Geoderma 455 (2025) 117210

Contents lists available at ScienceDirect

Geoderma

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

Optimization of soil hydraulic parameters within a constrained sampling space

Meijun Li [a], Wei Shao [a,b,*], Wenjun Yu [a], Ye Su [c,d], Qinghai Song [e], Yiping Zhang [e], Hongkai Gao [f], Yonggen Zhang [g], Jianzhi Dong [g]

[a] Key Laboratory of Hydrometeorological Disaster Mechanism and Warning, Ministry of Water Resources, School of Hydrology and Water Resources, Nanjing University of Information Science and Technology, Nanjing 210044, China

[b] Qinghai Provincial Meteorological Disaster Prevention and Defense Technology Center, Sining Qinghai 810001, China

[c] Department of Physical Geography and Geoecology, Faculty of Science, Charles University, Albertov 6, 128 43 Prague 2, Czech Republic

[d] Department of Physical Geography, and the Bolin Centre for Climate Research, Stockholm University, SE-106 91 Stockholm, Sweden

[e] Key Laboratory of Tropical Forest Ecology, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Mengla, Yunnan 666303, China

[f] School of Geographical Sciences, East China Normal University, Shanghai, China

[g] Institute of Surface-Earth System Science, Tianjin University, Tianjin, China

A R T I C L E I N F O

Keywords:
Soil hydraulic parameters
Particle Swarm Optimization (PSO)
Markov Chain Monte Carlo (MCMC)
Sequential Monte Carlo (SMC)
Rosetta 3 pedotransfer function

A B S T R A C T

The direct optimization of soil hydraulic parameters (SHP) in unconstrained parameter space introduces significant uncertainties in ecohydrological modeling, particularly when addressing the complex model structure of Richards' equation combined with Penman-Monteith equation. Pedotransfer functions (e.g., the latest version of Rosetta 3), which have been extensively trained using abundant soil hydraulic data, could potentially provide a physical constraint for sampling SHP. This study integrates optimization algorithms (Particle Swarm Optimization, PSO; Markov Chain Monte Carlo, MCMC; Sequential Monte Carlo, SMC; Generalized Likelihood Uncertainty Estimation, GLUE) with two sampling strategies – direct optimization of SHP and indirect optimization of SHP derived from soil particle composition (SPC) using Rosetta 3 – to evaluate their performance in ecohydrological modeling under predefined soil conditions. The results demonstrated that indirect optimization of SHP significantly enhances the accuracy in recovering predefined true parameters and states, and reduces the uncertainty of ecohydrological modeling compared to direct optimization of SHP. Specifically, the mean quartile deviation of biases in soil water content and evaporation were reduced from 0.0347 m^3/m^3 and 0.0027 m/h to 0.0061 m^3/m^3 and 0.0010 m/h, respectively. Furthermore, integration of the Rosetta 3 diminished the dimensionality of inverse modeling, thereby significantly enhancing algorithm convergence speed and precision. It is recommended to integrate Rosetta 3 with various optimization algorithms to enhance the accuracy of ecohydrological modeling.

1. Introduction

In the field of unsaturated hydrology, modeling soil moisture content (θ) and soil water potential (h) through Richards' equation is essential to support various applications, including agriculture management, geotechnical engineering, and earth system modeling (Cho, 2014; Liu and Wang, 2021; Xu et al., 2022). Soil hydraulic parameters (SHP), the description of the storage and transport properties of water within the soil profile (e.g. the soil water characteristic curve SWCC and the unsaturated hydraulic conductivity functions UHCF), are fundamental for comprehending and analyzing the soil hydrological processes (Vereecken et al., 2022). Various types of parametric equations have been employed for modeling soil hydrological processes. Notably, the parametric equation proposed by van Genuchten (VG) has been extensively utilized due to its simplicity and wide applicability across diverse soil types (Ippisch et al., 2006; Luo et al., 2019). A significant advantage of the VG model lies in its sharing of the same fitting parameters (α and n) between SWCC and UHCF, reducing the dimensionality in SHP to five parameters (α, n, residual water content θ_r, saturated water content θ_s and saturated conductivity K_s).

* Corresponding author at: Key Laboratory of Hydrometeorological Disaster Mechanism and Warning, Ministry of Water Resources, School of Hydrology and Water Resources, Nanjing University of Information Science and Technology, Nanjing 210044, China.
E-mail address: Shao@nuist.edu.cn (W. Shao).

https://doi.org/10.1016/j.geoderma.2025.117210
Received 18 January 2024; Received in revised form 31 December 2024; Accepted 4 February 2025
Available online 16 February 2025

Despite the utilization of numerous field and laboratory measurement techniques for the identification of SHP with advantage of precision (Bordoni et al., 2017), an obstacle persists owing to the challenge of measuring SHP with adequate spatial and temporal coverage (Vereecken et al., 2022). Furthermore, the measured SHP may exhibit only statistical significance rather than precise values that are suitable for soil hydrological modeling (Scharnagl et al., 2011). Inverse estimation of SHP through auto-calibration (using optimization algorithms) were commonly used in practical application (Scharnagl et al., 2011; Vereecken et al., 2022). The inversely estimated SHP, based on θ, is widely favored for its simplicity and the extensive availability of θ obtained from in-situ measurements and remote sensor technologies. However, the equifinality phenomenon, prevalent in highly nonlinear modeling systems, poses a threat to the reliability of SHP optimization and soil hydrological modeling (Shao et al., 2023a; Shao et al., 2023b).

The equifinality phenomenon predominantly arise from the compensation effects of implicit linked parameters and certain non-physical parameters that distort the physical relationship between SWCC and UHCF (Eberhart and Kennedy, 1995). In principle, the equifinality phenomenon can be alleviated by reducing the dimensionality of the parameter sampling space (Pollacco et al., 2008). Prior knowledge of soil properties can effectively constrain the parameter sampling space (Vereecken et al., 2022). In the realm of soil physics, the soil particle composition (SPC) is commonly utilized to estimate SHP (Chirico et al., 2007; Vereecken et al., 2010). For natural soils at shallow depths, their compressibility is likely to be similar, thereby rendering it reasonable to estimate SHP with SPC under comparable bulk density. Given that Rosetta 3, an artificial neural network-based pedotransfer function, incorporates extensive prior information on SHP for undisturbed soil samples (encompassing 2134 soil samples with water retention data and 1306 soil samples with measurements of saturated hydraulic conductivities), it could serve as a valuable instrument to constrain the parameter space of SHP (Zhang and Schaap, 2017). However, to the best of our knowledge, no comprehensive study has been conducted to investigate the statistical implication of optimized SHP within a constrained sampling space using Rosetta 3 pedotransfer function.

To evaluate the enhancement of Rosetta 3 in parameter optimization, this study conducted a systematic numerical experiment with two sampling strategies: direct optimization of SHP (Direct sampling strategy) and indirect optimization of SHP derived from soil particle composition (SPC) based on Rosetta 3 (Indirect sampling strategy). Specifically, three representative soil types (silty, loam, and sandy soil) were selected as study objects. Predefined true values for soil moisture state variables and evaporation rate were obtained under actual vegetation and meteorological conditions. Four prevalent parameter optimization algorithms, namely Particle Swarm Optimization (PSO), Markov Chain Monte Carlo (MCMC), Sequential Monte Carlo (SMC), and General Likelihood Uncertainty Estimation (GLUE), were employed to inversely estimate the optimal SHP for the unsaturated flow model. The study compared the results obtained using these four algorithms under two sampling strategies (Direct sampling strategy and Indirect sampling strategy) focusing on three key aspects: (1) whether the optimization algorithms could encompass the predefined true parameter values or manifest equifinality phenomena; (2) whether the posterior SWCC exhibited well-constrained results with a narrow uncertainty band under Indirect sampling strategy; and (3) the evaluation of the impact of parameter uncertainty on soil moisture and evaporation simulations.

2. Theory

In the soil hydrology model, the Mualem-van Genuchten (VG) model always used to characterized SWCC and UHCF (van Genuchten, 1980):

$$\Theta = \frac{\theta - \theta_r}{\theta_s - \theta_r} = \begin{cases} \dfrac{1}{\left(1 + |\alpha h|^{n}\right)^{m}}, h < 0 \\ 1, h \geq 0 \end{cases} \tag{1}$$

$$K = \begin{cases} K_s \Theta^l \left[1 - \left(1 - \Theta^{1/m}\right)^m\right]^2, h < 0 \\ K_s, h \geq 0 \end{cases} \tag{2}$$

where θ ($m^3\ m^{-3}$) is the soil moisture, the subscripts s and r represent the saturated and residual water content; α (m^{-1}) is the scaling parameter, its reciprocal is positively related to air entry pressure; n and m are the dimensionless parameter related to curve shape, where $m = 1 - 1/n$; Θ is the effective saturation; K_s ($m\ s^{-1}$) is the saturated hydraulic conductivity; l is an empirical parameter with typical default value of 0.5.

In the conventional sampling strategy, parameter samples within the optimization algorithm are both initialized and subsequently updated through random process confined to the same parameter space as forward models (e.g. direct sampling of SHP). This strategy is referred to as the Direct sampling strategy:

$$\boldsymbol{x}_{i,\tau}^{\mathrm{SHP}} = [\alpha, n, \theta_r, \theta_s, K_s] \tag{3}$$

Here, $\boldsymbol{x}_{i,\tau}^{\mathrm{SHP}}$ represents the i^{th} sample in τ^{th} iteration under Direct sampling strategy, where $i \in \{1, 2, \cdots, N_s\}$ and $\tau \in \{1, 2, \cdots, N_\tau\}$; N_s signifies the total number of samples, while the N_τ denotes the total number of iterations. The samples set at τ^{th} iteration of $\boldsymbol{x}_{i,\tau}^{\mathrm{SHP}}$ is named $\boldsymbol{X}_{\tau}^{\mathrm{SHP}} = \left[\boldsymbol{x}_{i,\tau}^{\mathrm{SHP}}\right], i \in \{1, 2, \cdots, N_s\}$. The last iteration of $\boldsymbol{X}_{\tau=N_\tau}^{\mathrm{SHP}}$ is denoted as the posterior parameters ($\boldsymbol{X}_{\mathrm{post}}^{\mathrm{SHP}}$) for statistical analysis of their probabilistic distribution. The entire possible range and probabilistic distributions of values for $\boldsymbol{x}_{i,\tau}^{\mathrm{SHP}}$ are denotes as $\Omega_{\mathrm{SHP}} = \{\alpha, n, \theta_r, \theta_s, K_s\}$, which represents 5-deminsional SHP space under Direct sampling strategy.

An innovative sampling strategy, termed the Indirect sampling strategy, is employed to sample SHP within a constrained parameter sampling space using the Rosetta 3 pedotransfer function. Under this strategy, the optimization algorithm initializes and updates samples in a constrained sampling space (SPC space):

$$\boldsymbol{x}_{i,\tau}^{\mathrm{SPC*}} = [f_{\mathrm{sand\%}}, f_{\mathrm{silt\%}}, f_{\mathrm{clay\%}}] \tag{4}$$

where $\boldsymbol{x}_{i,\tau}^{\mathrm{SPC*}}$ signifies the i^{th} sample in τ^{th} iteration; the superscript * denotes Indirect sampling strategy; the samples set at τ^{th} iteration of $\boldsymbol{x}_{i,\tau}^{\mathrm{SPC*}}$ was names $\boldsymbol{X}_{\tau}^{\mathrm{SPC*}} = \left[\boldsymbol{x}_{i,\tau}^{\mathrm{SPC*}}\right], i \in \{1, 2, \cdots, N_s\}$. The last iteration of $\boldsymbol{X}_{\tau=N_\tau}^{\mathrm{SPC*}}$ was the denotes as the posterior parameters ($\boldsymbol{X}_{\mathrm{post}}^{\mathrm{SPC*}}$). The terms $f_{\mathrm{sand\%}}$, $f_{\mathrm{silt\%}}$ and $f_{\mathrm{clay\%}}$ denote the fraction of sand, silt and clay particles, where $f_{\mathrm{clay\%}} + f_{\mathrm{silt\%}} + f_{\mathrm{sand\%}} = 100\%$; The entire possible range and probabilistic distributions of values for $\boldsymbol{x}_{i,\tau}^{\mathrm{SPC*}}$ are denotes as $\Omega_{\mathrm{SPC}} = \{f_{\mathrm{sand\%}}, f_{\mathrm{silt\%}}, f_{\mathrm{clay\%}}\}$, which represents the SPC sampling space under Indirect sampling strategy. Considering the summation of all three groups of particle equals 1 as a necessary constrain, the dimensionality of Ω_{SPC} is 2.

Rosetta 3, developed using the Python 3 programming language, conveniently supports the integration of newly developed numerical models (Zhang and Schaap, 2017). Detailed information about Rosetta 3 is available at https://github.com/usda-ars-ussl/rosetta-soil. Rosetta 3 incorporates five hierarchical models to estimate SHP. The first model (H1*w*) is a simple lookup table for 12 typical soil textures under USDA classifications. The second (H2w) through fifth models (H5w) are artificial neural networks, trained using bootstrap resampling of soil hydraulic data. Models H2*w* to H5*w* support detailed inputs of SPC, allowing for more accurate SHP predictions across the entire soil texture triangle. Specifically, H2w use SPC as its sole input, while other models (H3*w*, H4*w*, H5*w*) requiring additional information of bulk density (ρ_{BD}) and soil water content at 33 kPa (θ_{33}) and 1500 kPa (θ_{1500}) for models

H4*w* and H5*w*. This enables the description of SHP in soils with more distinct pore structures and provide a more accurate description for soil under special condition.

Commonly, the H2*w* model in Rosetta 3 suffices for describing the soil properties of the topsoil layer. Therefore, in this study, the H2*w* model was adopted to constrain the sampling space of SHP. As illustrated in Fig. 1, when using the Rosetta 3 H2*w* model to transfer the SPC randomly sampling within the soil texture triangle into SHP, the distribution of SHP exhibited bimodal characteristic (α and K_s) and skewed distribution (n, θ_r, and θ_s), rather than normal distribution or uniform distribution. This highlights the capability of Rosetta 3 to provide effective prior information for constraining the SHP parameter space.

Using Rosetta 3 pedotransfer function, $\boldsymbol{x}_{i,\tau}^{\mathrm{SPC}^*}$ can be transferred from SPC into SHP, denoted as $\boldsymbol{x}_{i,\tau}^{\mathrm{SHP}^*} = [\alpha, n, \theta_r, \theta_s, K_s]$:

$$\boldsymbol{x}_{i,\tau}^{\mathrm{SHP}^*} = \mathscr{R}\left(\boldsymbol{x}_{i,\tau}^{\mathrm{SPC}^*}\right) \tag{5}$$

where $\mathscr{R}$ represent the Rosetta 3 function. The transformed sample $\boldsymbol{x}_{i,\tau}^{\mathrm{SHP}^*}$ can be used in forward model. The samples set at τ^{th} iteration of $\boldsymbol{x}_{i,\tau}^{\mathrm{SHP}^*}$ is named as $\boldsymbol{X}_{\tau}^{\mathrm{SHP}^*} = \left[\boldsymbol{x}_{i,\tau}^{\mathrm{SHP}^*}\right], i \in \{1, 2, \cdots, N_s\}$. The last iteration of $\boldsymbol{X}_{\tau=N_\tau}^{\mathrm{SHP}^*}$ is denoted as the posterior parameters ($\boldsymbol{X}_{\mathrm{post}}^{\mathrm{SHP}^*}$).

The dimensionality of the constrained sampling space under Indirect sampling strategy diverges from conventional sampling space under the Direct sampling strategy. Specifically, Ω_{SHP} is 5-dimensional, as it simultaneously optimizes all five VG parameters. Conversely, Ω_{SPC}, used in this study, derives the five VG parameters from the sampling in a 2-dimensional SPC parameter space via Rosetta 3, thereby significantly reducing dimensionality of parameter space.

In this study, both sampling strategies are integrated into four algorithms, namely PSO, MCMC, SMC and GLUE. For the sake of clarity in the subsequent discussion of the update and resampling rules in Section 3, the parameters of both sampling strategies (denotes as $\boldsymbol{x}_{i,\tau}^{\mathrm{SHP}}, \boldsymbol{x}_{i,\tau}^{\mathrm{SPC}^*}$, and$\boldsymbol{x}_{i,\tau}^{\mathrm{SHP}^*}$) are simplified to $\boldsymbol{x}_{\tau}^{i}$. Similarly, the parameter sets (denotes as $\boldsymbol{X}_{\tau}^{\mathrm{SHP}}, \boldsymbol{X}_{\tau}^{\mathrm{SPC}^*}$, and$\boldsymbol{X}_{\tau}^{\mathrm{SHP}^*}$) are collectively referred to as $\boldsymbol{X}_{\tau}$.

3. Algorithm for optimizing soil hydraulic parameters

Within the framework of inverse modeling, each iteration of parameters is employed to advance forward model, thereby predicting the model's state:

$$\boldsymbol{y}_{\tau}^{i} = \mathscr{F}\left(\boldsymbol{x}_{\tau}^{i}, \widetilde{\boldsymbol{U}}_{\mathrm{forcing}}, \boldsymbol{\psi}_{\mathrm{initial}}\right) \tag{6}$$

$$\widetilde{\boldsymbol{y}}_{\mathrm{true}} = \boldsymbol{y}_{\tau}^{i} + \boldsymbol{\varepsilon} \tag{7}$$

where $\widetilde{\boldsymbol{y}}_{\mathrm{true}}$ represents the observation vector as times series ($\boldsymbol{t} = [1,2,\cdots, N_T]$), while $\boldsymbol{y}_{\tau}^{i}$ denotes the simulation vector estimated with parameter $\boldsymbol{x}_{\tau}^{i}$; The function $\mathscr{F}(\cdot)$ represents the simulator; $\boldsymbol{x}_{\tau}^{i}$ is the model param-

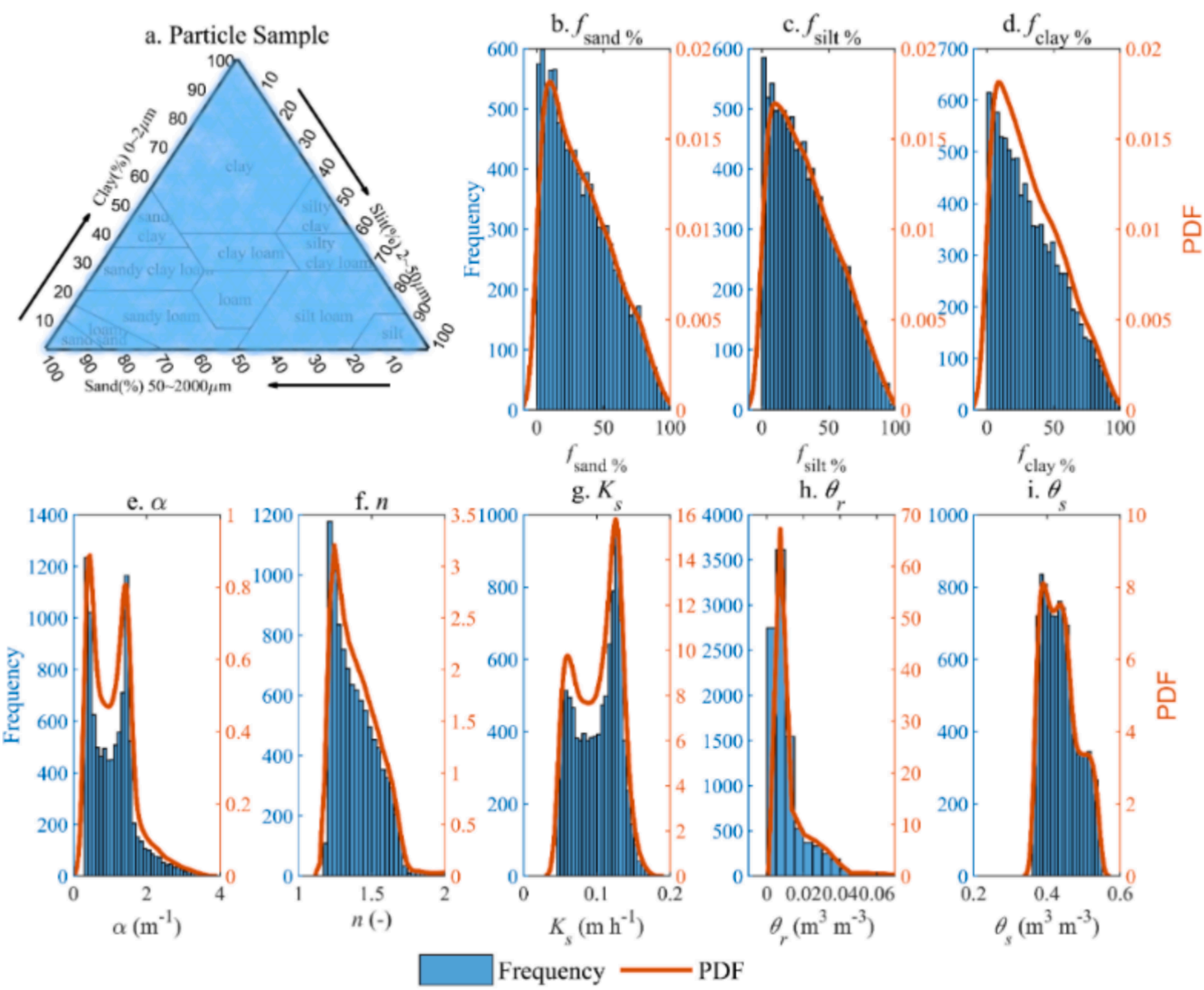

Fig. 1. The frequency histogram and probability distribution function (PDF) of the soil particle composition (SPC) and soil hydraulic parameters (SHP). The SPC were randomly sampled within the soil texture triangle, and the SHP were transferred from the SPC using the Rosetta 3 pedotransfer function.

eters with dimensionality of d; $\widetilde{\boldsymbol{U}}_{\text{forcing}}$ is the forcing variables (also called input data or boundary conditions); $\boldsymbol{\psi}_{\text{initial}}$ is the initial states, ε is the error including the observation error as well as modeling error.

The main model optimization process was shown in Fig. 2. In this study, the observation include soil moisture $\widetilde{\theta}_{\text{true}}$, soil water potential $\widetilde{\boldsymbol{h}}_{\text{true}}$ and evaporation $\widetilde{\boldsymbol{E}}_{\text{true}}$. $\mathscr{F}(\cdot)$ represents the forward model (the numerical model with details given in Section 3.1) for predicting simulated state variables $\boldsymbol{y}_\tau^i$ including soil moisture θ_τ^i, soil water potential $\boldsymbol{h}_\tau^i$ and evaporation $\boldsymbol{E}_\tau^i$; $\widetilde{\boldsymbol{U}}_{\text{Forcing}}$ stands for the meteorological forcing variables (detailed in Section 4.2), $\boldsymbol{\psi}_{\text{initial}}$ denotes the initial states including the initial soil moisture $\widetilde{\theta}_1$ and soil water potential $\widetilde{h}_1$, $\boldsymbol{\varepsilon}$ reflects the error between the observed and simulated state variables; $\boldsymbol{x}_\tau^i$ was parameters of VG model. Under Direct sampling strategy, $\boldsymbol{x}_\tau^i$ is denoted as $\boldsymbol{x}_{i,\tau}^{\text{SHP}}$, while under Indirect sampling strategy it is denoted as $\boldsymbol{x}_{i,\tau}^{\text{SHP}*}$. The true value of SHP was denoted as $\widetilde{\boldsymbol{x}}_{\text{true}}^{\text{SHP}} = [\widetilde{\alpha},\widetilde{n},\widetilde{\theta}_s,\widetilde{\theta}_r,\widetilde{K}_s]$. Given the challenge of accurately measuring the $\widetilde{\boldsymbol{x}}_{\text{true}}^{\text{SHP}}$ for natural soil, the true values of $\widetilde{\boldsymbol{x}}_{\text{true}}^{\text{SHP}}$ for three synthetic soils were pre-defined in synthetic numerical experiment (Section 4.1).

The implementation of any optimization algorithm is contingent upon forward modeling for the prediction of simulation time series. Upon executing the forward model $\mathscr{F}(\cdot)$, the simulation dataset $\boldsymbol{Y}_\tau = [\boldsymbol{y}_\tau^i]$, where $i \in \{1, 2, \cdots, N_s\}$ was generated. Subsequently, the objective function (Table 1) was utilized to assess the similarity between the simulation and observation. This estimation was then followed by an iterative procedure where the parameters were automatically updated/resampled from $\boldsymbol{X}_\tau$ to $\boldsymbol{X}_{\tau+1}$ through the application of optimization algorithms. The forward model is executed repeatedly until the simulated values closely approximate to the observed values. The final iteration ($\tau = N_\tau$) of the simulation represents the posterior parameters and results.

3.1. Forward modeling of unsaturated flow

In the context of a vertical coordinate system (positive in downward direction), the movement of unsaturated flow within soils can be mathematically represented by the one-dimensional Richards' equation (Richards, 1931):

$$C\frac{\partial h}{\partial t} = \frac{\partial}{\partial z}\left[K\left(\frac{\partial h}{\partial z} - 1\right) - S(z,t)\right] \tag{8}$$

where C (m^{-1}) is the differential water capacity ($= d\theta/dh$); h (m) is the soil water potential; z (m) is the vertical distance from soil surface; K (m s^{-1}) is the unsaturated hydraulic conductivity; $S(z, t)$ (m s^{-1}) is the sink term here accounting for root water uptake (see detail in supplemental material); and t (s) is the time.

3.2. Particle Swarm Optimization (PSO)

PSO algorithm is an adaptive optimization method inspired by social-psychological dynamics, where individuals (particles) interact within a social framework to collaboratively search for optimal solutions (Gholami et al., 2018). PSO conceptualizes the optimization process as particles navigating through the parameter space, with each particle adjusting its position based on both its own experience and the performance of the swarm as a whole.

In PSO, particles maintain two key records during the search process: their personal best position within the current generation ($\boldsymbol{x}_\tau^{p,\text{best}}$) and the global best position achieved by any particle across all generations ($\boldsymbol{x}_\tau^{g,\text{best}}$). These records guide the particles in updating their positions iteratively. Specifically, the position of a particle position $\boldsymbol{x}_\tau^i$ is updated based on its current velocity $\boldsymbol{v}_\tau^i$ and the influence of both personal and global best positions:

$$\boldsymbol{v}_{\tau+1}^i = w\boldsymbol{v}_\tau^i + c_p r_\tau^p\left(\boldsymbol{x}_\tau^{p,\text{best}} - \boldsymbol{x}_\tau^i\right) + c_g r_\tau^g\left(\boldsymbol{x}_\tau^{g,\text{best}} - \boldsymbol{x}_\tau^i\right) \tag{9}$$

$$\boldsymbol{x}_{\tau+1}^i = \boldsymbol{x}_\tau^i + \boldsymbol{v}_{\tau+1}^i \tag{10}$$

where $i \in \{1, 2, \cdots, N_s\}$ represent the index of a parameter sample, $\tau \in \{1, 2, \cdots, N_\tau\}$ represent the iteration step; r_τ^p and r_τ^g are randomly draw from uniform distribution $\mathscr{U}[0, 1]$. c_p and c_g are the constants known as the cognitive and social parameters, respectively. In this study, the value of c_p and c_g was set to 2. w is the inertia weigh parameter, specifying to be 0.6 in this study.

3.3. Differential Evolution Markov Chain Monte Carlo (DE-MCMC)

The basis of MCMC algorithm operates by constructing a Markov chain that performs a random walk through the parameter space. This chain sequentially visits solutions with frequencies determined by a stationary distribution, enabling the algorithm to approximate the target posterior distribution effectively. DE-MCMC algorithm builds upon MCMC by integrating differential evolution, a genetic algorithm-inspired approach, to evolve a population of solutions. DE-MCMC incorporates the Metropolis selection rule to decide whether proposed candidate points should replace their respective parent solutions. Typically, the initial population is drawn from the prior distribution, and DE-MCMC efficiently transforms these prior samples into posterior samples through iterative updates (Braak, 2006).

In DE-MCMC, observations are combined with prior knowledge of parameters to define the joint posterior probability distribution over a d-dimensional parameter space. The algorithm employs N independent Markov chains that run concurrently, proposing updates in parallel. The use of parallel chains and adaptive proposal generation makes DE-MCMC particularly effective for sampling from complex, high-dimensional parameter spaces, offering improved efficiency and convergence compared to traditional MCMC approaches. The multi-variate proposals $\boldsymbol{x}_p^i$ are dynamically generated using the differential evolution strategy, leveraging the diversity of the ensemble of chains:

$$\boldsymbol{x}_p^i = \gamma\left(\boldsymbol{x}_\tau^a - \boldsymbol{x}_\tau^b\right) + \zeta,\ a \neq b \neq i \tag{11}$$

where $\gamma = 2.38/\sqrt{2d}$ denotes jump rate, d is the dimension of $\boldsymbol{x}$ (ter Braak and Vrugt, 2008), a and b are integer values drawn without replacement from $\{1, \cdots, i-1, i+1, \cdots, N_s\}$, and $\zeta\ \mathscr{N}_d(0, c_*)$ is randomly disturbed from a normal distribution with small standard deviation, i.e., $c_* = 10^{-6}$(ter Braak and Vrugt, 2008). By accepting each proposal with Metropol probability:

$$p_{acc}(\boldsymbol{x}^i \rightarrow \boldsymbol{x}_p^i) = \begin{cases} min\left(\dfrac{p(\boldsymbol{x}^i)}{p\left(\boldsymbol{x}_p^i\right)}, 1\right), p\left(\boldsymbol{x}^i\right) > 0 \\ 1, p\left(\boldsymbol{x}^i\right) = 0 \end{cases} \tag{12}$$

the Markov chains are obtained, the stationary or limiting distribution of which is the posterior distribution. If $p_{acc}(\boldsymbol{x}^i \rightarrow \boldsymbol{x}_p^i)$ is larger than some uniform label drawn from U$(0, 1)$, then the candidate point is accepted and the i^{th} chain moves to the new position, that is $\boldsymbol{x}_{\tau+1}^i = \boldsymbol{x}_p^i$, otherwise $\boldsymbol{x}_{\tau+1}^i = \boldsymbol{x}_\tau^i$.

3.4. Sequential Monte Carlo (SMC)

SMC algorithm produces a population of weighted samples (or particles) from the targeted posterior distribution. The samplers are constructed through a sequence of sampling, disturbing, and resampling strategies (Del Moral et al., 2006). The SMC algorithm is initialized by sampling a population of N_s particles (i.e., model parameter vectors $\boldsymbol{x}_\tau^i$). Each particle is associated with a weight, W. At iterative step τ, assume

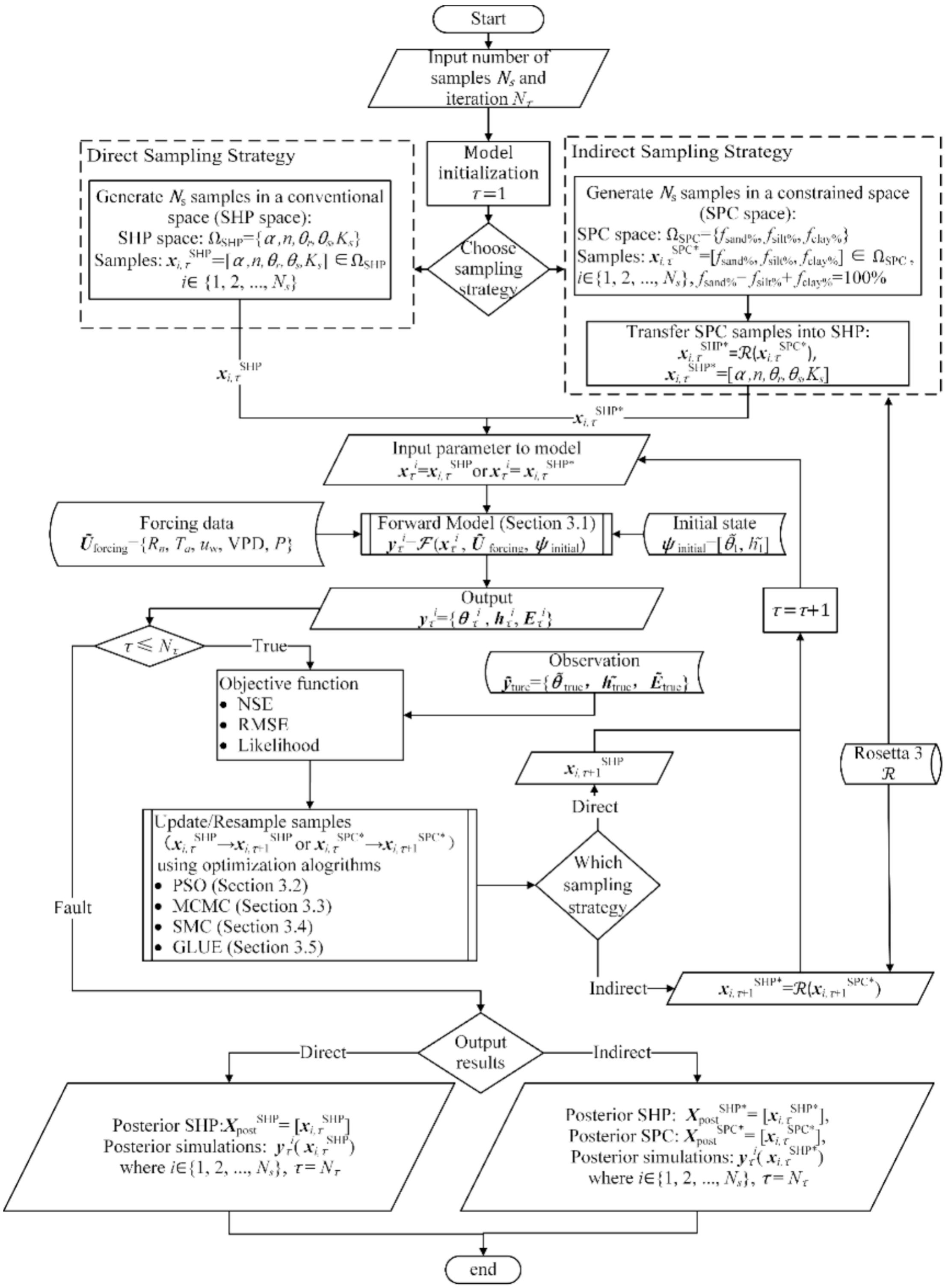

Fig. 2. The flow diagram illustrates the model optimization process, encompassing the following key components: (1) forward model $\mathscr{F}(\cdot)$, where the $\widetilde{\boldsymbol{U}}_{\text{forcing}}$ signifies the forcing variables, commonly referred to as input data to derive boundary conditions; $\widetilde{\boldsymbol{\psi}}_{\text{initial}}$ indicates initial states, while $\boldsymbol{x}_\tau^i$ symbolizes the model's parameters. (2) the model's output, also known as the simulation of state variable, $\boldsymbol{y}_\tau^i$, contains soil moisture θ_τ^i, soil water potential $\boldsymbol{h}_\tau^i$ and evaporation $\boldsymbol{E}_\tau^i$. (3) the observation state or the true value, $\widetilde{\boldsymbol{y}}_{\text{ture}}$, which incorporates the observation of soil moisture $\widetilde{\theta}_{\text{true}}$, soil water potential $\widetilde{\boldsymbol{h}}_{\text{true}}$, and evaporation $\widetilde{\boldsymbol{E}}_{\text{true}}$. (4) the objective function, which is computed by comparing θ_τ^i and $\widetilde{\theta}_{\text{true}}$, thus guiding the optimization process. (5) the optimization algorithm serves to optimize the model parameters.

Table 1
The definition of objective function and evaluation criteria.

Name	Definition	Range	Objective function	Evaluation criterion
NSE	$\mathrm{NSE}(\boldsymbol{y}_\tau^i) = 1 - \frac{\sum_{j=1}^{N_T}(y_j - \widetilde{y}_j)^2}{\sum_{j=1}^{m}(\widetilde{y}_j - \overline{y})^2}$	$(-\infty, 1]$	$\mathrm{NSE}(\theta_\tau^i)$ in GLUE	$\mathrm{NSE}(\theta_{N_\tau}^i)$
RMSE	$\mathrm{RMSE}(\boldsymbol{y}_\tau^i) = \sqrt{\frac{1}{N_T}\sum_{j=1}^{N_T}\left(y_j - \widetilde{y}_j\right)^2}$	$[0, +\infty)$	$\mathrm{RMSE}(\theta_\tau^i)$ in PSO and MCMC	$\mathrm{RMSE}(\theta_{N_\tau}^i)$
Likelihood $p(\boldsymbol{y}_\tau^i \vert \boldsymbol{x}_\tau^i)$	$p(\boldsymbol{y}_\tau^i \vert \boldsymbol{x}_\tau^i) = \prod_{j=1}^{m}\frac{1}{(2\pi)^{1/2}\det(\mathbf{R})^{1/2}}e^{\left[-0.5\left(y_j-\tilde{y}_j\right)^T\mathbf{R}^{-1}\left(y_j-\tilde{y}_j\right)\right]}$	$[0, 1]$	$p(\theta_\tau^i \vert \boldsymbol{x}_\tau^i)$ in SMC	–
Bias	$\mathrm{Bias}\ (\boldsymbol{y}_\tau^i) = \sum_{j=1}^{N_T}(y_j - \widetilde{y}_j)/N_T$	$(-\infty, +\infty)$	–	$\mathrm{Bias}(\theta_{N_\tau}^i)$, $\mathrm{Bias}(\boldsymbol{h}_{N_\tau}^i)$ and $\mathrm{Bias}(\boldsymbol{E}_{N_\tau}^i)$

Notes: $\mathbf{R}$ is the observed error covariance, $\det(\mathbf{R})$ is the determinate of $\mathbf{R}$, the $\overline{y}$ is the average value of the observation data. NSE denotes Nash-Sutcliffe Efficiency. RMSE signifies root mean square error.

that a set of weighted particles $\{W_\tau^i, \boldsymbol{x}_\tau^i\}, i \in \{1, 2, \cdots, N_s\}$ approximating probability distribution is available,

$$p_\tau^N(\mathrm{d}\boldsymbol{x}_{1:\tau}) = \sum_{i=1}^{N} W_\tau^i \delta_{\boldsymbol{x}_{1:\tau}^i}(\mathrm{d}\boldsymbol{x}_{1:\tau}) \tag{13}$$

$$W_\tau^i = \frac{w_\tau^i}{\sum_{i=1}^{N} w_\tau^i} \tag{14}$$

where the $p_\tau^N(\mathrm{d}\boldsymbol{x}_{1:\tau})$ is the approximating probability density $p_\tau^N(\boldsymbol{x}_{1:\tau})$ with respect to a finite dominating measure denoted $\mathrm{d}\boldsymbol{x}_{1:\tau}$, the W_τ^i is the normalized weight coefficient, the w_τ^i is the importance weight, and δ is Dirac delta function (Del Moral et al., 2006). According to the sequential importance sampling method, the w_τ^i can be updated as:

$$w_{\tau+1}^i \propto w_\tau^i p(\boldsymbol{y}_\tau^i \vert \boldsymbol{x}_\tau^i) \tag{15}$$

where w_τ^i is the unnormalized particle weight coefficient; $p(\boldsymbol{y}_\tau^i \vert \boldsymbol{x}_\tau^i)$ is Likelihood function (Table 1).

3.5. General Likelihood Uncertainty Estimation (GLUE)

Beven and Binley (1992) proposed a Generalized Likelihood Uncertainty Estimation (GLUE) approach based on the Bayesian framework, demonstrating its applicability in identifying the equifinality phenomenon in hydrological modeling. GLUE is based on a Monte Carlo analysis using a Latin Hypercube sampling method. The central idea behind GLUE entails the rejection of a model or a model parameter set as a possible simulator of the hydrological process if it fails to replicate the observed state within an acceptable threshold of measurement errors (Shao et al., 2023b). In this study, the Nash-Sutcliffe Efficiency (NSE) was used as an approximation of the likelihood to evaluate the model performance for all sampled parameter sets. The posterior parameter sets obtained from GLUE algorithm are selected based on an acceptable threshold of 0.85 (e.g., NSE $>$ 0.85). The objective function and evaluation criteria of other algorithms are shown in the Table 1.

4. Design of the numerical experiment

4.1. Predefined soil particle composition (SPC) and soil hydraulic parameters (SHP)

In this study, synthetic numerical experiments are conducted to simulate predefined time series of $\widetilde{\theta}_{\mathrm{true}}$, $\widetilde{\boldsymbol{h}}_{\mathrm{true}}$, and $\widetilde{\boldsymbol{E}}_{\mathrm{true}}$ using identical forward model and meteorological forcing data of parameter inverse process. In situ measured meteorological forcing variables and vegetation information are integrated with predefined true SHP (i.e., $\widetilde{\boldsymbol{x}}_{\mathrm{true}}^{\mathrm{SHP}} = \{\widetilde{\alpha}, \widetilde{n}, \widetilde{\theta}_r, \widetilde{\theta}_s, \widetilde{K}_s\}$). In this way, the numerical experiments conducted facilitate the exclusion of uncertainties stemming from model conceptualization and observation errors in input rainfall and soil moisture output, thereby facilitating a focused assessment of the impact of parameterization.

The numerical experiments encompass three representative soil textures, namely sandy, loam, and silty soils, as classified by the USDA (Fig. 3). The $\widetilde{x}_{\mathrm{true}}^{\mathrm{SHP}}$ is derived from SPC (i.e., $\widetilde{\boldsymbol{x}}_{\mathrm{true}}^{\mathrm{SPC}} = \left\{\widetilde{f}_{\mathrm{sand\%}}, \widetilde{f}_{\mathrm{silt\%}}, \widetilde{f}_{\mathrm{clay\%}}\right\}$, where $\widetilde{f}_{\mathrm{clay\%}} + \widetilde{f}_{\mathrm{silt\%}} + \widetilde{f}_{\mathrm{sand\%}} = 100\%$) using Rosetta 3 pedotransfer function. The value of $\widetilde{x}_{\mathrm{true}}^{\mathrm{SHP}}$ and $\widetilde{\boldsymbol{x}}_{\mathrm{true}}^{\mathrm{SPC}}$ have shown in Table 2.

4.2. Meteorological forcing

The meteorological forcing data $\widetilde{\boldsymbol{U}}_{\mathrm{Forcing}}$ are taken from a national ecosystem monitoring station situated within a tropical monsoon forest in Xishuangbanna, Yunnan, China. This station is characterized by a year-round tropical monsoon climate, exhibiting distinct wet and dry seasons. The study period spans from March 2004 to March 2005. Soil moisture exhibits a comprehensive wetting and drying cycle throughout designated period, indicative of a typical ecohydrological regime. The monitoring of net radiation R_n, air temperature T_a, wind speed u_w, vapor pressure deficit VPD and precipitation P are measured for 42 m, as illustrated in Fig. 4.

The daily mean meteorological forcing variable $\widetilde{\boldsymbol{U}}_{\mathrm{Forcing}} = \{R_n, T_a, u_w, \mathrm{VPD}, P\}$ demonstrate distinct seasonal pattens. The daily mean R_n ranged from 80.16 to 482.5 MJ m^{-2} d^{-1} with a small value during October to January in next year. Due to the smaller R_n, the air temperature decreased from 20.08 °C in October to 12.82 °C in January. The T_a in other months fluctuated stably with an average value of 22.13 °C. The u_w fluctuated with a high average value of 0.3814 m s^{-1} during March and April, and with a low average value of 0.2198 m s^{-1} at other periods. The VPD demonstrated a similar patten of u_w, with a high average value of 0.9182 kPa during March and April, and a low average value of 0.3553 kPa at other periods. The total annual of P was 1409.1 mm. Most of P was concentrated between May and October with cumulative total of 1177 mm.

4.3. Numerical implementation

The Darcy-Richards equation (Eq.8) is solved using the finite difference method and the Picard iteration format within Python 3.8 programming environment (Shao et al., 2023b; Shao et al., 2018). The boundary condition is capable of transitioning between a flux boundary (net rainfall) and a pressure head boundary (ponding depth) following the method proposed by van Dam and Feddes (2000). The lower boundary is specified as gravitational drainage. The thickness of soil is established at the upper 2 m. The domain is partitioned into 32 regions by non-uniform grid. The spatial step size escalates in accordance with

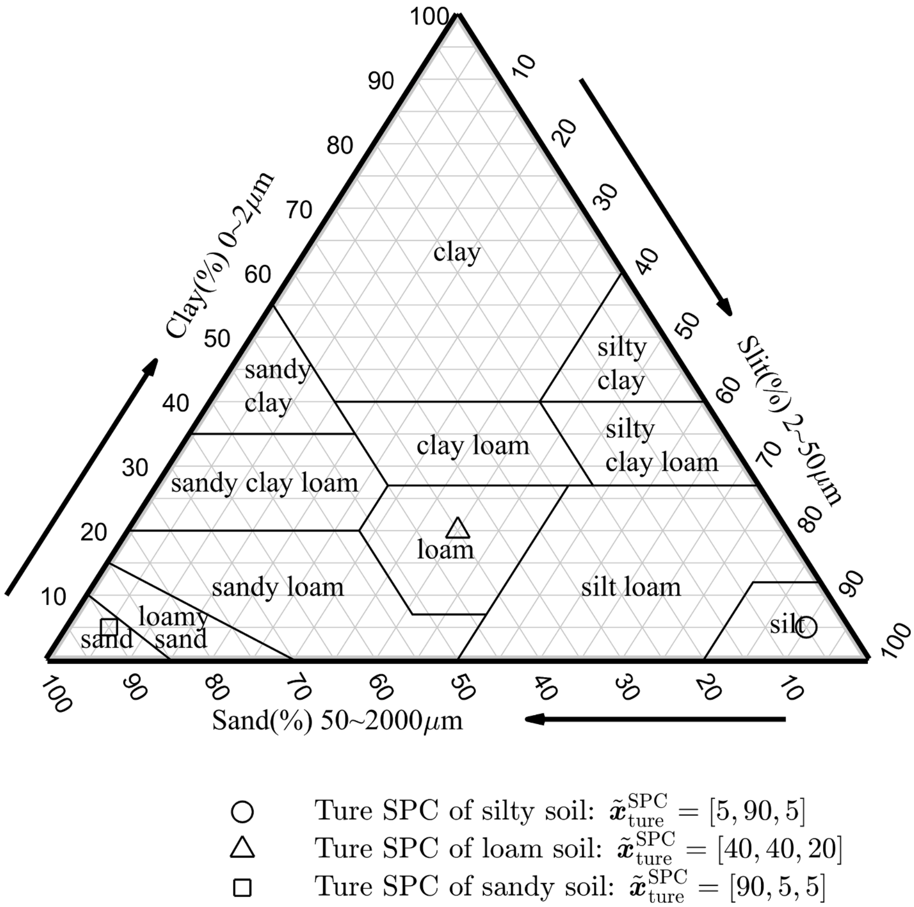

Fig. 3. The position of pre-defined true soil particle composition (SPC) $\tilde{\boldsymbol{x}}_{\text{true}}^{\text{SPC}}$ in soil texture triangle, where the black hollow "○", "△", "□" are the pre-defined true SPC of silty soil, loam soil and sandy soil, respectively.

Table 2
Pre-defined true soil particle composition ($\tilde{\boldsymbol{x}}_{\text{true}}^{\text{SPC}}$) and soil hydraulic parameters (..$\tilde{\boldsymbol{x}}_{\text{true}}^{\text{SHP}}$)

Soil texture	Soil particle composition (SPC) $\tilde{\boldsymbol{x}}_{\text{true}}^{\text{SPC}} = \{\tilde{f}_{\text{sand\%}}, \tilde{f}_{\text{silt\%}}, \tilde{f}_{\text{clay\%}}\}$			Soil hydraulic parameters (SHP) $\tilde{\boldsymbol{x}}_{\text{true}}^{\text{SHP}} = \{\tilde{\alpha}, \tilde{n}, \tilde{\theta}_r, \tilde{\theta}_s, \tilde{K}_s\}$				
	$\tilde{f}_{\text{sand\%}}$(%)	$\tilde{f}_{\text{silt\%}}$(%)	$\tilde{f}_{\text{clay\%}}$(%)	$\tilde{\alpha}$(m^{-1})	$\tilde{n}$(−)	$\tilde{\theta}_r$(m^3 m^{-3})	$\tilde{\theta}_s$(m^3 m^{-3})	$\tilde{K}_s$(m h^{-1})
Silty	5	90	5	0.4066	1.6436	0.0661	0.4585	0.0176
Loam	40	40	20	0.6674	1.4187	0.0887	0.4040	0.0051
Sandy	90	5	5	2.9057	2.3023	0.0550	0.3674	0.1332

Notes: $\tilde{f}_{\text{clay\%}} + \tilde{f}_{\text{silt\%}} + \tilde{f}_{\text{sand\%}} = 100\%$

the augmentation of soil depth, ranging from 0.01 m to 0.2 m. During the simulation, the tolerable error of θ is set as 0.00001 m^3 m^{-3}, and the time step is dynamically adjusted within the range of 0.234 to 60 min, ensuring both numerical precision and computational efficiency.

The water transport within the soil-vegetation system is governed by a dynamic coupling mechanism, as detailed in the supplementary material: (1) The presence of a vegetation canopy serves to intercept rainfall, thereby allowing only a fraction of net rainfall to infiltrate into the soil surface; (2) The actual transpiration occurring at leaf surface of the canopy is quantified using a single-layer Penman-Monteith equation under big-leaf conceptualization, with the rate being controlled by canopy conductance, subject to constraints of multiple environmental stresses, including solar radiation, VPD, T_a and h; (3) The E is partitioned into transpiration and evaporation from canopy interception; (4) The root uptake induced by transpiration exhibits an exponential distribution along the soil depth, while also incorporating a compensation mechanism.

In this study, the model describes water transport within a single-layer isotropic soil medium. The inverse model is specifically designed to encompass five SHP (α, n, θ_r, θ_s and K_s) of VG model. More comprehensive specifications pertaining to the model (e.g., dual-permeability model) and parameterization (e.g., hysteresis in the SWCC, anisotropic, layered soil structures) have been deliberately excluded. This exclusion is grounded in the rationale that the incorporation of these supplementary parameters could potentially exacerbate the dimensionality of parameter optimization problem.

4.4. Modeling strategy

This study employed four algorithms, namely PSO, MCMC, SMC and GLUE, to optimize SHP for the simulation of $\tilde{\theta}$ under two sampling strategies: Direct sampling strategy and Indirect sampling strategy.

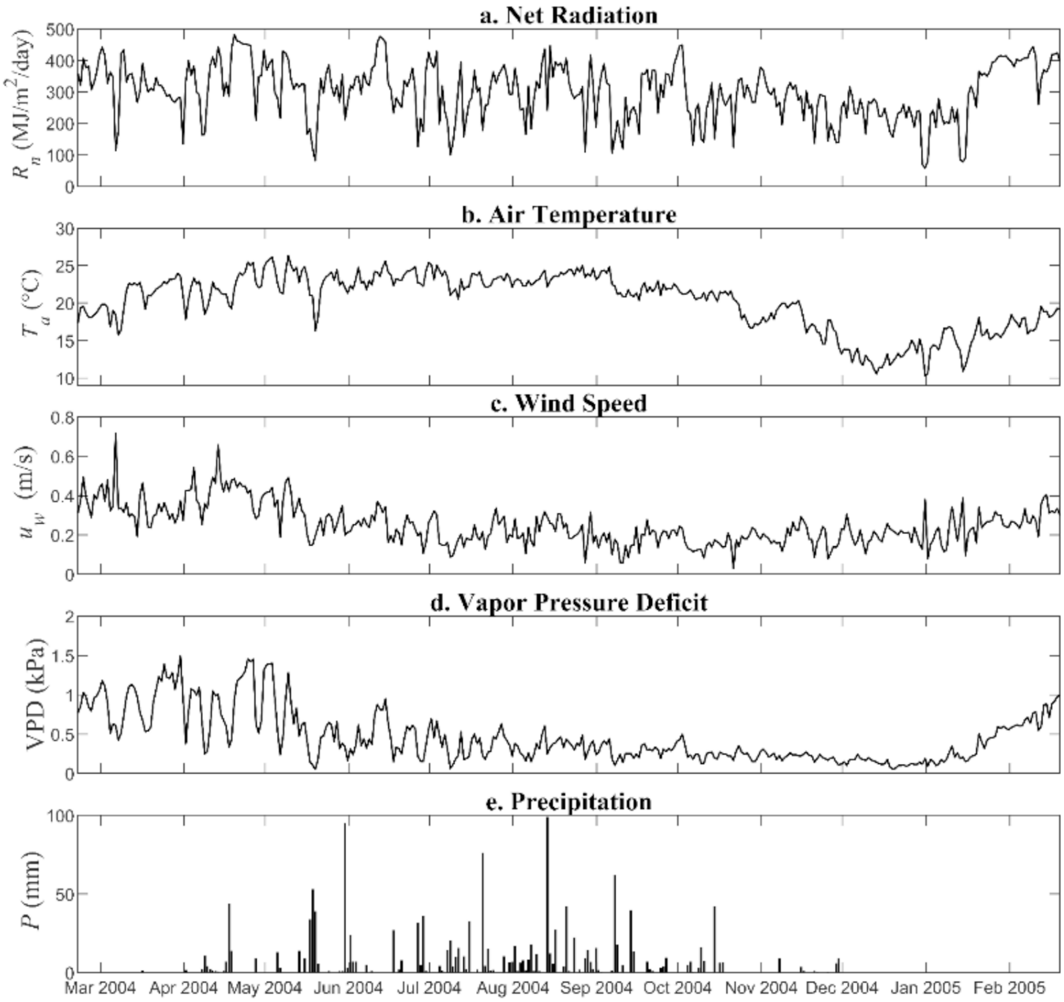

Fig. 4. The daily mean meteorological forcing variable: (a) net radiation R_n, (b) air temperature T_a, (c) wind speed u_w, (d) vapor pressure deficit, VPD, (e) precipitation..P

Within the Indirect sampling strategy, the algorithms initiate and subsequently update parameters within SPC sampling space Ω_{SPC} (refer to Section 2 for comprehensive details). Thereafter, the SPC samples $x_{i,\tau}^{\mathrm{SPC}*}$ are transformed into SHP samples $x_{i,\tau}^{\mathrm{SHP}*}$ using the Rosetta 3 pedotransfer function (Eq. 5). In the Direct sampling strategy, the algorithms directly initiate and update parameters within SHP sampling space Ω_{SHP}, where the samples are denotes as $x_{i,\tau}^{\mathrm{SHP}}$. The delineation of two sampling space (Ω_{SHP} and Ω_{SPC}) is provided in Table 3. The combinations of the four optimization algorithms with the two parameter sampling spaces yields a total of eight inverse modeling strategies, which symbolized as PSO-Indirect, PSO-Direct, MCMC-Indirect, MCMC-Direct, SMC-Indirect, SMC-Direct, GLUE-Indirect and GLUE-Direct.

The number of samples for the PSO and SMC algorithms is set as 50, with maximum iteration steps are 200. The MCMC algorithm employs a Markov chain to sample the stationary distribution, allowing the chain to thoroughly explore the model parameter space via continuous updates of sample data. To ensure a thorough exploration of the parameter space, MCMC algorithm requires a larger number of iteration steps to achieve convergence to a relative stationary distribution. Consequently, the number of samples and maximum iteration steps of MCMC are set as 20 and 500, respectively. The number of samples under GLUE algorithm is set as 10,000.

The critical variables and terminologies used in synthetic numerical experiment are delineated in Table 4. The synthetic numerical experiment was conducted with Python 3.8 and the resulting figure was generated via MATLAB 2022a. The results of inverse modeling were organized as follows: Section 5.1 provides a comparative analysis of posterior probability distribution function (PDF) of SHP and SPC, which were optimized by four algorithms (PSO, MCMC, SMC and GLUE) under

Table 3
The range of SHP space Ω_{SHP} and SPC space..Ω_{SPC}

	Soil particle composition space, SPC space: $\Omega_{\mathrm{SPC}}=\{f_{\mathrm{sand\%}}, f_{\mathrm{silt\%}}, f_{\mathrm{clay\%}}\}$			Soil hydraulic parameters space, SHP space: $\Omega_{\mathrm{SHP}}=\{\alpha, n, \theta_r, \theta_s, K_s\}$				
	$f_{\mathrm{sand\%}}$(%)	$f_{\mathrm{silt\%}}$(%)	$f_{\mathrm{clay\%}}$(%)	α(m^{-1})	n(−)	θ_r(m^3 m^{-3})	θ_s(m^3 m^{-3})	K_s(m h^{-1})
Lower bound	0	0	0	1	1.1	0.01	0.32	0
Upper bound	100	100	100	10	3.0	0.09	0.48	1

Notes:.$f_{\mathrm{clay\%}} + f_{\mathrm{silt\%}} + f_{\mathrm{sand\%}} = 100\%$

Table 4
Explication of critical variables and terminologies in synthetic numerical experiments.

Variable/Term	Explication
SWCC	Soil water characteristic curve, h vs θ relationship
UHCF	Unsaturated hydraulic conductivity functions, K_s vs θ relationship
SHP	Soil hydraulic parameters, it represents VG parameters in this study
VG	Mualem-van Genuchten model, including five parameters ($\alpha, n, \theta_r, \theta_s$ and K_s)
SPC	Soil particle composition (the fraction of sand $f_{\mathrm{sand\%}}$, silt $f_{\mathrm{silt\%}}$, and clay $f_{\mathrm{clay\%}}$ particles, where $f_{\mathrm{clay\%}} + f_{\mathrm{silt\%}} + f_{\mathrm{sand\%}} = 100\%$)
PSO	Particle Swarm Optimization (Section 3.2)
MCMC	Markov Chain Monte Carlo (Section 3.3)
SMC	Sequential Monte Carlo (Section 3.4)
GLUE	General Likelihood Uncertainty Estimation (Section 3.5)
Ω_{SHP}	Conventional sampling space or the sampling space of Direct sampling strategy (SHP space): $\Omega_{\mathrm{SHP}} = \{\alpha, n, \theta_r, \theta_s, K_s\}$
Ω_{SPC}	Constrained sampling space or the sampling space of Indirect sampling strategy (SPC space):$\Omega_{\mathrm{SPC}} = \{f_{\mathrm{sand\%}}, f_{\mathrm{silt\%}}, f_{\mathrm{clay\%}}\}$
$\widetilde{\boldsymbol{x}}^{\mathrm{SHP}}_{\mathrm{true}}$	The predefined true SHP,$\widetilde{\boldsymbol{x}}^{\mathrm{SHP}}_{\mathrm{true}} = [\widetilde{\alpha}, \widetilde{n}, \widetilde{\theta}_s, \widetilde{\theta}_r, \widetilde{K}_s]$
$\widetilde{\boldsymbol{x}}^{\mathrm{SPC}}_{\mathrm{true}}$	The predefined true SPC,$\widetilde{\boldsymbol{x}}^{\mathrm{SPC}}_{\mathrm{true}} = \{\widetilde{f}_{\mathrm{sand\%}}, \widetilde{f}_{\mathrm{silt\%}}, \widetilde{f}_{\mathrm{clay\%}}\}$
$\boldsymbol{x}^{\mathrm{SHP}}_{i,\tau}, \boldsymbol{X}^{\mathrm{SHP}}_{\tau}$	The i^{th} SHP samples and the SHP set sampled in Ω_{SHP} at τ^{th} iteration (Direct sampling strategy). If $\tau = N_\tau$ (i.e., final iteration), the SHP set represents the posterior SHP of Direct sampling strategy (i.e., $\boldsymbol{X}^{\mathrm{SHP}}_{\mathrm{post}}$).
$\boldsymbol{x}^{\mathrm{SPC*}}_{i,\tau}, \boldsymbol{X}^{\mathrm{SPC*}}_{\tau}$	The i^{th} SPC samples and the SPC set sampled in Ω_{SPC} at τ^{th} iteration (Indirect sampling strategy). If $\tau = N_\tau$ (i.e., final iteration), the SPC set represents the posterior SPC of Indirect sampling strategy (i.e., $\boldsymbol{X}^{\mathrm{SPC*}}_{\mathrm{post}}$).
$\boldsymbol{x}^{\mathrm{SHP*}}_{i,\tau}, \boldsymbol{X}^{\mathrm{SHP*}}_{\tau}$	The i^{th} SHP samples and the SHP set sampled in Ω_{SPC} at τ^{th} iteration (Indirect sampling strategy). If $\tau = N_\tau$ (i.e., final iteration), the SHP set represents the posterior SHP of Indirect sampling strategy (i.e., $\boldsymbol{X}^{\mathrm{SHP*}}_{\mathrm{post}}$).
$\widetilde{\theta}_{\mathrm{true}}, \widetilde{\boldsymbol{h}}_{\mathrm{true}}, \widetilde{\boldsymbol{E}}_{\mathrm{true}}$	The predefined true value of soil moisture, soil water potential and evaporation
$\theta, \boldsymbol{h}, \boldsymbol{E}$	The simulation of soil moisture, soil water potential and evaporation

the Direct and Indirect sampling strategies, respectively; In Section 5.2, the uncertainty bands of the posterior SWCC were compared, contingent upon the posterior SHP. In Section 5.3, the uncertainty bands of posterior soil moisture θ, soil water potential $\boldsymbol{h}$, and evaporation $\boldsymbol{E}$ were compared under Direct and Indirect sampling strategies. The Bias of posterior results was show in Section 5.4. Given that infrequent utilization of the GLUE algorithm in parameter optimization, Sections 5.3 and 5.4 only adopt PSO, MCMC, and SMC algorithms, which is sufficient to elucidate the improvement of the Indirect sampling strategy over Direct sampling strategy.

5. Result

5.1. The posterior distribution of samples

The uncertainty of SHP (soil hydraulic parameter: α, n, θ_r, θ_s, K_s), which optimized utilizing PSO, MCMC, SMC and GLUE algorithms, were compared under Direct and Indirect sampling strategy, respectively. Under the Indirect sampling strategy, SHP was transferred form SPC (soil particle composition: $f_{\mathrm{sand\%}}, f_{\mathrm{silt\%}}, f_{\mathrm{clay\%}}$), whereas under the Direct sampling strategy, the SHP was sampled directly. To assess the uncertainty and accuracy of optimal parameter sets, the posterior SPC under Indirect sampling strategies were compared with predefined true SPC ($\widetilde{\boldsymbol{x}}^{\mathrm{SPC}}_{\mathrm{true}}$) in a soil textural triangle, as depicted in Fig. 5. The posterior probability distribution function (PDF) of SHP sampled under Direct and Indirect sampling strategies were compared with predefined true SHP ($\widetilde{\boldsymbol{x}}^{\mathrm{SHP}}_{\mathrm{true}}$) in Fig. 6.

The posterior SPC under Indirect sampling strategies, as scatter dots in Fig. 5 shown, indicated varying degree of uncertainties across three different soil types. The GLUE algorithms using large samples covered almost all the possible SHP and SPC samples. Therefore, it can comprehensively manifest the uncertainty of sampling SHP. The posterior SPC and SHP associated with GLUE have been selected based on an acceptable threshold (NSE > 0.85), as illustrated in Fig. 5 d and Fig. 6 iv. The more comprehensive distribution of SHP and SPC samples, as associated with the GLUE algorithm, has been illustrated as scatter plots of NSE versus SHP and NSE versus SPC in supplementary material (Fig. S2). In Fig. 5 d, the posterior SPC optimized by GLUE-Indirect across all soil types was presented in a strip-shaped form, encompassing $\widetilde{\boldsymbol{x}}^{\mathrm{SPC}}_{\mathrm{true}}$. For sandy soil, the uncertainty of posterior SPC was minimal, with a strip concentrated in range of sandy soil. For the loam and silty soil, a strip of posterior SPC optimized by GLUE-Indirect spanned various soil types rather than concentrating on the predefined soil type.

Similar to the result of GLUE-Indirect, the uncertainties in the SPC of sandy soil, as optimized by PSO-Indirect, MCMC-Indirect, and SMC-Indirect, were minimal; conversely, larger uncertainties were found in the SPC for silty soil. Excluding SMC-Indirect, posterior SPC accurately encompassed $\widetilde{\boldsymbol{x}}^{\mathrm{SPC}}_{\mathrm{true}}$ (Fig. 5 a ~ c). SMC-Indirect exhibited the most significant deviation from $\widetilde{\boldsymbol{x}}^{\mathrm{SPC}}_{\mathrm{true}}$ for loam soil and silty soil. Nevertheless, all three algorithms yielded closer and more convergent approximations to $\widetilde{\boldsymbol{x}}^{\mathrm{SPC}}_{\mathrm{true}}$ in comparison to GLUE (Fig. 5 c). This discrepancy arises from GLUE's inclination to encompass all potential parameters, whereas other optimization algorithms prioritize the exploration of parameter spaces with higher probability densities.

As the PDF of posterior SHP given in Fig. 6, the parameter uncertainties under Indirect sampling strategies were significantly reduced when compared to those under Direct sampling strategies, especially for the MCMC and GLUE algorithm. Consistent with the finding illustrated in Fig. 5 a, b, the posterior SHP optimized by PSO-Indirect and MCMC-Indirect precisely matched to $\widetilde{\boldsymbol{x}}^{\mathrm{SHP}}_{\mathrm{true}}$. Although posterior SPC optimized by the SMC-Indirect and GLUE-Indirect did not converge to $\widetilde{\boldsymbol{x}}^{\mathrm{SPC}}_{\mathrm{true}}$ as shown in Fig. 5 c, the posterior SHP optimized by SMC-Indirect and GLUE-Indirect closely approximated the $\widetilde{\boldsymbol{x}}^{\mathrm{SHP}}_{\mathrm{true}}$ with substantial agreement (Fig. 6). Conversely, when integrating Direct sampling strategy with parameter optimizations, all posterior SHP failed to match the truth $\widetilde{\boldsymbol{x}}^{\mathrm{SHP}}_{\mathrm{true}}$. Overall, the Direct strategy resulted in an overestimation of the parameter values for α and K_s, and an underestimation of the parameter value for n, which exhibited a compensation effect.

5.2. Posterior SWCC

The posterior SWCC (Fig. 7) described with SHP under both Direct and Indirect sampling strategies was shown as uncertainty bands to express the h vs. θ relations ranging from residual to saturated water content. Consistence with the smaller errors and uncertainty bands of inversely estimated SHP in Fig. 6, the Indirect sampling strategy markedly reduced the uncertainty of SWCC. This was evident in two key aspects: the alignment with predefined true values, and the magnitude of the uncertainty bands of posterior SWCC.

Firstly, all posterior SWCC bands under Indirect sampling strategies accurately align with the predefined true SWCC. Conversely, posterior SWCC bands under Direct sampling strategies only partially overlap the predefined true SWCC. Especially for the silt soil, the posterior SWCC under Direct sampling strategy exhibited substantial deviations from the predefined SWCC. For sandy and loam soil, the posterior SWCC under Direct sampling strategies closely approximated the predefined SWCC within the soil moisture range of 0.1 m^3 m^{-3} to 0.3 m^3 m^{-3}. However, when soil moisture approaching θ_s and θ_r, the posterior SWCC exhibited significant discrepancies from predefined true SWCC of sandy and loam soil under Direct sampling strategies.

The magnitude of the uncertainty bands of posterior SWCC under Indirect sampling strategies is notably narrower than those under Direct sampling strategies. Under the Indirect sampling strategy, nearly all posterior SWCC converged into a single curve, closely approximating the predefined true SWCC. Conversely, under Direct sampling strategy,

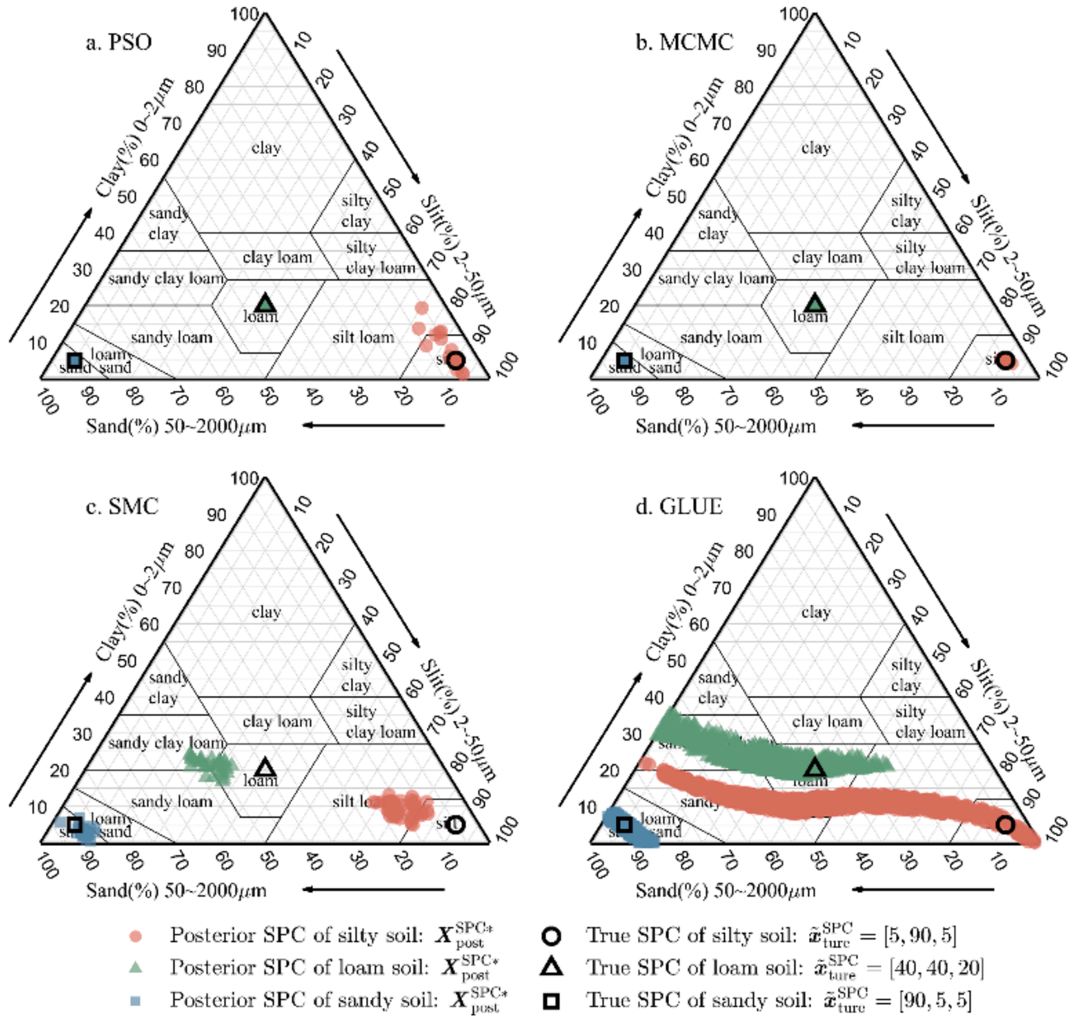

Fig. 5. The comparison between posterior SPC under Indirect sampling strategy X_{post}^{SPC*} and the predefined true SPC $\tilde{x}_{true}^{SPC}$: (a) posterior SPC optimized by PSO-Indirect; (b) posterior SPC optimized by MCMC-Indirect; (c) posterior SPC optimized by SMC-Indirect; (d) posterior SPC optimized by GLUE-Indirect. The red solid circle "●", green solid triangle "▲" and blue solid square "■" are the posterior SPC of silty soil, loam soil and sandy soil, respectively. The black hollow circle "○", triangle "△", square "□" are the predefined true SPC of silty soil, loam soil and sandy soil, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

posterior SWCC exhibited a broader uncertainty band, especially for the GLUE and MCMC algorithms.

5.3. Posterior simulated state variables

By using the optimization algorithms including PSO, MCMC and SMC, the posterior SHP of three soil types (silty, loam and sandy soil) were optimized under Direct and Indirect strategy, respectively (Fig. 6). The results of posterior soil moisture θ, soil water potential $\boldsymbol{h}$ and evaporation $\boldsymbol{E}$, simulated via the forward model (Fig. 8 ~Fig. 10), were compared against the predefined true value of $\widetilde{\theta}_{\text{true}}$, $\widetilde{\boldsymbol{h}}_{\text{true}}$ and $\widetilde{\boldsymbol{E}}_{\text{true}}$.

Consistent with the finding of posterior parameters depicted in Fig. 6, the uncertainty bands of θ under Indirect sampling strategy exhibited a significantly narrower range compared to those under Direct sampling strategy, particularly evident in the MCMC algorithm (Fig. 8). Especially, the ensemble means of θ under Indirect sampling strategy accurately simulated $\widetilde{\theta}_{\text{true}}$ (Fig. 8 a ~ c). However, under Direct sampling strategy, nearly all ensemble means of θ failed to accurately recover the $\widetilde{\theta}_{\text{true}}$ Fig. 8 d ~ f). The PSO-Direct and SMC-Direct algorithms exhibited superior accuracy in simulating θ than MCMC-Direct. The MCMC-Direct overestimated θ for sandy soil (Fig. 8 f) and underestimated θ for silty and loam soils (Fig. 8 d and e).

The uncertainty bands of $\boldsymbol{h}$ under Indirect sampling strategy is narrower than those under Direct sampling strategy. Under Indirect sampling strategy, the ensemble means of $\boldsymbol{h}$ accurately simulated $\widetilde{\boldsymbol{h}}_{\text{true}}$ (as shown in Fig. 9 a ~ c), which is consistent with the result shown in Fig. 7. Under Direct sampling strategy, posterior $\boldsymbol{h}$ significantly overestimated $\widetilde{\boldsymbol{h}}_{\text{true}}$ (Fig. 9 d ~ f), despite the ensemble means of θ under Direct sampling strategy being close to the $\widetilde{\theta}_{\text{true}}$ (Fig. 8 d ~ f). The larger uncertainty of h under the Direct sampling strategy may be attributed by the SWCC sampled failing to accurately recover the predefined true SWCC in Fig. 7.

The posterior evaporation $\boldsymbol{E}$ of silty, loam, and sandy soils is depicted in Fig. 10, exhibiting clear diurnal variations. The predefined true value of $\widetilde{\boldsymbol{E}}_{\text{true}}$ increased from 0 mm h^{-1} at 8 o'clock, peaked at 15 o'clock, and decreased to 0 mm h^{-1} by 20 o'clock. The average peak $\widetilde{\boldsymbol{E}}_{\text{true}}$ for sandy soil (0.23 mm h^{-1}) surpassed those of loam soil (0.20 mm h^{-1}) and silty soil (0.22 mm h^{-1}). The discrepancy in $\widetilde{\boldsymbol{E}}_{\text{true}}$ across different soil texture types under identical meteorological conditions primarily arises from variations in soil hydrological processes regulated by SHP. Under condition of sufficient water, $\widetilde{\boldsymbol{E}}_{\text{true}}$ is solely controlled by energy limitation, whereas as $\widetilde{\theta}_{\text{true}}$ decreases, the water limitation on $\widetilde{\boldsymbol{E}}_{\text{true}}$ intensifies progressively. In this unsaturated flow model, the water limitation of $\widetilde{\boldsymbol{E}}_{\text{true}}$

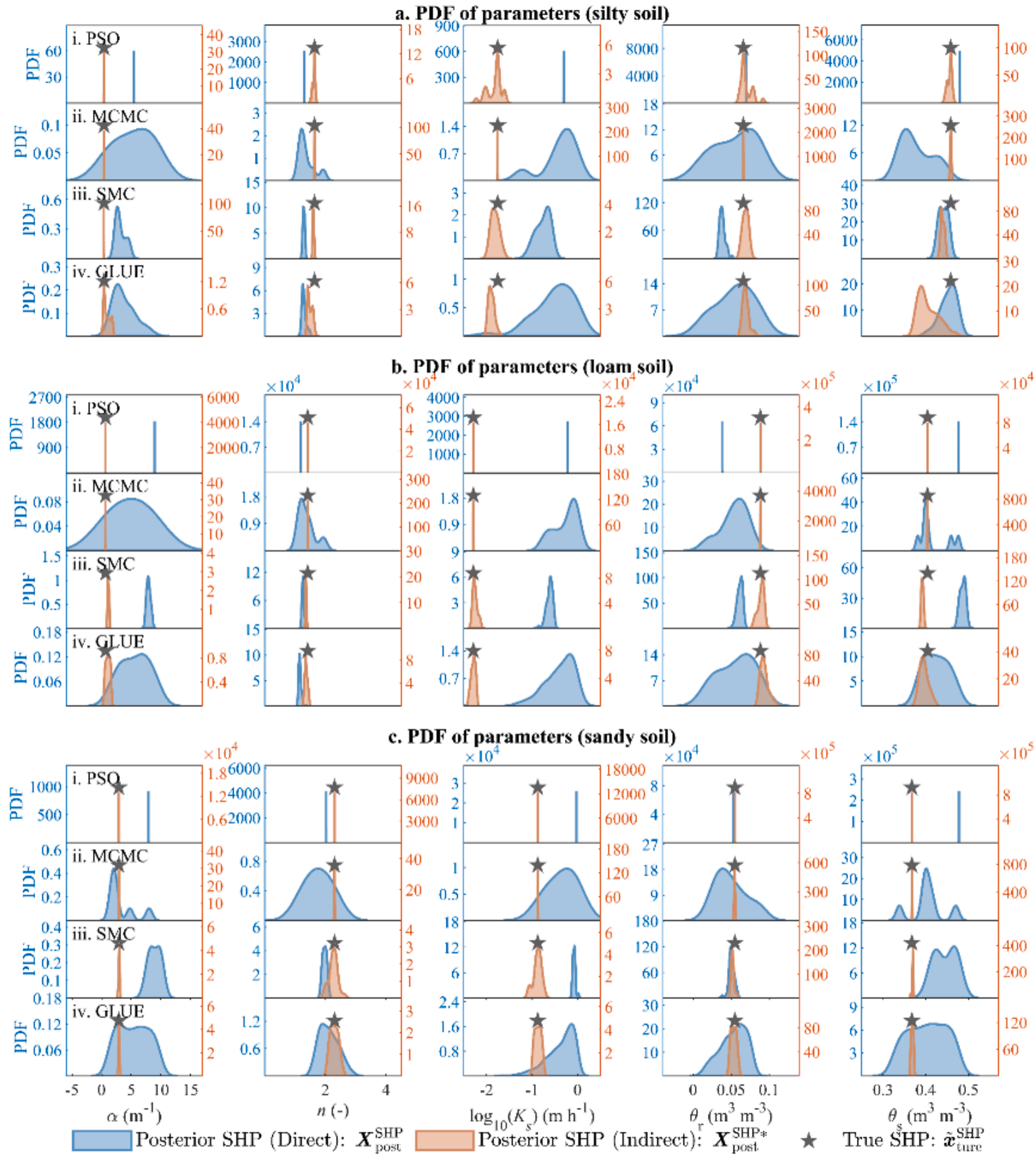

Fig. 6. The posterior probability distribution function (PDF) of optimized soil hydraulic parameters (SHP) under Direct sampling strategy $\boldsymbol{X}_{N_r}^{\mathrm{SHP}}$ and Indirect sampling strategy $\boldsymbol{X}_{N_r}^{\mathrm{SHP*}}$. $\widetilde{\boldsymbol{x}}_{\mathrm{true}}^{\mathrm{SHP}}$ is predefined true soil hydraulic parameters, which is shown as black solid star "★".

was positively correlative of $\widetilde{\boldsymbol{h}}_{\mathrm{true}}$, as detailed in the supplementary material. A steeper SWCC indicates a more pronounced decrease in $\widetilde{\boldsymbol{h}}_{\mathrm{true}}$ with an equivalent volume of water loss. And a more pronounced decrease of $\widetilde{\boldsymbol{h}}_{\mathrm{true}}$ result in a more significant limitation of $\widetilde{\boldsymbol{E}}_{\mathrm{true}}$. Consequently, sandy soil exhibits the flattest SWCC, leading to least decrease of the $\widetilde{\boldsymbol{h}}_{\mathrm{true}}$, thereby achieving the highest $\widetilde{\boldsymbol{E}}_{\mathrm{true}}$ among the soil types. Conversely, loam soil exhibits the steepest variation in SWCC, resulting in the lowest $\widetilde{\boldsymbol{E}}_{\mathrm{true}}$ among the soil types.

The uncertainty in SHP also propagated to the simulation of $\boldsymbol{E}$, mainly due to the effects of low water potential $\boldsymbol{h}$, which reducing root uptake and transpiration (see detail in supplementary material). The

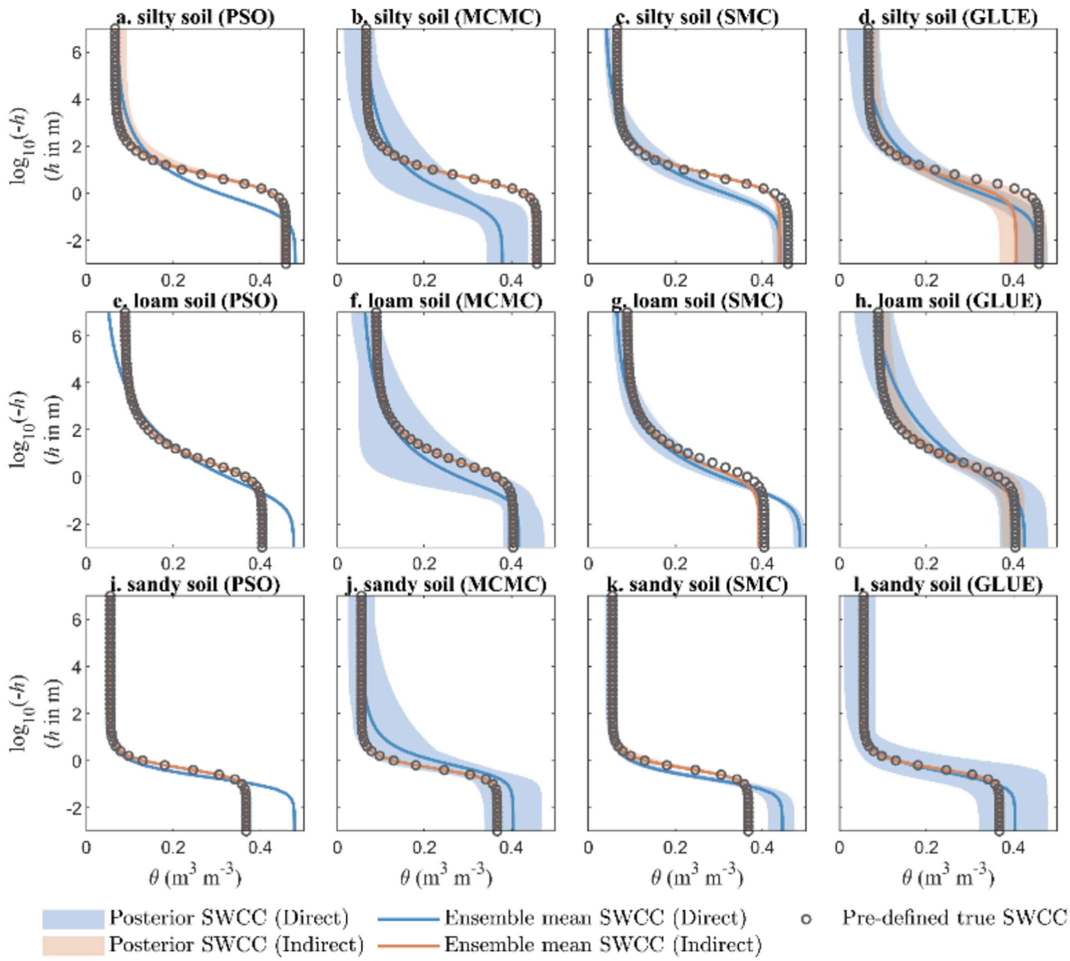

Fig. 7. The posterior SWCC calculated with optimized SHP (α, n, θ_r, θ_s and K_s of VG model) under Direct sampling strategy and Indirect strategy. The black hollow circle "○" represent predefined true h vs.θ relations to express SWCC under predefined true SHP.

Indirect sampling strategy, characterized by lower parameter uncertainty, more precisely simulated $\boldsymbol{E}$ compared with the Direct sampling strategy. The MCMC-Direct algorithm yielded the highest uncertainty in posterior $\boldsymbol{E}$. Furthermore, as the increasing $\boldsymbol{E}$, the width of the uncertainty bands associated with posterior $\boldsymbol{E}$ also augmented. As $\boldsymbol{E}$ values grew larger, the pattern of $\boldsymbol{E}$ underwent a transition from energy limitation to water limitation. Subsequently, the propagation from SHP to $\boldsymbol{E}$ exhibited an increasingly pronounced significant.

5.4. Propagation of error from parameters to simulated state variables

The Bias, representing the discrepancy between simulated and truth values of θ, $-\log_{10}(-\boldsymbol{h})$ and $\boldsymbol{E}$, is illustrated as a boxplot in Fig. 11. Additional evaluation criteria, including NSE and RMSE, are presented in supplemental material (Fig. S3). In alignment with the findings from Section 5.3, the broader fluctuation range of Bias in posterior soil moisture θ correlates with an equivalently enlarged range of Bias in posterior soil water potential $\boldsymbol{h}$ and evaporation $\boldsymbol{E}$. Generally, the posterior results exhibited a reduced fluctuation range of Bias under the Indirect sampling strategy, particularly when utilizing the MCMC algorithm. The optimization algorithms were designed to align with $\widetilde{\theta}_{\text{true}}$, and both Indirect and Direct sampling strategies successfully simulate θ, resulting in nearly zero median Bias values (Fig. 11. a ~ c). However, due to the parameter uncertainty illustrated in Fig. 6, a large Bias was apparent for $\boldsymbol{h}$ (Fig. 11. d ~ f). Under the Indirect strategy, the median Bias of $\boldsymbol{h}$ consistently oscillated around zero. Conversely, under the Direct strategy, the majority of median Bias values for $\boldsymbol{h}$ exceeded zero, signifying an overestimation of $\widetilde{\boldsymbol{h}}_{\text{true}}$.

The uncertainty inherent in $\boldsymbol{h}$ also propagated to the uncertainty in $\boldsymbol{E}$. Under the Indirect sampling strategy, the median Bias of $\boldsymbol{E}$ oscillated around zero, reflecting the performance of $\boldsymbol{h}$, with the exception of posterior $\boldsymbol{E}$ optimized by SMC-Indirect. In contrast, under the Direct sampling strategy, the median Bias of posterior $\boldsymbol{E}$ exhibited a more pronounced deviation from zero. For instance, under PSO-Direct, the median posterior $\boldsymbol{E}$ for silt soil oscillated around zero, whereas the median Bias of $-\log_{10}(-\boldsymbol{h})$ for silt soil was 0.341 m, and the median Bias of $\boldsymbol{E}$ for silt soil was -0.0039 mm h^{-1}.

Due to the positive correlation between $\boldsymbol{h}$ and $\boldsymbol{E}$, an overestimation of $\boldsymbol{h}$ was anticipated to result in an overestimation of $\boldsymbol{E}$. However, in the estimation of the soil hydrological process of silty soil using PSO-Direct, the overestimation of $\boldsymbol{h}$, characterized by a median value greater than zero, led to an underestimation of $\boldsymbol{E}$. This phenomenon was consistent across MCMC-Direct for three soil types, SMC-Indirect for loam soil, and SMC-Direct for silt soil. These discrepancies may stem from inaccurate parameter estimations that influence the soil hydrology dynamics, leading to varied performance in estimating $\boldsymbol{h}$ across different season. For instance, in the case of silty soil optimized by PSO-Direct, the $\boldsymbol{h}$ underestimated $\widetilde{\boldsymbol{h}}_{\text{true}}$ within the lower saturate range, yet overestimated it with the higher saturate range (Fig. 11 g). Due to the longer duration of the rainy season, characterized by a higher saturate range of $\boldsymbol{h}$, as compared to the dry season, featuring a lower saturate range of $\boldsymbol{h}$, the Bias of $\boldsymbol{h}$ for silt soil under PSO-Direct was larger than zero. According to the positive correlation between $\boldsymbol{h}$ and $\boldsymbol{E}$, $\boldsymbol{E}$ was underestimated during the dry season and overestimated during the rainy season. However, the

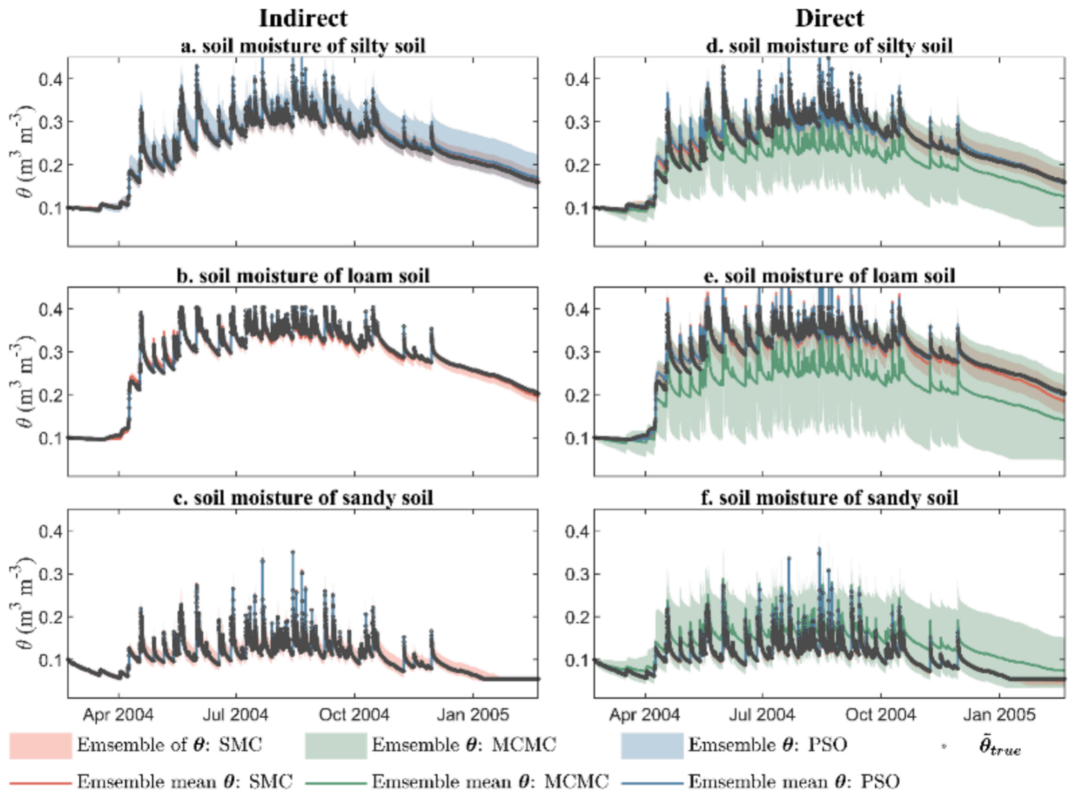

Fig. 8. The uncertainty bands of posterior soil moisture θ under Direct strategy and indirect strategy optimized by PSO (blue band), MCMC (green band) and SMC (red band). The black hollow circle "○" represent the predefined soil water content $\widetilde{\theta}_{\mathrm{true}}$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

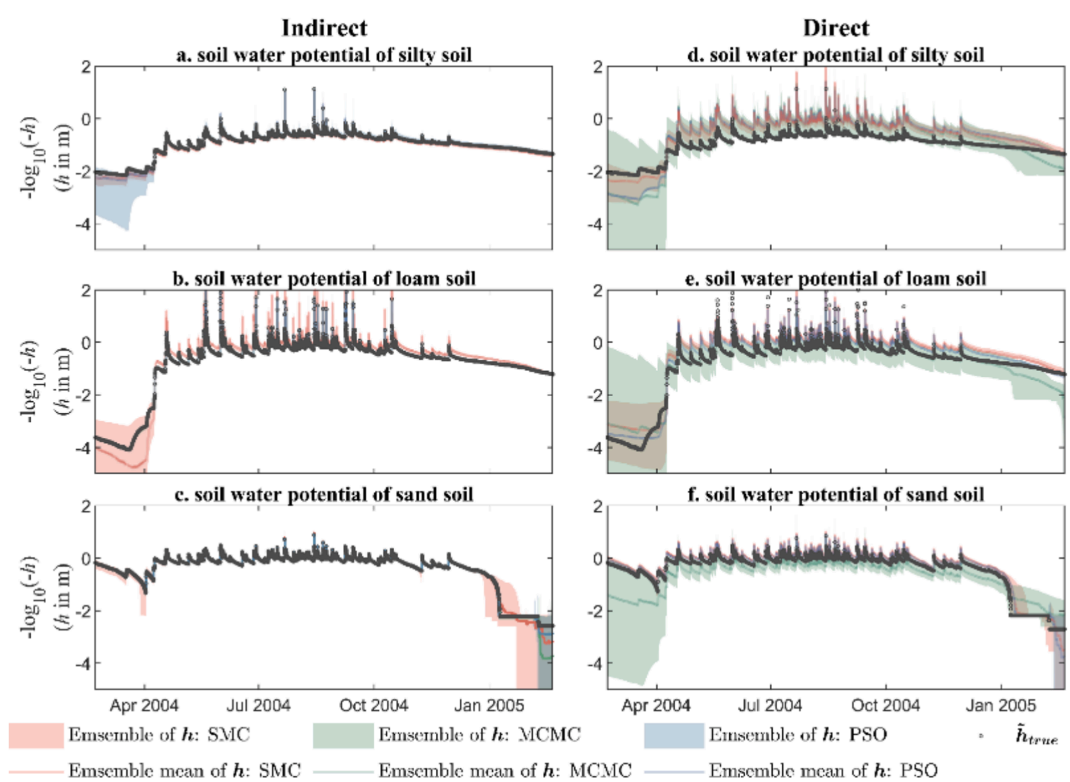

Fig. 9. The uncertainty bands of posterior soil water potential $\boldsymbol{h}$ under Indirect sampling strategy and soil water potential under Direct sampling strategy optimized by PSO (blue band), MCMC (green band) and SMC (red band). The black "○" is the predefined soil water potential $\widetilde{\boldsymbol{h}}_{\mathrm{true}}$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

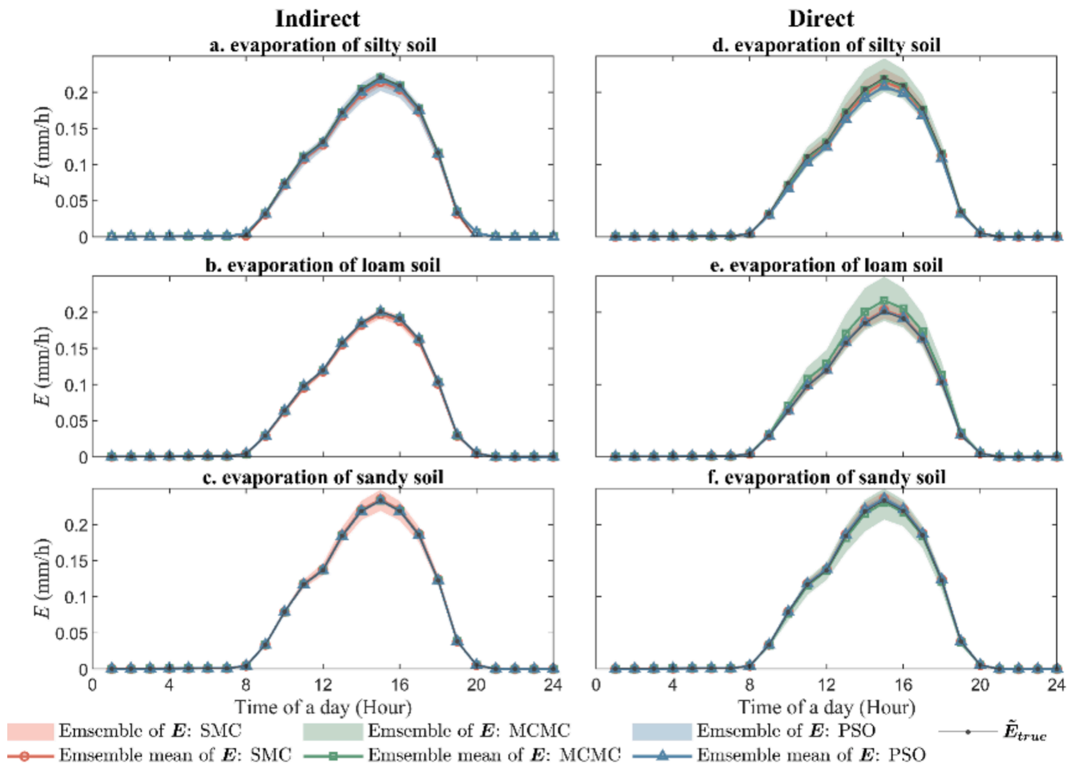

Fig. 10. The posterior uncertainty bands of diurnal variation of simulated evaporation using Direct sampling strategy and Indirect sampling strategy. The black "●" is the predefined soil water content $\widetilde{\boldsymbol{E}}_{\text{true}}$.

limitation of $\boldsymbol{h}$ on $\boldsymbol{E}$ was more significant during the dry season, thereby mitigating the overestimation of $\boldsymbol{E}$ during the rainy season. Therefore, the Bias of $\boldsymbol{E}$ for silt soil under PSO-Direct was less than zero.

6. Discussion

6.1. The characteristic of different optimization algorithms

Various algorithms for parameter optimization can be roughly classified into two categories (Zhou et al., 2014). The first category of inverse methods focuses on minimizing or maximizing an objective function that quantifies the discrepancy between simulated and observed values, exemplified by algorithms like PSO. This category generally employs a global optimization technique, yielding a single optimal parameter set. The second group of inverse methods are typically used for assessing parameter uncertainty by generating multiple sets of parameters based on probabilistic or sampling techniques, thus providing a range of possible solutions or parameter estimates rather than a single optimal set. Representative examples with this category include traditional rejection sampling (e.g., GLUE), importance resampling (e.g., SMC), and MCMC methods. Various algorithms exhibit distinct convergence characteristic. The iteration required for convergency and the corresponding computation times of these algorithms are display in the supplementary material as Fig. S4, Fig. S5 and Table 5.

PSO, as a representative optimization method, is known for its rapid convergence towards acceptable solutions. Regardless of whether Indirect or Direct sampling strategies are employed, PSO achieves a stable solution within approximately 20 iterations (Fig. S4 and Fig. S5). However, in scenarios with high-dimensional parameters, the rapid convergence provided by the PSO algorithm may encounter difficulties in accurately recovering the true parameter value due to equifinality phenomenon. Consequently, PSO converge to a local optimum in inverse modeling characterized by a high degree of parameter uncertainty.

MCMC, based on Bayesian theory, effectively explores parameter space through the stochastic traversal of Markov chain and converges towards high probability density regions. In Fig. 5, MCMC distinguishes itself as the sole algorithm that accurately converges to the predefined SPC of silty soil. However, MCMC is susceptible to the outlying chain, which can influence the jumping probability of the stationary distribution and make it impossible to reach convergence to a limiting distribution (Vrugt et al., 2008). Especially when confronted with high equifinality-related parameter uncertainty under Direct sampling strategy, MCMC trapped in a low probability region of parameter space. Using the Indirect sampling strategy, MCMC can accurately converging to the predefined true parameters (Fig. 5 and Fig. 6). Moreover, MCMC characterized by a slow mixing speed of chains, requires a considerable number of iterations to achieve convergence to a stationary distribution (Table 5, Fig. S4 and S5).

SMC, employing the importance sampling-resampling method, primarily relies on the posterior distribution rather than the exploration of parameter space. This approach enables the simultaneous computation of multiple particle states, facilitating rapid convergence (Fig. S4 and S5). Due to its swift convergence, SMC exhibits significant potential for adaptation to dynamically evolving systems, including alterations in the state transition function or observation function of the system. However, the rapid convergence may result in the substantial loss of particle information, reducing the accuracy of model simulations. In Fig. 6, the posterior SPC optimized by SMC is the only algorithm that doesn't encompass the predefined SPC of silty and loam soil.

Nearly all algorithms encounter challenges in parameter optimization in high-dimensional spaces, particularly when confronted with parameters that exhibit significant equifinality phenomena. Relying solely on soil texture information can further reduce the degrees of freedom in the parameter space, thereby potentially reducing the uncertainties inherent in estimated SHP. Consequently, this leads to a notable improvement in both the convergence speed (Table 5) and precision of the algorithm.

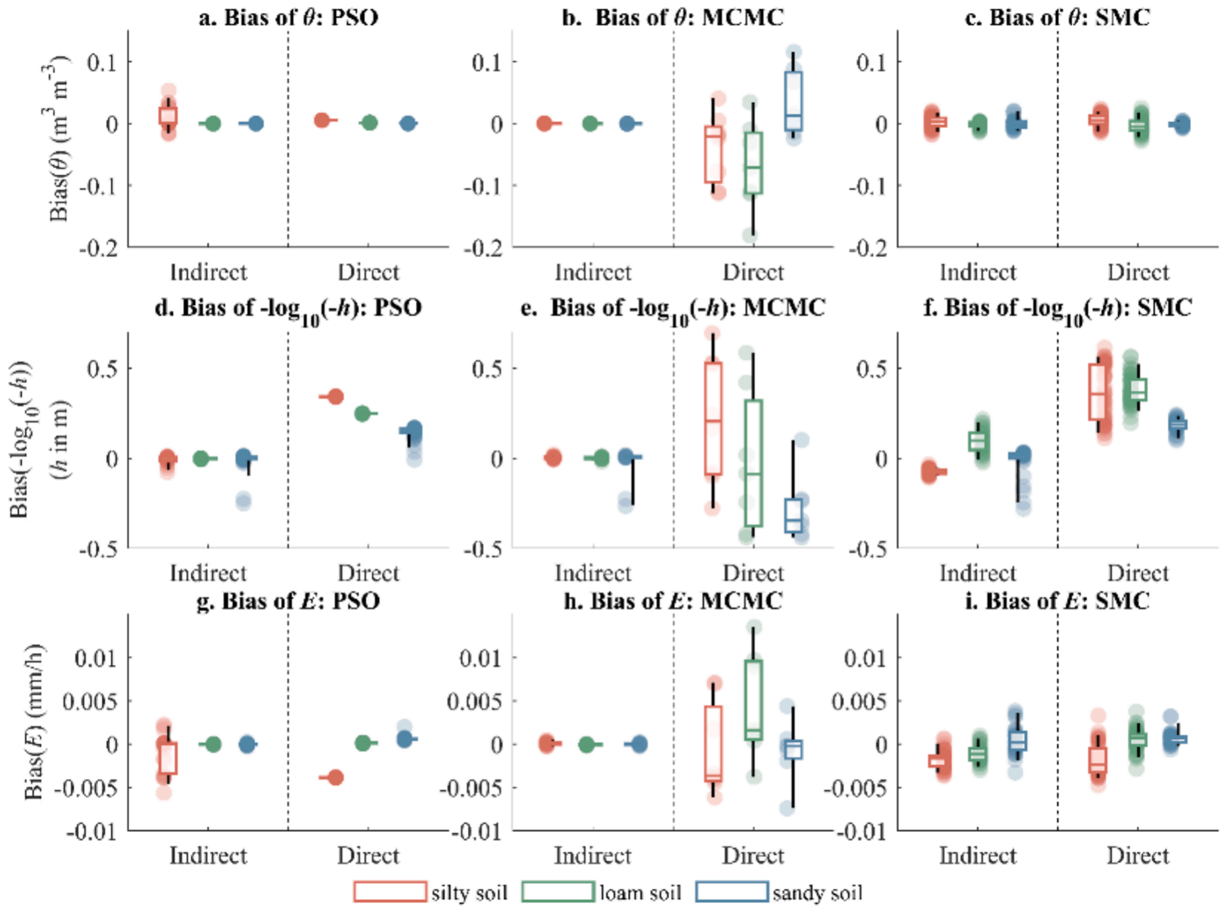

Fig. 11. The Bias of posterior soil moisture (θ), soil water potential (***h***) and evaporation (***E***) of silt (red box), loam (green box) and sand soil (blue box), when using three different algorithms (PSO, MCMC and SMC) and two different sampling strategies (Direct and Indirect). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 5
Computation times under single thread of Intel(R) Xeon(R) W-2145 CPU.

	PSO	MCMC	SMC	GLUE
Direct Sampling strategy	174.15 h	219.10 h	182.16 h	156 h
Indirect Sampling strategy	160.8 h	157.08 h	174.33 h	134 h

6.2. The adaptation of Indirect sampling strategy in different climate condition

With the development of computational power, the emergence of accurate and robust numerical solution schemes for the unsaturated flow equations, and the development of effective and efficient parameter optimization algorithms, the application of inverse modelling for determination of SHP has become increasingly popular over the last few decades(Scharnagl et al., 2011). However, due to the limited variability that yields a narrow spectrum of soil moisture under natural boundary conditions, the soil moisture observations lack the requisite information to support an accurate and precise estimation of all parameters (Bordoni et al., 2017). The findings of this study indicate that using the Rosetta 3 pedotransfer function to constrain the parameter sampling space can effectively improve parameter estimation. The meteorological forcing data used in this study covered the entire hydrological cycle, encompassing both wet and dry condition, thereby providing a sufficient information of soil moisture dynamics. However, in humid region and arid region, soil moisture may be insufficient to encapsulate comprehensive information on soil hydraulic properties. Consequently, it always displays significant uncertainty under converse conditions. Rosetta 3 incorporates abundant prior information on SHP correlations across a relatively comprehensive soil moisture range to effectively reduce the parameter uncertainty.

Taking the sandy soil as an example, a systematic numerical experiment was conducted to optimize SHP under wet and dry climatic condition. Under the dry condition, there were only five precipitation events, with a total precipitation of less than 33.7 mm. Conversely, under the wet condition, the precipitation events were more frequent, resulting in an accumulated precipitation of 609.8 mm. The GLUE algorithm was employed to assess the parameter uncertainty under Direct and Indirect sampling strategy, with an acceptance threshold specified to be 0.95. The posterior SWCC and θ were shown in Fig. 12. The results indicate that the simulated SWCC using the Direct sampling strategy under dry conditions exhibits reduced uncertainty in the low saturation region but exhibits larger uncertainty in the high saturation region. Adoption of the Indirect sampling strategy can notably reduce parameter uncertainties, thereby enhancing the accuracy of SWCC and θ simulations, regardless of dry or wet conditions.

6.3. Recommendation of optimizing soil hydraulic parameter using Rosetta 3

The Direct strategy revealed a significant equifinality phenomenon. As shown in Section 5.1, the results indicated that SHP sampled by Direct sampling strategy exhibited irregular posterior PDF and flat response surfaces (Fig. 6). The relatively wide parameter ranges implied a significant equifinality phenomenon, especially in the context of the MCMC algorithm. The uncertainty analysis underscores the challenges

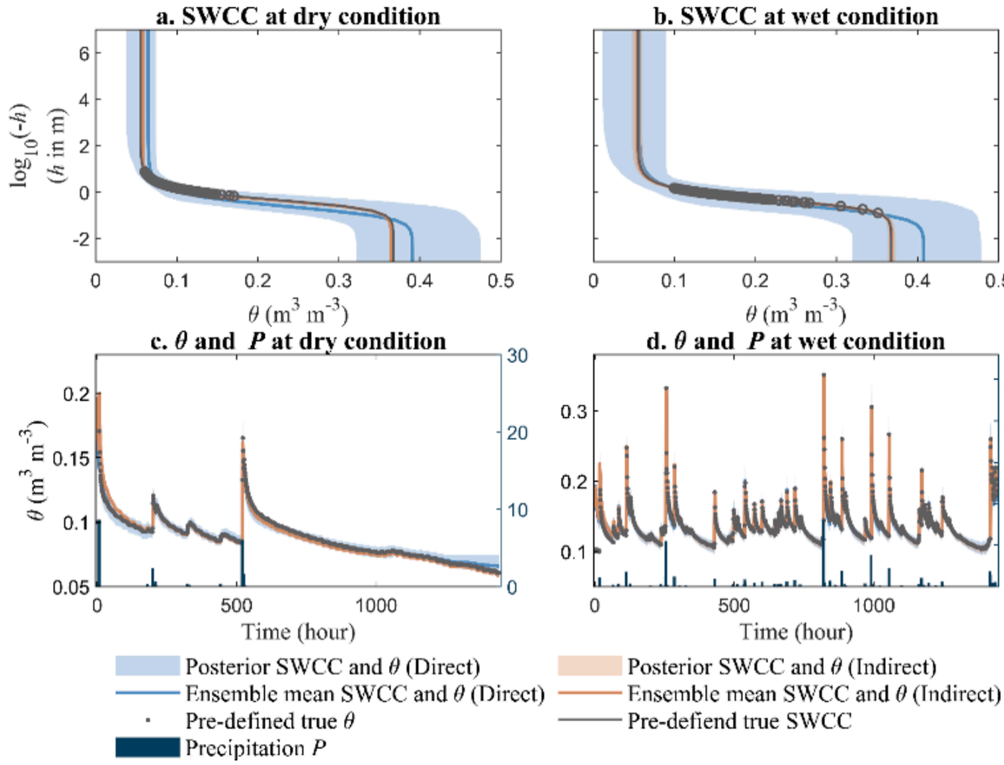

Fig. 12. The uncertainty bands of posterior SWCC and soil moisture θ under Indirect sampling strategy and under Direct sampling strategy.

in parameter optimization, particularly when using random sampling of the VG parameter (e.g., Direct sampling strategy) without utilizing any prior information. Without a predefined constrained range for sensitive SHP (i.e., K_s, θ_r ,θ_s, n, and α), randomly generated parameter sets could potentially yield a number of extreme combinations. It is plausible that a compensation effect may exist in SHP in the VG model. For instance, an overestimation of n combined with an underestimation of α may result in similar SWCC if the parameters θ_r and θ_s were also included in random sampling. If a parameter set comprised extreme values for each SHP, it may only describe an extreme condition.

Our results indicate that the Direct sampling strategy fails to provide a constrained description of hydraulic properties in natural soils, and the associated uncertainty is certainly inappropriate for ecosystem hydrological simulation. When taking the θ as the optimization objective, the significant equifinality phenomenon of parameters caused large uncertainty of estimation in h and E (Fig. 11). Considering the Rosetta 3 pedotransfer function incorporated prior information of more than 2000 groups of experimental data of undisturbed soils, SHP transferred through Rosetta 3 can significantly avoid the arbitrary combinations estimated through random parameter space (Fig. 1). Most importantly, employing the Rosetta 3 pedotransfer function to incorporate soil texture information reduced the dimensionality of parameter space, leading to a marked improvement in algorithm convergence speed and precision. Therefore, it is recommended to integrate various optimization algorithms and tools with the well-developed Rosetta 3 pedotransfer function for inverse modeling of ecohydrological processes.

7. Conclusion

The study conducted a systematic numerical experiment to optimize the SHP of three representative soils: silty, loam, and sandy soil. This optimization process was based on simulating the predefined true soil moisture dynamics under natural meteorological conditions. Four parameter optimization algorithms (PSO, MCMC, SMC, and GLUE) were employed to estimate the optimal SHP ($\alpha, n, \theta_r, \theta_s$ and K_s) for the unsaturated flow model. Two parameter sampling strategies were implemented in the optimization algorithms: the Indirect sampling strategy, which integrated soil texture information inherent in Rosetta 3, and the Direct sampling strategy, which randomly generated SHP values.

Upon evaluating the results derived from the Direct and Indirect sampling strategies across the three optimization algorithms, the following conclusions were formulated:

(1) Under the Indirect strategy, all four algorithms accurately estimated the predefined true SHP (Fig. 6). Conversely, when employing the Direct strategy, the posterior SHP exhibited deviation from the predefined true values. In comparison to Direct sampling strategy, the posterior PDF of SHP and the uncertainty bands of the SWCC demonstrated less uncertainty under the Indirect sampling strategy.

(2) The uncertainty in SHP propagated to soil moisture dynamics and evaporation (Fig. 8 ~Fig. 10). Under the Indirect strategy, all three algorithms accurately simulated $\widetilde{\theta}_{\text{true}}$, $\widetilde{\boldsymbol{h}}_{\text{true}}$, and $\widetilde{\boldsymbol{E}}_{\text{true}}$. However, under the Direct strategy, MCMC demonstrated significant uncertainty in θ, $\boldsymbol{h}$, and $\boldsymbol{E}$. PSO and SMC accurately captured the variation in θ but overestimated $\boldsymbol{h}$, consequently affecting the estimation of $\boldsymbol{E}$.

(3) The incorporation of soil texture information using Rosetta 3 reduced the degrees of freedom in the parameter space. This led to a significant improvement in the convergence speed and precision of the algorithm.

CRediT authorship contribution statement

Meijun Li: Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis. **Wei Shao:** Writing – review & editing, Validation, Software, Methodology, Funding acquisition, Conceptualization. **Wenjun Yu:** Writing – review & editing, Validation. **Ye Su:** Writing – review & editing, Funding acquisition. **Qinghai Song:** Data curation. **Yiping Zhang:** Data curation. **Hongkai Gao:** Writing – review & editing, Visualization. **Yonggen Zhang:** Software, Methodology, Conceptualization. **Jianzhi Dong:** Methodology, Conceptualization.

Declaration of competing interest

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.

Acknowledgement

This work was supported by National Natural Science Foundation of China (Grant No. 42471032); Postgraduate Research & Practice Innovation Program of Jiangsu Province (Grant No. KYCX24_1516); The Czech Science Foundation GAČR PIF OUT project (Grant No. 242-201284).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.geoderma.2025.117210.

Data availability

The data that support the findings of this study are available in China flux (http://www.cnern.org.cn/data/initDRsearch?classcode=SYC_A02&tdsourcetag=s_pcqq_aiomsg). The Rosetta 3 pedotransfer function is available at https://github.com/usda-ars-ussl/rosetta-soil.

References

Beven, K., Binley, A., 1992. The Future of Distributed Models: Model Calibration and Uncertainty Prediction. 6 (3), 279–298. https://doi.org/10.1002/hyp.3360060305.
Bordoni, M., Bittelli, M., Valentino, R., Chersich, S., Meisina, C., 2017. Improving the estimation of complete field soil water characteristic curves through field monitoring data. Journal of Hydrology 552, 283–305. https://doi.org/10.1016/j.jhydrol.2017.07.004.
Braak, C.J.F.T., 2006. A Markov Chain Monte Carlo version of the genetic algorithm Differential Evolution: easy Bayesian computing for real parameter spaces. Statistics and Computing 16 (3), 239–249. https://doi.org/10.1007/s11222-006-8769-1.
Chirico, G.B., Medina, H., Romano, N., 2007. Uncertainty in predicting soil hydraulic properties at the hillslope scale with indirect methods. Journal of Hydrology 334 (3), 405–422. https://doi.org/10.1016/j.jhydrol.2006.10.024.
Cho, S.E., 2014. Probabilistic stability analysis of rainfall-induced landslides considering spatial variability of permeability. Engineering Geology 171, 11–20. https://doi.org/10.1016/j.enggeo.2013.12.015.
Del Moral, P., Doucet, A., Jasra, A., 2006. Sequential Monte Carlo Samplers. Journal of the Royal Statistical Society Series b: Statistical Methodology 68 (3), 411–436. https://doi.org/10.1111/j.1467-9868.2006.00553.x %J Journal of the Royal Statistical Society Series B: Statistical Methodology.
Eberhart, R., Kennedy, J., 1995. A new optimizer using particle swarm theory, MHS'95. In: Proceedings of the Sixth International Symposium on Micro Machine and Human Science. Ieee, pp. 39–43.
Gholami, A., et al., 2018. Uncertainty analysis of intelligent model of hybrid genetic algorithm and particle swarm optimization with ANFIS to predict threshold bank profile shape based on digital laser approach sensing. Measurement 121, 294–303. https://doi.org/10.1016/j.measurement.2018.02.070.
Ippisch, O., Vogel, H.J., Bastian, P., 2006. Validity limits for the van Genuchten–Mualem model and implications for parameter estimation and numerical simulation. Advances in Water Resources 29 (12), 1780–1789. https://doi.org/10.1016/j.advwatres.2005.12.011.
Liu, X., Wang, Y., 2021. Bayesian selection of slope hydraulic model and identification of model parameters using monitoring data and subset simulation. Computers and Geotechnics 139, 104428. https://doi.org/10.1016/j.compgeo.2021.104428.
Luo, Z., et al., 2019. Evaluation and application of the modified van Genuchten function for unsaturated porous media. Journal of Hydrology 571, 279–287. https://doi.org/10.1016/j.jhydrol.2019.01.051.
Pollacco, J.A.P., Ugalde, J.M.S., Angulo-Jaramillo, R., Braud, I., Saugier, B., 2008. A Linking Test to reduce the number of hydraulic parameters necessary to simulate groundwater recharge in unsaturated soils. Advances in Water Resources 31 (2), 355–369. https://doi.org/10.1016/j.advwatres.2007.09.002.
Richards, L.A., 1931. Capillary conduction of liquids through porous mediums. Physics 1, 318–333.
Scharnagl, B., Vrugt, J.A., Vereecken, H., Herbst, M., 2011. Inverse modelling of in situ soil water dynamics: investigating the effect of different prior distributions of the soil hydraulic parameters. Hydrol. Earth Syst. Sci. 15 (10), 3043–3059. https://doi.org/10.5194/hess-15-3043-2011.
Shao, W., et al., 2023a. Reducing uncertainties in hydromechanical modeling with a recently developed Rosetta 3 podeotransfer function. Engineering Geology 324, 107250. https://doi.org/10.1016/j.enggeo.2023.107250.
Shao, W., et al., 2023b. Reduce uncertainty in soil hydrological modeling: A comparison of soil hydraulic parameters generated by random sampling and pedotransfer function. Journal of Hydrology 623, 129740. https://doi.org/10.1016/j.jhydrol.2023.129740.
Shao, W., Coenders-Gerrits, M., Judge, J., Zeng, Y., Su, Y., 2018. The impact of non-isothermal soil moisture transport on evaporation fluxes in a maize cropland. Journal of Hydrology 561. https://doi.org/10.1016/j.jhydrol.2018.04.033.
ter Braak, C.J.F., Vrugt, J.A., 2008. Differential Evolution Markov Chain with snooker updater and fewer chains. Statistics and Computing 18 (4), 435–446. https://doi.org/10.1007/s11222-008-9104-9.
van Dam, J.C., Feddes, R.A., 2000. Numerical simulation of infiltration, evaporation and shallow groundwater levels with the Richards equation. Journal of Hydrology 233 (1), 72–85. https://doi.org/10.1016/S0022-1694(00)00227-4.
van Genuchten, M.T., 1980. A Closed-form Equation for Predicting the Hydraulic Conductivity of Unsaturated Soils. Soil Sci. Soc. Am. J. 44 (5), 892–898. https://doi.org/10.2136/sssaj1980.03615995004400050002x.
Vereecken, H., et al., 2010. Using Pedotransfer Functions to Estimate the van Genuchten–Mualem Soil Hydraulic Properties: A Review. Vadose Zone Journal 9 (4), 795–820. https://doi.org/10.2136/vzj2010.0045.
Vereecken, H., et al., 2022. Soil hydrology in the Earth system. Nature Reviews Earth & Environment 3 (9), 573–587. https://doi.org/10.1038/s43017-022-00324-6.
Vrugt, J.A., ter Braak, C.J.F., Clark, M.P., Hyman, J.M., Robinson, B.A., 2008. Treatment of Input Uncertainty in Hydrologic Modeling: Doing Hydrology Backward with Markov Chain Monte Carlo Simulation. 44 (12). https://doi.org/10.1029/2007WR006720.
Xu, Q., et al., 2022. A novel frequency-domain integrated sensor for in-situ estimating unsaturated soil hydraulic conductivity. Journal of Hydrology 610, 127939. https://doi.org/10.1016/j.jhydrol.2022.127939.
Zhang, Y., Schaap, M.G., 2017. Weighted recalibration of the Rosetta pedotransfer model with improved estimates of hydraulic parameter distributions and summary statistics (Rosetta3). Journal of Hydrology 547, 39–53. https://doi.org/10.1016/j.jhydrol.2017.01.004.
Zhou, H., Gómez-Hernández, J.J., Li, L., 2014. Inverse methods in hydrogeology: Evolution and recent trends. Advances in Water Resources 63, 22–37. https://doi.org/10.1016/j.advwatres.2013.10.014.