Authors' responses to RC1' comments

Revision Notes

For

Manuscript ID: **egusphere-2026-2275**

Quantifying key parameter sensitivities for water table depth in hydrological schemes of CoLM-PSUADE

Dear Editor,

We thank the reviewers and the editor for their insightful comments, which have helped us substantially improve the manuscript. We have carefully considered all comments and revised the manuscript accordingly. Specifically, we have improved the results descriptions, refined several figures and captions, and strengthened the discussion of uncertainties, limitations, and implications. In particular, we further clarified the scope and objectives of the study. This manuscript aims to identify the parameters that exert strong influences on water table depth simulations at the study site. To achieve this goal, we established a framework that couples a land surface model with a parameter analysis and optimization toolkit, integrating qualitative screening, quantitative sensitivity analysis, and parameter optimization approaches.

We previously submitted a general response summarizing our overall revision strategy, and now provide a point-by-point response. Reviewers' comments are **Relisted in black**, followed by our **Responses in blue** and **Revisions in purple**.

Reviewer #1:

The authors have developed and applied a CoLM2024–PSUADE coupling framework to examine 56 hydrologically relevant parameters using multiple sensitivity-analysis methods. They further integrate parameter screening, quantitative sensitivity analysis, and parameter optimization to evaluate and improve the simulation of water table depth at a mountain forest site. Overall, I appreciate the considerable effort invested in model

coupling, experimental design, and the subsequent analyses. The study addresses an important topic and presents a potentially useful workflow for sensitivity-guided calibration in land surface models. However, several aspects of the experimental design and the interpretation of the results would benefit from further clarification and strengthening before publication.

Response:

We thank Reviewer #1 for his/her positive feedback and constructive comments. The comments have been carefully considered and **Relisted in black**, followed by our **Responses in blue** and **Revisions in purple**. According to the comments, we modified the manuscript, particularly on clarifying the study scope and limitations, improving the robustness assessment of sensitivity analysis, strengthening the optimization experiment design and convergence evaluation, and enhancing the interpretation of optimized WTD simulations. Major changes are summarized as follows:

No.	Major Revisions	Suggested Comments
1	Clarification of study scope, limitations, and representativeness	Major comment 1, 2
2	Enhancement of sensitivity-analysis robustness and uncertainty assessment	Major comment 3
3	Improvement of optimization experiment design and discussion	Major comment 5
4	Improved interpretation of optimized WTD simulations and event-scale evaluation	Major comment 6

(1) Clarification of study scope, limitations, and representativeness. We have revised the manuscript to clarify the scope and applicability of the proposed framework and sensitivity-analysis results. The title, conclusions, and discussion have been modified to clarify that the identified sensitive parameters are representative of the Gongga Mountain forest site under specific hydroclimatic, soil, vegetation, and groundwater conditions. We have also expanded the discussion of limitation regarding single-site simulations, sensitivity-analysis period, sampling uncertainty, and the use of WTD as a single calibration target. Moreover, repeated random-seed experiments were incorporated to evaluate the robustness of qualitative sensitivity rankings. ***These revisions improve the interpretation of the results, prevent overgeneralization, and emphasize the potential extension of the framework toward multi-site and multi-variable applications.***

(2) Enhancement of sensitivity-analysis robustness and uncertainty assessment. We have incorporated data from two additional sensitivity-analysis experiments using different random seeds to evaluate the robustness of the identified sensitive parameters. The consistency of sensitive parameters was assessed using selection frequency and mean sensitivity indices among repeated experiments. We further expanded the discussion on sampling-related uncertainty and parameter variability across different realizations, providing a more comprehensive interpretation of qualitative sensitivity rankings. ***These additional analyses clarify the influence of stochastic sampling uncertainty on sensitivity rankings and strengthen the reliability of the identified WTD-related sensitive parameters.***

(3) Improvement of optimization experiment design and discussion. We have strengthened the optimization analysis by clarifying the experimental design and convergence criteria. The revised manuscript further clarifies that the optimization experiments were intended to evaluate the effectiveness of the identified sensitive parameters for improving WTD simulations. We have also added the limitations associated with fixed-budget optimization comparisons in the discussion. The convergence criteria implemented in PSUADE for SCE, PSO, and BOBYQA have also been explicitly described, together with their mathematical formulations and corresponding control parameters. Additional experiments with 1000 objective-function evaluations were conducted to examine the overall conclusions. ***These revisions improve the interpretation of the optimization experiments.***

(4) Improved interpretation of optimized WTD simulations and event-scale evaluation. We have incorporated event-scale evaluation and additional diagnostic analyses to strengthen the interpretation of optimized WTD simulations. Complementary metrics, including KGE and event-based RMSE, have been used to assess whether the optimized parameters improved not only overall WTD accuracy but also temporal variability and groundwater fluctuation dynamics. In addition, we expanded the discussion regarding the unresolved WTD fluctuation events by analyzing potential contributions from objective-function selection, meteorological forcing uncertainty, freeze–thaw process representation, observational uncertainty, and simplified groundwater parameterization. ***These revisions provide a more comprehensive understanding of optimized model behavior.***

Major Comments:

1. The manuscript emphasizes two main contributions: the development of a CoLM2024–PSUADE coupling tool and the robust sensitivity analysis of water table depth-related parameters. The coupling framework is useful from a technical and practical perspective. However, the sensitivity-analysis results for key parameters are derived from a single-point simulation at one mountain forest site. Although this site-specific analysis may provide useful insights into WTD simulation under the local hydroclimatic and soil conditions, it is difficult to infer that the identified sensitive parameters or their optimal ranges are generally applicable to other regions, land-cover types, climate regimes, or groundwater conditions. Therefore, I recommend that the authors either expand the analysis to additional sites or grid cells. If additional multi-site experiments are not feasible, the conclusions should be refocused on the current watershed/site. Thus, the current title of the manuscript is inappropriate.

Response:

We thank the reviewer for this suggestion. We agree that the sensitivity of land surface model parameters may depend on hydroclimatic conditions, soil properties, vegetation characteristics, and groundwater regimes. Following the suggestion, we tried to conduct additional sensitivity-analysis experiments at another site. However, because establishing hydrological equilibrium at a new site in CoLM requires a relatively long spin-up period, extending the sensitivity analysis to additional sites would substantially increase the computational cost and was beyond the scope of the current study. Accordingly, we have revised the manuscript to clarify that the identified sensitive parameters are specific to the investigated hydroclimatic conditions and soil properties. Specifically, we have revised the title and conclusions to better reflect the scope of this study and highlight that further multi-site or regional-scale assessments are needed to evaluate the transferability of these findings.

Revision:

(1) Title

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

(2) Conclusions (Pages 46–47, Lines 761–764, 768–771, 787–788, 790–793, and 795–799)

The revised descriptions are indicated in *italics*:

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.

Three qualitative sensitivity 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 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 behavior across different sampling configurations. Building on the 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 providing the most stable and effective convergence.

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.*

2. The manuscript states that the study was conducted at a long-term integrated observation site, which is an important advantage of the dataset. However, the sensitivity analysis appears to be based on three years of data (2005-2007). The length and representativeness of the analysis period may be critical because WTD dynamics and parameter sensitivity may vary substantially. I recommend that the authors extend the sensitivity analysis to cover a longer observation period, or at least justify why the selected three-year period is representative of the long-term hydrological conditions at the site.

Response:

We thank the reviewer for this suggestion. As suggested, we have clarified the reason for selecting the 2005–2007 period and provided additional analysis of the hydroclimatic representativeness of the selected years in Sect. S1 of the supplementary material. We have also revised the manuscript to explicitly distinguish the periods used for model spin-up, sensitivity analysis and parameter optimization, and long-term validation.

Specifically, water year 2005–2006 was used as the spin-up period to establish a dynamically equilibrated model state, while 2006–2007 was selected for sensitivity analysis because it represented an active hydrological period with relatively strong groundwater dynamics. For parameter optimization, 2005–2006 was again used for spin-up, and the 2006–2007 period was used to derive the optimized parameter set. The optimized parameter set was subsequently applied to the longer validation period (2007–2014). This validation period covered a broader range of hydroclimatic conditions, including relatively wet and dry years, thereby providing an independent evaluation of the identified sensitive parameters. Nevertheless, we acknowledge that the sensitivity analysis was performed primarily based on one representative hydrological year may lead to uncertainty. We have also revised the conclusions to clarify this limitation.

Revision:

(1) Sect. 2.1.2 of the Data and methods (Page 6, Lines 139–147)

The revised descriptions are indicated in *italics*:

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.*

(2) Conclusions (Pages 46–47, Lines 790–791 and 798–799)

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.*

(3) Sect. S1 of the supplementary material (Page 2, Lines 6–21)

S1: Representativeness of selected water years

The period 2005–2007 was used for sensitivity analysis and parameter optimization. Specifically, the 2005–2006 water year was treated as the spin-up period to allow the model states to approach dynamic equilibrium before analysis. The climatic conditions during this year were close to the long-term mean at the GGF site, with an annual mean air temperature difference of approximately 0.09 °C and an annual precipitation difference of approximately −64 mm relative to the long-term average. The 2006–2007 water year was selected for analysis because it represented a hydrological period with relatively shallow observed WTD and enhanced groundwater dynamics. Compared with the long-term mean conditions, the annual mean air temperature and precipitation during the 2006–2007 water year differed by approximately −0.20 °C and +84 mm, respectively. Such conditions provided sufficient variability in hydrological processes, including infiltration, runoff generation, groundwater recharge, and drainage, facilitating the diagnosis of parameters controlling WTD simulations. The subsequent validation period (2007–2014) included diverse hydroclimatic conditions, including relatively wet and dry years and different groundwater states, allowing an independent evaluation of the identified parameters under varying environmental conditions.

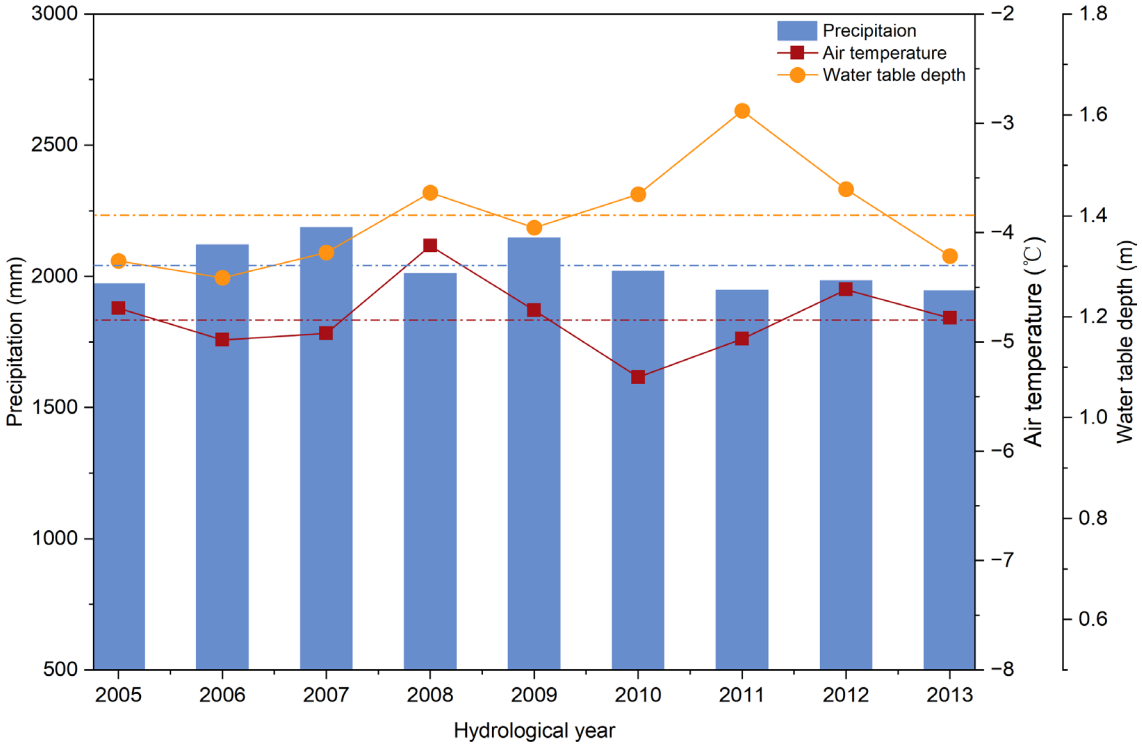


Figure S1: The total annual precipitation, annual average air temperature and annual average water table depth in the water years from 2005 to 2013. The dotted lines in the figure correspond to the average values over these years.

3. The qualitative sensitivity-analysis results show some variability across different methods, sampling techniques, sample sizes, and objective functions. In particular, Figures 7 and 8 suggest that 56-soil_alpha is the most consistently identified sensitive parameter, whereas the rankings of several other parameters vary among experimental configurations and with increasing sample size. This variability is understandable given the high dimensionality of the parameter space and the stochastic nature of the sampling procedures, but it would be helpful to further clarify the robustness of the identified sensitive parameters. I suggest that the authors consider adding repeated stochastic sensitivity-analysis experiments under the same settings, for example by using multiple random seeds for selected method–sampling technique–sample-size–objective-function combinations. The robustness of the sensitive parameters could then be evaluated using statistical indicators such as the mean and standard deviation of sensitivity indices, confidence intervals, coefficient of variation, selection frequency across repeated realizations.

Response:

We thank the reviewer for this suggestion. We agree that the stochastic nature of qualitative sensitivity analysis may introduce variability in the sensitivity rankings and that evaluating robustness across repeated realizations can improve the reliability of parameter screening. Following the suggestion, we performed two additional

independent sensitivity-analysis experiments using different random seeds while keeping all other experimental settings unchanged. Together with the original experiment, three independent realizations were obtained. These repeated experiments demonstrate that although moderate variations exist in the rankings of several secondary parameters, the dominant sensitive parameters remain highly stable across repeated realizations. In particular, 56-soil_alpha was consistently identified as the most sensitive parameter in all screening results. Other stable parameters identified in the original manuscript (45-krmax, 53-soil_n, 49-beta, 46-ck0, 10-rhol_nir, and 4-alpha_rain) also exhibited high selection frequencies across repeated experiments. However, parameter 40-rsubmax showed relatively lower selection consistency, indicating that its sensitivity is more dependent on stochastic sampling. Therefore, although qualitative sensitivity rankings exhibit some stochastic variability, the overall conclusions regarding the dominant sensitive parameters remain unchanged, confirming the robustness of the proposed screening strategy. The corresponding detailed descriptions have been added to the new Sect. 4.3 of the revised manuscript and Sect. S6 of the supplementary material.

Revision:

(1) Sect. 4.3 of the Discussion (Pages 42–44, Lines 699–725)

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.

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".

(2) Sect. S6 of the supplementary material (Pages 22–25, Lines 245–269)

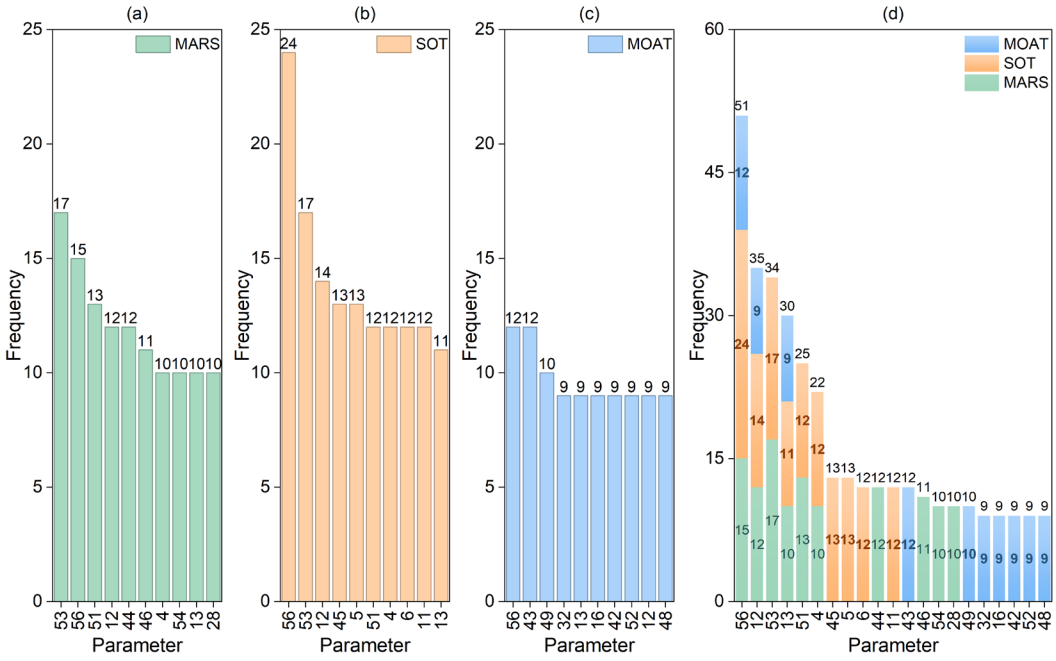


Figure S17: Sensitive parameters identified across various objectives and sample sizes under different sensitivity analysis methods based on experiment s1. The horizontal axis denotes parameter indices, while the vertical axis represents the frequency of occurrence. Labels above the bars indicate the 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.

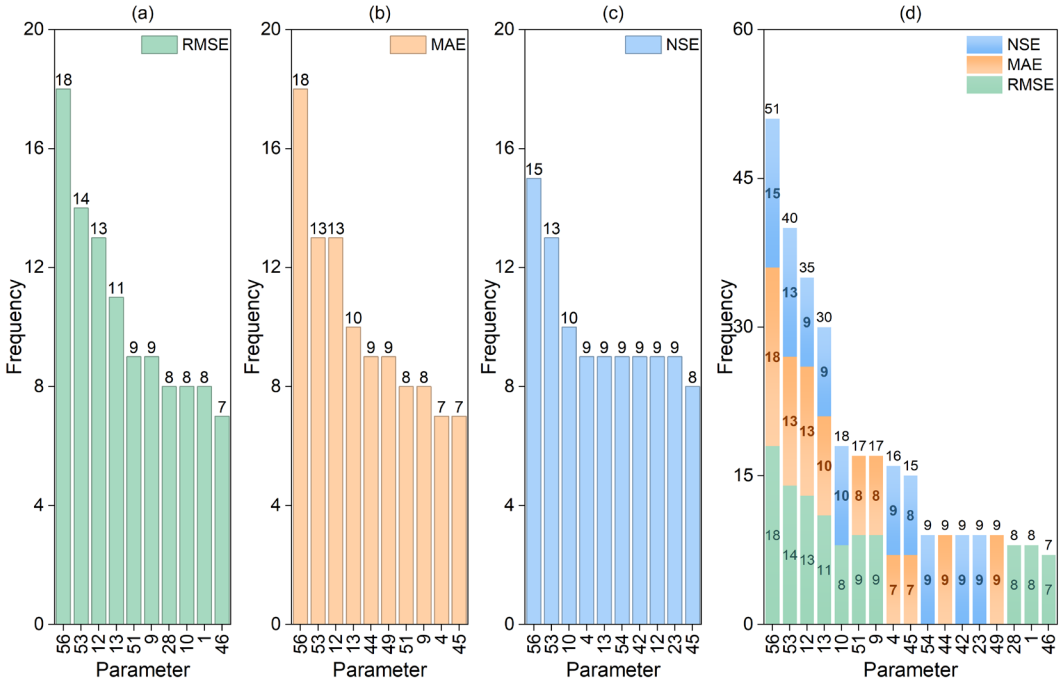


Figure S18: Sensitive parameters identified across various methods and sample sizes under different objective functions based on experiment s1. The horizontal axis denotes

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.

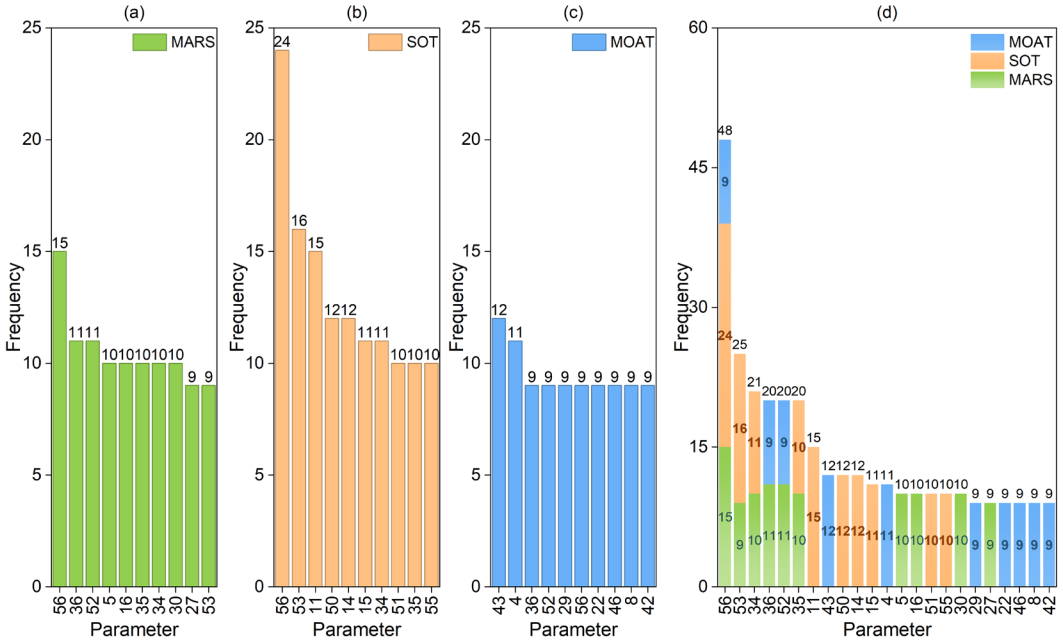


Figure S19: Sensitive parameters identified across various objectives and sample sizes under different sensitivity analysis methods based on experiment s2. The horizontal axis denotes parameter indices, while the vertical axis represents the frequency of occurrence. Labels above the bars indicate the 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.

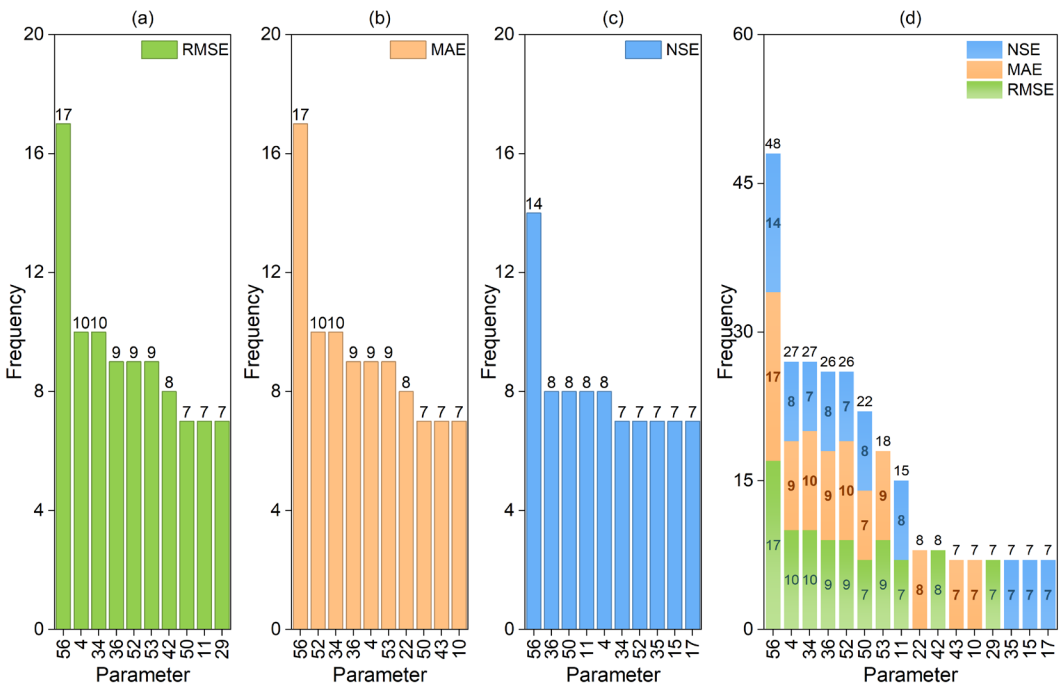


Figure S20: Sensitive parameters identified across various methods and sample sizes under different objective functions based on experiment s2. The horizontal axis denotes parameter indices, while the vertical axis represents the frequency of occurrence. Labels above the bars indicate the total frequency 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.

4. The role and notation of the subsurface runoff decay factor require further clarification. In Table 2, the 56 adjustable parameters include 38-fsatdcfa, which is described as a runoff decay parameter, but the later discussion refers to the decay factor as f_{drai}. It is therefore unclear whether fdrai is identical to 38-fsatdcfa or represents a separate parameter. I suggest that the authors clarify the relationship between 38-fsatdcfa and f_{drai}.

Response:

We thank the reviewer for pointing out this inconsistency. The parameter 38-fsatdcfa listed in Table 2 and the decay factor f_{drai} discussed in Sect. 4.2 represent the same subsurface runoff decay parameter in the SIMTOP scheme. The notation "38-fsatdcfa" corresponds to the original parameter name used in the CoLM model code, whereas f_{drai} was used in the later discussion to represent its physical meaning more clearly. To avoid ambiguity, we have revised the manuscript by explicitly clarifying this relationship in Sect. 4.2.

Revision:

Sect. 4.2 of the Discussion (Page 38, Line 635)

The revised descriptions are indicated in *italics*:

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 significantly enhanced simulation accuracy and captured WTD dynamics more effectively. 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.

5. The optimization experiment design and the interpretation of the optimized WTD simulations would benefit from further clarification. The manuscript compares SCE, PSO, and BOBYQA using a fixed budget of 500 evaluations, but it is not entirely clear whether this number is sufficient to assess optimizer convergence and robustness in a multi-parameter calibration problem. This is particularly relevant because the manuscript reports that PSO does not show a clear convergence trend, while BOBYQA exhibits stronger fluctuations after about 300 evaluations. I suggest that the authors provide a clearer justification for the 500-evaluation limit and, if computationally feasible, consider additional optimization experiments with larger evaluation budgets, for example 1000–2000 evaluations, or repeated runs using different random seeds. At minimum, the authors should discuss the potential limitations of comparing optimization methods under this relatively limited evaluation budget and report the convergence criteria more explicitly.

Response:

We thank the reviewer for this helpful suggestion. We agree that the optimization experiment should be described more explicitly, particularly regarding the computational budget, convergence criteria, and the limitations of comparing optimization algorithms under a fixed evaluation budget. In the revised manuscript, we have expanded the description of the optimization design, conducted additional experiments with an extended computational budget where feasible, and added a discussion regarding the interpretation and limitations of the optimization comparison. The major revisions are summarized as follows:

(1) Justification of the 500-evaluation budget. Because each objective-function evaluation requires one complete CoLM simulation, the optimization is computationally expensive. Therefore, we adopted a fixed-budget comparison, which represents a practical computational budget for model calibration. We further clarified

that the primary purpose of the optimization experiments was not to identify the global optimum of each optimization algorithm, but rather to evaluate whether the sensitive parameters identified through the screening procedure could substantially improve WTD simulations. Accordingly, a moderate evaluation budget was considered sufficient for validating the effectiveness of the sensitivity-analysis framework. The corresponding explanation has been added to Sect. 4.4 of the revised manuscript.

(2) Discussion of the limitations of a fixed evaluation budget. We have added a dedicated discussion describing the limitations of comparing optimization algorithms under a relatively limited computational budget. Specifically, we emphasize that optimization algorithms exhibit different convergence characteristics, and their relative performance may depend on the available number of objective-function evaluations. Algorithms that converge rapidly during the early search stage are not necessarily those achieving the best solutions after substantially larger computational budgets. Consequently, the present comparison should be interpreted as an assessment of the relative optimization efficiency of the three algorithms under a practically relevant computational budget, rather than as a definitive comparison of their ultimate optimization capability under unlimited objective-function evaluations. These limitations and their implications are now explicitly discussed in the Sect. 4.4 of the revised manuscript.

(3) Additional optimization experiments with a larger computational budget. Following the suggestion, we performed additional optimization experiments using 1000 objective-function evaluations for SCE and BOBYQA. For PSO, however, due to an unexpected interruption caused by a power outage during the additional experiment, the final number of valid objective-function evaluations did not reach 1000. In the original manuscript, the optimization trajectories were presented as the RMSE values obtained at each individual evaluation, which reflect not only the optimization progress but also the inherent search behaviors of different algorithms. To better characterize the convergence behavior of different optimization algorithms, we further presented the minimum RMSE obtained within each cumulative number of evaluations in the revised manuscript. SCE continues to exhibit the best optimization performance and converges to the lowest RMSE. PSO continues to exhibit oscillatory behavior during the search process, whereas BOBYQA achieves rapid improvement during the early search stage. Importantly, although larger evaluation budgets further refine the solutions, the overall ranking of the three optimization algorithms remains unchanged, indicating that the conclusions drawn from the original 500-evaluation experiments are qualitatively robust. The details are provided in the Sect. 3.3 of the revised manuscript.

(4) Explicit description of the convergence criteria. We have added a detailed description of the convergence criteria implemented in PSUADE. Specifically, we clarify that the optimization terminates when one of the following conditions is satisfied.

For the SCE algorithm, we now explicitly state that the optimization terminates when one of the following conditions is satisfied: (i) the maximum number of objective-function evaluations is reached; (ii) the normalized geometric range of the population decreases below the predefined parameter-space threshold; or (iii) the improvement in the best objective-function value becomes smaller than the user-defined optimization tolerance. Similarly, we have clarified the convergence criteria adopted by the BOBYQA. Specifically, the optimization was terminated when either (i) the maximum number of objective-function evaluations was reached, or (ii) the trust-region radius had been progressively reduced to the final trust-region radius. For the PSO algorithm, the optimization terminates when any of the following conditions is satisfied: (i) the maximum number of objective-function evaluations is reached; (ii) the current global-best objective-function value is within the user-specified tolerance of the prescribed target objective value; or (iii) the global-best objective-function value is not improved for consecutive particle evaluations. We further describe the corresponding PSUADE input parameters, provide the mathematical formulation of each convergence criterion, and explain how these parameters are implemented within the PSUADE optimization framework. The details are added to the Sect. S7 of the supplementary material.

Revision:

(1) Sect. 2.3.4 of the Data and methods (Page 17, Lines 299–303 and 310–312)

The revised descriptions are indicated in *italics*:

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.

(2) Sect. 3.3 of the Results (Page 29–30, Lines 499–517)

The revised descriptions are indicated in *italics*:

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. Overall, SCE outperformed PSO and BOBYQA in terms of convergence speed and solution quality, indicating that it is more suitable for parameter optimization in the CoLM.

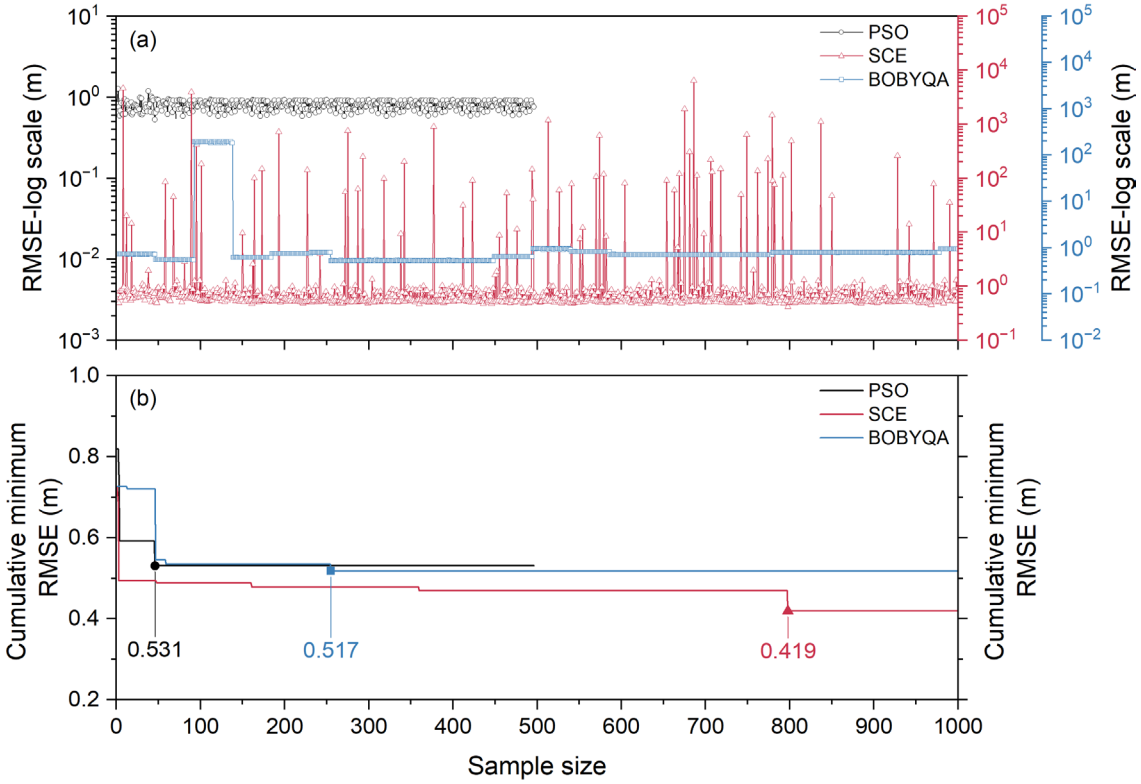


Figure 15. Evolution of (a) $RMSE$ and (b) cumulative minimum $RMSE$ with the number of evaluations for the SCE (1000 runs), PSO (496 runs), and BOBYQA (1000 runs) optimization methods. The vertical axis for subplot (a) is displayed on a logarithmic scale to accommodate the wide range of $RMSE$ values.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

(3) Sect. 4.4 of the Discussion (Page 45, Lines 726–743)

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 behaviors 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 behavior. 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).

(4) Sect. S7 of the supplementary material (Pages 25–27, Lines 270–337)

S7: Convergence criteria of the parameter optimization methods

Among the convergence criteria, the relevant user-defined settings in the PSUADE input file include *optimization max_feval*, *optimization* f_{min}, and *optimization tolerance*, which specify the maximum number of objective-function evaluations, the target objective-function value, and the convergence tolerance, respectively. Following the example settings provided in PSUADE and considering the actual performance of the WTD simulations, *optimization max_feval*, *optimization* f_{min}, and *optimization tolerance* were set to 10000, 0.4, and 1×10^{-4}, respectively.

S7.1 Implementation of SCE optimization convergence criteria in PSUADE

The convergence criteria for the SCE optimization are explicitly defined according to the implementation of the *SCEOptimizer.cpp* source code in PSUADE. Each SCE optimization run is terminated when any of the following conditions are satisfied:

(1) The maximum number of objective-function evaluations is reached:

$$N_{FE} \geq N_{FE,max}, \tag{S37}$$

where N_{FE} is the number of objective-function evaluations performed during the SCE optimization process, and $N_{FE,max}$ is the prescribed maximum allowable number of objective-function evaluations.

(2) The normalized parameter-space geometric range falls below an internal threshold:

$$GR < GR_{threshold}, \tag{S38}$$

$$GR = \exp\left(\frac{1}{m}\sum_{i=1}^{m} ln\left(\frac{x_{i,max}-x_{i,min}}{U_i-L_i}\right)\right), \tag{S39}$$

where GR denotes the normalized geometric range, $GR_{threshold}$ is the parameter space convergence threshold specified in PSUADE (0.0001), $x_{i,max}$ and $x_{i,min}$ are the maximum and minimum sampled values of parameter i, respectively, m is the number of parameters, and U_i and L_i are the upper bound and lower bound of the value of parameter i, respectively.

(3) The relative improvement of the best objective value over the recent evolution history becomes smaller than the user-specified tolerance:

$$C_f = \frac{100\left|f_{k-1}^{best} - f_{k-n}^{best}\right|}{max\left(1, \frac{1}{n}\sum_{j=k-n}^{j=k-1}\left|f_j^{best}\right|\right)} < \varepsilon, \quad \text{(S40)}$$

where f_k^{best} is the best objective value after the k-th evolution loop, k is the number of evolution loops, n is the number of recent evolution loops considered in the convergence test, and ε is the user-defined optimization tolerance.

In the PSUADE input file, *optimization max_feval* specifies the maximum number of evaluations for each SCE optimization run, whereas *optimization tolerance* defines the tolerance for the objective-function improvement criterion.

S7.2 Implementation of BOBYQA optimization convergence criteria in PSUADE

The convergence criteria for the BOBYQA optimization are explicitly defined according to the implementation of the *BobyqaOptimizer.cpp* and *bobyqa.f* source code in PSUADE. Each BOBYQA search is terminated when either of the following criteria is satisfied:

(1) The maximum number of objective-function evaluations is reached:

$$N_{FE} \geq N_{FE,max}, \quad \text{(S41)}$$

where N_{FE} is the number of objective-function evaluations performed during the BOBYQA optimization process, and $N_{FE,max}$ is the prescribed maximum allowable number of objective-function evaluations.

(2) The trust-region radius has been progressively reduced to the prescribed final trust-region radius:

$$\rho < \rho_{end}, \quad \text{(S42)}$$

$$\rho_{end} = \rho_{beg} \times \varepsilon, \quad \text{(S43)}$$

$$\rho_{beg} = \frac{1}{2}\min_i(U_i - L_i), \quad \text{(S44)}$$

where ρ is the current trust-region radius, ρ_{end} is the final trust-region radius, ε is the user-defined optimization tolerance specified in the PSUADE input file, ρ_{beg} is the initial trust-region radius, and U_i and L_i denote the upper and lower bounds of parameter i, respectively.

In the PSUADE input file, *optimization max_feval* specifies the maximum number of objective-function evaluations, whereas *optimization tolerance* controls the final trust-region resolution.

S7.3 Implementation of PSO optimization convergence criteria in PSUADE

The convergence criteria for the PSO optimization are explicitly defined according to the implementation of the *PSOptimizer.cpp* source code in PSUADE. The PSO iterations are continued only while all of the following conditions are satisfied:

(1) The maximum number of objective-function evaluations is reached:

$$tN_p \geq N_{FE,max}, \quad \text{(S45)}$$

where t is the number of completed epochs, N_p is the swarm size, and $N_{FE,max}$ is

the prescribed maximum allowable number of objective-function evaluations.

(2) The current global-best objective-function value is sufficiently close to the prescribed target value:

$$\left|f_{gbest} - f_{target}\right| \leq \varepsilon, \tag{S46}$$

$$f_{gbest} = \min_{\substack{0 \leq \tau \leq t \\ 1 \leq i \leq N_p}} f\left(x_i^{(\tau)}\right), \tag{S47}$$

where f_{gbest} is the current global optimal objective function value, f_{target} is the prescribed target value, ε is the prescribed tolerance, t is the current number of completed PSO iterations, τ is the iteration index, i is the particle number, and N_p is the total number of particles in the particle swarm.

(3) No new global-best solution is found for consecutive particle evaluations:

$$N_{stag} \geq 20N_p, \tag{S48}$$

$$N_{stag}^{(k+1)} = \begin{cases} N_{stag}^{(k)} + 1, & f(x_k) \geq f_{gbest}^{(k)} \\ 0, & f(x_k) < f_{gbest}^{(k)} \end{cases}, \tag{S49}$$

where N_{stag} is the stagnation counter, N_p is the number of particles, k denotes the k-th particle evaluation, $f(x_k)$ is the current particle's objective function value, and f_{gbest} is the current global optimal objective function value.

In the PSUADE input file, *optimization max_feval* specifies the maximum number of objective-function evaluations, whereas *optimization tolerance* controls the allowable absolute difference between the current global-best objective value and the prescribed target value. The prescribed target value is specified by *optimization* f_{min}, which defines the expected minimum objective-function value and is passed to the PSO optimizer. The stagnation multiplier of 20 is internally defined in *PSOptimizer.cpp*, while the number of particles is 50 by default and can be modified in optimization expert mode.

References added:

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.

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.

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.

6. In addition, Figure 16 suggests that although the SCE-optimized simulation improves the overall WTD level, it does not fully reproduce the two pronounced WTD declines observed in 2007. This point deserves further discussion, because capturing such event-scale or seasonal drawdown dynamics is important for evaluating whether the optimized parameters improve not only the mean WTD but also the temporal variability. The authors should clarify whether these discrepancies are related to limitations in the runoff/groundwater parameterization, freeze-thaw processes, meteorological forcing, observational uncertainty, or the optimization objective function. It would also be helpful to add event-based or seasonal performance metrics to quantify model performance during these pronounced drawdown periods.

Response:

We thank the reviewer for this valuable suggestion. We agree that the reason why the optimized WTD dynamic curve by SCE does not reproduce the observed WTD declines requires further discussion. During the additional analysis, we identified an inconsistency in the previous optimization simulations: the saturated water content, which was not included as an optimization parameter, was inadvertently assigned a large value. This inappropriate setting affected the simulated groundwater dynamics and prevented the WTD from exhibiting the expected fluctuations. We have corrected this issue and updated the optimization results accordingly. The updated results show improved WTD dynamics compared with the previous simulations; however, the optimized SCE simulation still does not fully reproduce the two pronounced WTD declines observed in 2007.

Following the suggestion, we have evaluated the existing SCE simulations using additional performance metrics, including KGE and event-based RMSE during the two major WTD fluctuation periods. We further discussed the potential causes of the discrepancies in reproducing these groundwater fluctuation events. The revised manuscript now highlights that the discrepancies may be related to several factors, including the choice of objective function, uncertainties in meteorological forcing and observations, freeze–thaw process representation, and simplified groundwater parameterization in land surface models. To avoid excessive length in the manuscript, details were provided in Sect. S8 of the supplementary material. Specifically, Sect. S8 includes (1) comparisons among global RMSE, KGE, and event-based performance metrics based on the SCE simulations; (2) analysis of the influence of TPMFD precipitation and temperature uncertainties; and (3) diagnosis of freeze–thaw-related limitations affecting groundwater recharge during the 2007 events. The corresponding discussion has been added to Sect. 4.4 of the revised manuscript, and the detailed analysis is provided in Sect. S8 of the supplementary material.

Revision:

(1) Sect. 4.4 of the Discussion (Page 45, Lines 744–758)

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.

(2) Sect. S8 of the supplementary material (Pages 28-31, Lines 338–408)

S8: Diagnosis of unresolved WTD fluctuation events

To further diagnose the factors contributing to the unresolved event-scale WTD fluctuations, the 1000 WTD simulations generated during the SCE search were additionally evaluated using the Kling–Gupta efficiency (KGE), followed by analyses of meteorological forcing uncertainty and freeze–thaw-related limitations. KGE was calculated as (Kling et al., 2012):

$$KGE = 1 - \sqrt{(r-1)^2 + (\beta-1)^2 + (\gamma-1)^2}, \tag{S50}$$

$$\beta = \frac{\mu_s}{\mu_o}, \tag{S51}$$

$$\gamma = \frac{CV_s}{CV_o} = \frac{\sigma_s/\mu_s}{\sigma_o/\mu_o}, \tag{S52}$$

where KGE is the modified KGE-statistic, r is the correlation coefficient between simulated and observed values, β is the bias ratio, γ is the variability ratio, μ is the mean value, CV is the coefficient of variation, σ is the standard deviation, and the

indices s and o represent simulated and observed values, respectively.

S8.1 Influence of objective functions on event-scale WTD dynamics

To diagnose the influence of the objective function, the 1000 WTD simulations generated during the SCE search were evaluated using additional performance metrics, including overall KGE and RMSE/KGE values during the two groundwater-rise periods (Fig. S21). KGE is commonly used during model calibration because it considers correlation, variability, and bias components, whereas RMSE is frequently used for evaluating and comparing model accuracy (Williams, 2025). Comparison between the minimum RMSE solution and the maximum KGE solution showed that the minimum RMSE solution achieved a KGE of 0.435 (Fig. S21a), whereas the maximum KGE among the 1000 simulations reached 0.529, with only a 0.123 m difference in RMSE. This indicates that parameter combinations with substantially improved temporal dynamics existed within the RMSE convergence region, but they were not favoured when RMSE was used as the sole optimization objective.

Similarly, the parameter combination achieving the lowest global RMSE did not correspond to the lowest RMSE during the two groundwater-rise events (Fig. S21b). This occurred because the event periods accounted for only a limited proportion of the entire simulation period. Therefore, optimization based on global RMSE preferentially reduced overall errors at the expense of event-scale deviations. When different evaluation metrics were used to select optimal simulations from the RMSE-driven parameter search results (Fig. S21c), RMSE tended to favour simulations with lower absolute errors and reduced variability. This finding is consistent with Gupta et al. (2009), who demonstrated that error-based metrics may favour models with underestimated variability. In contrast, KGE explicitly incorporates correlation and variability components and therefore better preserves temporal dynamics and overall fluctuations (Kling et al., 2012). It should be noted that the parameter sets selected according to KGE, event-scale RMSE, and event-scale KGE were obtained from the existing simulations rather than from independent optimization experiments using these metrics as objective functions. Therefore, the potential benefits of adopting multi-objective optimization strategies or alternative objective functions for improving event-scale groundwater dynamics require further investigation.

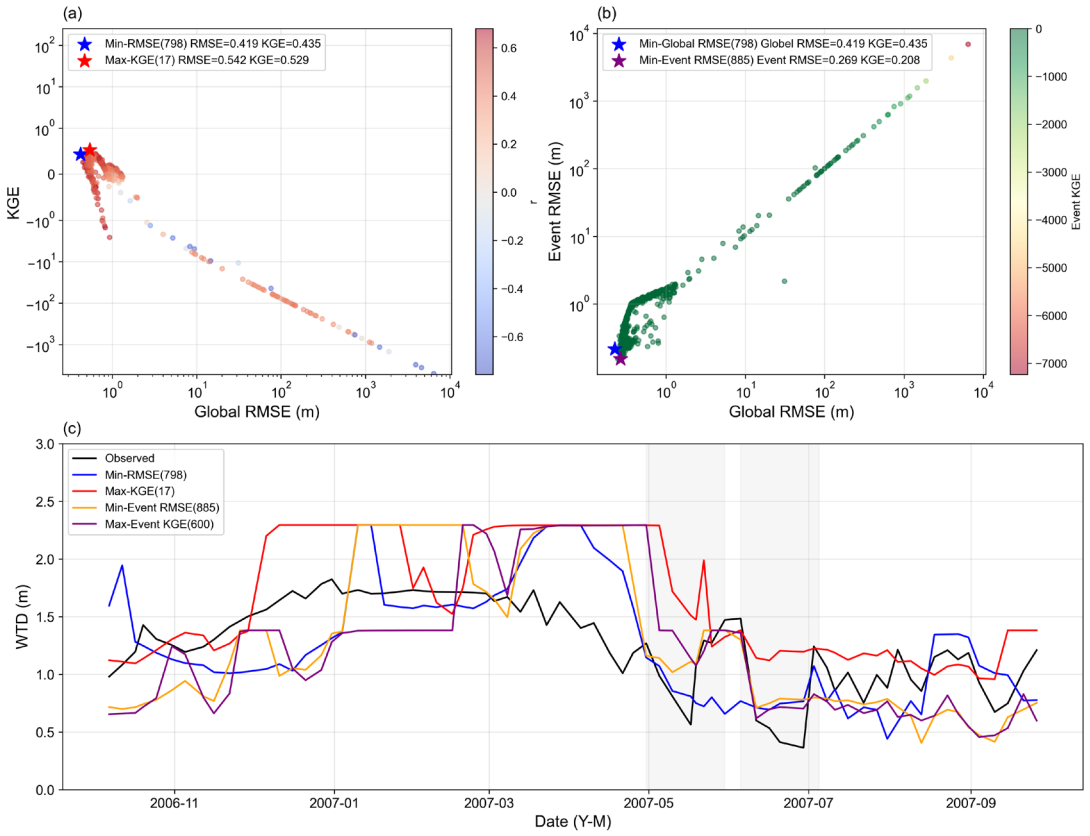


Figure S21: Comparison of model performance and simulated WTD hydrographs selected by different optimization metrics based on 1000 simulations: (a) KGE versus Global RMSE; (b) Event RMSE versus Global RMSE; and (c) Observed and simulated WTD time series under optimal parameter sets for different evaluation targets. The shaded areas in (c) indicate the periods of pronounced declines in the observed WTD, which were also used for calculating the event-based metric. The numbers in parentheses in the legends indicate the specific simulation ID (out of 1000 runs) that achieved the corresponding optimal metric.

S8.2 Influence of meteorological forcing uncertainty

Besides the influence of objective functions, uncertainties in meteorological forcing and observations may also contribute to the discrepancies in simulated WTD dynamics. As shown in Fig. S22a, during the months corresponding to major groundwater-rise events, the monthly precipitation from the TPMFD dataset was comparable with observations in June but lower in May (Yang et al., 2020). Although TPMFD has demonstrated good performance over the Tibetan Platea (Liu et al., 2025), discrepancies may still exist at individual sites, resulting in underestimated WTD responses under insufficient precipitation forcing. Furthermore, monthly air temperature from TPMFD was consistently lower than observations (Fig. S22b), which may delay soil thawing and reduce effective groundwater recharge during the freeze–thaw transition period.

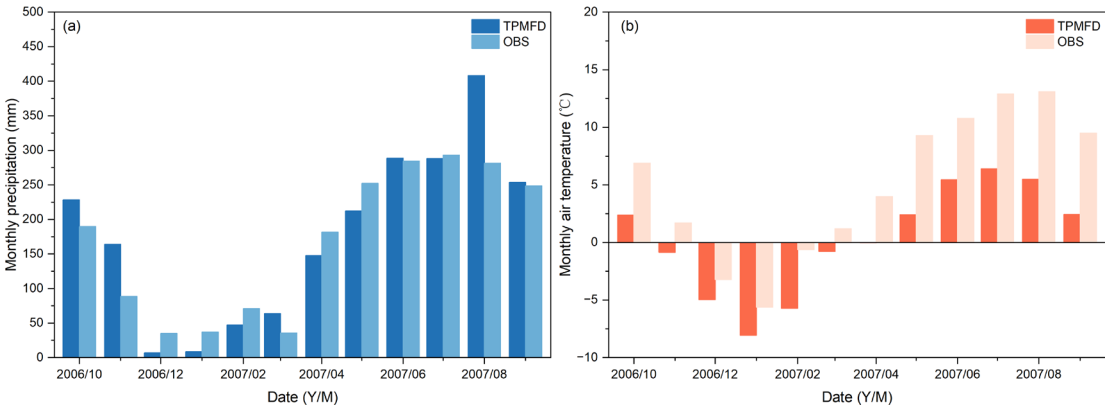


Figure S22: Comparison of monthly precipitation (a) and mean air temperature (b) between the TPMFD dataset and in-situ observations.

S8.3 Influence of freeze–thaw processes on groundwater recharge

Around the first observed groundwater-rise event, the optimal simulation indicated relatively large net infiltration (Fig. S23a). However, soil temperature in the eighth layer remained below the freezing point, resulting in considerable ice content and reduced liquid water availability. Consequently, soil matric potential and hydraulic conductivity remained low, limiting effective groundwater recharge. The subsurface runoff component, which is a deterministic function of WTD in CoLM, remained relatively stable when WTD was stable. Combined with evaporative losses, the soil–aquifer exchange flux remained positive, indicating water loss from the aquifer; however, the simulated stable WTD suggested a dynamic balance between limited recharge from the unsaturated zone and groundwater drainage. Across the soil profile (Fig. S23b), ice below approximately 0.6 m depth resulted in low hydraulic conductivity and weak water transport capacity. During the second groundwater-rise event, the soil profile gradually thawed, and liquid water content, matric potential, and hydraulic conductivity increased. However, net infiltration and liquid soil water at about 0–0.3 m remained relatively limited, resulting in insufficient groundwater recharge.

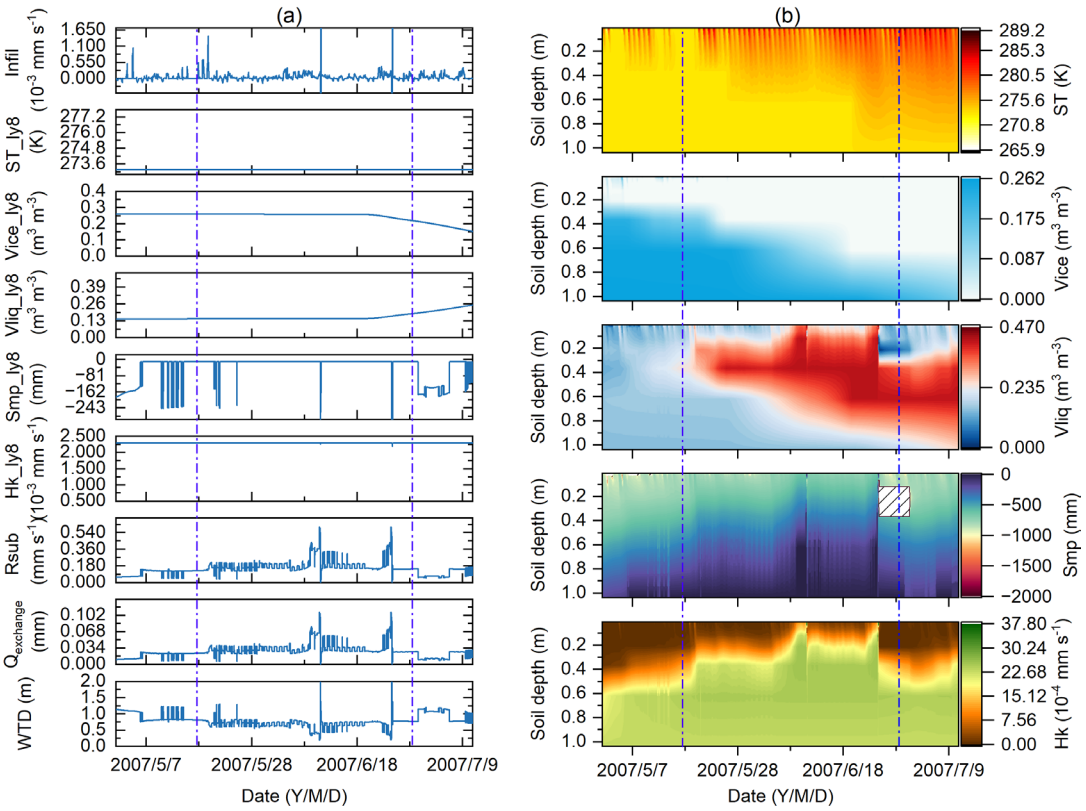


Figure S23: The key variables obtained through simulation corresponding to the optimal result of the SCE algorithm during the period when the measured WTD rose significantly: (a) Time series of infiltration (Infil), 8th-layer soil temperature (ST_ly8), 8th-layer partial volume of ice lens in layer (Vice_ly8), 8th-layer liquid volumetric water content (Vliq_ly8), 8th-layer soil matric potential (Smp_ly8), 8th-layer soil hydraulic conductivity (Hk_ly8), subsurface runoff (Rsub), water exchange between soil and aquifer (Qexchange), and water table depth (WTD); (b) Spatial and temporal variations of soil temperature (ST), partial volume of ice lens in layer (Vice), liquid volumetric water content (Vliq), soil matric potential (Smp) and soil hydraulic conductivity (Hk). The two dashed lines in the plots mark the timing of the observed WTD peaks.

References added:

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, J. Hydrol., 377, 80-91, https://doi.org/10.1016/j.jhydrol.2009.08.003, 2009.

Kling, H., Fuchs, M., and Paulin, M.: Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios, J. Hydrol., 424-425, 264-277, https://doi.org/10.1016/j.jhydrol.2012.01.011, 2012.

Liu, R., Su, J., Zheng, D., Lü, H., Zhu, Y., Pan, Y., Shi, C., Sun, S., and He, J.: Comprehensive assessment of various meteorological forcing datasets on the Tibetan Plateau: Insights from independent observations and multivariate comparisons, J.

Hydrol., 656, 133025, https://doi.org/10.1016/j.jhydrol.2025.133025, 2025.
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.
Williams, G. P.: Friends don't let friends use Nash-Sutcliffe Efficiency (NSE) or KGE for hydrologic model accuracy evaluation: A rant with data and suggestions for better practice, Environ. Modell. Softw., 194, 106665, https://doi.org/10.1016/j.envsoft.2025.106665, 2025.
Yang, Y., Wang, K., Hu, Z., and Wang, G.: Meteorological dataset observed by Alpine Ecosystem Observation and Experiment Station of Gongga Mountain, Chinese Academic of Sciences, during 1998–2018 [dataset], https://10.11922/sciencedb.945., 2020.

Minor Comments:

1. L16: In the Abstract, the phrase "sensitivity analysis" is repeated several times in the sentence

Response:

We thank the reviewer for pointing this out. We reduced the repeated use of the phrase "sensitivity analysis" in the Abstract. Specifically, "qualitative sensitivity analysis, quantitative sensitivity analysis, and parameter optimization techniques" has been revised to "qualitative screening, quantitative validation, and parameter optimization techniques" to better reflect the different roles of each analysis step.

Revision:

Abstract (Page 1, Lines 15–18)

The revised descriptions are indicated in *italics*:

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 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). Among them, the subsurface runoff parameter $r_{sub,max}$ 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.

2. L281: Please double-check the formula labelled as it appears to correspond to the Nash–Sutcliffe efficiency (NSE)

Response:

We thank the reviewer for catching this inconsistency. The formula previously labeled as R^2 is indeed the Nash–Sutcliffe efficiency (NSE), where the denominator is normalized by the variance of the observed values (Duc and Sawada, 2023). The mislabeling as R^2 was an error in the manuscript and source code. We have corrected the label throughout the manuscript, updated all related descriptions accordingly.

References added:

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.

Authors' responses to RC2' comments

Revision Notes

For

Manuscript ID: **egusphere-2026-2275**

Quantifying key parameter sensitivities for water table depth in hydrological schemes of CoLM-PSUADE

Dear Editor,

We thank the reviewers and the editor for their insightful comments, which have helped us substantially improve the manuscript. We have carefully considered all comments and revised the manuscript accordingly. The revised manuscript provides a clearer description of the sensitivity-analysis methods, a more comprehensive classification and interpretation of sensitive parameters, and a more explicit explanation of how optimization supports the validation of identified sensitive parameters.

We previously submitted provided a general response summarizing our overall revision strategy, and now provide a point-by-point response. Reviewers' comments are **Relisted in black**, followed by our **Responses in blue** and **Revisions in purple**.

Reviewer #2:

This study employed the Common Land Model (CoLM) to find the most sensitive parameters for water table depth (WTD) in the Gongga Mountain site, and to calibration the most sensitive parameters aimed at improving WTD simulation. This study suggested that CoLM-PSUADE was a useful method to find the most sensitive parameters, and the most sensitive parameters could be used to improve the temporal dynamics of WTD.

Response:

We thank Reviewer #2 for his/her positive feedback and constructive comments. The comments are all carefully considered and **Relisted in black**, followed by our **Responses in blue** and **Revisions in purple**. According to the comments, we modified

the manuscript, particularly regarding the clarity of methodological descriptions, the physical characterization and interpretation of model parameters, and the logical relationship between sensitivity analysis and parameter optimization. Major changes are summarized as follows:

No.	Major Revisions	Suggested Comments
1	Improvement of manuscript clarity and methodological descriptions	Comment 1, 2
2	Enhancement of parameter characterization and physical interpretation	Comment 3, 4
3	Clarification of the integration sensitivity analysis and parameter optimization	Comment 5

(1) Improvement of manuscript clarity and methodological descriptions. We have revised the manuscript to improve language consistency. The verb tense throughout the manuscript has been carefully standardized. In addition, the methodology section has been expanded to provide a more comprehensive description of the procedures for comparing sensitivity indices among different methods. ***These revisions improve the readability of the manuscript and interpretation of the proposed sensitivity-analysis framework.***

(2) Enhancement of parameter characterization and physical interpretation. We have improved the characterization and interpretation of model parameters by adding physical classifications and expanding the mechanistic explanation of sensitive parameters. The 56 adjustable parameters have been classified according to their physical meaning and associated physical processes. In addition, the discussion of sensitive parameters has been expanded to establish clearer links between parameter functions and WTD responses. ***These revisions improve the physical understanding and interpretation of the identified sensitive parameters.***

(3) Clarification of the integration of sensitivity analysis and parameter optimization. We have clarified the methodological relationship between sensitivity analysis and parameter optimization by emphasizing their sequential roles in the proposed framework. Specifically, sensitivity analysis was first conducted to identify stable influential parameters and reduce model dimensionality, whereas parameter optimization was subsequently performed on the screened parameters to evaluate whether these parameters could effectively improve WTD simulations. ***These revisions provide a clearer explanation of how optimization supports the validation of identified sensitive parameters.***

Comments:

1. The language in the text needs to be revised, particularly with regard to verb tense.

Response:

We thank the reviewer for this suggestion. We carefully revised the manuscript with particular attention to verb tense usage throughout the text. Specifically, we describe established knowledge, model descriptions, and general facts in the present tense; reported methodologies and results from our work in the past tense; and cumulative background or literature findings using the present perfect.

2. The methodology section of the paper needs to be expanded. Further details are required regarding the parameter sensitivity analysis methods.

Response:

We thank the reviewer for this valuable suggestion. In the original manuscript, we introduced the reason for selecting MOAT, MARS, SOT, and RSMSobol and the explanations regarding the methodological characteristics, while the detailed mathematical formulations of these methods were provided in Appendix. However, the description of how the sensitivity indices derived from qualitative methods were used for parameter ranking and comparison was insufficient. Following the suggestion, we have revised Section 2.3.1 by adding further explanations regarding the complementary roles of MOAT, MARS, and SOT in qualitative sensitivity screening.

Revision:

Sect. 2.3.1 of the Data and methods (Pages 15–16, Lines 254–256)

The revised descriptions are indicated in *italics*:

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.

3. Classifying the physical parameters in the model will aid the reader's understanding. For example, parameters could be categorised into soil parameters and vegetation parameters, and so on.

Response:

We thank the reviewer for this valuable suggestion. Following this suggestion, we have added a new column named "Category" in Table 2 to classify the 56 adjustable parameters according to their physical meaning and associated physical processes in CoLM2024, including snow, soil, canopy, runoff, and precipitation.

Revision:

Sect. 2.2.2 of the Data and methods (Pages 9–12, Lines 197–204)

The revised descriptions are indicated in *italics*:

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).*"

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	*Precipitation*	℃	[−5, 0]	(Ding et al., 2014)

		falls entirely as snow				
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 higi 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*	℃	[−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	*Runoff*	-	[0.1, 1.0]	(Niu et al., 2005)

		the topographic index's discrete CDF				
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	$[-8.6\times10^5, -6.6\times10^4]$	(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 to 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.

4. Please provide a more detailed explanation of the physical significance of the sensitive parameters.

Response:

We thank the reviewer for this valuable suggestion. We agree that a clearer explanation of the physical significance of the identified sensitive parameters improves the interpretation of the sensitivity analysis. In the original manuscript, the basic physical meaning of each parameter was already provided in Table 2. However, because several of the identified sensitive parameters are empirical or semi-empirical parameters whose physical roles are not fully evident from their names alone, we have further expanded the discussion in Sect. 4.1 by explaining how these parameters participate in the governing hydrological processes and how they influence WTD through the corresponding model formulations.

Revision:

Sect. 4.1 of the Discussion (Pages 36–37, Lines 595–597, 602–604, 609–611, 615, 619–620, and 628–629)

The revised descriptions are indicated in *italics*:

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 studied 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.*

5. A full explanation is required as to why the parameter sensitivity analysis and parameter optimisation are being carried out concurrently.

Response:

We thank the reviewer for this important comment. We agree that the relationship between sensitivity analysis and parameter optimization should be clarified more explicitly. In the original manuscript, we had indicated that optimization was performed based on the screened sensitive parameters and that its purpose was to evaluate whether these parameters could improve WTD simulations. Following the suggestion, we have revised Sect. 2.3.4 to further clarify that parameter optimization was a subsequent validation step of the sensitivity-analysis framework.

Revision:

Sect. 2.3.4 of the Data and methods (Page 17, Lines 299–303)

The revised descriptions are indicated in *italics*:

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.