 0.66 during calibration and 0.65 during validation, whereas for the Grid-XAJ model, the NSE values are 0.68 during calibration and 0.67 during validation. Additionally, the BIAS values for the two models are -10.7% and 1.7% during calibration, and -19.7% and 0.0% during validation (Table S3). These results indicate that the Grid-XAJ model outperforms the XAJ model.

4.3 Improved Streamflow Simulating with SM-DA

Figure 5 presents the results obtained from the Open Loop and SM-DA runs with 50 ensemble members. The ensemble predictions of streamflow before and after SM-DA for both the lumped and distributed model approaches were compared. The range between the 5th and 95th percentiles and the mean values represents ensembles. In both models, the SM-DA run (Fig. S2) reveals improved agreement between the overall shape of the streamflow hydrograph and the observed data in comparison to the Open Loop run. Furthermore, the streamflow ensemble simulations in the Open Loop run exhibit a wider variation range when compared to the SM-DA run, indicative of greater uncertainty. The period of peak flow exhibits the highest level of uncertainty (Fig. S2).

The quartile ranges of NSE during the Open Loop run range from 0.55 to 0.62 and 0.51 to 0.61, while during the SM-DA run, they are observed to be 0.64 to 0.66 and 0.69 to 0.70 for the XAJ and Grid-XAJ, respectively (Fig. 5a). Moreover, both models exhibit higher correlation coefficient (R) in the SM-DA run compared to the Open Loop simulation

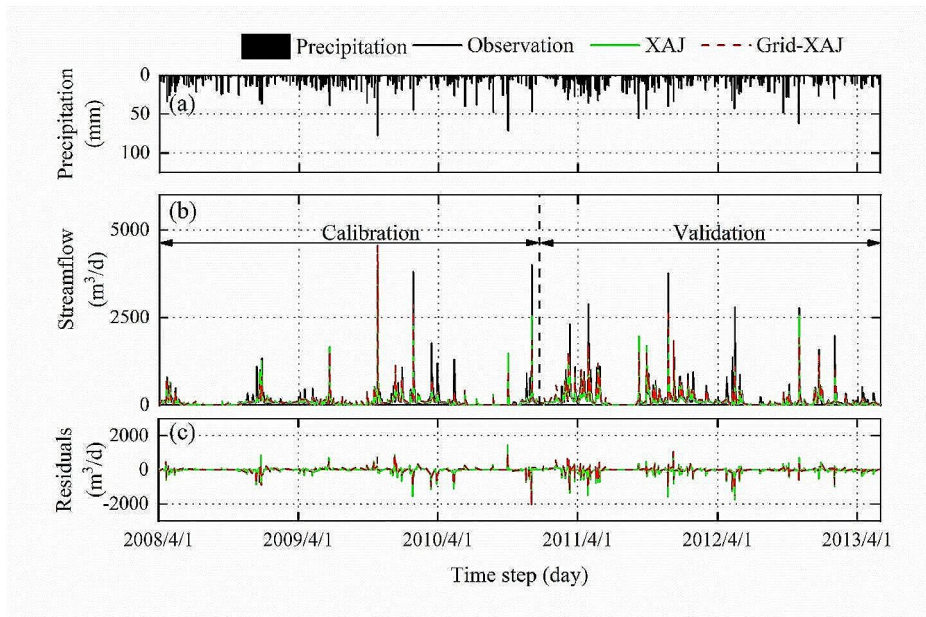


Fig. 4 Observed and simulated daily streamflow as well as their difference at the Shale Hills catchment from 2008 to 2013. **(a)** Daily precipitation time series. **(b)** Comparison of streamflow observations and simulations during the calibration and validation periods using XAJ and Grid-XAJ model. **(c)** Difference between simulated and observed daily streamflow

(Fig. 5b). Both the XAJ model and the Grid-XAJ model achieved statistically significant improvements in their NSE and R values, passing the significance test at the 0.01 level. This result underscores a significant enhancement in streamflow modeling before and after data assimilation. The Grid-XAJ approach demonstrates superior performance when employing SM-DA with the EnKF method in comparison to the lumped approach. The box plots in Fig. 5a further indicate that the NSE of the model simulations using SM-DA display narrower ranges, suggesting less uncertainty in streamflow modeling.

The simulated streamflow between the lumped and distributed model approach employing the assimilation method was compared. The results clearly demonstrate improvements achieved through assimilation, with the NSE increasing from 0.61 to 0.65 for the lumped model and from 0.62 to 0.70 for the distributed model (Table S4). Additionally, the R increased from 0.80 to 0.81 for the lumped approach and from 0.79 to 0.86 for the Grid-XAJ model, indicating that the assimilation methods enhanced the accuracy of runoff estimates by effectively integrating SM. Consequently, SM-DA proves to be effective in reducing uncertainty associated with streamflow simulations. Furthermore, the overall assimilation efficiency (EFF) for simulating streamflow is 11.6% for the XAJ model and 19.9% for the distributed approach. These findings underline the success of both lumped and distributed models in leveraging SM-DA to improve streamflow predictions, with the distributed model displaying a greater potential for enhancing streamflow prediction.

The similar analysis was conducted across different seasons. The improvements in streamflow modeling resulting from SM-DA were observed to be more pronounced during the wet-cold season (December-May) compared to the dry-warm season (June-November).

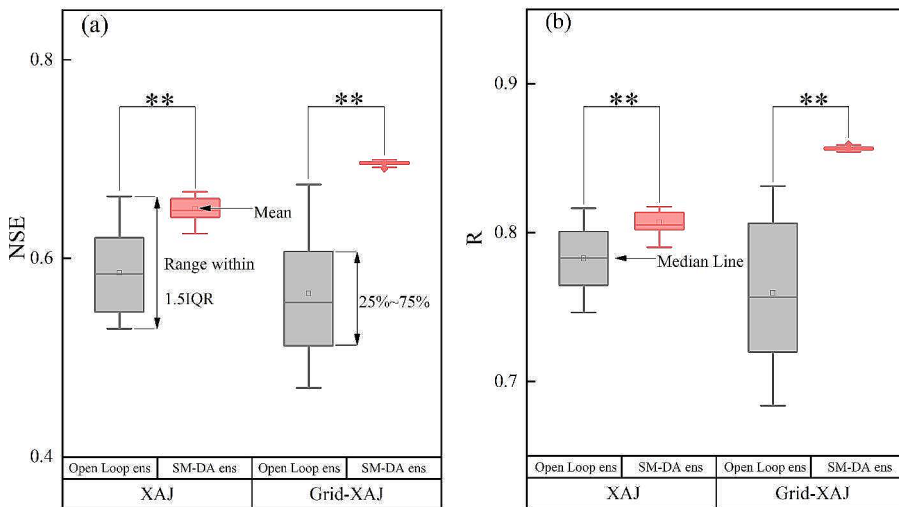


Fig. 5 Box plots showing the performance of simulated streamflow obtained from Open Loop and SM-DA runs using the ensemble members (Open Loop ens and SM-DA ens) in terms of (a) NSE, and (b) R of XAJ and Grid-XAJ model, and ** indicates a significance level of 0.01

Substantial enhancements in streamflow prediction were evident for wet-cold seasons, with the NSE increasing from 0.43 to 0.63 for the XAJ model and from 0.45 to 0.58 for the Grid-XAJ model (Table S5). However, during dry-warm seasons, SM-DA did not yield considerable improvements over the Open Loop simulation. This could be attributed to the fact that, during warm-dry seasons, streamflow is primarily influenced by precipitation and ET. Conversely, in wet-cold seasons, SM exerts a greater impact on runoff simulation, thus rendering SM-DA more effective in enhancing streamflow prediction specifically for wet-cold seasons.

4.4 Assimilating Surface SM and Subsurface SM

We investigated the assimilation of surface SM and subsurface SM to assess their respective benefits for simulating streamflow. When using subsurface SM in the assimilation process, the simulated streamflow more closely matched the observed streamflow than using surface SM. The assimilation of surface SM led to slight improvements in streamflow predictions, with average EFF values of 3.2% and 7.8% in the lumped and distributed models, respectively (Table 1). In contrast, the assimilation of subsurface SM had a much greater impact on streamflow (EFF=6.8% and 14.0%). This result can be attributed to the substantial difference in thickness between the two layers. Subsurface SM-DA has more relevance for the hydrological processes in general. As forests predominantly occupy the basin, water for transpiration is extracted from deeper layers. The percolation process (groundwater recharge) is deeper in the vadose zone, so sub-surface SM is more effective than the surface SM, which controls only the surface runoff process. This shows the physical robustness of

Table 1 Evaluation of the corresponding benefits on streamflow simulation using surface and subsurface SM for data assimilation

Assimilation	XAJ		Grid-XAJ	
	NSE	EFF (%)	NSE	EFF (%)
Surface SM	0.62	3.2	0.64	7.8
Subsurface SM	0.64	6.8	0.66	14.0

the model. Accurately estimating SM in the subsoil is critical for simulating streamflow, as it accounts for a significant proportion of soil water in the vadose zone.

5 Discussion

5.1 Comparison of Streamflow Modelling between Lumped and Distributed Hydrological Models with SM-DA

The results of this study demonstrate the effective simulation of daily streamflow in the study catchment by both the lumped and distributed XAJ models, effectively capturing the dynamic rainfall-runoff processes. However, various statistical analyses indicate the superior performance of the distributed approach over the lumped approach. This improved performance can mainly be attributed to considering the spatial distribution of SM data and catchment attributes, such as soil properties and topography, in the distributed model. The spatial heterogeneity of these catchment attributes significantly influences the patterns of SM. The Shale Hills catchment exhibits complex terrain, with steep slopes and clear spatial variability of soil types (Lin et al. 2006) also see Fig. 1. The valley geomorphological feature dominated by Ernest soils generally exhibits higher moisture content, and ridges dominated by Weikert soils are the driest areas in the catchment. These distinct characteristics can lead to different flow generation mechanisms, with the distributed models providing the means to account for distributed runoff generation (Xiao et al. 2019). In contrast, the lumped approach represents runoff mechanisms by disregarding the spatial heterogeneity of catchment attributes, relying instead on a single average soil water storage value for the entire area (Alvarez-Garreton et al. 2015). Our study highlights that this approach poses significant limitations for calibration schemes and serves as a major source of uncertainty. Consequently, our results underscore the advantages of utilizing the distributed approach, which demonstrates higher accuracy and lower uncertainty in comparison to the lumped model (Tables S3 and S4).

The spatial distribution of catchment attributes plays a significant role in streamflow generation. Further, the simulated streamflow by the lumped and distributed was divided into surface runoff and subsurface runoff, encompassing interflow and groundwater. The proportions of streamflow results from the lumped (applied to the entire catchment) and the distributed approaches (applied to six soil-terrain units) are depicted after the assimilation of SM (Fig. S3). The surface runoff simulated by both models accounts for a small portion, constituting only 16% for the lumped model and ranging from 20 to 36% for the distributed model (Fig. S3). Notably, the proportion of surface runoff varies significantly among all six soil-terrain units in the distributed model. The valley, characterized by a smaller slope (3.2–18.4%), exhibits the lowest surface runoff proportion (20%). In contrast, the S-Slope and Ridge units display the highest proportions of surface runoff (36% and 32%, respectively) attributed to greater steepness (20.4–36% and 7.5–22.3%). The XAJ and Grid-XAJ models

employ the concept of a free water reservoir to segregate different types of runoff. In these models, incoming rainfall first compensates for any deficit in soil moisture, contributing to the formation of interflow. This process is followed sequentially by subsurface runoff and then by surface runoff. The relationship between rainfall events and surface flow is provided (Fig. S4). Rainfall exhibits a strong correlation with both surface and subsurface runoff across six distinct soil-terrain units, which underscores its critical role in recharging runoff systems. Typically, rainfall is the primary contributor to surface runoff. During smaller precipitation events, limited surface runoff is generated as a greater proportion of the rainfall infiltrates the soil. Conversely, during intense rainfall events, the infiltration capacity is overwhelmed by the volume of rainfall, resulting in more substantial surface runoff. On the other hand, subsurface flows exhibited dominance and sustained for extended periods, demonstrating relative steadiness (Fig. S4).

5.2 Uncertainties and Limitations of the SM-DA Approach

This study provides compelling evidence supporting the effectiveness of integrating in situ SM measurements into the XAJ model to enhance streamflow prediction. However, it is important to acknowledge that both models exhibit underestimation of the highest streamflow peaks. This discrepancy can be attributed to the limited density of rainfall gauges in the catchment area, which emerges as a prominent factor contributing to these prediction errors. The neglect of spatial heterogeneity of rainfall distribution may further contribute to the underestimation of peak flows. Another crucial consideration is the significant influence of preferential flow within the study catchment, which has not been accounted for in the XAJ model. Previous studies have documented evidence of preferential flow paths through macropore networks in forested hillslopes (Guo et al. 2018). Subsurface preferential flow also occurs at interfaces between soil horizons and soil-rock transitions within the study catchment (Lin et al. 2006; Liu et al. 2020). In addition, Yu et al. (2014) highlighted the successful performance of hydrological models incorporating macropore effects in predicting peak flow within the Shale Hills catchment, emphasizing the critical role of macropore flow in determining recharge thresholds and runoff response.

The underlying structure of hydrological models highly influences the incorporation of SM data for streamflow prediction (Massari et al. 2015; Nayak et al. 2021). This study utilized the XAJ model, a two-layer soil water balance model consisting of an upper and lower layer. However, it is crucial to note that the depth of SM considered by the XAJ model does not always align precisely with the in situ measurements. Additionally, the hydrological partitioning in the XAJ model is based on the SM deficit in each layer, representing the amount of moisture required to reach saturation level Zhao (1992). The XAJ model assumes that soil water movement to the lower layer occurs only after the upper layer is saturated. This leads to higher variations in the upper layer and less variation in the deeper layer, contrary to actual soil water movement. Similar limitations are also present in other models, such as SWAT (Chen et al. 2010; Patil and Ramsankaran 2017). These discrepancies can introduce errors when assimilating surface and subsurface SM data into the XAJ model. These limitations may affect the model's ability to accurately describe the soil moisture state, which in turn affects the model's ability to understand and predict hydrological processes. Previous studies have highlighted the influence of SM-DA on streamflow prediction, emphasizing the importance of reliable SM information obtained from in-situ measurements compared

to that derived from meteorological data during simulation (Nayak et al. 2021). It is worth noting that the reliability of SM information is a critical factor affecting the effectiveness of SM-DA. This discrepancy in data reliability is one of the primary reasons that SM-DA did not consistently improve or even resulted in degraded streamflow simulation accuracy in certain studies (Massari et al. 2015).

5.3 Future Outlook

This study evaluates the effects of assimilating in situ SM data into lumped and distributed XAJ models in a forested catchment with complex terrain. Several factors influencing the improvement in streamflow prediction with SM-DA need to be addressed in future studies to investigate the remaining uncertainties. For instance, our findings suggest that the model structure generally influences the accuracy of SM-DA schemes. Therefore, developing of hydrological models incorporating a multilayer soil water sub-model capable of better characterizing vertical soil water movement is crucial. Additionally, the representation of preferential flow in catchment hydrological modeling emerges as another important issue (Xiao et al. 2019). In forested catchments, preferential flow through macropores can result in significant and rapid infiltration and deep percolation (Sidle et al. 2001). Previous studies have demonstrated the improvements achieved by incorporating preferential flow concepts compared to models without them (Yu et al. 2014). The XAJ model lacks the capability to effectively incorporate the impacts of preferential flow. To address this issue and more accurately represent the hydrological process and runoff generation mechanism in the Shale Hills catchment, either a different hydrological model that includes preferential flow could be utilized, or the XAJ model could be enhanced by integrating a module specifically designed to account for preferential flow.

Our findings demonstrate that soil properties and topographic characteristics contribute additional improvements to distributed flow simulation in catchments with complex terrains. Therefore, it is crucial to employ a distributed model to enhance data assimilation performance (Yu et al. 2014). Although distributed physics-based modeling has been utilized to explore the impacts of spatial catchment characteristics on streamflow and overall catchment model performance, further investigation is necessary. Furthermore, it should be noted that our study was conducted within a relatively steep catchment, and the results of SM-DA were inherently model (and site) specific. To validate this potential, further studies will be conducted in catchments with flatter terrain in the future. Moreover, while the EnKF method was employed for SM-DA in this study, conducting in-depth analysis to evaluate alternative assimilation techniques would be worthwhile.

6 Conclusions

This work presents an evaluation of assimilating in situ SM into hydrological models to reduce streamflow prediction uncertainty in a well-monitored catchment. The study explores the advantages of SM spatial distribution within the catchment using both lumped and distributed hydrological models. The results demonstrate that the SM data in the study catchment exhibits a strong response to precipitation and streamflow. Comparing the simulations of the SM-DA runs with the Open Loop runs, a general improvement is observed

in both the lumped and distributed models. Specifically, the distributed model ($NSE=0.70$, $R=0.86$) outperforms the lumped model ($NSE=0.65$, $R=0.81$) by considering the spatial heterogeneity of SM and catchment attributes, thus providing more accurate and robust streamflow simulations. SM updates in the wet-cold season primarily enhance the performance of streamflow simulation compared to the dry-warm season. Furthermore, the assimilation of subsurface SM yields greater improvements in streamflow simulation than the surface SM. This highlights the importance of incorporating deeper SM information for accurate streamflow prediction. However, it should be noted that the updated simulation in the catchment remains limited by the model structure and the quality of forcing data prior to assimilation. SM-DA alone cannot address the systematic errors present in the model pre-assimilation. In conclusion, this study contributes to understanding of the connection between SM, streamflow, and hydrological connectivity in headwater catchments, and also offers crucial methodological and theoretical support for accurate simulation of hydrological processes in catchments and real-time calibration of hydrological models.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11269-024-03895-9>.

Acknowledgements Logistical support was supported by the USA NSF-supported Susquehanna Shale Hills Critical Zone Observatory.

Author contribution **Hongxia Li**: Conceptualization, Data curation, Methodology, Funding acquisition, Writing-Review & editing. **Yuanyuan Huang**: Data curation, Formal analysis, Methodology, Software, Investigation, Writing-original draft. **Yongliang Qi**: Data curation, Investigation, Methodology. **Yanjia Jiang**: Data curation, Investigation. **Xuan Tang**: Data curation. **Elizabeth W. Boyer**: Visualization. **Carlos R. Mello**: Visualization. **Ping Lan**: Validation, Data curation. **Li Guo**: Conceptualization, Data curation, Methodology supervision, Writing-Review & editing.

Funding This work was supported by the National Natural Science Foundation of China (Grant No. 51979177 and 42201033).

Data Availability Data can be provided upon request.

Declarations

Ethical Approval The manuscript has not been submitted to any other journal. The proposed work is original and has not been published anywhere else.

Consent to Participate Not applicable.

Consent to Publish The research is scientifically consent to be published.

Competing Interests The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- Alvarez-Garreton C, Ryu D, Western AW et al (2015) 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

- Avellaneda PM, Ficklin DL, Lowry CS, Knauft JH, Hall DM (2020) Improving hydrological models with the assimilation of crowdsourced data. *Water Resour Res* 56
- Beven K (2016) Facets of uncertainty: epistemic uncertainty, non-stationarity, likelihood, hypothesis testing, and communication. *Hydrol Sci J* 61:1652–1665. <https://doi.org/10.1080/02626667.2015.1031761>
- Bournas A, Baltas E (2021) Increasing the efficiency of the sacramento model on event basis in a mountainous river basin. *Environ Process* 8:943–958. <https://doi.org/10.1007/s40710-021-00504-4>
- Brocca L, Moramarco T, Melone F et al (2012) Assimilation of surface-and root-zone ascats soil moisture products into rainfall-runoff modeling. *IEEE Trans Geosci Remote Sens* 50:2542–2555
- Cai J, Liu Y, Lei T, Pereira LS (2007) Estimating reference evapotranspiration with the FAO Penman-Monteith equation using daily weather forecast messages. *Agric Meteorol* 145:22–35
- Caldwell TG, Bongiovanni T, Cosh MH et al (2019) The texas soil observation network: a comprehensive soil moisture dataset for remote sensing and land surface model validation. *Vadose Zone J* 18
- Chen F, Crow W, Moriasi DN, Starks P (2010) Assimilating remotely sensed surface soil moisture into swat using ensemble kalman filter, Proceedings of the 2010 Watershed Management Conference
- Duan Q, Sorooshian S, Gupta V (1992) Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resour Res* 28:1015–1031
- Evensen G (1994) Sequential data assimilation with a nonlinear quasi-geostrophic model using monte-carlo methods to forecast error statistics. *J Geophys Res: Oceans* 99:10143–10162
- Fan BH, Tao WH, Qin GH et al (2020) Soil micro-climate variation in relation to slope aspect, position, and curvature in a forested catchment. *Agric Meteorol* 290. <https://doi.org/10.1016/j.agrformet.2020.107999>
- Gan YJ, Zhang Y, Liu YQ, Kongoli C, Grassotti C (2022) Assimilation of blended in situ-satellite snow water equivalent into the National Water Model for improving hydrologic simulation in two US river basins. *Sci Total Environ* 838. <https://doi.org/10.1016/j.scitotenv.2022.156567>
- Gottinger J, Bardossy A (2008) Generic error model for calibration and uncertainty estimation of hydrological models. *Water Resour Res* 44
- Guo L, Fan B, Zhang J, Lin H (2018) Occurrence of subsurface lateral flow in the Shale Hills Catchment indicated by a soil water mass balance method. *Eur J Soil Sci* 69:771–786. <https://doi.org/10.1111/ejss.12701>
- Ireson AM, Sanchez-Rodriguez I, Basnet S et al (2022) Using observed soil moisture to constrain the uncertainty of simulated hydrological fluxes. *Hydrol Process* 36. <https://doi.org/10.1002/hyp.14465>
- Khaki M, Hoteit I, Kuhn M, Forootan E, Awange J (2019) Assessing data assimilation frameworks for using multi-mission satellite products in a hydrological context. *Sci Total Environ* 647:1031–1043
- Koohi S, Azizian A, Brocca L (2022) Calibration of a distributed hydrological model (VIC-3L) based on global water resources reanalysis datasets. *Water Resour Manage* 36:1287–1306. <https://doi.org/10.1007/s11269-022-03081-9>
- Laiolo P, Gabellani S, Campo L et al (2016) Impact of different satellite soil moisture products on the predictions of a continuous distributed hydrological model. *Int J Appl Earth Obs Geoinf* 48, 131–145
- Li HX, Zhang YQ, Chiew FHS, Xu SG (2009) Predicting runoff in ungauged catchments by using Xinanjiang model with MODIS leaf area index. *J Hydrol* 370:155–162. <https://doi.org/10.1016/j.jhydrol.2009.03.003>
- Li HX, Zhang YQ, Zhou XY (2015) Predicting surface runoff from catchment to large region. *Adv Meteorol* 2015. <https://doi.org/10.1155/2015/720967>
- Li ZL, Leng P, Zhou C, Chen KS, Shang GF (2021) Soil moisture retrieval from remote sensing measurements: current knowledge and directions for the future. *Earth Sci Rev* 218:103673
- Lin H (2006) Temporal stability of soil moisture spatial pattern and subsurface preferential flow pathways in the shale hills catchment. *Vadose Zone J* 5:317–340. <https://doi.org/10.2136/vzj2005.0058>
- Lin H, Zhou X (2008) Evidence of subsurface preferential flow using soil hydrologic monitoring in the Shale Hills catchment. *Eur J Soil Sci* 59:34–49. <https://doi.org/10.1111/j.1365-2389.2007.00988.x>
- Lin HS, Kogelmann W, Walker C, Bruns MA (2006) Soil moisture patterns in a forested catchment: a hydrogeological perspective. *Geoderma* 131:345–368
- Liu H, Lin H (2015) Frequency and control of subsurface preferential flow: from pedon to catchment scales. *Soil Sci Soc Am J* 79:362–377
- Liu H, Yu Y, Zhao WZ et al (2020) Inferring subsurface preferential flow features from a wavelet analysis of hydrological signals in the shale hills catchment. *Water Resour Res* 56. <https://doi.org/10.1029/2019wr026668>
- Liu YW, Cui W, Ling Z et al (2024) The impact of satellite soil moisture data assimilation on the hydrological modeling of swat in a highly disturbed catchment. *Remote Sens* 16
- Loizu J, Massari C, Alvarez-Mozos J et al (2018) On the assimilation set-up of ASCAT soil moisture data for improving streamflow catchment simulation. *Adv Water Resour* 111:86–104. <https://doi.org/10.1016/j.advwatres.2017.10.034>

- Massari C, Brocca L, Tarpanelli A, Moramarco T (2015) Data assimilation of satellite soil moisture into rainfall-runoff modelling: a complex recipe? *Remote Sens* 7:11403–11433. <https://doi.org/10.3390/rs70911403>
- Matgen P, Fenicia F, Heitz S et al (2012) Can ASCAT-derived soil wetness indices reduce predictive uncertainty in well-gauged areas? A comparison with in situ observed soil moisture in an assimilation application. *Adv Water Resour* 44:49–65. <https://doi.org/10.1016/j.advwatres.2012.03.022>
- Mello CR, Vieira NPA, Guzman JA et al (2021) Climate change impacts on water resources of the largest hydropower plant reservoir in southeast Brazil. *Water* 13. <https://doi.org/10.3390/w13111560>
- Nayak AK, Biswal B, Sudheer KP (2021) Role of hydrological model structure in the assimilation of soil moisture for streamflow prediction. *J Hydrol* 598. <https://doi.org/10.1016/j.jhydrol.2021.126465>
- Patil A, Ramsankaran R (2017) Improving streamflow simulations and forecasting performance of swat model by assimilating remotely sensed soil moisture observations. *J Hydrol*, S0022169417307357
- Primka EJ, Adams TS, Buck A, Eissenstat DM (2021) Topographical shifts in fine root lifespan in a mixed, mesic temperate forest. *PLoS ONE* 16
- Samuel J, Coulibaly P, Dumedah G, Moradkhani H (2014) Assessing model state and forecasts variation in hydrologic data assimilation. *J Hydrol* 513:127–141
- Shi YN, Davis KJ, Zhang FQ, Duffy CJ, Yu X (2014) Parameter estimation of a physically based land surface hydrologic model using the ensemble Kalman filter: a synthetic experiment. *Water Resour Res* 50:706–724
- Sidle RC, Noguchi S, Tsuboyama Y, Laursen K (2001) A conceptual model of preferential flow systems in forested hillslopes: evidence of self-organization. *Hydrol Process* 15:1675–1692. <https://doi.org/10.1002/hyp.233>
- Tekeli AE, Fouli H (2017) Reducing false flood warnings of trmm rain rates thresholds over riyadh city, Saudi Arabia by utilizing amsr-e soil moisture information. *Water Resour Manag* 31:1243–1256
- Venegas-Cordero N, Cherrat C, Kundzewicz ZW, Singh J, Piniewski M (2023) Model-based assessment of flood generation mechanisms over Poland: the roles of precipitation, snowmelt, and soil moisture excess. *Sci Total Environ* 891:164626–164626
- Wang QJ (1991) The genetic algorithm and its application to calibrating conceptual rainfall-runoff models. *Water Resour Res* 27:2467–2471
- Wu WY, Yang ZL, Zhao L, Lin PR (2022) The impact of multi-sensor land data assimilation on river discharge estimation. *Remote Sens Environ* 279. <https://doi.org/10.1016/j.rse.2022.113138>
- Xiao DC, Shi YN, Brantley SL et al (2019) Streamflow generation from catchments of contrasting lithologies: the role of soil properties, topography, and catchment size. *Water Resour Res* 55:9234–9257
- Yao C, Li Z, Yu Z, Zhang K (2012) A priori parameter estimates for a distributed, grid-based Xinanjiang model using geographically based information. *J Hydrol* 468:47–62. <https://doi.org/10.1016/j.jhydrol.2012.08.025>
- Yin ZR, Qin GH, Guo L et al (2022) Coupling antecedent rainfall for improving the performance of rainfall thresholds for suspended sediment simulation of semiarid catchments. *Sci Rep* 12. <https://doi.org/10.1038/s41598-022-08342-6>
- Yu X, Duffy C, Baldwin DC, Lin H (2014) The role of macropores and multi-resolution soil survey datasets for distributed surface-subsurface flow modeling. *J Hydrol* 516:97–106. <https://doi.org/10.1016/j.jhydrol.2014.02.055>
- Zhao RJ (1992) The Xinanjiang model applied in China. *J Hydrol* 135:371–381. [https://doi.org/10.1016/0022-1694\(92\)90096-e](https://doi.org/10.1016/0022-1694(92)90096-e)
- Ziliani MG, Ghostine R, Ait-El-Fquih B, McCabe MF, Hoteit I (2019) Enhanced flood forecasting through ensemble data assimilation and joint state-parameter estimation. *J Hydrol* 577. <https://doi.org/10.1016/j.jhydrol.2019.123924>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Authors and Affiliations

Hongxia Li¹ · Yuanyuan Huang¹ · Yongliang Qi¹ · Yanjia Jiang¹ · Xuan Tang¹ · Elizabeth W. Boyer² · Carlos R. Mello³ · Ping Lan¹ · Li Guo¹

✉ Li Guo
liguo01@scu.edu.cn

¹ State Key Laboratory of Hydraulics and Mountain River Engineering, College of Water Resource and Hydropower, Sichuan University, Chengdu 610065, China

² Department of Ecosystem Science and Management, Pennsylvania State University, University Park, PA 16802, USA

³ Water Resources Department, Universidade Federal de Lavras, CP 3037, Lavras 37200- 900, MG, Brazil