Quantifying key parameter sensitivities for water table depth in hydrological schemes of CoLM-PSUADE at a mountain forest site

Tingting Wu[1,2], Shupeng Zhang[3], Xiaofan Yang[1,2], and Yongjiu Dai[3]

[1]State Key Laboratory of Earth Surface Processes and Disaster Risk Reduction, Faculty of Geographical Science, Beijing Normal University, Beijing, 100875, China
[2]Guangdong Provincial Observation and Research Station for Coupled Human and Natural Systems in Land–ocean Interaction Zone, Beijing Normal University, Zhuhai, 519087, China
[3]School of Atmospheric Sciences, Sun Yat-sen University, Zhuhai, 519000, China

Correspondence to: Xiaofan Yang (xfyang@bnu.edu.cn)

Abstract. Accurately representing groundwater dynamics in land surface models (LSMs) is crucial for understanding water-energy cycles and assessing water resources. However, most LSMs lack systematic evaluations of parameter sensitivities governing water table depth (WTD) simulations. This study coupled the Common Land Model (CoLM) with Problem Solving environment for Uncertainty Analysis and Design Exploration (PSUADE) in a single-point framework to facilitate systematic parameter analysis and optimization for improving WTD simulation. The CoLM-PSUADE framework was then applied to evaluate groundwater-related parameters using WTD observations from the Gongga Mountain site. A comprehensive analysis integrating qualitative screening, quantitative validation, and parameter optimization techniques was conducted to evaluate the sensitivities of 56 parameters associated with key hydrological processes and identify their optimal ranges. The results identified eight sensitive parameters in the initial screening, ~~indicated that eight parameters can be identified as robustly sensitive,~~ including those controlling unsaturated soil water movement (56-soil_alpha, 53-soil_n), subsurface runoff (40-rsubmax), plant hydraulic processes (49-beta, 45-krmax, 46-ck0), and net surface water infiltration (4-alpha_rain, 10-rhol_nir). Additional experiments with independent random seeds revealed varying sensitivity rankings. ~~Among them~~In particular, the subsurface runoff parameter $r_{sub,max}$ exhibited sampling-dependent sensitivity but showed relatively high importance in the analysis. It exhibited a well-defined optimal range (on the order of 10^{-4}) and regulated both the magnitude of subsurface runoff and its decay with increasing WTD when combined with another empirical parameter in the SIMTOP (Simple TOPMODEL-based) scheme, f_{drai}, thereby exerting strong control on WTD. The soil hydraulic parameter α showed the highest sensitivity. It regulated unsaturated hydraulic conductivity and soil water retention, thereby exerting a dominant influence on the variability and lagged response of WTD. Based on these findings, a stepwise calibration strategy was recommended, in which the subsurface runoff parameters ($r_{sub,max}$ and f_{drai}) are first adjusted to constrain the mean WTD, followed by optimization of other key parameters, such as α, to improve the temporal dynamics of WTD. This study demonstrated that CoLM-PSUADE provides a useful tool for sensitivity-guided parameter optimization in high-dimensional LSMs and hydrological models.

1 Introduction

Groundwater, as a key component of the hydrological cycle, modulates water and energy balances by regulating soil moisture and evapotranspiration, which in turn affects the climate system and terrestrial ecosystem (Vogelbacher et al., 2024). Hence, groundwater processes of varying complexity are increasingly represented in land surface models (LSMs), such as Variable Infiltration Capacity (VIC;) (Liang et al., 2003), Noah-MP ((Niu et al., 2007; Niu et al., 2011; Yang et al., 2011), Community Land Model (CLM) (; Gulden et al., 2007; Maxwell and Miller, 2005), and Common Land Model (CoLM) (; Dai et al., 2019a), all of which simulate land surface processes and land–atmosphere interactions. These developments enable investigations into the important role of groundwater in regulating soil moisture (Akhter et al., 2025), evapotranspiration (Maina et al., 2022), and vegetation dynamics (Gou et al., 2018), as well as in moderating global warming and extreme weather events (Marchionni et al., 2020). In most LSMs, groundwater processes are typically solved within a variably saturated flow framework using pressure head or soil moisture profiles (Dai et al., 2019a; Yang et al., 2026), and are tightly linked with infiltration, runoff, and evapotranspiration processes (Jia et al., 2026). In addition, new modules such as plant hydraulics (Kennedy et al., 2019; Li et al., 2021) or new parameterizations of the soil water characteristic curve (Silwimba et al., 2025) have been progressively integrated into LSMs to investigate coupled interactions between water cycles and atmospheric feedback. Against this backdrop, LSMs have continued to evolve, with developments such as CoLM2004 (Dai et al., 2003), CoLM2014 (Dai et al., 2019b), and CoLM2024 (Bai et al., 2024). As a result, the simulation of groundwater dynamics has been increasingly dependent on new parameterization schemes. This trend highlights the need to accurately simulate water table depth (WTD) to better understand and predict groundwater behaviour in a changing climate.

During the model calibration process, parameter uncertainty is widely recognized as one of the greatest sources of uncertainty in process-based LSMs (Raoult et al., 2025). Consequently, the appropriate specification of model parameters is essential for ensuring the accurate representation of land surface processes (Zheng et al., 2019; Tsai et al., 2021). Conventional hydrological modeling typically relies on manual calibration, in which simulated outputs are compared against observations to adjust parameter values. However, as LSMs continue to evolve and increasingly incorporate key components of the hydrological cycle, the energy balance, and biogeochemical processes (Dai et al., 2019c), the number of adjustable parameters has expanded substantially (from O(10) to O(100)) (Li et al., 2013). This growth renders parameter calibration a computationally expensive and time-consuming task, particularly for LSMs characterized by high-dimensional systems, strong nonlinear parameter interactions, and pronounced parameter space heterogeneity. Therefore, identifying the parameters that exert the dominant influence on model outputs is crucial for improving efficiency while maintaining accuracy. By fixing insensitive parameters or those that can be directly observed, the dimensionality of the adjustable parameters can be reduced, thereby lowering computational costs and enabling effective enhancement of model performance. In this process, many parameters without clear physical meaning cannot be directly measured, while they are essentially effective parameters (Raoult et al., 2025), such as empirical parameters used to control subsurface runoff. For these

parameters, sensitivity analysis becomes a critical tool for identifying the most influential ones (Ratto et al., 2012). Among available approaches, global sensitivity analysis methods have gained widespread application in nonlinear process-based models because they quantify both individual parameter effects and interaction effects on model outputs (Zhan et al., 2013; Erdal and Cirpka, 2019). Qualitative analysis methods provide intuitive rankings of relative parameter importance with fewer model runs, making them suitable for efficiently screening out insensitive parameters in high-dimensional systems (Guerrero et al., 2017). However, they do not explain how much of the performance variance is due to a given parameter. In contrast, quantitative approaches evaluate the contribution of each parameter to the total output variance, thereby enabling a more rigorous and comprehensive assessment (Gan et al., 2014). Nevertheless, such methods typically require a large number of model runs and thus incur substantial computational costs, especially for complex models with high-dimensional parameter spaces. Therefore, an integrated strategy that combines qualitative screening with subsequent quantitative assessment has increasingly become a common practice (Gan et al., 2014; Li et al., 2013). To further evaluate optimal parameter values, sensitivity analysis is commonly combined with parameter optimization techniques (Duan et al., 1994). In this context, integrated analysis platforms play an increasingly important role. In particular, the Problem Solving environment for Uncertainty Analysis and Design Exploration (PSUADE) software package offers a complete workflow that includes parameter sampling, sensitivity analysis, and parameter optimization (Tong, 2022b). PSUADE supports various sampling strategies and analysis methods, thereby enabling effective parameter screening and optimization in complex LSMs.

Previous studies have successfully coupled PSUADE with the CoLM2014 (Li et al., 2013), the NCAR community atmosphere model (Pathak et al., 2020), the Sacramento Soil Moisture Accounting (SAC-SMA) model (Gan et al., 2014), and the Canadian Small Lake Model (CSLM; Guerrero et al., 2017) to evaluate the performance and computational efficiency of various sensitivity analysis methods and sampling strategies. However, these studies have largely focused on sensible and latent heat fluxes (Pathak et al., 2020), net radiation, soil temperature, and soil moisture (Li et al., 2013; Guerrero et al., 2017), while systematic sensitivity analysis targeting WTD remain scarce. Notably, most existing studies focus only on sensitivity analysis and uncertainty quantification, with limited efforts to further integrate parameter optimization based on sensitivity analysis results. Therefore, exploring the combined use of sensitivity analysis and parameter optimization within PSUADE can enable effective parameter estimation and evaluation. This combined approach improves the accuracy of WTD simulations and is of great significance for understanding the dynamic response of WTD to hydrological fluxes within land–atmosphere interactions.

This study aims to improve WTD simulation by coupling CoLM2024 with PSUADE (CoLM-PSUADE). Through the integration of qualitative and quantitative sensitivity analyses, together with parameter optimization, key parameters affecting WTD simulations at the site scale, including those governing infiltration, runoff, evapotranspiration, and freeze-thaw processes, are systematically identified, evaluated, and constrained. These results provide a scientific basis and practical guidance for improving WTD simulation and parameter calibration in LSMs.

2 Data and methods

2.1 Study area and datasets

2.1.1 Study area

In this study, we selected a long-term integrated observation site located in a mature Abies fabri forest in the Gongga Mountain region on the eastern Qinghai-Tibet Plateau (hereafter referred to as the GGF site) as the research domain. The GGF site is situated within the Yangtze River Basin (Fig. 1), at 29°34' N and 101°59' E, with an elevation of 3100 m. The site was chosen due to its shallow WTD, generally less than 2 m (see Sect. 3.4 for details), which enhances the interaction between key hydrological processes such as infiltration, runoff, evaporation, and freeze-thaw. In such shallow groundwater systems, the water table is located within or close to the rooting zone, and capillary rise can further extend the influence of groundwater into the unsaturated zone (Rohde et al., 2024). This strong coupling between groundwater and hydrological processes allows for a comprehensive sensitivity analysis of WTD on these critical parameters. Shallow groundwater systems, defined as those with WTD < 2 m (O'connor et al., 2019), are widespread globally (Fan et al., 2013), making this site particularly relevant for studying their behaviour.

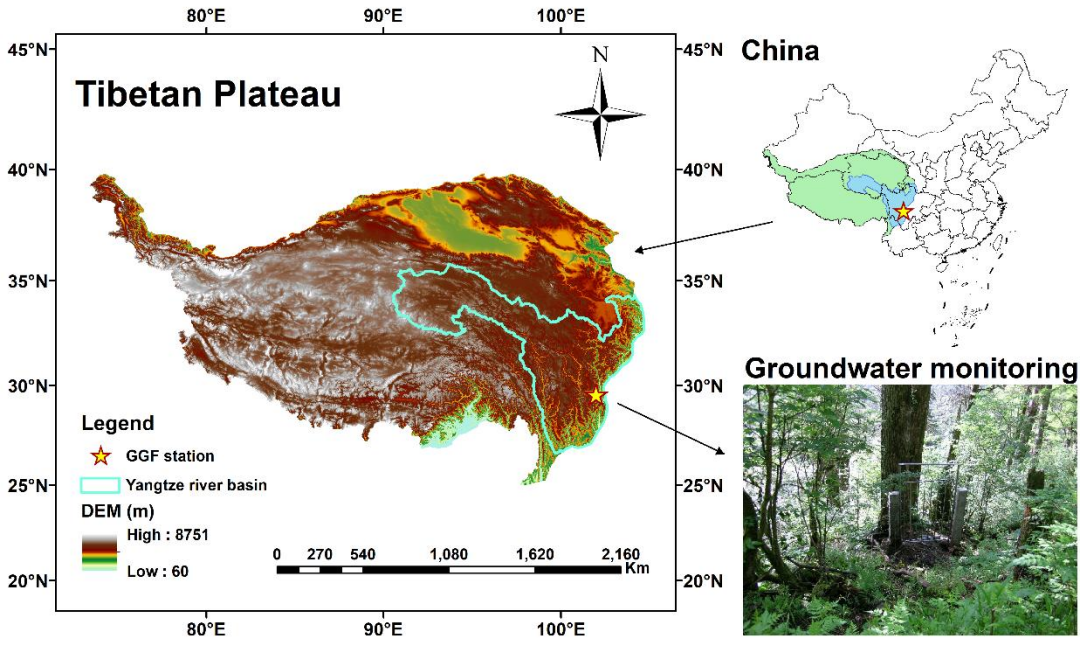


Figure 1: The location and digital elevation model (DEM) of study area.

The GGF site is characterized by a cold and humid climate, with a mean annual air temperature of approximately 4 ℃ and mean annual precipitation of about 1900 mm, most of which occurs between June and September (Zhao et al., 2017). This site, with its mature Abies fabri forest community, represents the typical altitudinal zonal vegetation type found between 2800 m and 3600 m in the Gongga Mountain region (Sun et al., 2013; Zhao et al., 2017). The soil is classified as brown coniferous forest soil with the subclass albic brown coniferous forest soil. Soil texture characterization was based on the measured particle size compositions at five depth intervals (0–10 cm, 10–20 cm, 20–40 cm, 40–60 cm, and 60–100 cm), which were collected on 25 August 2005 and 27 September 2015 (Table 1;) (Yang et al., 2020). As the particle size fractions were determined according to the United States Department of Agriculture (USDA) classification system, which defines sand (0.05–2 mm), silt (0.002–0.05 mm), and clay (<0.002 mm), the soil texture classification was consistently interpreted using the USDA system. Accordingly, the soil was predominantly identified as loamy sand or sandy loam, exhibiting a relatively uniform vertical distribution throughout the profile (Fig. 2).

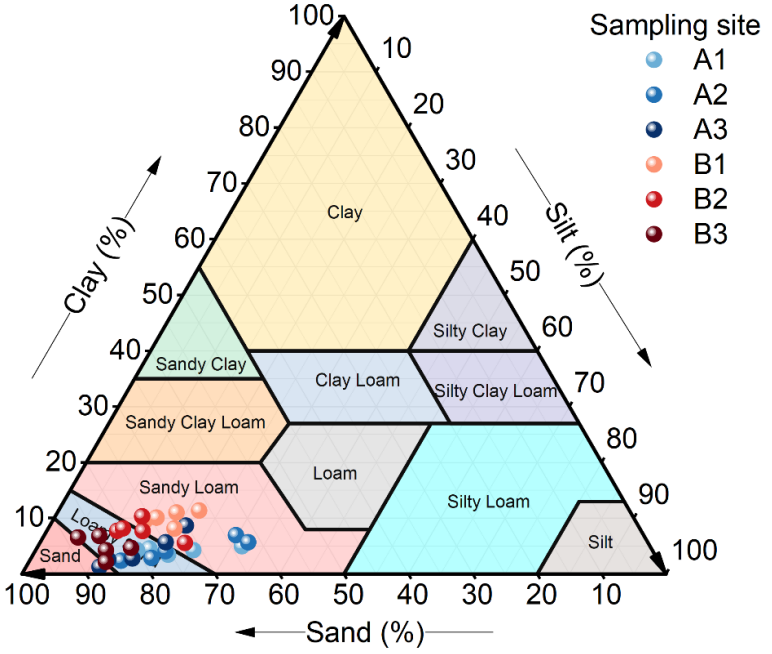


Figure 2: The soil texture classifications for samples collected at the GGF site in 2005 (A1, A2, A3) and 2015 (B1, B2, B3). A and B represent different sampling years, while 1, 2, and 3 denote three distinct plots. Each sampling point contains five measurements, representing soil depths of 0–10 cm, 10–20 cm, 20–40 cm, 40–60 cm, and 60–100 cm.

The CoLM-PSUADE was used to conduct sensitivity analysis and parameter optimization for WTD simulation, providing insights into the coupled hydrological processes at this representative site.

2.1.2 Forcing data and validation data

The forcing and validation data are summarized in Table 1. Meteorological forcing for CoLM was obtained from the meteorological forcing dataset for the Third Pole (TPMFD), including precipitation, 2 m air temperature, 2 m specific

humidity, 10 m wind speed, near-surface air pressure, downward longwave radiation, and downward shortwave radiation (Jiang et al., 2025; Jiang et al., 2023; Shao et al., 2022). Soil hydrothermal parameters were sourced from the Global Soil Dataset for Earth System Modeling (GSDE;) ((Dai et al., 2019d; Shangguan et al., 2014a), which provides bulk density, thermal conductivity, volumetric and gravimetric fractions of soil particles, and key parameters of the van Genuchten model. The dataset is vertically discretized into eight layers corresponding to depths of 0–0.045 m, 0.045–0.091 m, 0.091–0.166 m, 0.166–0.289 m, 0.289–0.493 m, 0.493–0.829 m, 0.829–1.383 m, and 1.383–2.296 m (Dai et al., 2019d). Leaf area index (LAI) data were derived from the reprocessed MODIS Version 6.1 LAI dataset (Yuan et al., 2011; Lin et al., 2023). WTD measurements at the GGF site were sourced from the dataset of groundwater levels in the Chinese Ecosystem Research Network (CERN;) ((Zhu et al., 2017b). The simulation periods were assigned different purposes for spin-up, sensitivity analysis, parameter optimization, and long-term validation of the optimized parameter set. The sensitivity analysis covered the 2005–2007 period, whereas the parameter optimization and validation covered the longer 2005–2014 period. Specifically, the model was spun up during 2005–2006 to achieve a dynamically equilibrated state, after which the 2006–2007 period was used for sensitivity analysis and parameter optimization. The selected 2006–2007 period exhibited relatively representative WTD dynamics at the GGF site while maintaining computational feasibility for the repeated sensitivity experiments. The temporal generalization of optimized parameters was subsequently validated using the simulation spanning 2007–2014, which included diverse hydroclimatic conditions for further evaluation of model performance. Additional discussion on the representativeness of the selected period is provided in Sect. S1 of the supplementary material.

Table 1: The datasets used in this study.

Data Type	Source	Coverage	Resolution
Soil texture	National Ecosystem Science Data Center (https://www.nesdc.org.cn/)	Point; Aug 2005, Sep 2015	1 day
Atmosphere	National Tibetan Plateau Data Center Third Pole Environment Data Center (https://data.tpdc.ac.cn/)	Tibetan Plateau; 2005–201407	1/30°; Hourly
LAI	Land-Atmosphere Interaction Research Group at Sun Yat-sen University (http://globalchange.bnu.edu.cn)	Global; 2005–201407	15 "; Monthly
Soil hydraulic and thermal parameters	Land-Atmosphere Interaction Research Group at Sun Yat-sen University (http://globalchange.bnu.edu.cn)	Global	15 "
WTD	National Ecosystem Science Data Center (https://www.nesdc.org.cn/)	Point; 2005–2014	Daily[a]

[a]WTD observations were recorded at irregular intervals ranging from 1–9 days, with most measurements spaced 4–6 days apart. During model evaluation, simulated WTD values were matched with the corresponding observation dates.

2.2 CoLM-PSUADE coupling framework

2.2.1 PSUADE

PSUADE software package, first released in 2007 and developed by Charles Tong, has been widely used for over 18 years and provides a comprehensive suite of tools for uncertainty analysis, sensitivity analysis, response surface analysis, and parameter optimization (Tong, 2022a, 2016). Notably, PSUADE is well suited for models with a large number of parameters and complex parameter interdependencies.

Core components of PSUADE include the sampling module, a simulator execution environment, and the analysis module (Fig. 3). The sampling module generates random samples based on user-specified parameters, their ranges, and the number of samples. The simulator execution environment, after the parameter samples have been constructed, propagates these sampled values to the simulator to produce corresponding outputs. The integrated environment in PSUADE enables simulator execution and results collection. In this study, the simulator is implemented as a Python script (Wu et al., 2026). When invoked by PSUADE, this script is responsible for launching CoLM and handling data exchange between PSUADE and CoLM, thereby achieving the coupling. The analysis module then evaluates the relationships between the sampled parameters and the resulting objective functions. In this study, qualitative sensitivity analysis, quantitative sensitivity analysis, and parameter optimization methods were employed (see Sect. 2.3 for details).

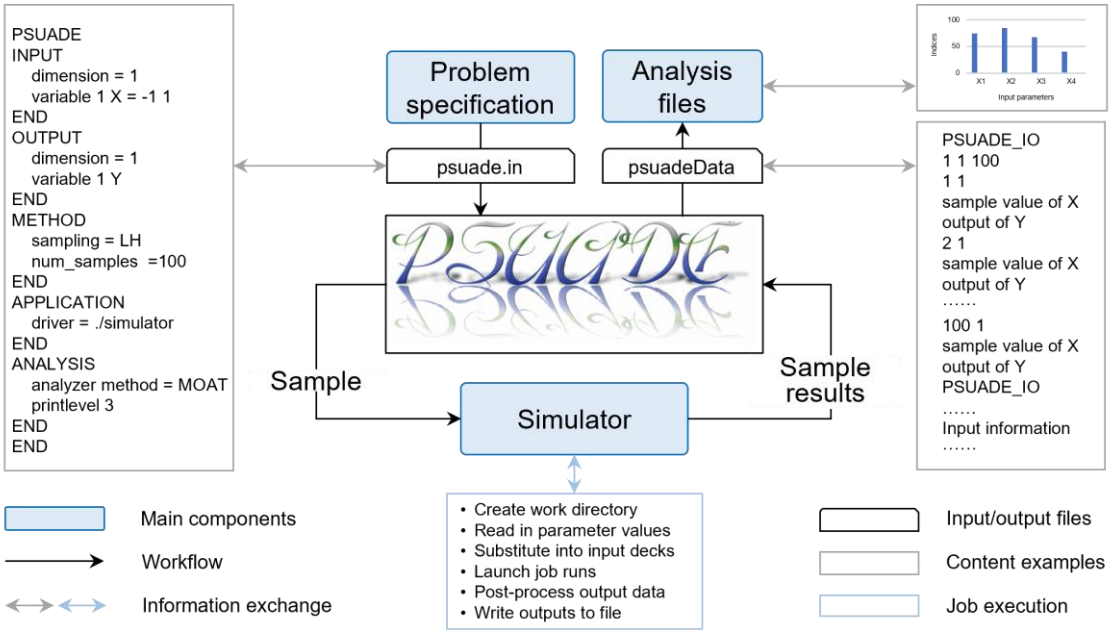


Figure 3: Execution framework of PSUADE (modified from https://www5.in.tum.de/~neckel/siamuq18_slides_minisymp/2018_PSUADE_Tong.pdf).

2.2.2 CoLM and adjustable parameters identification

CoLM is built upon the foundations of the Bonan LSM (Bonan, 1996), the Biosphere-Atmosphere Transfer Scheme (BATS;) (Dickinson et al., 1993), and the 1994 version of the LSM from the Institute of Atmospheric Physics, Chinese Academy of Sciences (IAP94;) (Dai and Zeng, 1997; Dai et al., 2003). It can be applied in both offline mode and coupled configurations with global or regional climate models (Zhang et al., 2025). Following major updates in 2004 and 2014, the model has advanced to the CoLM2024 (Bai et al., 2024). In the latest version, substantial improvements have been implemented in the model grid structure (Huang et al., 2022; Dai et al., 2019c), vegetation subgrid representation (Dai et al., 2019c), and fundamental datasets (Dai et al., 2019b; Dai et al., 2019d). Process parameterizations have also been substantially improved, including surface energy fluxes (Chen et al., 2024), hydrology (Dai et al., 2019a; Zhu et al., 2017a), and human disturbances (Zhang et al., 2025). CoLM2024 has also been successfully coupled with a general circulation model (GCM;) (Xin et al., 2019).

The computational framework of CoLM is illustrated in Fig. 4. The execution of CoLM typically follows a sequential workflow of making surface data, making initial condition data, and executing the main program. Making surface data involves constructing the model grid and subgrid units, as well as aggregating high-resolution raw data to derive surface properties at the model grid scale. Making initial data defines the model's initial states. The main program performs time integration of key land surface processes, including physical, chemical, biological, and anthropogenic processes. When the model is run in offline mode, atmospheric forcing data must be provided as input. In addition, depending on the model configuration, auxiliary inputs such as aerosol, nitrogen deposition, and ozone data may also be required and are read during model execution. The model outputs generally consist of surface property data, model state variables, and historical output data.

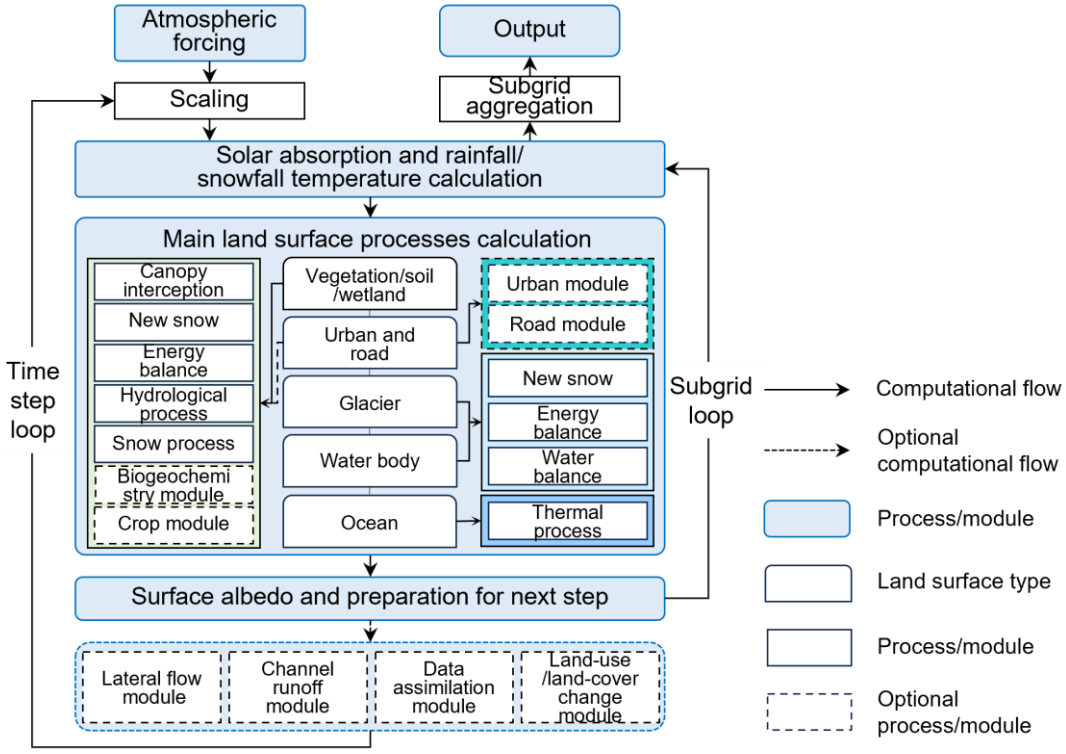


Figure 4: Computational framework of CoLM (modified from https://github.com/CoLM-SYSU/CoLM-doc/).

When coupled with PSUADE, the execution of CoLM is initiated by the simulator specified within PSUADE, which is implemented as a Python script in this study (Wu et al., 2026). Prior to execution, the script replaces the corresponding parameters in the CoLM source files using the sampled parameter values, after which the model is compiled and executed. The selection of adjustable model parameters is primarily guided by their relevance to WTD simulations, with careful consideration of the strong linkages between the simulation process and key hydrological components such as infiltration, evaporation, runoff, freeze-thaw dynamics, and the vertical movement of soil moisture. Accordingly, a total of 56 key parameters associated with these processes were identified. Their physical meanings, categories, value ranges, and the references for determining these ranges are summarized in Table 2. The parameter categories were determined based on the physical processes, with some classifications adapted from Li et al. (2013). The parameter ranges were determined based on the default parameter values implemented in CoLM, the values reported in previous studies, and the need to provide a broad parameter space for analysis. Therefore, the ranges in Table 2 were interpreted as the parameter space explored in this study rather than as universal physically valid limits.

Table 2: Adjustable parameters and their ranges.

NO.	Parameter	Physical meaning	Category	Unit	Range	Ref.
1	ssi	Irreducible water saturation of snow	*Snow*	-	[0.03, 0.07]	(Colbeck, 1974; Li et al., 2013)

2	wimp	A factor for controlling whether water is Impermeable	*Soil*	-	[0.01, 0.1]	(Li et al., 2013)
3	dewmx	Maximum ponding of leaf area	*Canopy*	mm	[0.05, 0.15]	(Li et al., 2013)
4	alpha_rain	Coefficient of interception of rain	*Canopy*	-	[0.25, 1.00]	(Lawrence et al., 2007)
5	all_snow_t	Temperature at which all precip falls entirely as snow	*Precipitation*	℃	[−5, 0]	(Ding et al., 2014)
6	all_rain_t	Temperature at which precip falls entirely as rain	*Precipitation*	℃	[1, 8]	(Ding et al., 2014)
7	cnfac	Crank Nicholson factor between 0 and 1	*Canopy*	-	[0.25, 0.50]	(Li et al., 2013)
8	rhol_vis	Shortwave reflectance of living leaf	*Canopy*	-	[0.070, 0.105]	(Li et al., 2013)
9	rhos_vis	Shortwave reflectance of dead leaf	*Canopy*	-	[0.16, 0.36]	(Li et al., 2013)
10	rhol_nir	Longwave reflectance of living leaf	*Canopy*	-	[0.35, 0.58]	(Li et al., 2013)
11	rhos_nir	Longwave reflectance of dead leaf	*Canopy*	-	[0.39, 0.58]	(Li et al., 2013)
12	taul_vis	Shortwave transmittance of living leaf	*Canopy*	-	[0.04, 0.08]	(Li et al., 2013)
13	taus_vis	Shortwave transmittance of dead leaf	*Canopy*	-	[0.1, 0.3]	(Li et al., 2013)
14	taul_nir	Longwave transmittance of living leaf	*Canopy*	-	[0.1, 0.3]	(Li et al., 2013)
15	taus_nir	Longwave transmittance of dead leaf	*Canopy*	-	[0.3, 0.5]	(Li et al., 2013)
16	chil	Leaf angle distribution factor	*Canopy*	-	[−0.3, 0.1]	(Li et al., 2013)
17	zlnd	Roughness length for soil surface	*Soil*	m	[0.005, 0.015]	(Li et al., 2013)
18	zsno	Roughness length for snow	*Snow*	m	[0.0012, 0.0036]	(Li et al., 2013)
19	htop0	Canopy top height	*Canopy*	m	[20, 40]	(Gao et al., 2003)
20	sqrtdi	Inverse sqrt of leaf dimension	*Canopy*	$m^{-0.5}$	[2.5, 7.5]	(Li et al., 2013)
21	Cd	Leaf drag coefficient	*Canopy*	-	[0.05, 2.00]	(Shaw and Pereira, 1982)
22	cd1	Free parameter	*Canopy*	-	[1, 10]	(Raupach, 1994; Dai et al., 2019d)
23	Cs	Drag coefficient of the substrate surface at height h in the absence of roughness elements	*Canopy*	-	[0.001, 0.003]	(Raupach, 1994, 1992)

24	Cr	Drag coefficient of an isolated roughness element mounted on the surface	*Canopy*	-	[0.25, 0.40]	(Raupach, 1994, 1992)
25	gradm	Slope of conductance-photosynthesis model	*Canopy*	-	[4, 9]	(Li et al., 2013)
26	binter	Intercept of conductance-photosynthesis model	*Canopy*	-	[0.125, 0.375]	(Li et al., 2013)
27	vmax25	Maximum carboxylation rate at 25°C	*Canopy*	-	[1×10^{-5}, 2×10^{-4}]	(Li et al., 2013)
28	effcon	Quantum efficiency of rubp regeneration	*Canopy*	mol CO2 mol quanta^{-1}	[0.035, 0.350]	(Li et al., 2013)
29	shti	Slope of high temperature inhibition function	*Canopy*	-	[0.15, 0.45]	(Li et al., 2013)
30	slti	Slope of low temperature inhibition function	*Canopy*	-	[0.1, 0.3]	(Li et al., 2013)
31	hlti	1/2 point of low temperature inhibition function	*Canopy*	K	[278, 288]	(Li et al., 2013)
32	hhti	1/2 point of high temperature inhibition function	*Canopy*	K	[305, 315]	(Li et al., 2013)
33	trdm	Temperature coefficient of conductance-photosynthesis model	*Canopy*	K	[300, 350]	(Li et al., 2013)
34	trda	Temperature coefficient of conductance-photosynthesis model	*Canopy*	-	[0.65, 1.95]	(Li et al., 2013)
35	tfrz	Freezing temperature	*Soil*	°C	[−0.63, −0.19]	(Pardo Lara et al., 2020)
36	fsatmax	Maximum saturated fractional area	*Runoff*	-	[0.28, 0.46]	(Niu et al., 2005)
37	sur_coef	Exponential-fit coefficient for the topographic index's discrete CDF	*Runoff*	-	[0.1, 1.0]	(Niu et al., 2005)
38	fsatdcf[a]	Runoff decay parameter that determines recession curve	*Runoff*	m^{-1}	[1, 10]	(Niu et al., 2005)
39	sur_imped	Impedance parameter	*Soil*	-	[1, 10]	(Lundin, 1990)
40	rsubmax	Maximum subsurface runoff when the grid cell mean water table depth is zero	*Runoff*	mm s^{-1}	[1×10^{-5}, 5×10^{-3}]	(Niu et al., 2005)
41	sub_alpha	Adjustable scale-dependent parameter	*Soil*	-	[1, 10]	(Niu and Yang, 2006; Swenson and Lawrence, 2012)
42	soil_theta_s[b]	Soil saturated water content	*Soil*	mm^3 mm^{-3}	[0.25, 0.75]	(Li et al., 2013)
43	soil_k_s[c]	Hydraulic conductivity at saturation	*Soil*	cm day^{-1}	[8.64, 8640.00]	(Li et al., 2013)

44	kmax_sun0[d]	Maximum sun branch and shaded branch conductance; stem conductivity	*Canopy*	s^{-1}; m s^{-1}	[1×10^{-8}, 9×10^{-8}]	(Kennedy et al., 2019)
45	krmax	Maximum root conductivity	*Canopy*	m s^{-1}	[1×10^{-9}, 9×10^{-8}]	(Kennedy et al., 2019)
46	ck0	Shape-fitting parameter for vulnerability curve	*Canopy*	-	[1, 5]	(Neufeld et al., 1992)
47	psi50_sun0	Water potential at 50% loss of sunlit/shaded leaf tissue conductance	*Canopy*	mmH_20	[-8.6×10^{5}, -6.6×10^{4}]	(Choat et al., 2012; He et al., 2025)
48	d50	Depth at 50% roots	*Canopy*	cm	[10, 30]	(Schenk and Jackson, 2002; Fan et al., 2016)
49	beta	Dimensionless Shape-parameter	*Canopy*	-	[−3, −1]	(Fan et al., 2016; Schenk and Jackson, 2002)
50	capr	Tuning factor to turn first layer T into surface T	*Soil*	-	[0.17, 0.51]	(Li et al., 2013)
51	soil_psi_s[e]	Saturated capillary potential	*Soil*	mm	[50, 500]	(Li et al., 2013)
52	soil_theta_r	Residual soil moisture	*Soil*	mm^3 mm^{-3}	[0.01, 0.10]	(Ippisch et al., 2006; Schaap and Van Genuchten, 2006)
53	soil_n_vgm	Dimensionless coefficient that characterizes the shape of the retention curve	*Soil*	-	[1.2, 10]	(Ippisch et al., 2006; Tian et al., 2018)
54	k_accum	Accumulation parameter	*Snow*	-	[0.02, 0.26]	(Swenson and Lawrence, 2012)
55	m	Melting factor	*Snow*	-	[1.0, 2.0]	(Swenson and Lawrence, 2012)
56	soil_alpha_vgm	Inverse of the air entry value	*Soil*	mm^{-1}	[0.0001, 0.0100]	(Tian et al., 2018; Ippisch et al., 2006)

[a]Parameter 38 was applied to replace the decay factors in both surface runoff and subsurface runoff calculations. [b]Parameter 42 was also used to update porosity values. [c]For parameter 43, unit conversion was performed by converting the original values reported in the literature from mm s^{-1} ~~mm/s~~ to cm day^{-1} ~~cm/day~~; [d]Parameter 44 simultaneously updates the maximum conductance of sun branch, shaded branch, and stems. [e]Parameter 51 is prescribed as a fixed small value in the new model version. Variations in this parameter have negligible effects on model outputs and are therefore serve as an indirect check within the sensitivity analysis.

The key parameterization schemes adopted in CoLM, corresponding to the selection of adjustable parameters, are summarized in Table 3. The SIMTOP runoff scheme was used because it is well suited for large-scale runoff generation (Decker, 2015), thereby facilitating future regional-scale applications. For freeze-thaw processes, the supercooled-water scheme newly implemented in CoLM2024 was activated, while all other parameterizations remained consistent with default settings.

Table 3: Summary of key parameterization schemes configured in CoLM.

Parameterization schemes	Configuration
DEF_Interception_scheme	1: CoLM

DEF_USE_SUPERCOOL_WATER	true
DEF_Runoff_SCHEME	0: scheme from SIMTOP model
DEF_SPLIT_SOILSNOW	false
DEF_VEG_SNOW	false
DEF_USE_VariablySaturatedFlow	true
DEF_USE_PLANTHYDRAULICS	true
DEF_RSS_SCHEME	1: SL14, Swenson and Lawrence
DEF_THERMAL_CONDUCTIVITY_SCHEME	4: Balland and Arp

In the simulations, CoLM was operated at a 30-min temporal resolution on a water-year basis, where a water year is defined from 1 October to 30 September of the following year. Because the TPMFD is provided in coordinated universal time (UTC), the water-year configuration was also specified in UTC time. For simulations spanning both 2005–2007 and 2005–2014, the 2005–2006 water year was consistently repeated 10 times as a spin-up period to ensure the model reached dynamic equilibrium. The subsequent years (i.e. 2006–2007 or 2006–2014) were then used for model evaluation and analysis. The spin-up behaviour of WTD is illustrated in Fig. 5. The spin-up period of 10 years (n = 10) satisfies the criterion that the difference in annual mean WTD between two consecutive spin-up years is less than 0.1% of the annual mean (Hu et al., 2023), as expressed in Eq. (1):

$$|Var^{n+1} - Var^{n}| < 0.001 \cdot |Var^{n}|, \quad (1)$$

where n is the year for spin-up, Var is the variable (i.e. WTD), and the spin-up values are the ratios of the variable difference to the annual mean.

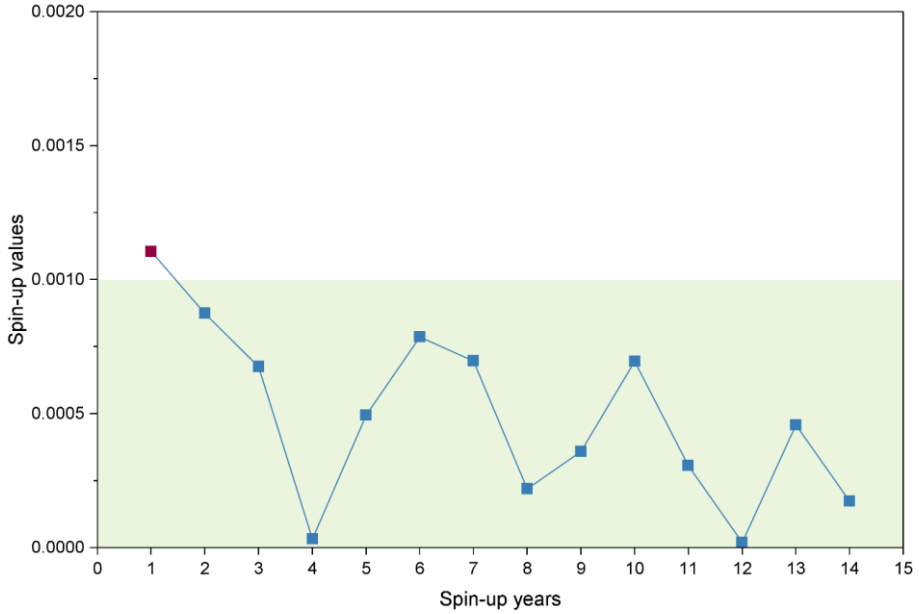
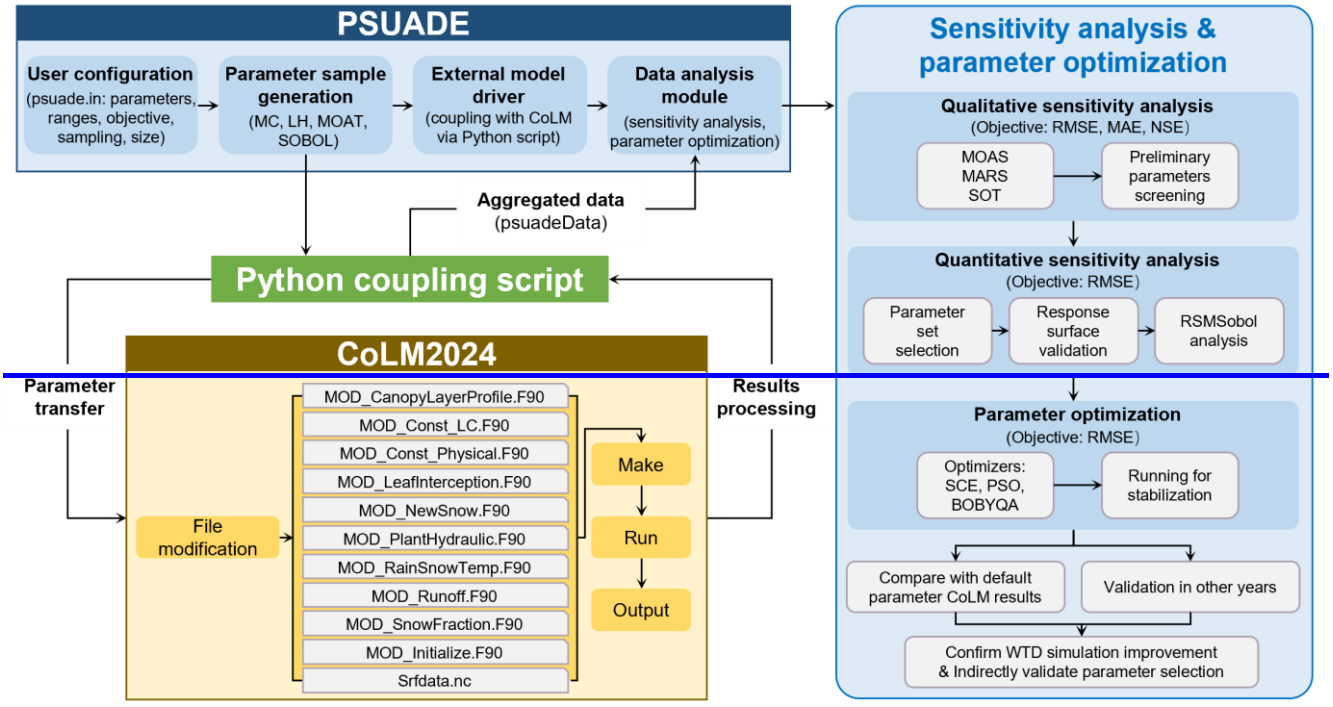


Figure 5: The model spin-up process for WTD, which pass the spin-up test when the curves fall into the green areas.

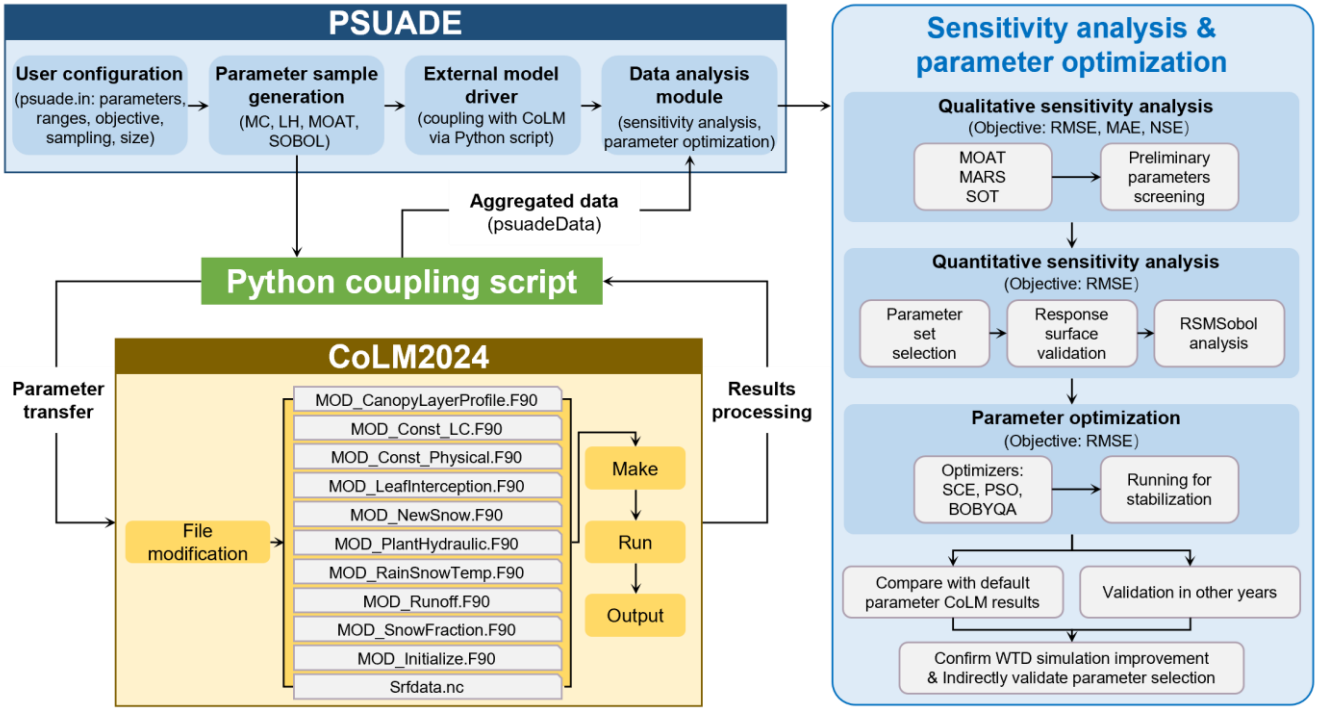
2.2.3 CoLM-PSUADE coupling framework

Coupling between CoLM and PSUADE was implemented by a Python-based linking script (Wu et al., 2026). Within this coupling framework, PSUADE is employed to randomly generate input parameter sets, control the execution of CoLM simulations through the linking script, and perform subsequent data analysis. CoLM is used to simulate the corresponding WTD for each input parameter set. The linking script served as an interface to transfer input parameters from PSUADE to CoLM and return analysis metrics calculated from the CoLM simulation results back to PSUADE.

The detailed procedure is shown in Fig. 6. Firstly, the PSUADE input file (psuade.in) is configured to include key settings (such as the number of parameters, their ranges, the objective function, the sampling method, and the sample size) and the designation of a driver script for coupling with the external model (i.e. CoLM). Secondly, a custom script is developed to sequentially perform a series of tasks, including reading the sampled values of the 56 parameters, replacing the corresponding default values in the CoLM source files, compiling and executing CoLM, post-processing simulation outputs to calculate objective function values, and writing the aggregated results to the PSUADE output file (psuadeData). This procedure is repeated iteratively until the number of runs reaches the prescribed sample size. Subsequently, the built-in statistical methods in PSUADE are employed to analyze the relationships between parameter combinations and the evaluation metrics. Finally, the internal data analysis modules of PSUADE enable further investigations, such as statistical characterization of input-output relationships, sensitivity analysis, response surface construction, and parameter optimization.


PSUADE
User configuration
(psuade.in: parameters, ranges, objective, sampling, size)
Parameter sample generation
(MC, LH, MOAT, SOBOL)
External model driver
(coupling with CoLM via Python script)
Data analysis module
(sensitivity analysis, parameter optimization)
Aggregated data
(psuadeData)
Python coupling script
CoLM2024
Parameter transfer
Results processing
File modification
MOD_CanopyLayerProfile.F90
MOD_Const_LC.F90
MOD_Const_Physical.F90
MOD_LeafInterception.F90
MOD_NewSnow.F90
MOD_PlantHydraulic.F90
MOD_RainSnowTemp.F90
MOD_Runoff.F90
MOD_SnowFraction.F90
MOD_Initialize.F90
Srfdata.nc
Make
Run
Output
Sensitivity analysis & parameter optimization
Qualitative sensitivity analysis
(Objective: RMSE, MAE, NSE)
MOAS
MARS
SOT
Preliminary parameters screening
Quantitative sensitivity analysis
(Objective: RMSE)
Parameter set selection
Response surface validation
RSMSobol analysis
Parameter optimization
(Objective: RMSE)
Optimizers: SCE, PSO, BOBYQA
Running for stabilization
Compare with default parameter CoLM results
Validation in other years
Confirm WTD simulation improvement & Indirectly validate parameter selection
Major component of framework
Process/method
Modified file of CoLM
Computational flow

246


PSUADE
User configuration
(psuade.in: parameters, ranges, objective, sampling, size)
Parameter sample generation
(MC, LH, MOAT, SOBOL)
External model driver
(coupling with CoLM via Python script)
Data analysis module
(sensitivity analysis, parameter optimization)
Aggregated data
(psuadeData)
Python coupling script
CoLM2024
Parameter transfer
Results processing
File modification
MOD_CanopyLayerProfile.F90
MOD_Const_LC.F90
MOD_Const_Physical.F90
MOD_LeafInterception.F90
MOD_NewSnow.F90
MOD_PlantHydraulic.F90
MOD_RainSnowTemp.F90
MOD_Runoff.F90
MOD_SnowFraction.F90
MOD_Initialize.F90
Srfdata.nc
Make
Run
Output
Sensitivity analysis & parameter optimization
Qualitative sensitivity analysis
(Objective: RMSE, MAE, NSE)
MOAT
MARS
SOT
Preliminary parameters screening
Quantitative sensitivity analysis
(Objective: RMSE)
Parameter set selection
Response surface validation
RSMSobol analysis
Parameter optimization
(Objective: RMSE)
Optimizers: SCE, PSO, BOBYQA
Running for stabilization
Compare with default parameter CoLM results
Validation in other years
Confirm WTD simulation improvement & Indirectly validate parameter selection
Major component of framework
Process/method
Modified file of CoLM
Computational flow

247

Figure 6: Coupling framework of CoLM-PSUADE.

2.3 Sensitivity analysis and parameter optimization

2.3.1 Sensitivity analysis methods

Sensitivity analysis often combines multiple analytical methods to screen sensitive parameters, particularly during the qualitative screening stage (Gan et al., 2014; Li et al., 2013). Previous studies have shown that Multivariate Adaptive Regression Splines (MARS) method performs better in identifying insensitive parameters, whereas the Sum-of-Trees (SOT) method is more effective in detecting highly sensitive parameters (Gan et al., 2014). Generally, MARS, SOT and Morris One-At-a-Time (MOAT) method yield comparable screening results (Li et al., 2013). Specifically, the MOAT method evaluates not only the overall importance of input parameters but also their interaction effects (Gan et al., 2014). According to the PSUADE handbook, MOAT is effective when the number of input parameters ranges from 10 to 100. The SOT method performs well with large random sample sizes, while MARS is a low-cost response surface modeling approach. Based on these considerations, this study selected MOAT (Morris, 1991), SOT (Breiman et al., 1984), and MARS (Friedman, 1991) as global sensitivity analysis methods to qualitatively screen 56 parameters. To enable comparison among different methods, all sensitivity indices were normalized. The consistent identification of sensitive parameters among these methods was considered as stronger evidence of parameter importance than the results obtained from a single method.

Following qualitative sensitivity screening, quantitative sensitivity analysis should be performed on the reduced parameter set to assess parameter sensitivity by computing the impact of the parameter on the total variance of model outputs. The Sobol method is often used for this purpose by providing robust variance-based sensitivity measures (Sobol, 1990; Sobol, 2001). However, it requires a large sample size for high-dimensional problems, leading to high computational cost. To reduce the number of model runs, Sobol analysis can be conducted on a surrogate response surface model. Once the performance of the response surface model is validated, the response surface model-based Sobol method (RSMSobol) can be applied efficiently (Li et al., 2013). Therefore, RSMSobol was employed to validate the results of the qualitative screening in this study. Specifically, the Gaussian Process (GP) response surface model was adopted following the recommendations in the PSUADE and previous studies demonstrating its superior performance in approximating complex model responses (Gong et al., 2015).

A comprehensive description of these methodologies, including their theoretical principles, min-max normalization, and response surface validation metrics, is provided in Appendix A.

2.3.2 Sampling methods and sample size

Based on the work of Li et al. (2014) and Gan et al. (2014), Monte Carlo (MC) and Latin Hypercube (LH) sampling were adopted for the SOT and MARS methods, respectively. In addition, the MOAT and SOBOL methods utilized their inherent sampling schemes, respectively. Regarding sample size, Li et al. (2014) concluded that sensitive parameters can be

effectively screened when the sample size is 10 times the total number of parameters. Therefore, four sampling sizes, corresponding to 5, 10, 15, and 20 times the total number of parameters, were adopted to evaluate the qualitative sensitivity analysis results comprehensively. For quantitative sensitivity analysis, model runs were set between 20 and 150 times the number of different input dimensions to align with the approximately 1000–3000 runs reported in the studies (Li et al., 2013; Gan et al., 2014; Guerrero et al., 2017). Notably, unlike MARS and SOT, the MOAT and SOBOL methods require (n+1) and (n+2) runs per sample, respectively, where n represents the number of parameters. The detailed experimental design is summarized in Table 4.

Table 4: Experimental design of the sensitivity analysis methods.

Sensitivity analysis method	Sampling technique	Sample size
MOAT	MOAT	285, 570, 855, 1140
MARS	LH, MC	280, 560, 840, 1120
SOT	LH, MC	280, 560, 840, 1120
SOBOL	SOBOL	Depends on the input parameter dimension; see Sect. 3.2 for details.

All experiments were run on a workstation with a 128-cores AMD EPYC 7C13 processor (2.0 GHz base clock) and 64 GB DDR4 RAM. Each sampling run typically requires approximately 30–40 minutes.

2.3.3 Objective function

The root mean square error ($RMSE$), mean absolute error (MAE) and Nash-Sutcliffe Efficiency (NSE) (Duc and Sawada, 2023) of the simulated WTD were adopted as objective functions for analysis. The corresponding equations (Eq. (2)–(5)) are given as follows.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(O_i - S_i)^2}\,, \tag{2}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|O_i - S_i|, \tag{3}$$

$$NSE = 1 - \frac{\sum_{i=1}^{N}(O_i - S_i)^2}{\sum_{i=1}^{N}(O_i - \bar{O})^2}, \tag{4}$$

$$\bar{O} = \frac{1}{N}\sum_{i=1}^{N} O_i, \tag{5}$$

where N is the number of data points, O_i and S_i denote the observed and simulated values at the i-th time step, respectively, and $\bar{O}$ is the mean of observed values.

When calculating these evaluation metrics, model outputs originally recorded in Coordinated Universal Time (UTC) are first converted to local time, after which daily mean values are computed. The resulting daily averages were then temporally aligned with the observations, and the evaluation metrics are finally calculated based on the aligned data.

2.3.4 Parameter optimization

In this study, parameter optimization was performed on the previously screened sensitive parameters to reduce model complexity and mitigate computational costs (Gong et al., 2015). This sequential design was used not only to estimate reasonable parameter combinations but also to provide an independent evaluation of the effectiveness of the selected sensitive parameters. Specifically, after identifying the stable sensitive parameters through qualitative screening and quantitative verification, optimization and validation were conducted to examine whether adjusting these parameters could effectively improve WTD simulations. The optimization methods within the PSUADE framework were selected based on a comprehensive evaluation of the problem types (global or local), variable types (continuous, discrete, or categorical), and range constraints of variables. Consequently, Shuffled Complex Evolution (SCE), Particle Swarm Optimization (PSO), and Bound Optimization BY Quadratic Approximation (BOBYQA) were identified as suitable optimizers for bound-constrained continuous variables, with $RMSE$ serving as the objective function. As the primary goal of the optimization was to demonstrate the validity of the identified sensitive parameters by enhancing WTD simulation accuracy, the comparison was restricted to results obtained within a fixed number of runs. Specifically, three optimization methods were compared within about 500–1000 runs, a duration during which the $RMSE$ reached and maintained a stable state. The SCE and BOBYQA methods were evaluated for 1000 objective-function evaluations, whereas the PSO method included only 496 valid evaluations due to an unexpected interruption caused by a power outage at the workstation during the additional experiment. In addition, LH sampling was employed due to its effective stochastic properties, which allow for superior optimization performance even with smaller sample sizes (Gong et al., 2015). This characteristic ensures the feasibility of using the stable minimum $RMSE$ attained within 500–1000 runs as the final evaluation index. Then, the parameter combination corresponding to the lowest $RMSE$ was fed back into the CoLM to generate optimized WTD time series, which were then compared against the baseline results obtained using default parameters. The specific principles of the selected optimization methods are summarized in Appendix B.

3 Results

3.1 Qualitative parameter screening

Qualitative sensitivity scores for 56 parameters across the MOAT, MARS, and SOT methods, obtained using different sampling techniques and sample sizes relative to the $RMSE$ objective function, are shown in Fig. 7. In the figure, longer pointers indicate higher parameter sensitivity. Notably, only the MARS and SOT methods consistently identified the same

highly sensitive parameter (56-soil_alpha) at the sample size of 560. In other cases, results exhibited significant discrepancies in ranking. This variability indicated that sensitivity conclusions depend strongly on the analysis method and sampling design, and therefore should not be inferred from a single method or solely from the highest sampling level. Although previous studies on soil temperature and soil moisture have often adopted sensitivity results obtained at the maximum sample size (Li et al., 2013), such a practice may not be sufficient for WTD, as the present results revealed pronounced inter-method variability that could mask critical sensitivity information. In addition, the quality of the generated samples was quantitatively evaluated using distance metrics, and detailed validation is provided in Sect. S2 of the supplementary material.

Based on a cross-method evaluation, the number of significantly sensitive parameters, represented by prominently longer pointers in the radar chart, ranged from 1 to 16. Comparing different sampling techniques (column 2 vs 3 and column 4 vs 5) revealed that MC sampling tended to identify a broader set of sensitive parameters than LH sampling at equivalent sample sizes. Furthermore, the sensitive parameters identified using MAE (Fig. S2) and NSE (Fig. S3) as objective functions were largely consistent with those based on $RMSE$, except for the MARS-LH-280 and MARS-MC-280 cases. This consistency indicated that the dominant sensitivity structure was largely preserved across different objective functions.

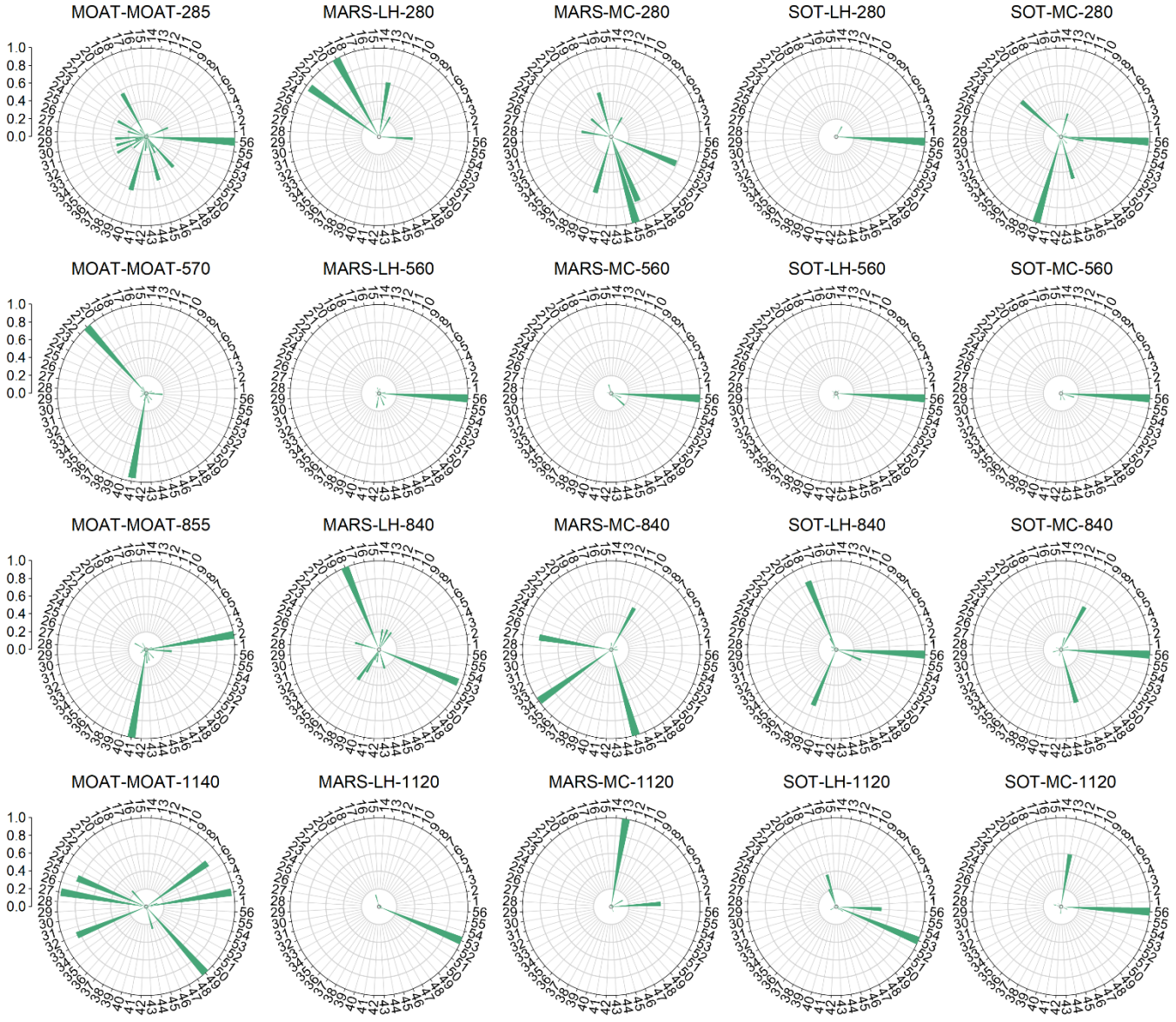


Figure 7: Qualitative sensitivity analysis results based on the *RMSE* objective function. The numbers on the outer ring represent parameter indices (refer to Table 2 for the corresponding parameter definitions). The length of each pointer indicates the normalized sensitivity index, ranging from 0 to 1, with longer pointers denoting higher sensitivity. Each subplot title follows the format "sensitivity analysis method-sampling technique-sample size".

The results were further compared using a heat map (Fig. 8). As the sample size increased, the sensitivities of several parameters changed substantially. Compared with the case where the sample size was five times the total number of parameters, the sensitivities of parameters 1-ssi, 2-wimp, 6-allrain_t, 13-taus_vis, 17-zlnd, 25-gradm, 27-vmax25, 33-trdm, 49-beta, and 53-soil_n increased markedly, whereas those of parameters 22-cd1 and 40-rsubmax decreased. For most parameters, sensitivity exhibited unstable and non-monotonic responses to increasing sample size. In contrast, parameter 56-

soil_alpha consistently remained highly sensitive across different sample sizes, with normalized sensitivity values exceeding 0.5 in most cases. Overall, parameter sensitivities varied irregularly with sample size, further highlighting the necessity of jointly analyzing results from multiple experiment designs to robustly identify stably high-sensitivity parameters.

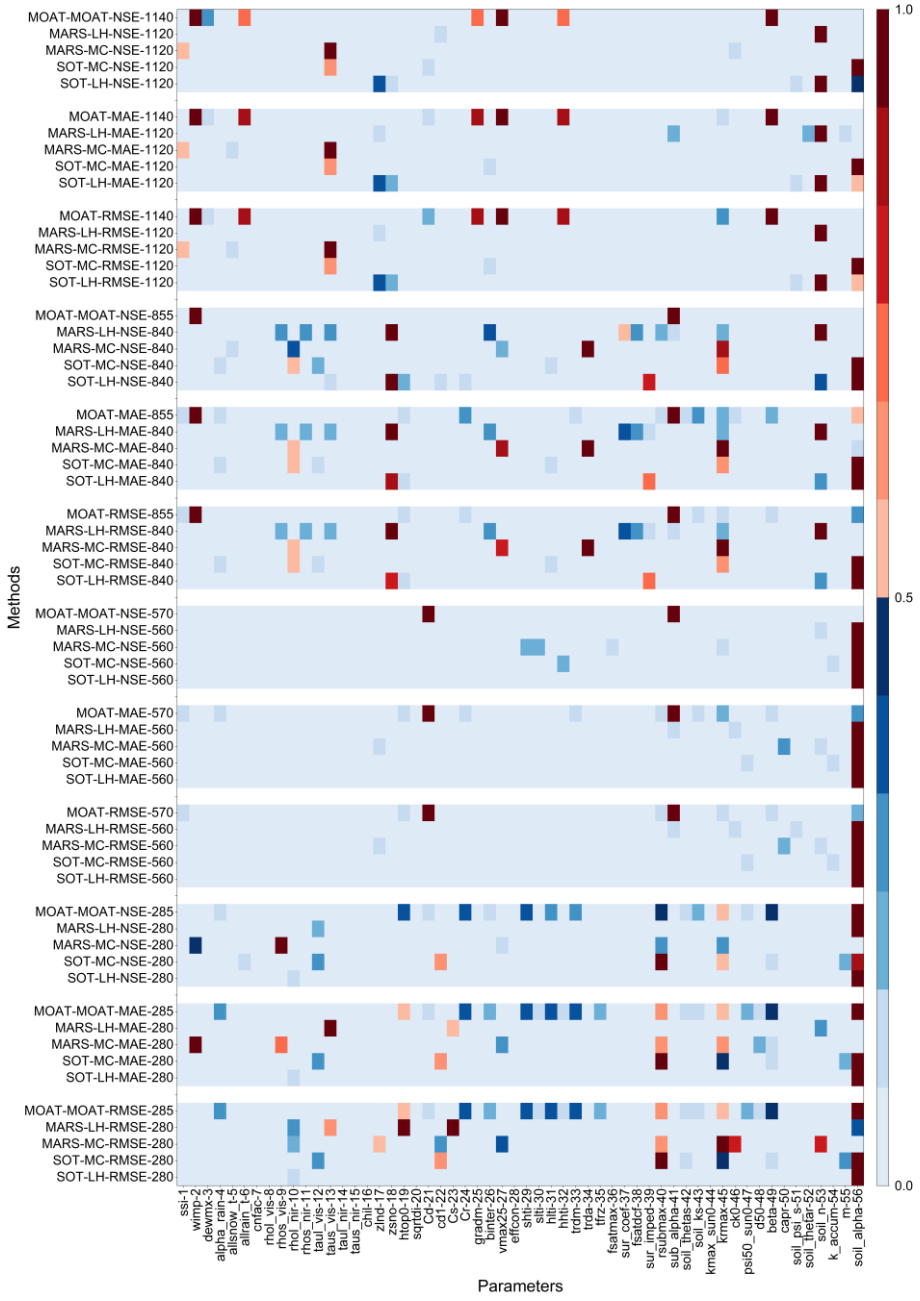


Figure 8: Normalized qualitative sensitivity analysis results (where 1 and 0 represent the most and least sensitive, respectively). The horizontal axis denotes parameter indices, while the vertical axis represents different analysis methods and sampling techniques, following the format "sensitivity analysis method-sampling method-objective function-sample size".

To ensure a comprehensive sensitivity analysis while accounting for variations in methods, objective functions, and sample sizes, sensitive parameters were selected from two complementary perspectives: (1) integrating different analysis methods and sample sizes under a fixed objective function, and (2) integrating different objective functions and sample sizes under a fixed analysis method.

Taking the first perspective as an example, the screening procedure is described as follows. Based on the earlier radar chart results, the number of notably sensitive parameters ranged from 1 to 16. For each combination of analysis method and sample size, the top 15 parameters in the sensitivity ranking were included in an initial candidate set. These parameters were then sorted primarily by their frequency of occurrence across all combinations, with higher frequency ranked first. In cases of equal frequency, their sensitivity rankings across different results were compared, and the final order was determined by the cumulative ranking value in ascending order. The top 10 parameters in the comprehensive ranking were ultimately identified as highly sensitive parameters. The combined set of highly sensitive parameters obtained under different objective functions constituted the final parameter set for this perspective. Parameters classified as highly sensitive under most objective functions were further regarded as stably sensitive parameters. The second perspective followed a similar procedure: with the analysis method fixed, sensitivity rankings from different objective functions and sample sizes were integrated to derive another set of highly sensitive parameters.

From the perspective of fixed objective functions (Fig. 9), parameter 56-soil_alpha was consistently identified as the most sensitive parameter across all objective functions, demonstrating high stability and significance. Within the set of highly sensitive parameters derived by integrating the screening results of the three objective functions (Fig. 9d), aside from 56-soil_alpha, parameters such as 45-krmax, 4-alpha_rain, 53-soil_n, 49-beta, 10-rhol_nir, 40-rsubmax, and 46-ck0 were also consistently classified as highly sensitive across multiple objective functions and are therefore considered as stably sensitive parameters. In contrast, parameters 19-htop0, 42-soil_thetas, 2-wimp, and 1-ssi were identified as highly sensitive under only one or two objective functions and are thus regarded as unstable highly sensitive parameters.

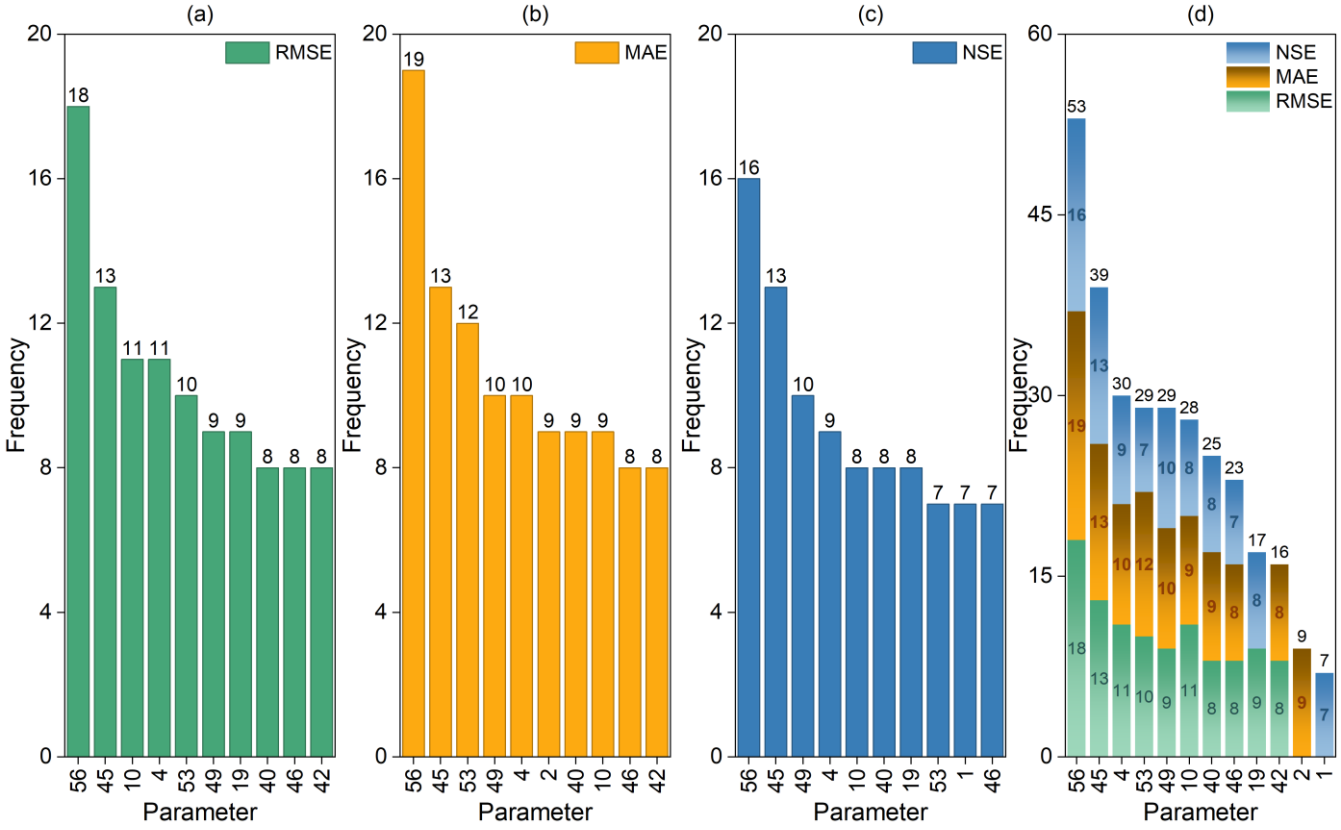


Figure 9: Highly sensitive parameters identified across various methods and sample sizes under different objective functions. The horizontal axis denotes parameter indices, while the vertical axis represents the frequency of occurrence. Labels above the bars indicate the total frequency for each parameter. Subplots (a)–(c) display the 10 most frequently identified sensitive parameters under the *RMSE*, *MAE*, and *NSE* objective functions, respectively. Subplot (d) shows the integrated results of (a)–(c), where labels within the bars indicate the frequency counts for each specific objective function.

From the perspective of fixed analysis methods (Fig. 10), parameter 56-soil_alpha remained the most sensitive parameter. In addition to 56-soil_alpha, 45-krmax was the only parameter identified as the most sensitive across all analysis methods. Although the stably sensitive parameters selected under this perspective differed from those obtained under the previous perspective, parameters 53-soil_n, 10-rhol_nir, 49-beta, 4-alpha_rain, and 40-rsubmax followed 45-krmax in descending order of frequency. Parameter 46-ck0 ranked relatively lower, preceded by parameters 42-soil_thetas, 18-zsno, 24-Cr, and 13-taus_vis.

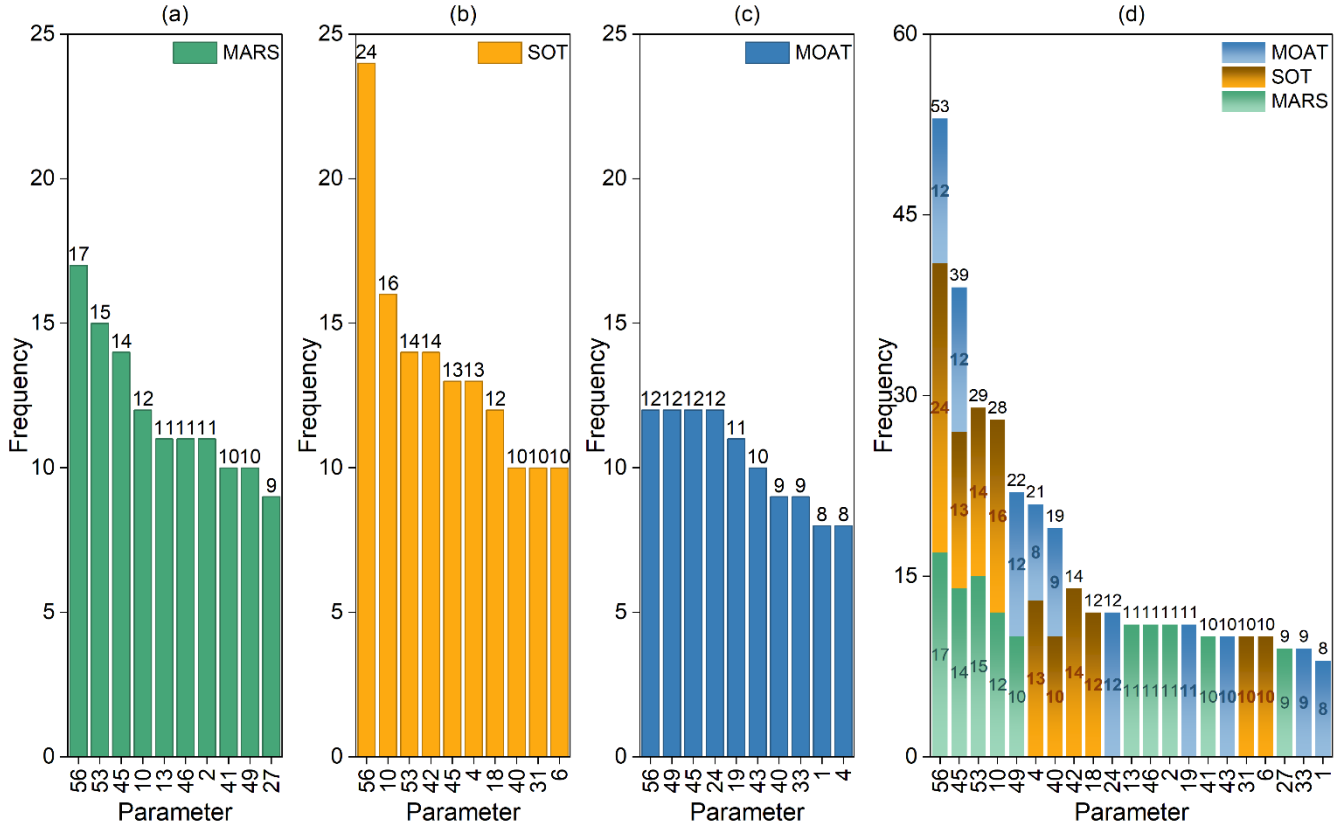


Figure 10: Highly sensitive parameters identified across various objectives and sample sizes under different sensitivity analysis methods. The horizontal axis denotes parameter indices, while the vertical axis represents the frequency of occurrence. Labels above the bars indicate the total frequency for each parameter. Subplots (a)–(c) display the 10 most frequently identified sensitive parameters under the MARS, SOT, and MOAT analysis methods, respectively. Subplot (d) shows the integrated results of (a)–(c), where labels within the bars indicate the frequency counts for each specific analysis method.

Based on the comprehensive ranking results from the two aforementioned perspectives, eight parameters in this screening were identified as stable highly sensitive parameters: 56-soil_alpha, 45-krmax, 4-alpha_rain, 53-soil_n, 49-beta, 10-rhol_nir, 46-ck0, and 40-rsubmax. The remaining highly sensitive parameters were classified as unstable. This process effectively identified key parameters that significantly influence WTD from the total set of 56 parameters. Given that the number of stably sensitive parameters was notably small relative to the total parameter count, the screening procedure demonstrated good applicability for this study.

The aforementioned sensitivity ranking reflected the overall effect of the parameters, while the MOAT method can further reveal the characteristics of parameter interactions. In this approach, the importance of a parameter is indicated by the modified mean (μ_i^*), with larger values indicating higher sensitivity. Meanwhile, the standard deviation (σ_i^*) characterizes the strength of its interactions, where larger values imply stronger interactions with other parameters. Therefore, in the scatter plots of (μ_i^*, σ_i^*), parameters that are clearly separated from the main cluster, typically with large μ_i^* and σ_i^*, can be identified as sensitive. Their distribution further facilitates analysis of the relationship between parameter sensitivity and

interaction strength. Taking the experiment with a sample size of 285 as an example (Fig. 11), parameter 56-soil_alpha exhibited the maximum values for both μ_i^* and σ_i^*, indicating that it was not only the most sensitive parameter but also the one with the strongest interaction effects. For other sensitive parameters, although interaction strength generally increased with sensitivity, certain exceptions were observed. For instance, the sensitivity rankings of parameters 45-krmax, 19-htop0, and 40-rsubmax followed the order 40>19>45 (Fig. 11a), whereas their interaction strengths followed the order 45>19>40. In other words, the most sensitive parameter among them exhibited the weakest interactions, suggesting that higher sensitivity does not necessarily imply stronger interactions. Under other sample sizes (Fig. S4–S6), similar non-monotonic relationships between sensitivity and interaction strength were observed, further confirming the sample dependence of qualitative results and the necessity of comprehensive multi-perspective analysis.

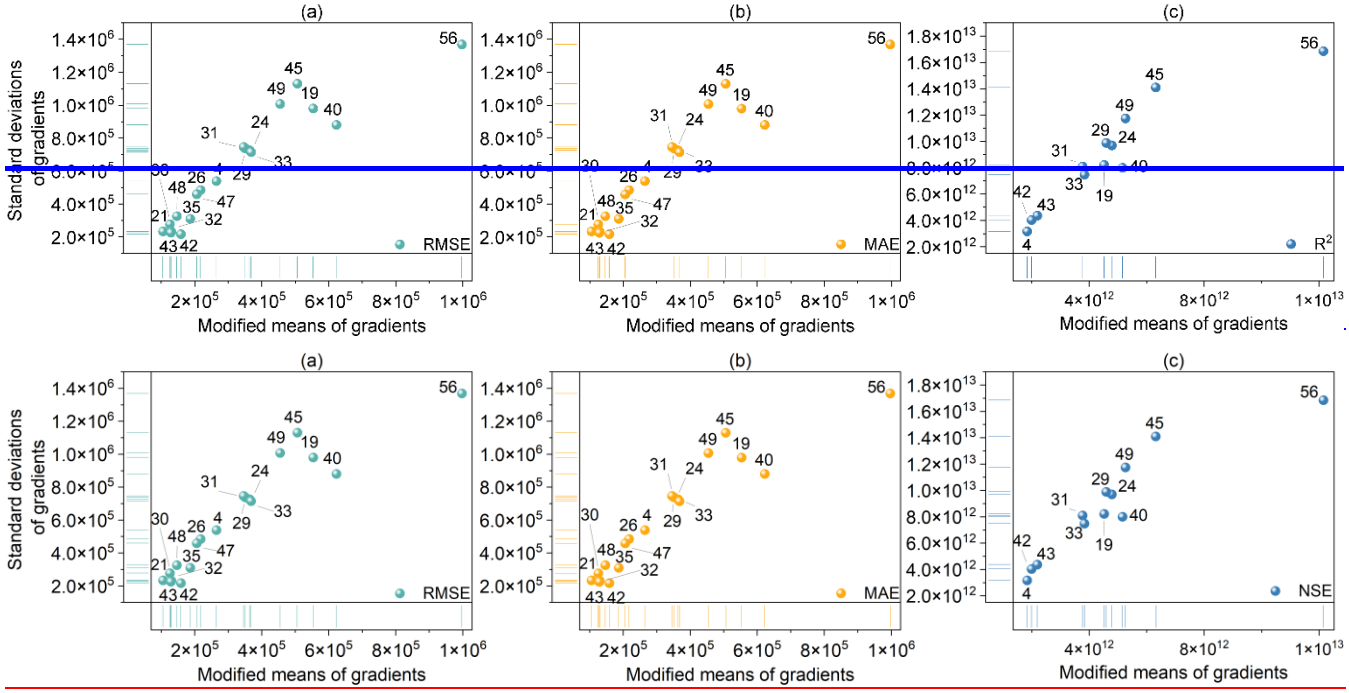


Figure 11: Sensitivity and interaction results from the MOAT method with a sample size of 285. The horizontal axis denotes the modified mean (μ_i^*), and the vertical axis denotes the standard deviation (σ_i^*). Subplots (a)–(c) correspond to the *RMSE*, *MAE*, and *NSE* objective functions, respectively.

3.2 Quantitative validation of key sensitivity parameters

To validate the eight stably sensitive parameters influencing WTD identified through the previous multi-perspective comprehensive screening, the RSMSobol method was employed for quantitative sensitivity analysis. This approach, based on response surfaces, accounted for varying input dimensions and sample sizes. The Sobol' total-order indices (TSI) reflect the overall influence of each parameter on the objective function, where higher TSI values indicate a more significant impact on the output.

Using all 56 parameters as inputs would lead to high dimensionality and substantial computational cost. Moreover, most response surface analysis methods are recommended for input dimensions below 20. Therefore, three parameter sets with different input dimensions (8, 12, and 21) were selected for analysis: the stably sensitive parameters (8 parameters), the sensitive parameters obtained from fixed-objective-function screening (12 parameters), and the sensitive parameters obtained from fixed-analysis-method screening (21 parameters). Detailed experimental configurations are provided in Table 5.

Table 5: Experimental design of the RSMSobol method.

Input dimension	Response surface	Sample size	
		Multiples of input dimension	Model runs
8	GP/logGP	50 times	500
		100 times	1000
		150 times	1500
12	GP/logGP	40 times	560
		80 times	1120
		100 times	1400
21	GP/logGP	20 times	460
		40 times	920
		100 times	2300

Because the earlier qualitative sensitivity analysis showed consistent sensitivity structures across different objective functions, only $RMSE$ was used as the objective function in the RSMSobol quantitative analysis. A GP was used as the response surface model in all experiments. Additionally, due to the wide range of $RMSE$ (Table S2), $RMSE$ was log-transformed (natural logarithm) in PSUADE, and a GP-based response surface was then constructed using the transformed data for quantitative analysis, hereafter referred to as logGP. Detailed validations of response surface models are provided in Sect. S3 of the supplementary material.

In the quantitative sensitivity analysis for an input dimension of 8 (corresponding to the set of stably sensitive parameters; Fig. 12), parameter 56-soil_alpha exhibited a significantly higher TSI than the other parameters, contributing individually 32%–72% to the relative importance. This was consistent with its stable and pronounced sensitivity observed in the qualitative sensitivity analysis. Among the remaining parameters, 53-soil_n, 49-beta, 45-krmax, 40-rsubmax, and 4-alpha_rain also showed relatively high sensitivity (Fig. S10). Across experiments with different sample sizes, the cumulative TSI proportion stabilized after 2–7 parameters, reaching values above 90%. Compared with GP, the cumulative proportion based on logGP tended to stabilize more rapidly.

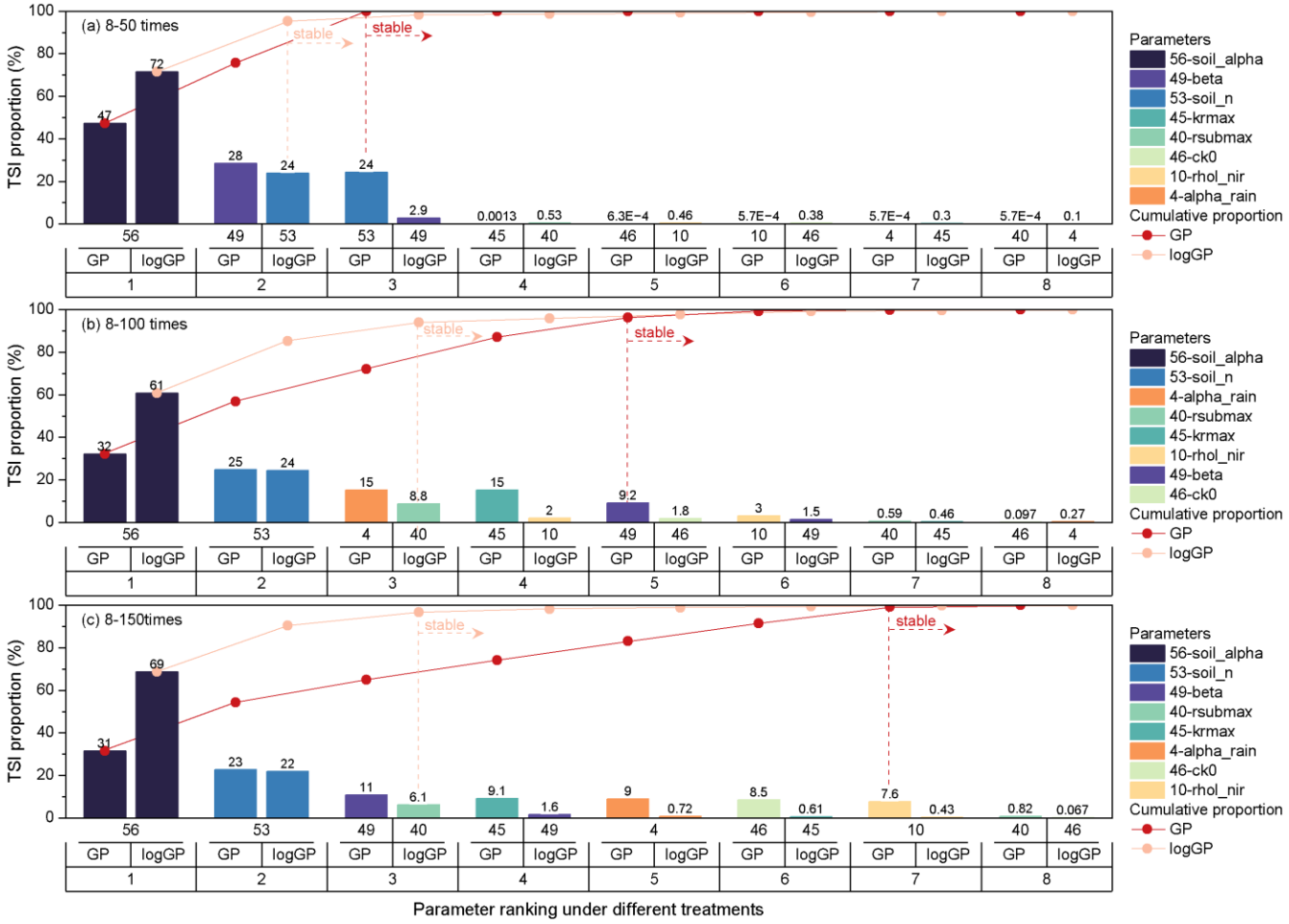


Figure 12: Quantitative sensitivity ranking for an input dimension of 8 (corresponding to the number of stably sensitive parameters). The vertical axis denotes the relative importance based on TSI. The horizontal axis displays the parameter number, response surface method, and ranking number (from top to bottom). The curves represent the cumulative relative importance of parameters, ranked in descending order of importance. Note: "8–k times" in the subplot title denotes an experiment with 8 input parameters and a sample size equal to k times the number of parameters.

When the input dimension is 12 (corresponding to the set of sensitive parameters obtained from fixed-objective-function screening), the stably sensitive parameters 56-soil_alpha, 53-soil_n, 49-beta, 45-krmax, 40-rsubmax, and 4-alpha_rain still retained high relative importance (Fig. 13). Among them, the relative importance of 56-soil_alpha alone reached 41%–57%. In terms of cumulative TSI proportion, the eight stably sensitive parameters accounted for 79%–89% of the total. This result further confirmed the reliability of the qualitative sensitivity screening. Other parameters, such as 1-ssi, 19-htop0, and 42-soil_thetas, which, although not classified as stably sensitive, ranked among the top 8 (Fig. S11), explain the slightly lower cumulative TSI proportion. Notably, among the eight stably sensitive parameters, the cumulative importance showed a clear tendency to level off after the first 3–5 parameters.

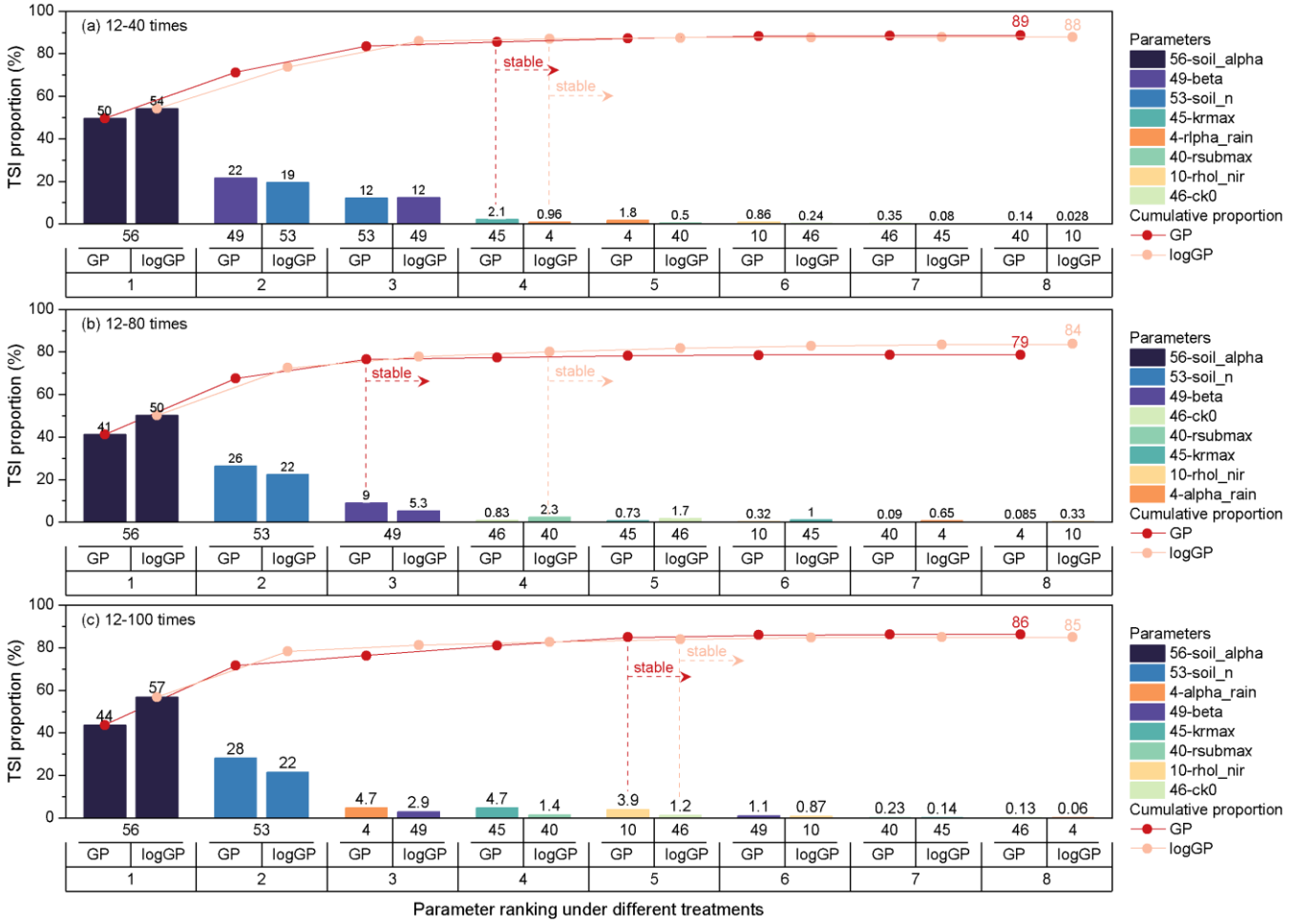


Figure 13: Quantitative sensitivity ranking for an input dimension of 12 (corresponding to the number of sensitivity parameters obtained by screening with fixed objective functions). The vertical axis denotes the relative importance based on TSI. The horizontal axis displays the parameter number, response surface method, and ranking number (from top to bottom). The curves represent the cumulative relative importance of parameters ranked in descending order of importance. Note: "12-k times" in the subplot title denotes an experimental configuration with 12 input parameters and a sample size equal to k times the number of parameters.

When the input dimension was 21 (corresponding to the set of sensitive parameters identified using a fixed screening method), the high-ranking parameters among the eight stably sensitive ones remained concentrated on 56-soil_alpha, 53-soil_n, 49-beta, 45-krmax, 40-rsubmax, and 4-alpha_rain (Fig. 14). Among these, the relative importance of 56-soil_alpha alone accounted for ~~13~~3.3%–60%. The cumulative importance of the eight stably sensitive parameters within the total set of 21 parameters was 36%–77%, which was lower than that obtained for an input dimension of 12. This decrease was mainly due to the relatively higher rankings of other parameters, such as 6-all–rain_t–e, 43-soil_ks, 31-hlti, 42-soil_thetas, 19-htop0, 1-ssi, 24-Cr, 33-trdm, and 27-vmax25 (Fig. S12). Similarly, the cumulative importance of the eight stably sensitive parameters tended to stabilize after the first 2–5 parameters.

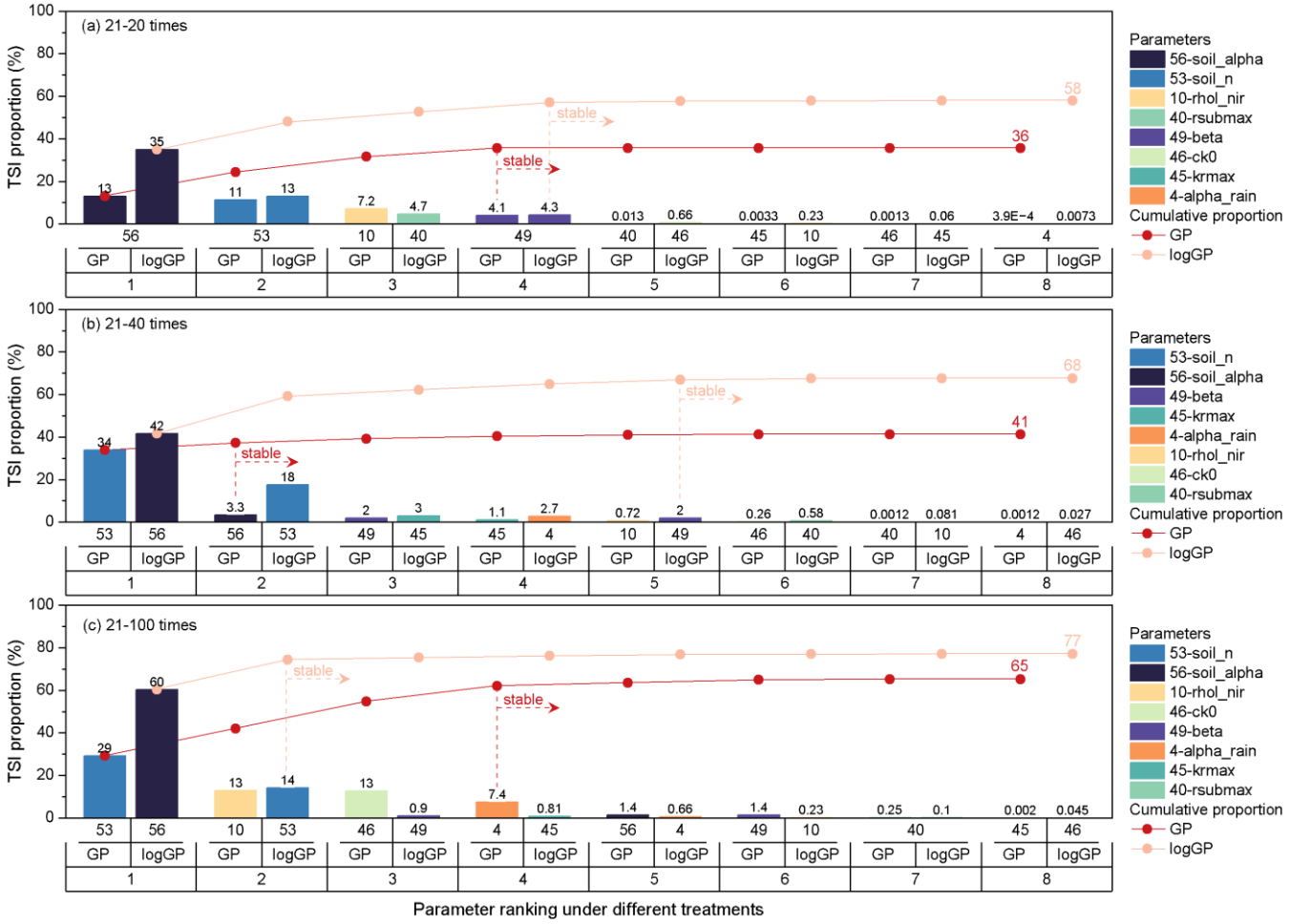


Figure 14: Quantitative sensitivity ranking for an input dimension of 21 (corresponding to the number of sensitivity parameters obtained by screening with fixed screening method). The vertical axis represents the relative importance based on TSI. The horizontal axis displays the parameter number, response surface method, and ranking number (from top to bottom). The curves represent the cumulative relative importance of parameters ranked in descending order of importance. Note: "21-k times" in the subplot title denotes an experimental configuration with 21 input parameters and a sample size equal to k times the number of parameters.

An integrated examination of the quantitative sensitivity analysis results across different input dimensions showed that, as the input dimension increased, the cumulative importance of the eight stably sensitive parameters decreased. This indicated that some non-stably sensitive parameters exerted a non-negligible influence on WTD dynamics in higher-dimensional parameter spaces. Nevertheless, within the stably sensitive parameter set, the importance distribution remained highly concentrated. In particular, 56-soil_alpha, 53-soil_n, 49-beta, 45-krmax, 40-rsubmax, and 4-alpha_rain consistently exhibited dominant contributions across different experimental settings. The subset composed of these six parameters accounted for a cumulative importance close to that of the entire stably sensitive set, indicating that this parameter combination captured the dominant controls on WTD.

3.3 Optimal estimation of key sensitivity parameters

Parameter optimization experiments were performed based on the eight stably sensitive parameters identified in the previous screening to minimize the $RMSE$ of the simulated WTD, aiming to determine the optimal parameter combination and its reasonable value ranges. It should be noted that, within a limited set of 500–1000 optimization runs, the primary purpose of performing parameter optimization was to indirectly validate the effectiveness of the parameter screening process by comparing model performance before and after optimization.

A comparison of the SCE, PSO, and BOBYQA methods revealed clear differences in their convergence characteristics and optimization performance (Fig. 15). In terms of overall behaviour (Fig. 15a), most $RMSE$ values obtained by the SCE method were concentrated between 0.4 and 1 m, with occasional sharp peaks corresponding to failed solutions. The low-$RMSE$ range was stable across evaluations. For the PSO method, $RMSE$ values were primarily clustered between 0.6 and 0.9 m and exhibited evident high-frequency oscillations across evaluations when multiple particles were simultaneously searched. The BOBYQA method showed a markedly different changing pattern. Specifically, during the first 90–140 evaluations, $RMSE$ values fell into the relatively poor range because the change of trust region. However, after approximately 200 evaluations, $RMSE$ remained stable within the range of 0.5–1.0 m. Regarding optimization performance (Fig. 15b), SCE found a better solution during the early stage. As the number of evaluations increased, population of parameter sets evolved in the solution space were progressively constrained to good-quality zones. Ultimately, the SCE method achieved the lowest $RMSE$ value of 0.419 m, demonstrating the best overall performance. Although the $RMSE$ obtained by PSO exhibited considerable variability, its fluctuation range remained within a reasonable interval (0.5–1.2 m), and it was able to identify an optimal solution (0.531 m) within the early stage of the optimization process (around the first 50 evaluations), suggesting relatively fast initial exploration capability. The minimum $RMSE$ obtained by BOBYQA was 0.517 m, lying between the results of the other two methods. The cumulative minimum RMSE curves exhibited relatively stable over the respective evaluation ranges of the three methods, although the comparison was conducted under unequal computational budgets (1000 evaluations for SCE and BOBYQA and 496 valid evaluations for PSO). Overall, SCE outperformed PSO and BOBYQA in terms of the speed of identifying the best solution and the minimum RMSE achieved~~convergence speed and solution quality~~, indicating that it is more suitable for parameter optimization in the CoLM under present experimental configuration.

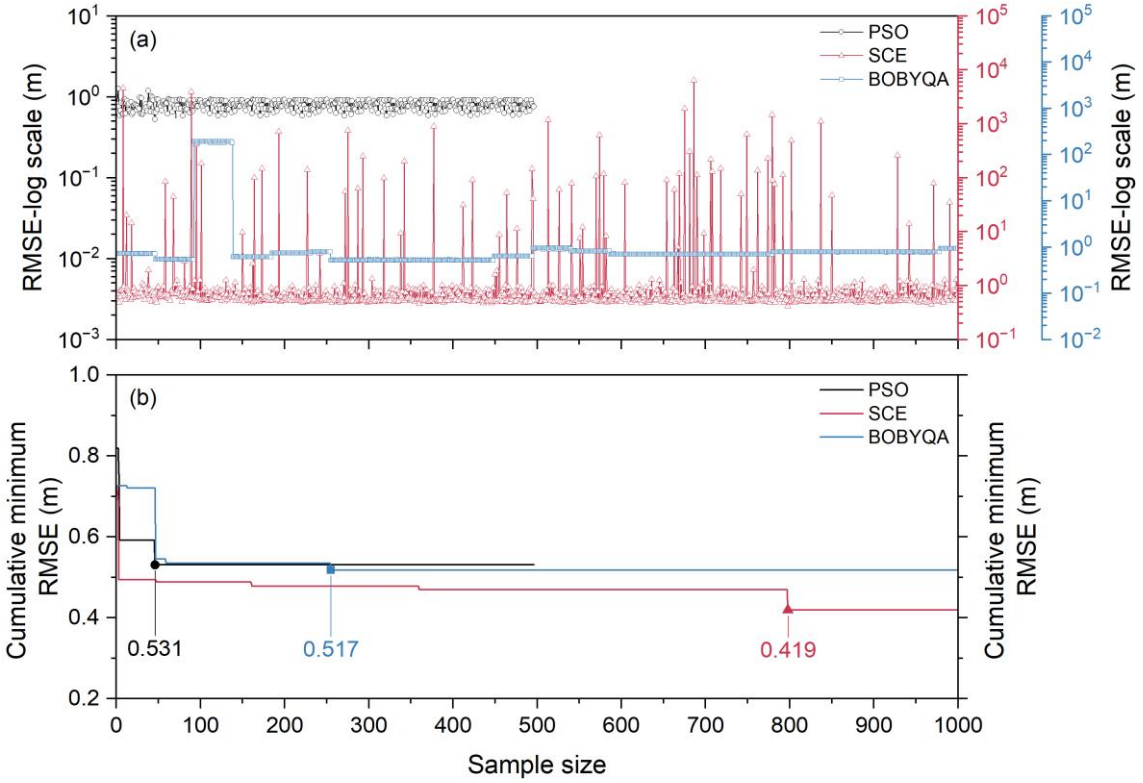


Figure 15. Evolution of (a) $RMSE$ and (b) cumulative minimum $RMSE$ with the number of evaluations for the SCE (1000 runs), PSO (496 runs), and BOBYQA (1000 runs) optimization methods. The vertical axis for subplot (a) is displayed on a logarithmic scale to accommodate the wide range of $RMSE$ values. The vertical axis for subplot (b) represents the lowest $RMSE$ value obtained up to each model evaluation during the optimization process.

Based on the comparison of simulated WTD time series (Fig. 16), the optimal parameter set derived from the SCE method successfully captured the average temporal patterns of the observed WTD. In contrast, the simulations obtained using the PSO and BOBYQA methods showed a systematic overestimation. This result further highlighted differences in parameter search capability among the three optimization methods.

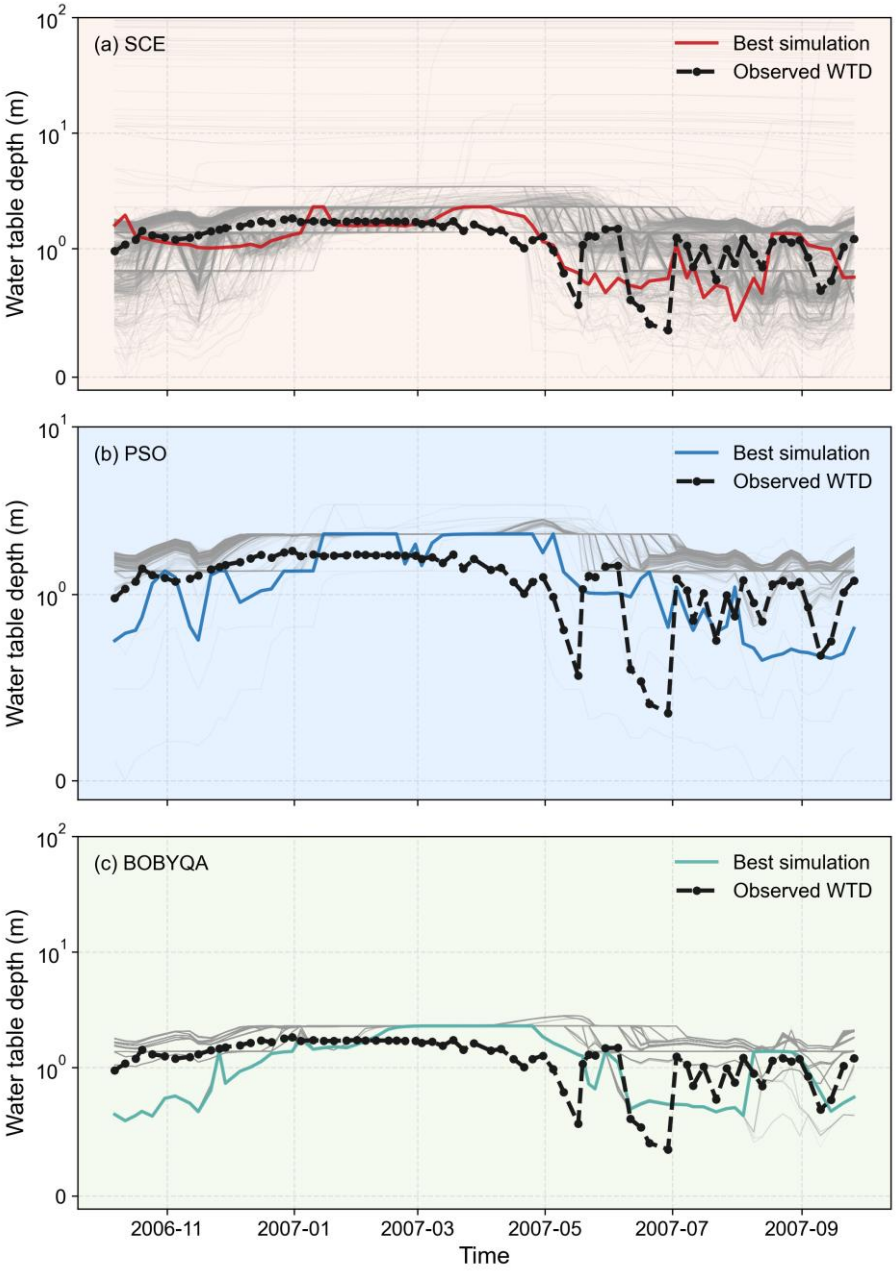


Figure 16: Comparison between observed WTD and simulated time series generated by (a) SCE, (b) PSO, and (c) BOBYQA optimization methods after 1000, 496, 1000 evaluations, respectively. The vertical axis is displayed on a symlog scale to accommodate the wide range of values. The grey lines in the background show the simulation results excluding the best run.

To further analyze the distribution characteristics of parameters associated with favourable solutions, the top 20% of samples ranked by $RMSE$ from low to high (hereafter referred to as low-$RMSE$ samples) and their associated parameter sets were selected for kernel density estimation (KDE) analysis. The KDE results (Fig. 17) showed that, for the SCE method, the

parameter distributions of the low-$RMSE$ samples were markedly more concentrated than those of the full sample set, indicating that optimal solutions were confined to specific parameter ranges. In particular, parameters 40-rsubmax (Fig. 17c) and 56-soil_alpha (Fig. 17h) exhibited the most pronounced convergence. The optimal values of 40-rsubmax were primarily clustered at the order of 10^{-4}, whereas those of 56-soil_alpha were mainly distributed between 10^{-3} to 10^{-2}. In contrast, for the PSO method (Fig. S13), the distributions of favourable solutions showed minimal deviation from those of the full sample set for most parameters, except for the parameter 40-rsubmax, suggesting a relatively limited optimization capability. The BOBYQA method (Fig. S14) failed to sufficiently explore the parameter space within 1000 evaluations.

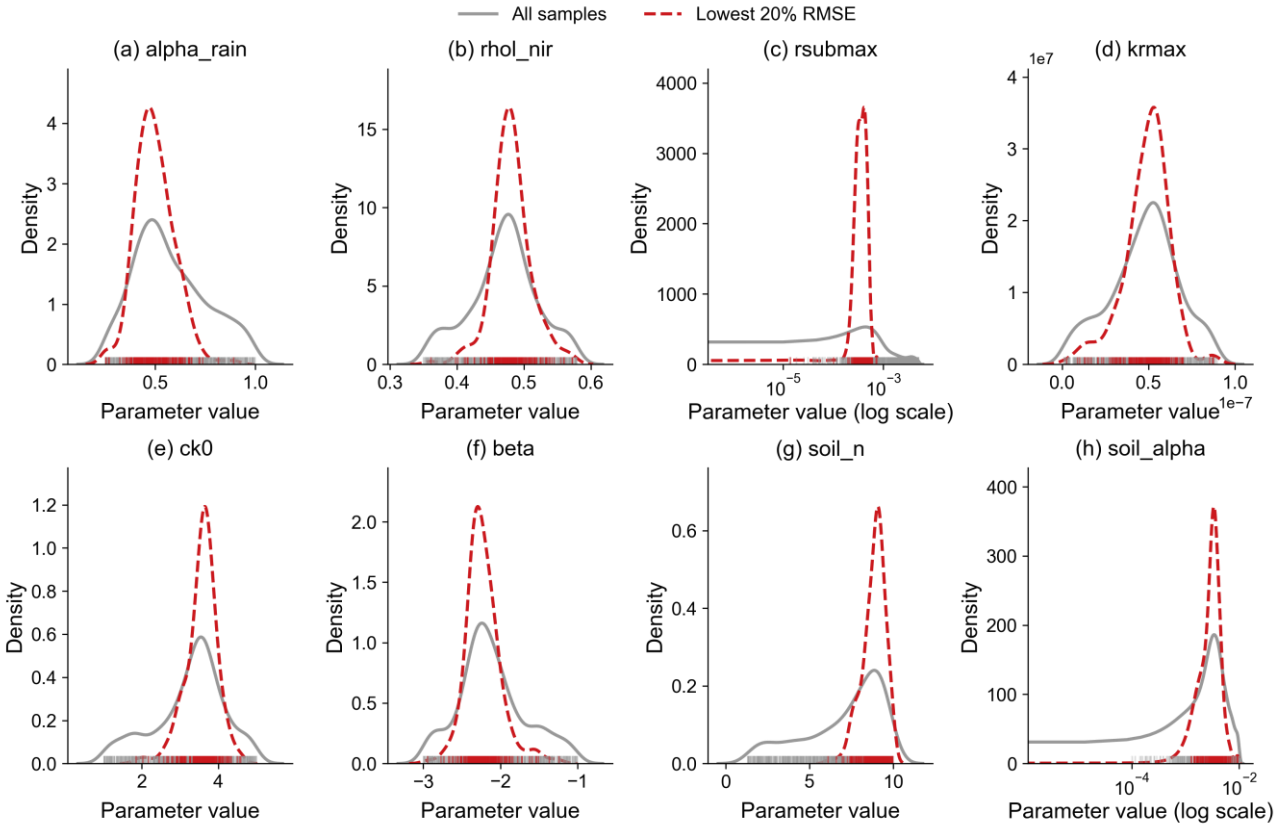


Figure 17: Comparison of KDE for the top 20% low-$RMSE$ parameter sets (red) versus the entire set of 1000 evaluations (grey) obtained by the SCE method: (a) alpha_rain, (b) rhol_nir, (c) rsubmax, (d) krmax, (e) ck0, (f) beta, (g) soil_n, and (h) soil_alpha. Note that the horizontal axes for rsubmax and soil_alpha are plotted on a logarithmic scale to better visualize their distributions.

The scatter plots of $RMSE$ against parameter values further revealed the influence of key parameters on the optimization outcomes. For the SCE method (Fig. 18), parameter 40-rsubmax (Fig. 18c) exhibited a clear V-shaped pattern in its relationship with $RMSE$, suggesting a well-defined optimal region. A similar feature appeared as a distinct branching pattern in the PSO method (Fig. S15). In contrast, no obvious trend was observed for the BOBYQA method (Fig. S16), likely due to its limited exploration of the parameter space. These results demonstrated that WTD simulations in the CoLM were highly

sensitive to parameter 40-rsubmax. Consequently, constraining this parameter to its optimal range prior to fine-tuning the remaining parameters could significantly improve simulation accuracy.

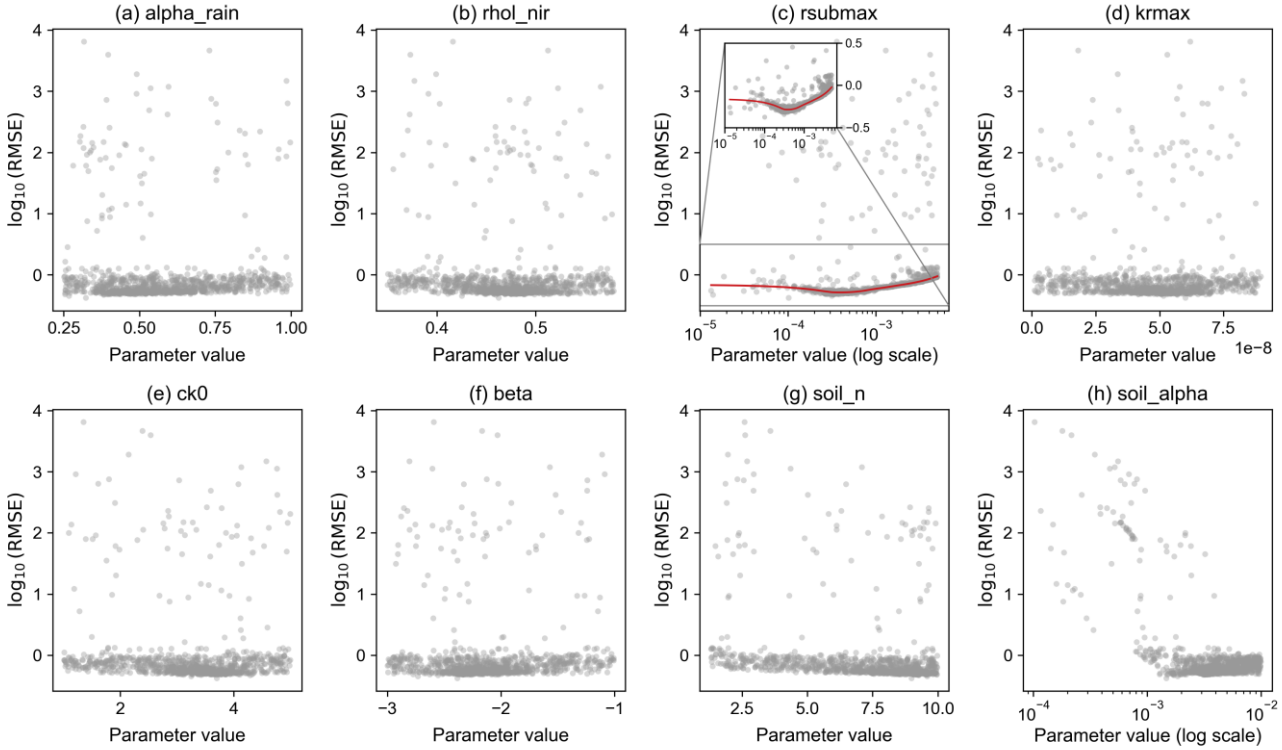


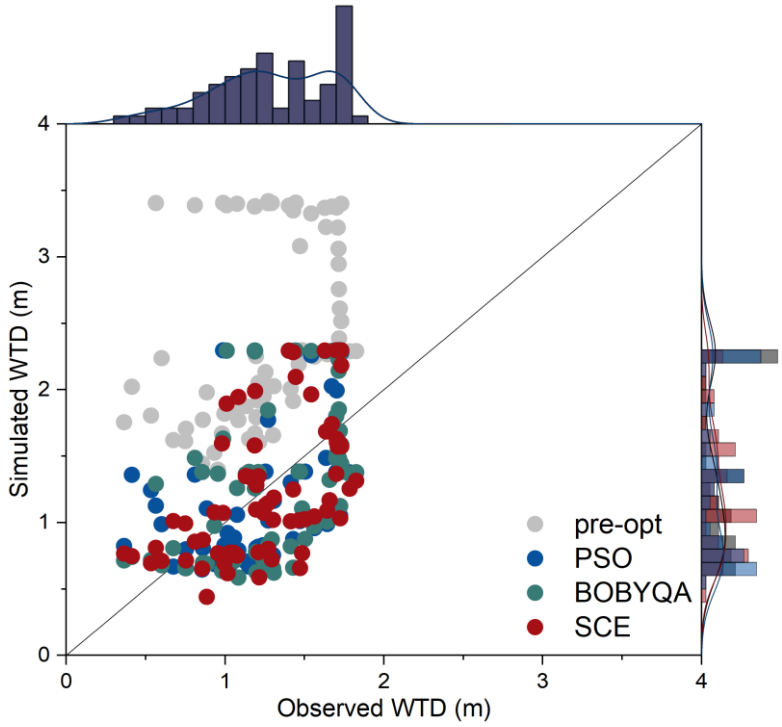
Figure 18: Scatter plots of *RMSE* against parameter values based on 1000 evaluations obtained during the SCE optimization process: (a) alpha_rain, (b) rhol_nir, (c) rsubmax, (d) krmax, (e) ck0, (f) beta, (g) soil_n, and (h) soil_alpha.

Based on the optimal parameter combinations (Table 6), only three of the eight stably sensitive parameters, namely 56-soil_alpha, 45-krmax, and 40-rsubmax, exhibited order-of-magnitude changes relative to their default values, further highlighting their pivotal roles in WTD simulations. The optimal parameter set obtained from the SCE method markedly improved the WTD simulation, reducing the *RMSE* by 68%, with simulated values clustering more closely around the 1:1 diagonal (Fig. 19). The PSO and BOBYQA methods achieved improvements of 59% and 60%, respectively. It is worth noting, however, that pronounced “horizontal bands” were evident in the scatter plots both before and after optimization, where the simulated WTD remained nearly constant while the observed WTD varied. This phenomenon primarily occurred during the freeze-thaw period, indicating that the representation of freeze-thaw processes and their interactions with groundwater in the CoLM warrants further investigation.

572 **Table 6: Comparison of the default and optimized values for the eight stably sensitive parameters.**

Parameter	Default value	Optimized value		
		SCE	PSO	BOBYQA
alpha_rain	2.500×10^{-1}	3.232×10^{-1}	7.170×10^{-1}	3.310×10^{-1}
rhol_nir	3.500×10^{-1}	4.626×10^{-1}	5.245×10^{-1}	4.116×10^{-1}
rsubmax[a]	**5.500×10^{-3}**	**1.020×10^{-4}**	**4.286×10^{-4}**	**2.451×10^{-4}**
krmax	**3.981×10^{-9}**	**6.013×10^{-8}**	**4.185×10^{-8}**	**7.008×10^{-8}**
ck0	3.950	3.701	2.724	2.602
beta	-1.623	-2.383	-2.864	-1.028
soil_n[b~~a~~]	1.255	8.994	9.609	5.556
soil_alpha[b~~a~~]	**2.245×10^{-2}**	**1.475×10^{-3}**	**2.408×10^{-3}**	**1.847×10^{-3}**

573 [a]Default values represent the original parameter settings implemented in CoLM and are provided as the baseline for comparison. The optimized values were obtained
574 by searching within the parameter ranges specified in Table 2. For rsubmax, the CoLM default value (5.500×10^{-3}) slightly exceeds the prescribed upper bound
575 (5.000×10^{-3}), because the latter was manually rounded; [b~~a~~]The original values of the two parameters are different across~~at~~ the 10-layer soil profile. Here, they are
576 represented as the average values.



577

Figure 19: Scatter plots of observed versus simulated WTD using default (before optimization) and optimized parameter sets.

3.4 Validation of the optimized parameters

The preceding analysis indicated that the SCE method achieved the most significant results in parameter optimization. To assess the robustness of the optimized parameters and indirectly validate the effectiveness of the selected parameter set, long-term WTD dynamics at the GGF site from 2005 to 2014 were simulated using the SCE-derived optimal parameter set (i.e. the set corresponding to the lowest $RMSE$).

A comparison of simulation results across different water years (Fig. 20) revealed that the optimal parameter set derived from the 2006–2007 water year reasonably captured the seasonal and interannual temporal trends of WTD in the subsequent years (2007–2014), with $RMSE$ values ranging from 0.42 to 1.13 m (Fig. 21). The optimized simulations reduced the overestimation and achieved better agreement with observations compared to original simulations. However, although the optimization can enhance the overall simulation, there were still relatively lagging responses in capturing the dynamic changes of WTD (Fig. 20a–c) and abrupt changes or deviations from the observations (Fig. 20c, h). Notably, parameter 40-rsubmax exhibited a distinct turning point in the aforementioned parameter– $RMSE$ scatter relationship. When only this parameter was updated to its optimized value while keeping others at their default values, the $RMSE$ ranged from 0.59 to 0.91 m. The overall simulation value of WTD has improved in this situation and there has been no significant change in the dynamics of WTD.

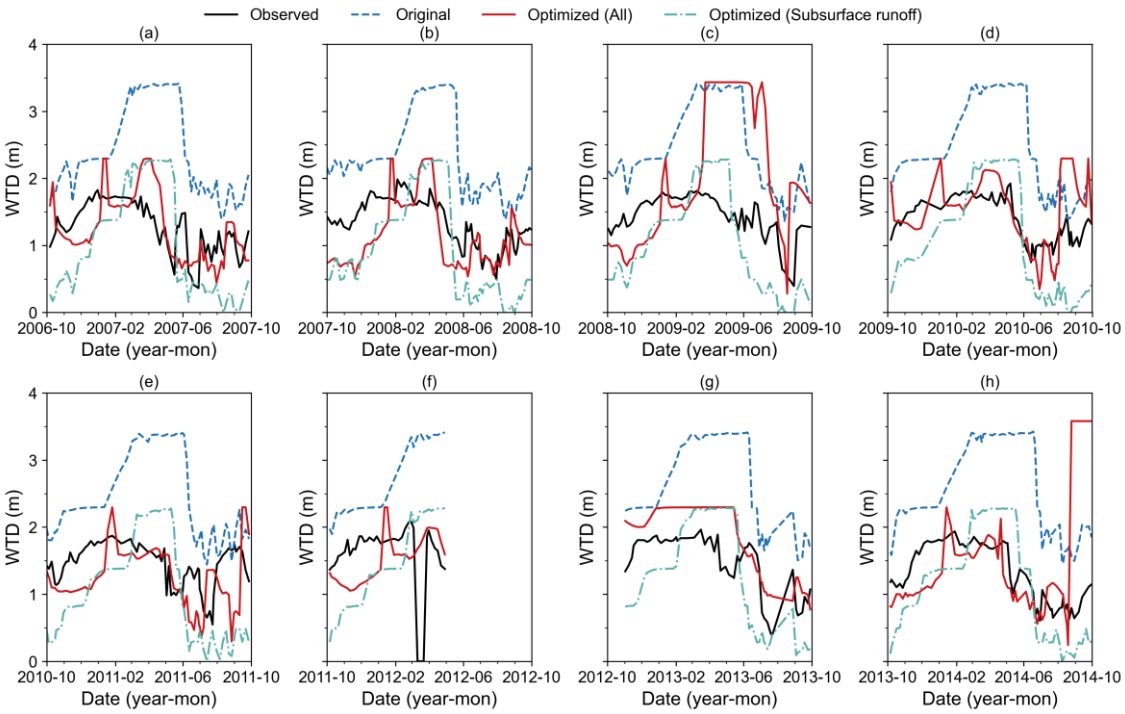


Figure 20: WTD dynamics simulations for the 2006–2014 water years based on the default parameters (blue), the SCE-derived optimal parameter set (red), and the optimized subsurface runoff parameter (40-rsubmax) alone (green).

In contrast, simulations using the default parameter set yielded higher $RMSE$ values, ranging from 1.05 to 1.35 m (Fig. 21). Regardless of whether the complete optimal parameter set or only the optimized subsurface runoff parameter (40-rsubmax) was used, the $RMSE$ decreased significantly, with reductions of 3%–68% and 28%–56%, respectively. These results indirectly confirm the effectiveness of the selected parameters.

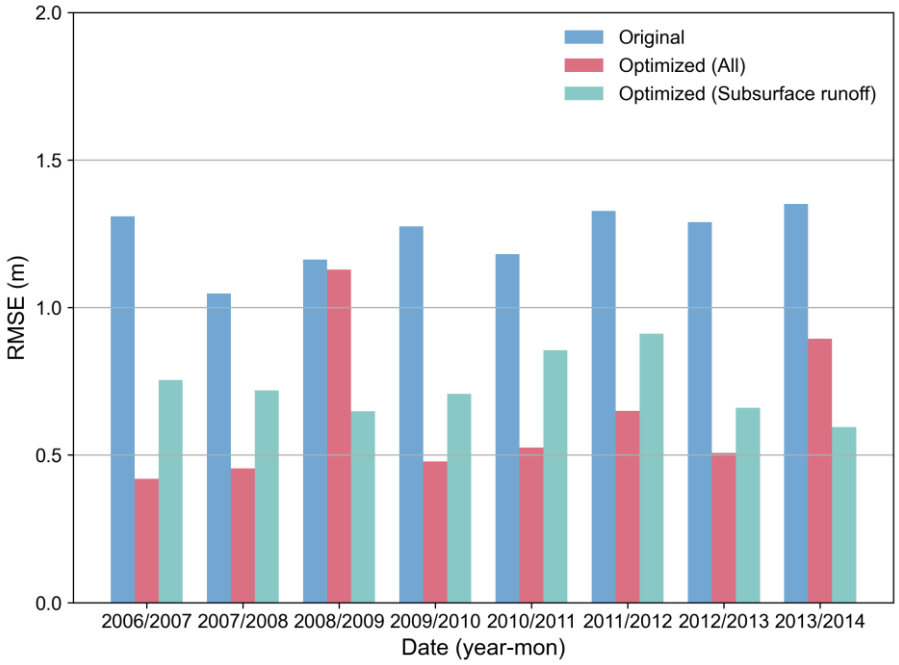


Figure 21: *RMSE* of simulated WTD across different water years (2006–2014) based on the default parameters (blue), the SCE-derived optimal parameter set (red), and the optimized subsurface runoff parameter (40-rsubmax) alone (green).

4 Discussion

4.1 The consistency of the screening results and physical interpretations

The eight stably sensitive parameters identified from the previous screening including 56-soil_alpha, 53-soil_n, 49-beta, 40-rsubmax, 45-krmax, 4-alpha_rain, 46-ck0 and 10-rhol_nir for WTD simulations were comprehensively identified through both qualitative and quantitative sensitivity analyses. The key processes and calculation formulas related to the parameters are detailed in Sect. S5 of the supplementary material.

From physical interpretations, parameters 56-soil_alpha and 53-soil_n govern the soil-water characteristic curve, where 56-soil_alpha is related to the inverse of the air-entry pressure and 53-soil_n represents the pore-size distribution of the soil (Schaap and Van Genuchten, 2006). By regulating critical thresholds such as field capacity and wilting point, they directly control soil water content in the unsaturated zone, thereby exerting a pronounced influence on WTD dynamics (Muñoz-Vega et al., 2025). The high sensitivity aligns with previous studies, in which both the ISBA land surface model and Hydrus-1D identified these parameters as primary drivers of soil water dynamics (Sobaga et al., 2023; Rezaei et al., 2016). Parameter 40-rsubmax is involved in subsurface runoff parameterization, which characterizes the exponential decay relationship between subsurface runoff and WTD. Physically, 40-rsubmax controls the maximum intensity of subsurface runoff. Changes in this parameter directly alter the equilibrium state of groundwater and consequently influence the simulated WTD. Existing

studies have confirmed that this scheme influences the overall depth of WTD (Akhter et al., 2025). Moreover, the critical role of parameter 40-rsubmax itself is also reflected by the distinct turning points observed in the parameter– $RMSE$ scatter plots. Parameter 49-beta controls the root fraction within the soil profile, thereby influencing root distribution and, subsequently, the calculation of soil and root conductance, which determines the plant transpiration rate. Since root distribution is a key factor in groundwater contribution to the root zone, and this contribution occurs when WTD is no deeper than 3.9 to 4.1 m (Zhu et al., 2018), it follows that parameter 49-beta plays a crucial role in modulating WTD. This indicates that root water uptake represents an important pathway linking groundwater dynamics with root water redistribution, particularly in forest ecosystems. Parameters 45-krmax and 46-ck0 are directly involved in computing the hydraulic conductance of plant segments (e.g. leaves, stems, and roots). Field studies~~d~~ have investigated how plants modify hydraulic conductance to maintain water status (Rickard et al., 2025). By modulating transpiration rates, these parameters indirectly regulate water redistribution within soil–vegetation systems and thus influence WTD (O'connor et al., 2019; Luo et al., 2016). Therefore, their sensitivity reflects the coupling between plant hydrological processes and groundwater dynamics. Parameter 4-alpha_rain is an empirical coefficient representing the interception capacity of the canopy. It is used to estimate throughfall after subtracting canopy interception, thereby determining the net water input reaching the land surface. Rainfall interception loss has been recognized as a vegetation influence on groundwater recharge and consequently on WTD (Fan et al., 2014; Smerdon et al., 2008). Thus, the sensitivity of 4-alpha_rain highlights the importance of accurately representing plant water redistribution processes when simulating groundwater responses. Parameter 10-rhol_nir is used in surface albedo calculations, affecting surface energy balance and snowmelt processes, which indirectly affects infiltration and groundwater recharge. Previous LSM-based studies have reported the direct impact of leaf optical properties on surface solar radiation partitioning (Dong et al., 2021) and the influence of canopy radiation on the dynamics of snowmelt dynamics (Todt et al., 2019).

In summary, the sensitive parameters controlling WTD can be classified into four categories: (1) parameters governing unsaturated water movement (56-soil_alpha and 53-soil_n); (2) those affecting subsurface runoff process (40-rsubmax); (3) those regulating plant hydraulic processes (49-beta, 45-krmax, and 46-ck0); and (4) those determining net surface water input and infiltration (4-alpha_rain and 10-rhol_nir). These sensitive parameters represent different but interconnected hydrological processes within the soil–vegetation–atmosphere system.

4.2 Different roles of subsurface runoff and soil hydraulic parameters in controlling WTD dynamics

Based on the preceding analysis, the subsurface runoff parameter $r_{sub,max}$ (40-rsubmax) exhibited two key characteristics: (1) a significant turning point in the parameter– $RMSE$ scatter plot during the optimization process, indicating a well-defined optimal parameter range; and (2) the optimization of $r_{sub,max}$ alone improved the simulations, while avoiding abrupt changes in WTD dynamics. In the SIMTOP parameterization, subsurface runoff is jointly governed by $r_{sub,max}$ and the decay factor

f_{drai} (i.e. parameter 38-fsatdcfa in Table 2). Although f_{drai} was initially screened as less sensitive, its synergetic interaction with $r_{sub,max}$ necessitates a more detailed investigation into their combined influence on WTD.

Optimization results indicated that the optimal $r_{sub,max}$ was on the order of 10^{-4}. Consequently, five representative values (3×10^{-4}, 4×10^{-4}, 5×10^{-4}, 6×10^{-4}, 7×10^{-4}) were selected for sensitivity analysis. Within this range of $r_{sub,max}$, where subsurface runoff intensity remains reasonable, the decay coefficient f_{drai} (default value 2.5) was further varied across five levels (1, 2, 3, 4, 5), resulting in 25 parameter combinations (Fig. 22). The results showed that with a fixed $r_{sub,max}$, f_{drai} primarily influenced the mean level, amplitude, and intra-annual variability of the WTD. Lower f_{drai} values yielded larger seasonal amplitudes and generally deeper WTD. Conversely, increasing f_{drai} accelerated the decay of subsurface runoff, thereby dampening WTD fluctuations. In contrast, when f_{drai} was fixed, $r_{sub,max}$ serves as the primary control for the mean WTD. Higher $r_{sub,max}$ values resulted in a systematic deepening of the simulated WTD, accompanied by intensified fluctuations in some cases.

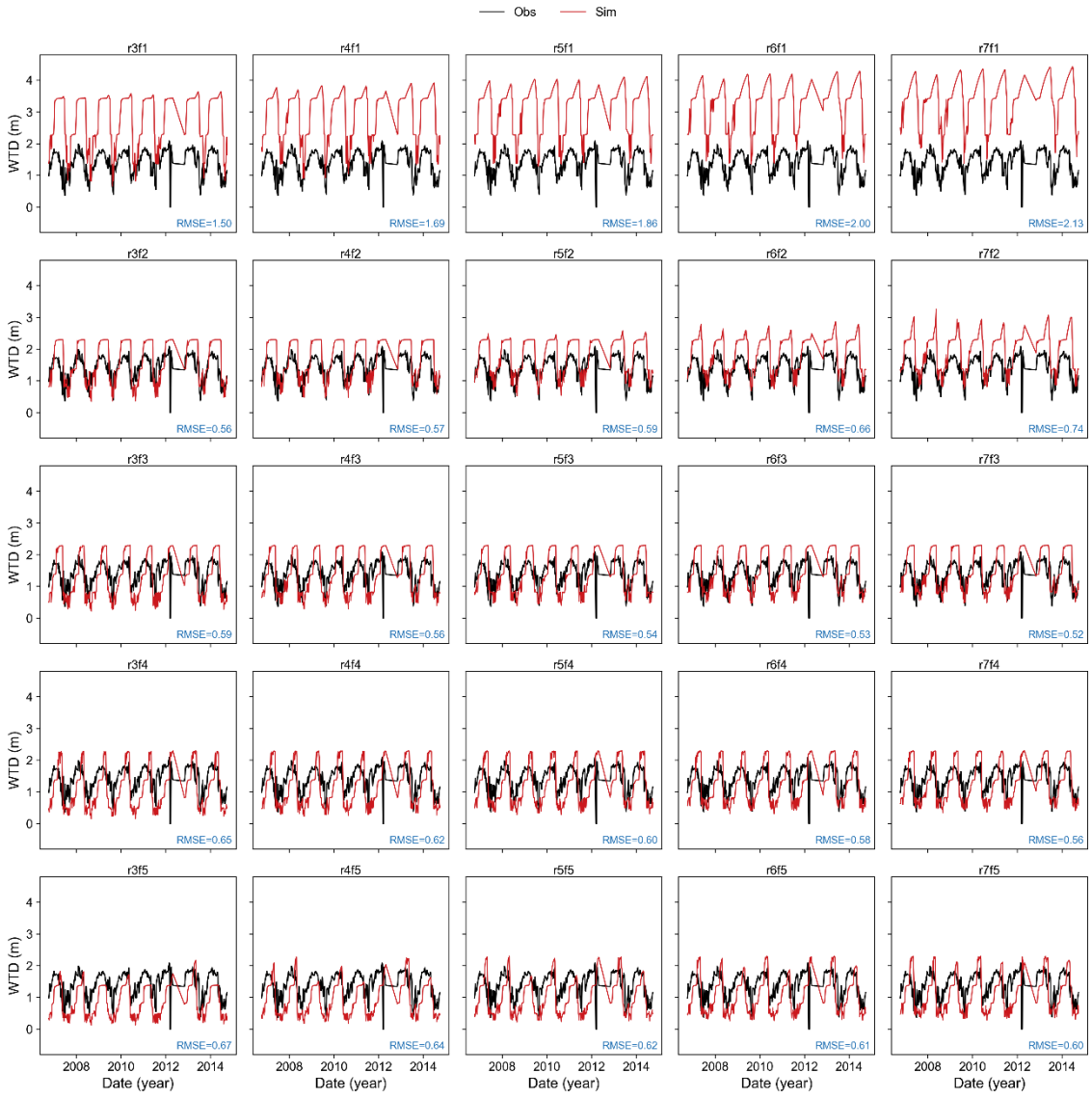


Figure 22: Comparison of observed and simulated daily WTD dynamics across 25 parameter ensembles for the 2006–2014 water years. Each panel is labeled as 'rXfY', where X corresponds to $r_{sub,max}$ (3×10^{-4}, 4×10^{-4}, 5×10^{-4}, 6×10^{-4}, and 7×10^{-4}) and Y corresponds to f_{drai} (1, 2, 3, 4, and 5).

Regarding the multi-year average (Fig. 23), when $r_{sub,max}$ was fixed, the mean WTD decreased markedly (approaches the surface) as f_{drai} increases. This response was more sensitive at lower parameter values and weakened as f_{drai} increases further. Conversely, when f_{drai} was fixed, increasing $r_{sub,max}$ leaded to a deeper WTD due to enhanced groundwater discharge (Fig. 24). Compared with f_{drai}, the response of WTD to variations in $r_{sub,max}$ was relatively moderate within the tested ranges. Specifically, variations in f_{drai} led to WTD changes exceeding 2 m, while $r_{sub,max}$ accounted for changes of

approximately 1 m. This suggested that once $r_{sub,max}$ establishes a reasonable runoff baseline, f_{drai} can further regulate both the mean level and dynamic amplitude of WTD.

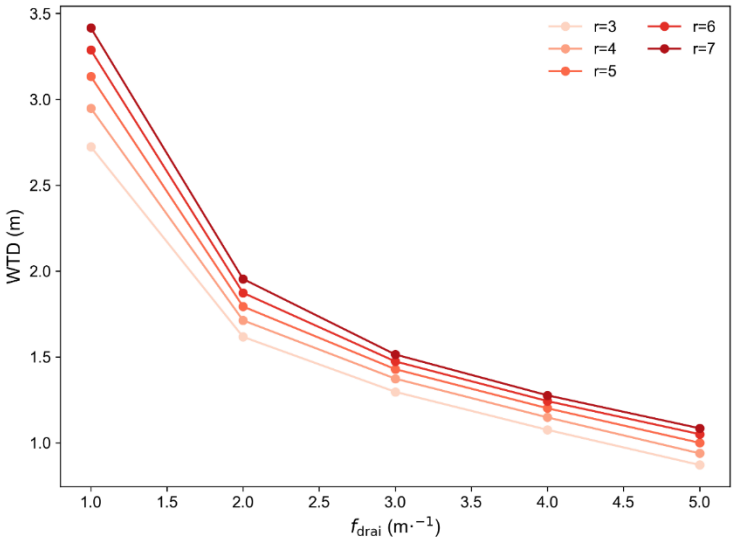


Figure 23: Sensitivity of multi-year mean WTD to the decay factor f_{drai} under various constant $r_{sub,max}$ values. The legend labels 3, 4, 5, 6, and 7 correspond to $r_{sub,max}$values of 3×10^{-4}, 4×10^{-4}, 5×10^{-4}, 6×10^{-4}, and 7×10^{-4}, respectively.

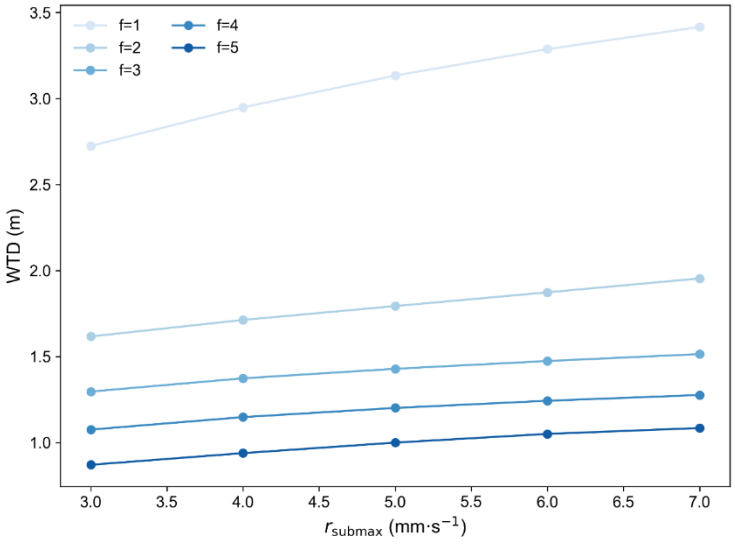


Figure 24: Sensitivity of multi-year mean WTD to $r_{sub,max}$ under various constant f_{drai} values. The horizontal axis labels 3, 4, 5, 6, and 7 correspond to $r_{sub,max}$values of 3×10^{-4}, 4×10^{-4}, 5×10^{-4}, 6×10^{-4}, and 7×10^{-4}, respectively.

According to the SIMTOP subsurface runoff parameterization (Niu et al., 2005), $r_{sub,max}$ and f_{drai} control the intensity and decay rate of subsurface runoff, respectively. This theoretical structure explains our result that $r_{sub,max}$ facilitates systematic

deepening of the WTD when increased, while f_{drai} dictates the system's response. Consequently, $r_{sub,max}$ and f_{drai} primarily regulate the mean level and response characteristics of WTD. When $r_{sub,max}$ is small, subsurface runoff discharge is reduced, leading to an overall shallower WTD. In contrast, larger values of $r_{sub,max}$ result in a systematic deepening of the WTD. When f_{drai} is small, subsurface runoff decreases more slowly with increasing WTD, allowing groundwater discharge to remain relatively strong, which enables the pronounced intra-annual variability in WTD. Conversely, when f_{drai} is large, subsurface runoff declines rapidly as WTD increases, creating a "buffering effect" that smooths WTD fluctuations and substantially reduces the intra-annual fluctuation range.

Compared with the distinctive behaviour of subsurface runoff parameters, the van Genuchten parameter α was identified as the most sensitive variable overall. This is consistent with previous literature, which has shown that α exerts a significant influence on evapotranspiration and vertical soil water transport, thereby affecting WTD (Maina and Siirila-Woodburn, 2020). As shown in Fig. 25, when other parameters were held at their default values, varying α across several orders of magnitude reshaped both the mean state and the transient response of the WTD. Specifically, smaller α values corresponded to deeper WTD and more pronounced fluctuations, whereas increasing α led to shallower WTD, a more lagged groundwater response, and smoother variations.

Regarding the relationship between α and WTD, α is a scale parameter inversely related to the mean pore size (Maina and Siirila-Woodburn, 2020). Thus, lower α values indicate a greater capacity of the medium to retain water, resulting in relatively higher evapotranspiration and a deeper WTD, particularly in forested regions with high potential evapotranspiration. In addition, as the WTD deepens, the importance of α increases. This is because, with a thicker unsaturated zone, ET is primarily controlled by soil moisture and atmospheric conditions, making it more sensitive to soil hydraulic parameters in the unsaturated zone. As for the relationship between α and WTD variability, α is also the inverse of the air-entry value (Peters and Durner, 2006). A smaller α implies a higher air-entry value, allowing water to more easily fill pore spaces and increasing hydraulic conductivity, which accelerates downward water movement and leads to stronger WTD fluctuations (Corona and Ge, 2022). Conversely, larger α values imply smaller air-entry values, meaning that water has more difficulty entering pore spaces relative to air, thereby reducing unsaturated hydraulic conductivity. Lower conductivity results in slower and more delayed downward flow toward the water table. Over time, this attenuated percolation leads to a slightly lagged groundwater response and smoother WTD variations.

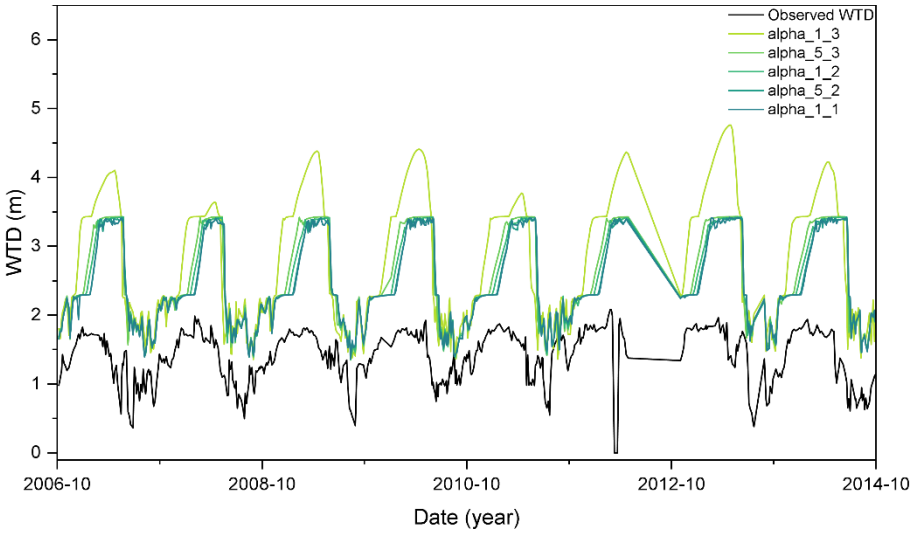


Figure 25: Sensitivity of simulated WTD to the van Genuchten parameter α with all other parameters held at their default values. The legend labels in the format 'alpha_X_Y' denote α values of X×10^{-Y} (specifically: 1×10^{-3}, 5×10^{-3}, 1×10^{-2}, 5×10^{-2}, 1×10^{-1}).

Comparing the effects of different parameters on WTD, changes in subsurface runoff parameters $r_{sub,max}$ and f_{drai} can substantially alter the depth of WTD, bringing it closer to observations. In contrast, α played a more prominent role in shaping the fluctuation and lag characteristics of WTD. This suggested a practical calibration strategy in which subsurface runoff parameters are first adjusted to maintain WTD at a reasonable depth, followed by fine-tuning of other sensitive parameters to improve the temporal dynamics and variability of WTD.

4.3 Robustness of qualitative sensitivity analysis

Because qualitative sensitivity analysis involves stochastic sampling, additional experiments using two independent random seeds were conducted to evaluate the robustness of the identified sensitive parameters. Together with the original realization, three repeated experiments were available. The original experiment presented in the manuscript was designated as Seed 3 (s3), while the two newly generated realizations were designated as Seed 1 (s1) and Seed 2 (s2). For all realizations, failed simulations caused by model non-convergence were excluded before calculating sensitivity indices. Then, the screening procedure mentioned in Sect. 3.1 was applied to the additional two experiments. The robustness was evaluated using selection frequency and the mean sensitivity index among repeated realizations.

The screening results of the additional random seed experiments are shown in Table 7. In this table, 1 indicates that a parameter was selected as a sensitive parameter in the corresponding experiment, whereas 0 indicates that it was not selected. The detailed screening results are provided in Sect. S6 of the supplementary material. Overall, the repeated experiments showed that the qualitative sensitivity rankings were highly consistent for the dominant parameters but exhibited moderate variability. Among all parameters, 56-soil_alpha was identified as the most sensitive parameter in every repeated experiment and exhibited the highest selection frequency, indicating excellent robustness (Fig. 26). Similarly, parameters (i.e. 53-soil_n,

45-krmax, 49-beta, 46-ck0, 10-rhol_nir, and 4-alpha_rain) were repeatedly selected as sensitive across different random seeds (Table 7), although their detailed rankings varied. In contrast, parameter 40-rsubmax showed relatively larger variability among repeated realizations. Although it was identified as highly sensitive in the original experiment and exhibited important effects on WTD dynamics, its selection frequency decreased in additional realizations (Table 7). This suggested that the sensitivity of 40-rsubmax may be more dependent on the specific sampling trajectories compared with the dominant parameters such as 56-soil_alpha.

These results also highlight a potential limitation of qualitative sensitivity screening itself. Different screening methods characterize input-output relationships in different ways and may produce different sensitivity classifications to the same parameter, including cases where a parameter is classified as sensitive by one method but insensitive by another, or conversely. For MARS and SOT, the resulting importance measures may not represent the isolated contribution of a parameter, which can complicate the attribution of sensitivity to individual parameters and contribute to differences among screening methods (Gan et al., 2014). By contrast, MOAT changes one parameter at a time while keeping the other parameters unchanged, which may reduce such attribution ambiguity. Therefore, the use of multiple screening methods and random seeds provides an important means of cross-checking their robustness.

Table 7: Parameter selection results under different random seed experiments.

Parameter	s1-fixed method	s1-fixed objective function	s2-fixed method	s2-fixed objective function
4-alpha_rain	1	1	1	1
10-rhol_nir	0	1	0	1
40-rsubmax	0	0	0	0
45-krmax	1	1	0	0
46-ck0	1	1	1	0
49-beta	1	1	0	0
53-soil_n	1	1	1	1
56-soil_alpha	1	1	1	1

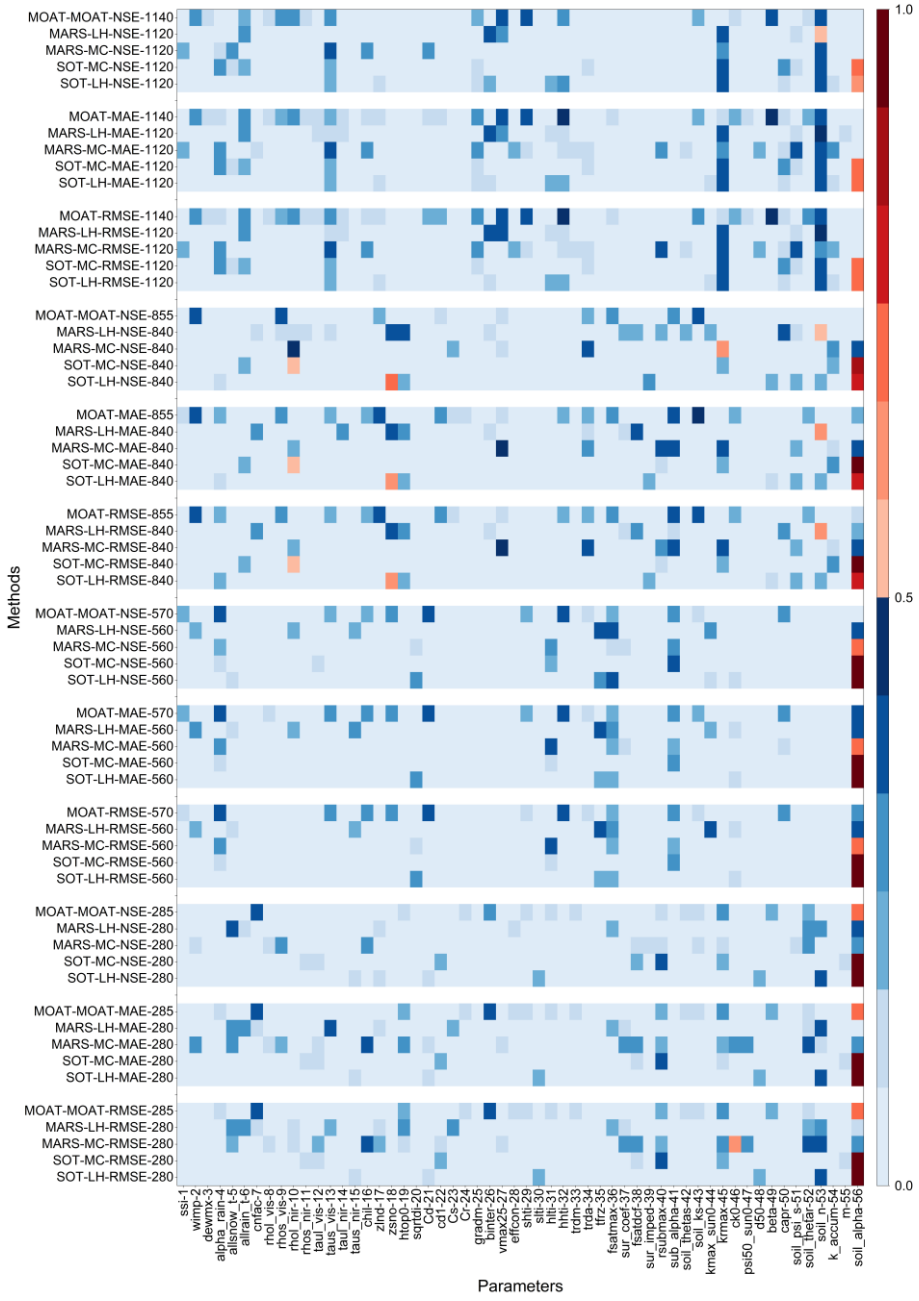


Figure 26: Mean normalized qualitative sensitivity analysis results across three independent random seed realizations (s1–s3). The normalized sensitivity index ranges from 0 to 1, where 1 and 0 represent the most and least sensitive parameters, respectively. The horizontal axis denotes parameter indices, while the vertical axis represents different analysis methods and sampling techniques, following the format "sensitivity analysis method-sampling method-objective function-sample size".

4.4 Performance and limitations of parameter optimization for WTD simulation

Considering the cost of computation, a fixed optimization cost (about 500–1000 runs) was adopted for the evaluation. Such comparisons of fixed costs have also been conducted in previous studies (Kavetski et al., 2018), and they are divided into two types of computational budgets. The evaluations ranging from 200 to 500 times are classified as the "tight" type, while those ranging from 1000 to 5000 times are classified as the "large" type (Qin et al., 2022). The convergence behaviours of different methods are different. Some methods can still find satisfactory solutions even when the budget is limited. However, for some methods, even in the case of a large budget, there is still a significant variability in the convergence behaviour. This might be related to the complexity of the model, especially when dealing with highly parameterized models, and the speed of convergence also depends on various factors such as the sampling strategy of the algorithm (Qin et al., 2022). In addition, the optimization process was also controlled by the convergence criteria. The detailed stopping criteria of PSO, SCE, and BOBYQA implemented in PSUADE are provided in Sect. S7 of the supplementary material. The primary purpose of the optimization experiments in this study was not to benchmark the absolute global optimization capability of different algorithms, but rather to evaluate whether the sensitive parameters identified from the sensitivity-analysis framework could effectively improve WTD simulations. Therefore, a moderate computational budget in this study was considered appropriate for assessing the effectiveness of the selected parameters. Nevertheless, the use of fixed-cost evaluation for parameter optimization still has limitations. Specifically, performance metrics can be subject to sampling error under the low number of algorithm invocations (Qin et al., 2022). In addition, the relative performance may vary under larger evaluation budgets (Zhang et al., 2009). Thus, the comparison of the three optimization methods should be interpreted in the context of their practical but unequal computational budgets used in this study. Further experiments with matched evaluation budgets would be useful for a more rigorous comparison of their performance.

Although the SCE-optimized simulation substantially improved the overall WTD performance, Fig. 16 showed that the two groundwater fluctuation events between May and July in 2007 were not fully reproduced. To further evaluate whether the selection of the objective function influences the performance, the existing SCE simulations were additionally assessed using Kling-Gupta efficiency (KGE, Eq. (S50)) and event-based $RMSE$ metrics (Sect. S8.1). The results demonstrated that the parameter set achieving the lowest global $RMSE$ was not necessarily optimal for reproducing temporal variability (Fig. S21a). Specifically, the solution with the maximum KGE achieved a higher KGE value (0.529) than the minimum $RMSE$ solution (0.435), while its $RMSE$ increased by only 0.123 m compared with the global minimum $RMSE$ solution. This indicated that optimization based solely on global $RMSE$ may reduce overall errors while compromising event-scale dynamics. Besides the influence of objective functions, meteorological forcing uncertainty (Sect. S8.2) and freeze–thaw representation (Sect. S8.3) may also contribute to difficulties in reproducing rapid groundwater fluctuations. In addition, previous studies have shown that simplified groundwater parameterization schemes may limit the ability of land surface models to represent groundwater recharge dynamics, particularly during intense precipitation events (Sobaga et al., 2023). Furthermore, uncertainties associated with groundwater observations may also introduce deviations. Overall, accurately

reproducing event-scale groundwater fluctuations requires further improvements in groundwater process parameterization and the integration of multi-objective optimization approaches in future studies.

5 Conclusions

In this study, a CoLM-PSUADE coupling framework was developed to integrate parameter sensitivity diagnosis, dimensionality reduction, and parameter optimization within a unified workflow for groundwater-related modeling. Using in-situ water table depth (WTD) observations at the GGF mountain forest site, the framework provided a systematic approach to identify influential parameters and reduce the dimensionality of the calibration problem under the studied site conditions.

The initial~~Three~~ qualitative sensitivity analyses identified eight stably sensitive parameters (56-soil_alpha, 45-krmax, 4-alpha_rain, 53-soil_n, 49-beta, 10-rhol_nir, 46-ck0, and 40-rsubmax) ~~analysis methods (MOAT, MARS, and SOT), combined with multiple performance metrics ($RMSE$, MAE, and NSE), consistently identified eight stable sensitive parameters (56-soil_alpha, 45-krmax, 4-alpha_rain, 53-soil_n, 49-beta, 10-rhol_nir, 46-ck0, and 40-rsubmax)~~ governing soil water movement, subsurface runoff, plant hydraulics, and surface infiltration processes. ~~Additional experiments using independent random seeds further demonstrated that qualitative sensitivity rankings were generally consistent for the dominant parameters but exhibited variability among different sampling realizations. Uniquely among all parameters, 56-soil_alpha maintained its top sensitivity ranking across every repeated experiment, indicating its dominant role in controlling WTD simulations.~~ Subsequent RSMSobol-based quantitative analysis provided a complementary assessment of these eight parameters and identified a core group (56-soil_alpha, 53-soil_n, 49-beta, 45-krmax, 40-rsubmax, and 4-alpha_rain).~~further confirmed the robustness of this reduced parameter set and indicated that a core group of parameters (56-soil_alpha, 53-soil_n, 49-beta, 45-krmax, 40-rsubmax, and 4-alpha_rain) exerts dominant control on model behaviour across different sampling configurations.~~ Building on ~~the~~ this reduced parameter space,~~reduced dimensional space defined by the eight identified sensitive parameters,~~ parameter optimization using SCE, PSO, and BOBYQA demonstrated that targeted calibration can substantially improve WTD simulation performance, with SCE identifying the lowest-RMSE solution most rapidly~~providing the most stable and effective convergence~~. To further assess the robustness of the initial qualitative sensitivity results, two additional experiments using independent random seeds were conducted. These experiments demonstrated that qualitative sensitivity rankings were generally consistent for the dominant parameters but exhibited variability among different random-seed realizations. In particular, 40-rsubmax was identified in the initial screening but was not selected as sensitive in the summarized screening results of either additional random-seed realization. In contrast, 56-soil_alpha maintained its top sensitivity ranking across all experiments, confirming its robust and dominant influence on WTD simulations.

Among identified sensitive parameters, the subsurface runoff parameter $r_{sub,max}$ was unique in exhibiting a clear inflection point in the parameter-$RMSE$ relationship, and the soil hydraulic parameter α was identified as the most sensitive parameter.

From a mechanistic perspective, subsurface runoff parameters ($r_{sub,max}$ and f_{drai}) primarily control the mean state of WTD by regulating runoff generation and decay, while the soil hydraulic parameter α governs soil water retention and vertical transport processes, thereby controlling variability and temporal response. Based on these results, a stepwise calibration strategy was recommended, prioritizing subsurface runoff parameters to constrain mean WTD, followed by other sensitive parameters to improve temporal dynamics.

It should be noted that although the GGF site selected in this study is representative for studying interactions between WTD dynamics and key hydrological processes such as runoff, infiltration, evaporation, and freeze-thaw, several limitations remain. The sensitivity results obtained in this study are specific to the GGF mountain forest site and reflect the combined effects of its hydroclimatic conditions, soil properties, vegetation characteristics, and groundwater regime. Single-point simulations cannot fully capture the spatial heterogeneity of parameter sensitivities under different climate conditions and land-surface characteristics. In addition, the relatively short sensitivity-analysis period may limit the assessment of temporal variations in parameter sensitivities under different hydrological conditions. Although repeated random seed experiments improved the assessment of sampling uncertainty, variations in sensitivity rankings indicate that qualitative sensitivity analysis remain partly dependent on sampling trajectories. Moreover, considering WTD as a single target variable may lead to parameter equivalence. This issue is exacerbated by the limited availability of observational data, resulting in insufficiently constrained parameter uncertainty. Nevertheless, the proposed research framework provides a systematic approach for sensitivity analysis and sensitivity-guided parameter optimization, offering a foundation for broader applications in groundwater-related land surface modeling. Future studies should extend the framework to multi-site and regional-scale applications. Incorporating longer simulation periods and multi-variable observations will further improve parameter identifiability.

Appendix A: Principles and formulations of sensitivity analysis methods

A.1 MOAT method

MOAT is a gradient-based sensitivity analysis method. The sensitivity of model performance to an input parameter is quantified by elementary effects, which approximate local partial derivatives with respect to the parameter (Guerrero et al., 2017). The modified mean and standard deviation of these elementary effects represent overall effects and interaction effects of input parameters, respectively. Larger values of the modified mean indicate stronger parameter sensitivity, while larger values of the standard deviation suggest more significant interactions with other input parameters (Li et al., 2013). Because MOAT makes no assumptions about the relationships between model parameters and model performance, it is considered a robust screening method (Guerrero et al., 2017). The main equations are given as follows (Gan et al., 2014).

First, the range of each input parameter is normalized to [0,1]. Each parameter may take on values from:

$$X = \left\{0, \frac{1}{p-1}, \dots, 1\right\}, \quad \text{(A1)}$$

where X is the set of discretized levels of the input parameter, p is the number of levels into which the parameter space is divided.

For each input parameter, the elementary effect is calculated as an approximation of the first-order derivative. This metric quantifies the variation in model output caused by a change in a single parameter while all other input parameters are fixed:

$$EE_i = \frac{f(x_1,\dots,x_i+\Delta,\dots,x_k)-f(x_1,\dots,x_i,\dots,x_k)}{\Delta}, \quad \text{(A2)}$$

where EE_i is the elementary effect of the i-th parameter, Δ is a predetermined multiple of $\frac{1}{p-1}$, x is the parameter, and k is the number of parameters.

To ensure statistical robustness, each input parameter is sampled r times, which is equals to r different paths. Elementary effects EE_i are calculated for each x_i along each path, from which the sensitivity indices are derived as follows:

$$\mu_i^* = \frac{1}{r}\sum_{j=1}^{r}\left|EE_i^{(j)}\right|, \quad \text{(A3)}$$

$$\sigma_i^* = \sqrt{\frac{1}{r-1}\sum_{j=1}^{r}\left(\left|EE_i^{(j)}\right| - \mu_i^*\right)^2}, \quad \text{(A4)}$$

where μ_i^* is modified mean of EE_i, and σ_i^* is modified standard deviation. The r-1 is used for σ_i^*, consistent with the implementation.

A.2 MARS method

MARS is a regression model that extends linear modeling by automatically incorporating piecewise linear basis functions to capture nonlinearities and variable interactions in high-dimensional data. The relative importance of input parameters is evaluated by re-assessing model performance after removing each parameter from the regression function (Guerrero et al., 2017). A larger reduction in the model performance indicates a higher importance of the corresponding parameter. However, this approach is less robust because it assumes a fixed functional relationship between model performance and input parameters (Guerrero et al., 2017). The regression function is composed of a series of basic functions and can be expressed as follows (Gan et al., 2014):

$$Y = a_0 + \sum_{m=1}^{M} a_m \prod_{k=1}^{K_m}\left[s_{km}\left(x_{v(k,m)} - t_{km}\right)\right]_+, \quad \text{(A5)}$$

where a_0 is a constant term, a_m denotes the fitting coefficient of the m-th basis function, M is the total number of basic functions, K_m is the number of knots, s_{km} takes values of 1 or -1 and represents the right or left sense of step function, $v(k,m)$ is the label of independent value, and t_{km} is the knot location.

MARS first constructs an overfitted model by subsequently adding all input parameters. The model is then modified by removing one parameter at a time to generate a series of modified models. Both the overfitted and the modified models are used to calculate the generalized cross validation (GCV) index:

$$GCV(M) = \frac{1}{N}\frac{\sum_{k=1}^{N}(Y_k - \widehat{Y_k})^2}{\left[1-\frac{C(M)}{N}\right]^2}, \quad \text{(A6)}$$

$$C(M) = 1 + c(M)d, \quad \text{(A7)}$$

where M is the model, N is the number of observations, Y_k is the k-th observation, $\widehat{Y_k}$ is the estimated value of Y_k, $C(M)$ is the number of effective parameters, d is the effective degrees of freedom, and $c(M)$ is a penalty term associated with adding a basis function.

The increase in the GCV value between the modified and the overfitted model is considered as the importance index for the removed parameter. A larger increase in GCV suggests a higher importance of that parameter. Accordingly, the sensitivity score of the *i*-th parameter is defined as:

$$score(i) = \frac{\Delta g(i)}{max\{\Delta g(1), \Delta g(2), \ldots, \Delta g(n)\}} \times 100, \quad \text{(A8)}$$

where $\Delta g(i)$ indicates the increase in the GCV value resulting from the removal of the i-th parameter.

A.3 SOT method

The SOT method is a tree-based method. It represents model performance as the sum of multiple binary decision trees constructed from model parameters. However, it is less robust, as this approach hinges on the assumption that the relationship between input parameters and model performance can be captured by such a tree structure (Guerrero et al., 2017). Each parameter splits the parameter space into two regions with distinct responses. Overall model performance is then obtained by aggregating the simulated results across all partitioned regions. The relative importance of a parameter is measured by the total number of splits associated with that parameter, with parameters appearing in more splits regarded as more important (Li et al., 2013).

The structure of a single tree consists of decision nodes and terminal nodes (Chipman et al., 2010). Each decision node partitions the input parameter space $X = \{x_1, \ldots, x_p\}$ according to a splitting rule. Through a sequence of decisions from the root node to a leaf node, each input parameter is assigned to a terminal node associated with a prediction value $\mu_i \in M = \{\mu_1, \mu_2, \ldots, \mu_b\}$. The SOT method is then defined as:

$$Y = \sum_{j=1}^{m} g(X;\, T_j, M_j) + \varepsilon, \quad \text{(A9)}$$

where T_j is the j-th binary decision tree, M_j is the set of terminal node parameters, $g(X;\, T_j, M_j)$ is the function which assigns μ_i to X, m is the total number of trees, ε is a random noise term with $\varepsilon \sim \mathcal{N}(0, \sigma^2)$.

In the above splitting procedure, the residual sum of squares (RSS) is used as the splitting criterion. The parameter that produces the largest reduction in RSS is selected to split the node until each terminal node includes minimum number of data points (Gan et al., 2014). Subsequently, the total number of splits associated with each parameter is considered as a score

metric (Gan et al., 2014). A larger number of splits indicates higher sensitivity of the corresponding input parameter. The score of the i-th input parameter is defined as:

$$score(i) = \frac{p(i)}{max\{p(1),p(2),\dots,p(n)\}} \times 100, \quad \text{(A10)}$$

where $p(i)$ denotes the number of splits associated with the i-th input parameter.

A.4 Sobol method

The Sobol method is a variance-based method. It quantifies parameter sensitivity by decomposing the total variance into contributions from individual input parameters and their interactions (Li et al., 2013). This method assumes that the model output can be decomposed into terms of increasing dimensionality:

$$f(X) = f_0 + \sum_{i=1}^{n} f_i(x_i) + \sum_{1\le i<j\le n} f_{ij}(x_i, x_j) + \cdots + f_{1,2,\cdots,n}(x_1, x_2, \cdots, x_n), \quad \text{(A11)}$$

where $Y = f(X) = f(x_1, x_2, \dots, x_n)$ is the model output, $x_i \in [0,1]$, and f_0 is the constant term.

Accordingly, the total variance of the model output can be decomposed as：

$$V = \sum_{i=1}^{n} V_i + \sum_{1\le i<j\le n} V_{ij} + \cdots + V_{1,2,\cdots,n}, \quad \text{(A12)}$$

where V is the total variance, V_i denotes the main effect variance associated with x_i, V_{ij} denotes the interaction variance between x_i and x_j, and $V_{1,2,\cdots,n}$ is the variance due to interactions among all input factors.

By dividing each variance component by total variance V, the normalized Sobol sensitivity indices are quantified as:

$$\sum_{i=1}^{n} S_i + \sum_{1\le i<j\le n} S_{ij} + \cdots + S_{1,2,\cdots,n} = 1, \quad \text{(A13)}$$

where $S_i = V_i/V$ is the first-order Sobol sensitivity index of x_i and represents its main effect. $S_{ij} = V_{ij}/V$ is the second-order Sobol index, accounting for the interaction effect between x_i and x_j.

The total effect sensitivity index of parameter x_i is defined as:

$$S_{T_i} = \sum S_{(i)}, \quad \text{(A14)}$$

where $S_{(i)}$ is the sensitivity of x_i, (i) denotes all subsets of indices containing i.

To facilitate comparison of the results across different experiments, the concepts of relative importance and cumulative importance were introduced. The relative importance of the i-th parameter in parameter set A is defined as:

$$C(i) = \frac{S_{T_i}}{\sum_{k=1}^{n} S_{T_k}}, \quad \text{(A15)}$$

where $C(i)$ denotes the relative importance of parameter i within set A; n is the total number of parameters in set A; and S_{T_i} represents the Sobol' total-order index of parameter i. The cumulative importance of a parameter set is then defined as the sum of the relative importance values.

A.5 Min-max Normalization

To enable a consistent comparison of parameter screening results across different qualitative sensitivity analysis methods, the sensitivity metrics derived from the MOAT, MARS, and SOT methods are scaled using min-max normalization. The original sensitivity metrics are linearly mapped into the range [0, 1] using the following function:

$$X^* = \frac{X - X_{min}}{X_{max} - X_{min}}, \tag{A16}$$

where X^* is the normalized sensitivity index, X is the original sensitivity metric, and X_{min} and X_{max} represent the minimum and maximum values of the metrics within each method, respectively.

A.6 Response surface validation metrics

After loading the parameter inputs and objective function outputs into PSUADE, the quality of the response surface trained on these samples can be evaluated. A comprehensive assessment is performed based on cross-validation errors and the goodness of fit. Assuming the true output (i.e. the corresponding PSUADE output metric) is y and the predicted output from the response surface is $\hat{y}$, the evaluation metrics are defined as follows:

$$err_{avgs} = \frac{1}{n}\sum_{i=1}^{n}\frac{\hat{y}_i - y_i}{|y_i|}, \tag{A17}$$

$$err_{rmss} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{|\hat{y}_i - y_i|}{|y_i|}\right)^2}, \tag{A18}$$

$$err_{maxs} = \max_{1\le i\le n}\left(\frac{|\hat{y}_i - y_i|}{|y_i|}\right), \tag{A19}$$

$$err_{maxs,base} = |y_{i^*}|, i^* = \operatorname*{argmax}_{1\le i\le n}(\hat{y}_i - y_i), \tag{A20}$$

$$err_{maxs2} = \max_{1\le i\le n}\left(\frac{|\hat{y}_i - y_i|}{y_{max} - y_{min}}\right), \tag{A21}$$

$$R^2_{rsm} = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y_i})^2}, \tag{A22}$$

where n is the total number of samples, y_i and $\hat{y}_i$ are the true and predicted values of the i-th sample, err_{avgs} is the scaled mean error, err_{rmss} is the scaled root mean square error, err_{maxs} is the scaled maximum error, $err_{maxs,base}$ is the absolute true value of the sample corresponding to the scaled maximum error, err_{maxs2} is the maximum scaled error based on the output range, and R^2_{rsm} is the coefficient of determination, indicating the goodness of fit of the response surface in approximating the original input-output relationship.

For cross-validation, the number of groups K is set to 20 (as recommended by the PSUADE manual). The training set for the x-th group consists of samples outside the interval $[x \cdot \frac{n}{K},\ (x+1) \cdot \frac{n}{K}]$, while the validation set comprises the samples within this interval. The final error metrics are not simple averages of the individual groups. Instead, the overall evaluation metrics are calculated as follows:

$$Err_{avgs} = \frac{1}{N}\sum_{j=1}^{K}\sum_{i\in G_j}\left(\frac{\hat{y}_i - y_i}{|y_i|}\right), \tag{A23}$$

$$Err_{rmss} = \sqrt{\frac{1}{N}\sum_{j=1}^{K}\sum_{i\in G_j}\left(\frac{|\hat{y}_i - y_i|}{|y_i|}\right)^2}, \tag{A24}$$

$$Err_{maxs} = \max_{1\le j\le K}\left(err_{maxs,j}\right), \tag{A25}$$

$$Err_{maxsb} = \max_{1\le j\le K}\left(err_{maxs,base,j}\right), \tag{A26}$$

$$Err_{maxs2} = \max_{1\le j\le K}\left(err_{maxs2}\right), \tag{A27}$$

where N is the total number of samples, K is the number of groups, and G_j represents the set of samples in the j-th group. If Err_{rmss} deviates significantly from 0, it indicates the presence of systematic bias. If Err_{maxs} is greater than 1 and Err_{maxsb} is non-negative, it suggests a relatively significant fitting error. However, if this error value is large while the variation range of y_i is also substantial, further analysis in conjunction with the magnitude of Err_{maxs2} is required. In such cases, a smaller Err_{maxs2} is considered acceptable. Theoretically, the predicted and actual values in cross-validation should align along the 1:1 diagonal line, and the overall fitting performance can be judged by combining this visual assessment with the R^2_{rsm}.

Appendix B: Principles of parameter optimization methods

B.1 SCE method

The SCE method is a global optimization algorithm developed by Duan et al. 1992 (Duan et al., 1994; Duan et al., 1992; Duan et al., 1993), which is widely utilized for calibrating large-scale watershed models. This method integrates the advantages of competitive biological evolution, deterministic search, and random search, thereby reducing parameter uncertainty and accelerating the search for the global optimum. Its core idea involves complex shuffling and the systematic evolution of a 'complex' of points spanning the parameter space. During the shuffling step, the initial population is randomly partitioned into several complexes. Within each complex, points are evolved through competitive mechanisms to generate new populations. This process enhances diversity in the search space and prevents the algorithm from being trapped in local optima. The evolution step employs selection, crossover, and mutation operations iteratively to update the points in each complex until a solution satisfying the stopping criterion is achieved.

B.2 PSO method

The PSO method is a global optimization algorithm developed by Kennedy and Eberhart (Kennedy and Eberhart, 1995). As a heuristic optimization technique rooted in collective intelligence, it is inspired by the social foraging behaviour of bird flocks. Compared to adaptive random search algorithms, PSO utilizes a set of potential solutions, commonly known as a population, to achieve global or near-global optima through mechanisms of collaboration and competition (Gill et al., 2006). Within this framework, each potential solution is abstracted as a particle moving through the search space. The position of particle represents a candidate solution, while its velocity determines the search direction and rate. To balance global search and local optimization, particles iteratively adjust their velocity and position guided by both their individual historical optima and the collective historical optima.

B.3 BOBYQA method

The BOBYQA method is a derivative-free algorithm developed by Powell solving bound-constrained optimization problems (Powell, 2009). The core principle of this algorithm is to construct a quadratic model that approximates the objective function near the current search point. Rather than seeking the global optimum directly, BOBYQA explores a trust region, defined as a neighbourhood focused on the current point with a predefined radius, to determine a descent step based on the local quadratic model. When generating candidate points, BOBYQA ensures that both the search direction and step length do not violate the parameter boundaries. If a predicted optimum lies outside these limits, the algorithm projects it back onto the feasible domain. As a derivative-free method, BOBYQA functions as a 'black-box' optimizer and does not require gradient information from the objective function.

Code and data availability

PSUADE version 3.0.0 (released in 2024) was used, which is publicly available at: https://github.com/LLNL/psuade. The source code of CoLM2024 is publicly available at: https://github.com/CoLM-SYSU/CoLM202X. In this study, we used a pre-release version downloaded on 13 November 2024. The code and data of CoLM-PSUADE in this study is publicly available at: https://doi.org/ 10.5281/zenodo.23080567~~10.5281/zenodo.19641291~~ (Wu et al., 2026).

The meteorological variables from the TPMFD can be freely accessed from National Tibetan Plateau/Third Pole Environment Data Center (http://data.tpdc.ac.cn; Yang et al., 2023). The soil characteristics datasets (GSDE) can be freely accessed from http://globalchange.bnu.edu.cn/research/data/ (Shangguan et al., 2014b). WTD and soil structure dataset (Zhu et al., 2017b; Yang et al., 2020) is provided by National Ecosystem Science Data Center, National Science & Technology Infrastructure of China (http://www.nesdc.org.cn). The DEM data (Su, 2025) and the boundary data of the Yangtze river basin (Zhang, 2019; Zhang et al., 2013) are provided by National Tibetan Plateau/Third Pole Environment Data Center (http://data.tpdc.ac.cn). The long-term sampling information for the GGF site is available at: https://ggf.cern.ac.cn/.

Supplement link

The link to the supplement will be included by Copernicus, if applicable.

Author contributions

TW: Methodology, Formal analysis, Visualization, Data curation, and Writing (original draft preparation). XY: Conceptualization, Funding acquisition, Project administration, Supervision, Resources, and Writing (review and editing). SZ: Conceptualization, Software, Resources. YD: Software, Resources.

Competing interests

The authors declare that they have no conflict of interest.

Disclaimer

Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.

Acknowledgements

The authors wish to thank other members of the Land-Atmosphere Interaction Research Group at Sun Yat-sen University for their valuable feedback about CoLM applications and simulations. Special thanks are also extended to Charles Tong for his valuable guidance on the implementation of PSUADE.

Financial support

This work was supported by Guangdong Provincial Observation and Research Station for Coupled Human and Natural Systems in Land–Ocean Interaction Zone (2024B1212040003) and the Western Light-Key Laboratory Cooperative Research Cross-Team Project of Chinese Academy of Sciences (xbzg-zdsys-202309).

References

Akhter, T., Pokhrel, Y., Felfelani, F., Ducharne, A., Lo, M.-H., and Reinecke, R.: Implications of Lateral Groundwater Flow Across Varying Spatial Resolutions in Global Land Surface Modeling, Water Resour. Res., 61, e2024WR038523, https://doi.org/10.1029/2024WR038523, 2025.

Bai, F., Wei, Z., Wei, N., Lu, X., Yuan, H., Zhang, S., Liu, S., Zhang, Y., Li, X., and Dai, Y.: Global Assessment of Atmospheric Forcing Uncertainties in The Common Land Model 2024 Simulations, J. Geophys. Res.-Atmos., 129, e2024JD041520, https://doi.org/10.1029/2024JD041520, 2024.

Bonan, G. B.: Land surface model (LSM version 1.0) for ecological, hydrological, and atmospheric studies: Technical description and user`s guide, University Corporation for Atmospheric Research, https://doi.org/10.5065/D6DF6P5X, 1996.

Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J.: Classification and regression trees (1st ed.), The Wadsworth and Brooks-Cole statistics-probability series, Taylor & Francis, https://doi.org/10.1378/chest.104.5.1359, 1984.

Chen, S., Li, L., Wei, Z., Wei, N., Zhang, Y., Zhang, S., Yuan, H., Wei, S., Zhang, S., Li, Q., and Dai, Y.: Exploring Topography Downscaling Methods for Hyper-Resolution Land Surface Modeling, J. Geophys. Res.-Atmos., 129, e2024JD041338, https://doi.org/10.1029/2024jd041338, 2024.

Chipman, H. A., George, E. I., and McCulloch, R. E.: BART: Bayesian additive regression trees, Ann. Appl. Stat., 4, 266-298, https://doi.org/10.1214/09-AOAS285, 2010.

Choat, B., Jansen, S., Brodribb, T. J., Cochard, H., Delzon, S., Bhaskar, R., Bucci, S. J., Feild, T. S., Gleason, S. M., Hacke, U. G., Jacobsen, A. L., Lens, F., Maherali, H., Martínez-Vilalta, J., Mayr, S., Mencuccini, M., Mitchell, P. J., Nardini, A., Pittermann, J., Pratt, R. B., Sperry, J. S., Westoby, M., Wright, I. J., and Zanne, A. E.: Global convergence in the vulnerability of forests to drought, Nature, 491, 752-755, https://doi.org/10.1038/nature11688, 2012.

Colbeck, S. C.: The capillary effects on water percolation in homogeneous snow, Journal of Glaciology, 13, 85-97, https://doi.org/10.3189/S002214300002339X, 1974.

Corona, C. R. and Ge, S.: Examining subsurface response to an extreme precipitation event using HYDRUS-1D, Vadose Zone J., 21, e20189, https://doi.org/10.1002/vzj2.20189, 2022.

Dai, Y. and Zeng, Q.: A land surface model (IAP94) for climate studies part I: Formulation and validation in off-line experiments, Adv. Atmos. Sci., 14, 433-460, https://doi.org/10.1007/s00376-997-0063-4, 1997.

Dai, Y., Zhang, S., Yuan, H., and Wei, N.: Modeling Variably Saturated Flow in Stratified Soils With Explicit Tracking of Wetting Front and Water Table Locations, Water Resour. Res., 55, 7939-7963, https://doi.org/10.1029/2019wr025368, 2019a.

Dai, Y., Wei, N., Yuan, H., Zhang, S., Shangguan, W., Liu, S., Lu, X., and Xin, Y.: Evaluation of Soil Thermal Conductivity Schemes for Use in Land Surface Modeling, J. Adv. Model. Earth Syst., 11, 3454-3473, https://doi.org/10.1029/2019MS001723, 2019b.

Dai, Y., Yuan, H., Xin, Q., Wang, D., Shangguan, W., Zhang, S., Liu, S., and Wei, N.: Different representations of canopy structure—A large source of uncertainty in global land surface modeling, Agric. For. Meteorol., 269-270, 119-135, https://doi.org/10.1016/j.agrformet.2019.02.006, 2019c.

Dai, Y., Xin, Q., Wei, N., Zhang, Y., Shangguan, W., Yuan, H., Zhang, S., Liu, S., and Lu, X.: A Global High-Resolution Data Set of Soil Hydraulic and Thermal Properties for Land Surface Modeling, J. Adv. Model. Earth Syst., 11, 2996-3023, https://doi.org/10.1029/2019MS001784, 2019d.

Dai, Y., Zeng, X., Dickinson, R. E., Baker, I., Bonan, G. B., Bosilovich, M. G., Denning, A. S., Dirmeyer, P. A., Houser, P. R., Niu, G., Oleson, K. W., Schlosser, C. A., and Yang, Z.-L.: The Common Land Model, Bull. Amer. Meteorol. Soc., 84, 1013-1024, https://doi.org/10.1175/BAMS-84-8-1013, 2003.

Decker, M.: Development and evaluation of a new soil moisture and runoff parameterization for the CABLE LSM including subgrid-scale processes, J. Adv. Model. Earth Syst., 7, 1788-1809, https://doi.org/10.1002/2015MS000507, 2015.

Dickinson, R. E., Henderson-Sellers, A., and Kennedy, P. J.: Biosphere atmosphere Transfer Scheme (BATS)Version 1e as Coupled to the NCAR Community Climate Model, University Corporation for Atmospheric Research, https://doi.org/10.5065/D67W6959, 1993.

Ding, B., Yang, K., Qin, J., Wang, L., Chen, Y., and He, X.: The dependence of precipitation types on surface elevation and meteorological conditions and its parameterization, J. Hydrol., 513, 154-163, https://doi.org/10.1016/j.jhydrol.2014.03.038, 2014.

Dong, W., Yuan, H., Zhang, R., Li, H., Huang, L., Zhu, S., Peng, J., and Dai, Y.: Effects of Incorporating Measured Leaf Optical Properties in Land Surface Models, Frontiers in Earth Science, 9, , https://doi.org/10.3389/feart.2021.663917, 2021.
Duan, Q., Sorooshian, S., and Gupta, V.: Effective and efficient global optimization for conceptual rainfall-runoff models, Water Resour. Res., 28, 1015-1031, https://doi.org/10.1029/91WR02985, 1992.
Duan, Q., Sorooshian, S., and Gupta, V. K.: Optimal use of the SCE-UA global optimization method for calibrating watershed models, J. Hydrol., 158, 265-284, https://doi.org/10.1016/0022-1694(94)90057-4, 1994.
Duan, Q. Y., Gupta, V. K., and Sorooshian, S.: Shuffled complex evolution approach for effective and efficient global minimization, J. Optim. Theory Appl., 76, 501-521, https://doi.org/10.1007/BF00939380, 1993.
Duc, L. and Sawada, Y.: A signal-processing-based interpretation of the Nash–Sutcliffe efficiency, Hydrol. Earth Syst. Sci., 27, 1827-1839, https://doi.org/10.5194/hess-27-1827-2023, 2023.
Erdal, D. and Cirpka, O. A.: Global sensitivity analysis and adaptive stochastic sampling of a subsurface-flow model using active subspaces, Hydrol. Earth Syst. Sci., 23, 3787-3805, https://doi.org/10.5194/hess-23-3787-2019, 2019.
Fan, J., McConkey, B., Wang, H., and Janzen, H.: Root distribution by depth for temperate agricultural crops, Field Crop. Res., 189, 68-74, https://doi.org/10.1016/j.fcr.2016.02.013, 2016.
Fan, J., Oestergaard, K. T., Guyot, A., and Lockington, D. A.: Estimating groundwater recharge and evapotranspiration from water table fluctuations under three vegetation covers in a coastal sandy aquifer of subtropical Australia, J. Hydrol., 519, 1120-1129, https://doi.org/10.1016/j.jhydrol.2014.08.039, 2014.
Fan, Y., Li, H., and Miguez-Macho, G.: Global Patterns of Groundwater Table Depth, Science, 339, 940-943, https://doi.org/10.1126/science.1229881, 2013.
Friedman, J. H.: Multivariate Adaptive Regression Splines, Ann. Stat., 19, 1-67, https://doi.org/10.1214/aos/1176347963, 1991.
Gan, Y., Duan, Q., Gong, W., Tong, C., Sun, Y., Chu, W., Ye, A., Miao, C., and Di, Z.: A comprehensive evaluation of various sensitivity analysis methods: A case study with a hydrological model, Environ. Modell. Softw., 51, 269-285, https://doi.org/10.1016/j.envsoft.2013.09.031, 2014.
Gao, J., Zhang, D., Niu, J., Yu, X., and Zhang, W.: Structural characteristics of Abies fabri forests at the upper reach of Yangtze River (in Chinese), Chinese J. Appl. Ecol., 20-24, https://www.cjae.net/EN/Y2003/V/I1/20, 2003.
Gill, M. K., Kaheil, Y. H., Khalil, A., McKee, M., and Bastidas, L.: Multiobjective particle swarm optimization for parameter estimation in hydrology, Water Resour. Res., 42, https://doi.org/10.1029/2005WR004528, 2006.
Gong, W., Duan, Q., Li, J., Wang, C., Di, Z., Dai, Y., Ye, A., and Miao, C.: Multi-objective parameter optimization of common land model using adaptive surrogate modeling, Hydrol. Earth Syst. Sci., 19, 2409-2425, https://doi.org/10.5194/hess-19-2409-2015, 2015.
Gou, S., Miller, G. R., Saville, C., Maxwell, R. M., and Ferguson, I. M.: Simulating groundwater uptake and hydraulic redistribution by phreatophytes in a high-resolution, coupled subsurface-land surface model, Adv. Water Resour., 121, 245-262, https://doi.org/10.1016/j.advwatres.2018.08.008, 2018.
Guerrero, J.-L., Pernica, P., Wheater, H., Mackay, M., and Spence, C.: Parameter sensitivity analysis of a 1-D cold region lake model for land-surface schemes, Hydrol. Earth Syst. Sci., 21, 6345-6362, https://doi.org/10.5194/hess-21-6345-2017, 2017.
Gulden, L. E., Rosero, E., Yang, Z.-L., Rodell, M., Jackson, C. S., Niu, G.-Y., Yeh, P. J. F., and Famiglietti, J.: Improving land-surface model hydrology: Is an explicit aquifer model better than a deeper soil profile?, Geophys. Res. Lett., 34, https://doi.org/10.1029/2007GL029804, 2007.
He, P., Ye, Q., Yu, K., Liu, X., Liu, H., Liang, X., Zhu, S., Wang, H., Yan, J., Wang, Y.-P., and Wright, I. J.: Relationship between wind speed and plant hydraulics at the global scale, Nat. Ecol. Evol., 9, 273-281, https://doi.org/10.1038/s41559-024-02603-5, 2025.
Hu, W., Ma, W., Yang, Z. L., Ma, Y., and Xie, Z.: Sensitivity Analysis of the Noah-MP Land Surface Model for Soil Hydrothermal Simulations Over the Tibetan Plateau, J. Adv. Model. Earth Syst., 15, https://doi.org/10.1029/2022ms003136, 2023.
Huang, L., Zhang, S., Niu, G.-Y., Wei, N., Yuan, H., Wei, Z., Lu, X., Peng, J., Li, W., and Dai, Y.: A Catchment-Based Hierarchical Spatial Tessellation Approach to a Better Representation of Land Heterogeneity for Hyper-Resolution Land Surface Modeling, Water Resour. Res., 58, e2021WR031589, https://doi.org/10.1029/2021WR031589, 2022.

Ippisch, O., Vogel, H. J., and Bastian, P.: Validity limits for the van Genuchten–Mualem model and implications for parameter estimation and numerical simulation, Adv. Water Resour., 29, 1780-1789, https://doi.org/10.1016/j.advwatres.2005.12.011, 2006.
Jia, Z., Chen, S., Fu, Y. H., Martín Belda, D., Wårlind, D., Olin, S., Xu, C., and Tang, J.: Advancing ecohydrological modelling: coupling LPJ-GUESS with ParFlow for integrated vegetation and surface-subsurface hydrology simulations, Geosci. Model Dev., 19, 1727-1747, https://doi.org/10.5194/gmd-19-1727-2026, 2026.
Jiang, Y., Tang, W., Yang, K., He, J., Shao, C., Zhou, X., Lu, H., Chen, Y., Li, X., and Shi, J.: Development of a high-resolution near-surface meteorological forcing dataset for the Third Pole region, Science China Earth Sciences, 68, 1274-1290, https://doi.org/10.1007/s11430-024-1507-6, 2025.
Jiang, Y., Yang, K., Qi, Y., Zhou, X., He, J., Lu, H., Li, X., Chen, Y., Li, X., Zhou, B., Mamtimin, A., Shao, C., Ma, X., Tian, J., and Zhou, J.: TPHiPr: a long-term (1979–2020) high-accuracy precipitation dataset (1⁄30°, daily) for the Third Pole region based on high-resolution atmospheric modeling and dense observations, Earth Syst. Sci. Data, 15, 621-638, https://doi.org/10.5194/essd-15-621-2023, 2023.
Kavetski, D., Qin, Y., and Kuczera, G.: The Fast and the Robust: Trade-Offs Between Optimization Robustness and Cost in the Calibration of Environmental Models, Water Resour. Res., 54, 9432-9455, https://doi.org/10.1029/2017WR022051, 2018.
Kennedy, D., Swenson, S., Oleson, K. W., Lawrence, D. M., Fisher, R., Lola da Costa, A. C., and Gentine, P.: Implementing Plant Hydraulics in the Community Land Model, Version 5, J. Adv. Model. Earth Syst., 11, 485-513, https://doi.org/10.1029/2018MS001500, 2019.
Kennedy, J. and Eberhart, R.: Particle swarm optimization, Proceedings of the IEEE International Conference on Neural Networks, 27 Nov.-1 Dec. 1995, 1942-1948, http://dx.doi.org/10.1109/ICNN.1995.488968,
Lawrence, D. M., Thornton, P. E., Oleson, K. W., and Bonan, G. B.: The Partitioning of Evapotranspiration into Transpiration, Soil Evaporation, and Canopy Evaporation in a GCM: Impacts on Land–Atmosphere Interaction, J. Hydrometeorol., 8, 862-880, https://doi.org/10.1175/JHM596.1, 2007.
Li, H., Lu, X., Wei, Z., Zhu, S., Wei, N., Zhang, S., Yuan, H., Shangguan, W., Liu, S., Zhang, S., Huang, J., and Dai, Y.: New Representation of Plant Hydraulics Improves the Estimates of Transpiration in Land Surface Model, Forests, 12, 722, https://doi.org/10.3390/f12060722, 2021.
Li, J., Duan, Q. Y., Gong, W., Ye, A., Dai, Y., Miao, C., Di, Z., Tong, C., and Sun, Y.: Assessing parameter importance of the Common Land Model based on qualitative and quantitative sensitivity analysis, Hydrol. Earth Syst. Sci., 17, 3279-3293, https://doi.org/10.5194/hess-17-3279-2013, 2013.
Liang, X., Xie, Z., and Huang, M.: A new parameterization for surface and groundwater interactions and its impact on water budgets with the variable infiltration capacity (VIC) land surface model, J. Geophys. Res.-Atmos., 108, https://doi.org/10.1029/2002JD003090, 2003.
Lin, W., Yuan, H., Dong, W., Zhang, S., Liu, S., Wei, N., Lu, X., Wei, Z., Hu, Y., and Dai, Y.: Reprocessed MODIS Version 6.1 Leaf Area Index Dataset and Its Evaluation for Land Surface and Climate Modeling, Remote Sens., 15, 1780, https://doi.org/10.3390/rs15071780, 2023.
Lundin, L.-C.: Hydraulic properties in an operational model of frozen soil, J. Hydrol., 118, 289-310, https://doi.org/10.1016/0022-1694(90)90264-X, 1990.
Luo, X., Liang, X., and Lin, J.-S.: Plant transpiration and groundwater dynamics in water-limited climates: Impacts of hydraulic redistribution, Water Resour. Res., 52, 4416-4437, https://doi.org/10.1002/2015WR017316, 2016.
Maina, F. Z. and Siirila-Woodburn, E. R.: The Role of Subsurface Flow on Evapotranspiration: A Global Sensitivity Analysis, Water Resour. Res., 56, e2019WR026612, https://doi.org/10.1029/2019WR026612, 2020.
Maina, F. Z., Siirila-Woodburn, E. R., and Dennedy-Frank, P.-J.: Assessing the impacts of hydrodynamic parameter uncertainties on simulated evapotranspiration in a mountainous watershed, J. Hydrol., 608, 127620, https://doi.org/10.1016/j.jhydrol.2022.127620, 2022.
Marchionni, V., Daly, E., Manoli, G., Tapper, N. J., Walker, J. P., and Fatichi, S.: Groundwater Buffers Drought Effects and Climate Variability in Urban Reserves, Water Resour. Res., 56, e2019WR026192, https://doi.org/10.1029/2019WR026192, 2020.
Maxwell, R. M. and Miller, N. L.: Development of a Coupled Land Surface and Groundwater Model, J. Hydrometeorol., 6, 233-247, https://doi.org/10.1175/JHM422.1, 2005.

Morris, M. D.: Factorial Sampling Plans for Preliminary Computational Experiments, Technometrics, 33, 161-174, https://doi.org/10.1080/00401706.1991.10484804, 1991.
Muñoz-Vega, E., Bogena, H. R., Richard-Cerda, J. C., Maxwell, R. M., and Schulz, S.: Influence of Hydraulic Conductivity Conceptualization and Unsaturated Flow Parameters for an Integrated Hydrological Model, Water Resour. Res., 61, e2025WR040785, https://doi.org/10.1029/2025WR040785, 2025.
Neufeld, H. S., Grantz, D. A., Meinzer, F. C., Goldstein, G., Crisosto, G. M., and Crisosto, C.: Genotypic Variability in Vulnerability of Leaf Xylem to Cavitation in Water-Stressed and Well-Irrigated Sugarcane 1, Plant Physiol., 100, 1020-1028, https://doi.org/10.1104/pp.100.2.1020, 1992.
Niu, G.-Y. and Yang, Z.-L.: Effects of Frozen Soil on Snowmelt Runoff and Soil Water Storage at a Continental Scale, J. Hydrometeorol., 7, 937-952, https://doi.org/10.1175/JHM538.1, 2006.
Niu, G.-Y., Yang, Z.-L., Dickinson, R. E., Gulden, L. E., and Su, H.: Development of a simple groundwater model for use in climate models and evaluation with Gravity Recovery and Climate Experiment data, J. Geophys. Res.-Atmos., 112, https://doi.org/10.1029/2006JD007522, 2007.
Niu, G.-Y., Yang, Z.-L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., Kumar, A., Manning, K., Niyogi, D., Rosero, E., Tewari, M., and Xia, Y.: The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements, J. Geophys. Res.-Atmos., 116, https://doi.org/10.1029/2010JD015139, 2011.
Niu, G. Y., Yang, Z. L., Dickinson, R. E., and Gulden, L. E.: A simple TOPMODEL-based runoff parameterization (SIMTOP) for use in global climate models, J. Geophys. Res.-Atmos., 110, https://doi.org/10.1029/2005jd006111, 2005.
O'Connor, J., Santos, M. J., Rebel, K. T., and Dekker, S. C.: The influence of water table depth on evapotranspiration in the Amazon arc of deforestation, Hydrol. Earth Syst. Sci., 23, 3917-3931, 10.5194/hess-23-3917-2019, 2019.
Pathak, R., Sahany, S., and Mishra, S. K.: Uncertainty quantification based cloud parameterization sensitivity analysis in the NCAR community atmosphere model, Sci Rep, 10, https://doi.org/10.1038/s41598-020-74441-x, 2020.
Peters, A. and Durner, W.: Improved estimation of soil water retention characteristics from hydrostatic column experiments, Water Resour. Res., 42, https://doi.org/10.1029/2006WR004952, 2006.
Pardo Lara, R., Berg, A. A., Warland, J., and Tetlock, E.: In Situ Estimates of Freezing/Melting Point Depression in Agricultural Soils Using Permittivity and Temperature Measurements, Water Resour. Res., 56, e2019WR026020, https://doi.org/10.1029/2019WR026020, 2020.Powell, M. J. D.: The BOBYQA algorithm for bound constrained optimization without derivatives, Department of Applied Mathematics and Theoretical Physics, Cambridge University, 2009.
Qin, Y., Kavetski, D., Kuczera, G., McInerney, D., Yang, T., and Guo, Y.: Can Gauss-Newton Algorithms Outperform Stochastic Optimization Algorithms When Calibrating a Highly Parameterized Hydrological Model? A Case Study Using SWAT, Water Resour. Res., 58, e2021WR031532, https://doi.org/10.1029/2021WR031532, 2022.
Raoult, N., Douglas, N., MacBean, N., Kolassa, J., Quaife, T., Roberts, A. G., Fisher, R., Fer, I., Bacour, C., Dagon, K., Hawkins, L., Carvalhais, N., Cooper, E., Dietze, M. C., Gentine, P., Kaminski, T., Kennedy, D., Liddy, H. M., Moore, D. J. P., Peylin, P., Pinnington, E., Sanderson, B., Scholze, M., Seiler, C., Smallman, T. L., Vergopolan, N., Viskari, T., Williams, M., and Zobitz, J.: Parameter Estimation in Land Surface Models: Challenges and Opportunities With Data Assimilation and Machine Learning, J. Adv. Model. Earth Syst., 17, e2024MS004733, https://doi.org/10.1029/2024ms004733, 2025.
Ratto, M., Castelletti, A., and Pagano, A.: Emulation techniques for the reduction and sensitivity analysis of complex environmental models, Environ. Modell. Softw., 34, 1-4, https://doi.org/10.1016/j.envsoft.2011.11.003, 2012.
Raupach, M. R.: Drag and drag partition on rough surfaces, Bound.-Layer Meteor., 60, 375-395, https://doi.org/10.1007/BF00155203, 1992.
Raupach, M. R.: Simplified expressions for vegetation roughness length and zero-plane displacement as functions of canopy height and area index, Bound.-Layer Meteor., 71, 211-216, https://doi.org/10.1007/BF00709229, 1994.
Rezaei, M., Seuntjens, P., Joris, I., Boënne, W., Van Hoey, S., Campling, P., and Cornelis, W. M.: Sensitivity of water stress in a two-layered sandy grassland soil to variations in groundwater depth and soil hydraulic parameters, Hydrol. Earth Syst. Sci., 20, 487-503, https://doi.org/10.5194/hess-20-487-2016, 2016.
Rickard, W., Hossain, I., Zhang, X., Cooper, H. V., Mooney, S. J., Hawkesford, M. J., and Whalley, W. R.: Field plants strategically regulate water uptake from different soil depths by spatiotemporally adjusting their radial root hydraulic conductivity, New Phytologist, 247, 546-561, https://doi.org/10.1111/nph.70013, 2025.

Rohde, M. M., Stella, J. C., Singer, M. B., Roberts, D. A., Caylor, K. K., and Albano, C. M.: Establishing ecological thresholds and targets for groundwater management, Nat. Water, 2, 312-323, https://doi.org/10.1038/s44221-024-00221-w, 2024.
Schaap, M. G. and van Genuchten, M. T.: A modified Mualem-van Genuchten formulation for improved description of the hydraulic conductivity near saturation, Vadose Zone J., 5, 27-34, https://doi.org/10.2136/vzj2005.0005, 2006.
Schenk, H. J. and Jackson, R. B.: THE GLOBAL BIOGEOGRAPHY OF ROOTS, Ecological Monographs, 72, 311-328, https://doi.org/10.1890/0012-9615(2002)072[0311:TGBOR]2.0.CO;2, 2002.
Shangguan, W., Dai, Y., Duan, Q., Liu, B., and Yuan, H.: A global soil data set for earth system modeling, J. Adv. Model. Earth Syst., 6, 249-263, https://doi.org/10.1002/2013MS000293, 2014a.
Shangguan, W., Dai, Y., Duan, Q., Liu, B., and Yuan, H.: The Global Soil Dataset for Earth System Modeling, The Land-Atmosphere Interaction Research Group [dataset], http://globalchange.bnu.edu.cn/research/data, 2014b.
Shao, C., Yang, K., Tang, W., He, Y., Jiang, Y., Lu, H., Fu, H., and Zheng, J.: Convolutional neural network-based homogenization for constructing a long-term global surface solar radiation dataset, Renew. Sust. Energ. Rev., 169, 112952, https://doi.org/10.1016/j.rser.2022.112952, 2022.
Shaw, R. H. and Pereira, A. R.: Aerodynamic roughness of a plant canopy: A numerical experiment, Agricultural Meteorology, 26, 51-65, https://doi.org/10.1016/0002-1571(82)90057-7, 1982.
Silwimba, K., Flores, A. N., Cionni, I., Billings, S. A., Sullivan, P. L., Ajami, H., Hirmas, D. R., and Li, L.: Soil parameterization in land surface models drives large discrepancies in soil moisture predictions across hydrologically complex regions of the contiguous United States, Geosci. Model Dev., 18, 7707-7734, https://doi.org/10.5194/gmd-18-7707-2025, 2025.
Smerdon, B. D., Mendoza, C. A., and Devito, K. J.: Influence of subhumid climate and water table depth on groundwater recharge in shallow outwash aquifers, Water Resour. Res., 44, https://doi.org/10.1029/2007WR005950, 2008.
Sobaga, A., Decharme, B., Habets, F., Delire, C., Enjelvin, N., Redon, P. O., Faure-Catteloin, P., and Le Moigne, P.: Assessment of the interactions between soil–biosphere–atmosphere (ISBA) land surface model soil hydrology, using four closed-form soil water relationships and several lysimeters, Hydrol. Earth Syst. Sci., 27, 2437-2461, https://doi.org/10.5194/hess-27-2437-2023, 2023.
Sobol, I. M.: Estimation of the sensitivity of nonlinear mathematical models, Mat. Model., 2, 112-118, 1990.
Sobol, I. M.: Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, Institute for Mathematical Modelling, Russian Academy of Sciences, 4 Miusskaya Square, Moscow 125047, Russian Federation, 55, 271-280, https://doi.org/10.1016/s0378-4754(00)00270-6, 2001.
Su, X.: 90m elevation data of Qinghai Tibet Plateau, National Tibetan Plateau / Third Pole Environment Data Center [dataset], 2025.
Sun, H. Y., Wu, Y. H., Yu, D., and Zhou, J.: Altitudinal Gradient of Microbial Biomass Phosphorus and Its Relationship with Microbial Biomass Carbon, Nitrogen, and Rhizosphere Soil Phosphorus on the Eastern Slope of Gongga Mountain, SW China, PLOS ONE, 8, https://doi.org/10.1371/journal.pone.0072952, 2013.
Swenson, S. C. and Lawrence, D. M.: A new fractional snow-covered area parameterization for the Community Land Model and its effect on the surface energy balance, J. Geophys. Res.-Atmos., 117, https://doi.org/10.1029/2012JD018178, 2012.
Tian, Z., Gao, W., Kool, D., Ren, T., Horton, R., and Heitman, J. L.: Approaches for Estimating Soil Water Retention Curves at Various Bulk Densities With the Extended Van Genuchten Model, Water Resour. Res., 54, 5584-5601, https://doi.org/10.1029/2018wr022871, 2018.
Todt, M., Rutter, N., Fletcher, C. G., and Wake, L. M.: Simulated single-layer forest canopies delay Northern Hemisphere snowmelt, The Cryosphere, 13, 3077-3091, https://doi.org/10.5194/tc-13-3077-2019, 2019.
Tong, C.: Problem Solving Environment for Uncertainty Analysis and Design Exploration, in: Handbook of Uncertainty Quantification, edited by: Ghanem, R., Higdon, D., and Owhadi, H., Springer International Publishing, Cham, 1-37, https://doi.org/10.1007/978-3-319-11259-6_53-1, 2016.
Tong, C.: Problem Solving Environment for Uncertainty Analysis and Design Exploration, PSUADE Short Manual (Version 2), Center for Applied Scientic Computing Lawrence Livermore National Laboratory Livermore, 1695-1731 pp., https://doi.org/10.1007/978-3-319-12385-1_53, 2022a.
Tong, C.: PSUADE Short Manual (Version 2), Lawrence Livermore National Laboratory, Livermore, CA, 2022b.

Tsai, W., Feng, D., Pan, M., Beck, H., Lawson, K., Yang, Y., Liu, J., and Shen, C.: From calibration to parameter learning: Harnessing the scaling effects of big data in geoscientific modeling, Nat. Commun., 12, 5988, https://doi.org/10.1038/s41467-021-26107-z, 2021.
Vogelbacher, A., Aminzadeh, M., Madani, K., and Shokri, N.: An Analytical Framework to Investigate Groundwater-Atmosphere Interactions Influenced by Soil Properties, Water Resour. Res., 60, e2023WR036643, https://doi.org/10.1029/2023WR036643, 2024.
Wu, T., Zhang, S., Yang, X., and Dai, y.: Quantifying key parameter sensitivities for water table depth in hydrological schemes of CoLM-PSUADE at a mountain forest site (v1), Zenodo [code], https://doi.org/10.5281/zenodo.23080567~~10.5281/zenodo.19641291~~, 2026.
Xin, Y., Dai, Y., Li, J., Rong, X., and Zhang, G.: Coupling the Common Land Model to ECHAM 5 Atmospheric General Circulation Model, J. Meteorol. Res., 33, 251-263, https://doi.org/10.1007/s13351-019-8117-y, 2019.
Yang, C., Sun, A., Zhang, S., Dai, Y., Kollet, S., and Maxwell, R.: 20 years of trials and insights: bridging legacy and next generation in ParFlow and Land Surface Model Coupling, Geosci. Model Dev., 19, 1849-1866, https://doi.org/10.5194/gmd-19-1849-2026, 2026.
Yang, K., Jiang, Y., Tang, W., He, J., Shao, C., Zhou, X., Lu, H., Chen, Y., Li, X., and Shi, J.: A high-resolution near-surface meteorological forcing dataset for the Third Pole region (TPMFD, 1979-2023), National Tibetan Plateau / Third Pole Environment Data Center [dataset], https://doi.org/10.11888/Atmos.tpdc.300398, 2023.
Yang, Y., Li, W., and Wang, G.: A dataset of soil physical properties and elements of Abies fabri in Gongga Mountain, southeastern Qinghai-Tibet Plateau, during 2005-2015 [dataset], https://doi.org/10.11922/csdata.2020.0003.zh, 2020.
Yang, Z.-L., Niu, G.-Y., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., Longuevergne, L., Manning, K., Niyogi, D., Tewari, M., and Xia, Y.: The community Noah land surface model with multiparameterization options (Noah-MP): 2. Evaluation over global river basins, J. Geophys. Res.-Atmos., 116, https://doi.org/10.1029/2010JD015140, 2011.
Yuan, H., Dai, Y., Xiao, Z., Ji, D., and Shangguan, W.: Reprocessing the MODIS Leaf Area Index products for land surface and climate modelling, Remote Sens. Environ., 115, 1171-1187, https://doi.org/10.1016/j.rse.2011.01.001, 2011.
Zhan, C.-s., Song, X.-m., Xia, J., and Tong, C.: An efficient integrated approach for global sensitivity analysis of hydrological model parameters, Environ. Modell. Softw., 41, 39-52, https://doi.org/10.1016/j.envsoft.2012.10.009, 2013.
Zhang, G.: Dataset of river basins map over the TP (2016), National Tibetan Plateau / Third Pole Environment Data Center [dataset], https://doi.org/10.11888/BaseGeography.tpe.249465.file, 2019.
Zhang, G., Yao, T., Xie, H., Kang, S., and Lei, Y.: Increased mass over the Tibetan Plateau: From lakes or glaciers?, Geophys. Res. Lett., 40, 2125-2130, https://doi.org/10.1002/grl.50462, 2013.
Zhang, S., Liang, H., Li, F., Lu, X., and Dai, Y.: Representation of a two-way coupled irrigation system in the Common Land Model, Hydrol. Earth Syst. Sci., 29, 3119-3143, https://doi.org/10.5194/hess-29-3119-2025, 2025.
Zhang, X., Srinivasan, R., Zhao, K., and Liew, M. V.: Evaluation of global optimization algorithms for parameter calibration of a computationally intensive hydrologic model, Hydrol. Process., 23, 430-441, https://doi.org/10.1002/hyp.7152, 2009.
Zhao, G., Liu, G., Zhu, W., Zhao, J., Wang, X., Wang, Y., and Jia, M.: Stem CO2 efflux of Abies fabri in subalpine forests in the Gongga Mountains, Eastern Tibetan Plateau, Journal of Plant Ecology, 10, 1001-1011, https://doi.org/10.1093/jpe/rtw100, 2017.
Zheng, H., Yang, Z., Lin, P., Wei, J., Wu, W., Li, L., Zhao, L., and Wang, S.: On the Sensitivity of the Precipitation Partitioning Into Evapotranspiration and Runoff in Land Surface Parameterizations, Water Resour. Res., 55, 95-111, https://doi.org/10.1029/2017wr022236, 2019.
Zhu, S., Chen, H., Zhang, X., Wei, N., Shangguan, W., Yuan, H., Zhang, S., Wang, L., Zhou, L., and Dai, Y.: Incorporating root hydraulic redistribution and compensatory water uptake in the Common Land Model: Effects on site level and global land modeling, J. Geophys. Res.-Atmos., 122, 7308-7322, https://doi.org/10.1002/2016JD025744, 2017a.
Zhu, Y., Ren, L., Horton, R., Lü, H., Wang, Z., and Yuan, F.: Estimating the Contribution of Groundwater to the Root Zone of Winter Wheat Using Root Density Distribution Functions, Vadose Zone J., 17, 170075, https://doi.org/10.2136/vzj2017.04.0075, 2018.
Zhu, Z., Tang, X., Yuan, G., Zhang, X., Sun, X., Chang, X., Cheng, Y., Chu, G., Dai, G., Dou, S., Du, J., Fu, W., Guan, Y., Guo, Y., He, Q., Jiang, J., Jiang, Z., Lai, J., Li, M., Li, S., Li, W., Li, X., Lin, J., Lin, Y., Liu, W., Liu, X., Liu, Y., Lu, Z., Lu, Y., Sheng, Y., Su, H., Tang, J., Yang, F., Yin, C., Zhang, F., Zhang, Z., Zhao, C., Zhao, Y., and Zhu, M.: A dataset of

groundwater level in Chinese Ecosystem Research Network (2005–2014) [dataset], https://doi.org/10.11922/csdata.170.2016.0099, 2017b.