Response to comments on Manuscript egusphere-2025-6543

Title: What can hydrological modelling gain from spatially explicit parameterization and multi-gauge calibration?

Authors: Xudong Zheng, Dengfeng Liu*, Hao Wang, Chuanhui Ma, Hui Liu, Guanghui Ming, Qiang Li, Mohd Yawar Ali Khan, and Fiaz Hussain

Manuscript ID: egusphere-2025-6543

Dear Editor and Reviewers,

Please find enclosed our responses to the manuscript assessment entitled "What can hydrological modelling gain from spatially explicit parameterization and multi-gauge calibration?".

We would like to express our sincere gratitude to the Editor and the two anonymous reviewers for their invaluable support and constructive suggestions, as well as for the opportunity afforded to us to revise our work. We have given full attention to all comments and suggestions and made all revisions accordingly. It has resulted in an improved manuscript that fully addresses all concerns.

Concerning the revision of the manuscript, the following major changes have been made:

(1) **Structural Refinement:** The Abstract and Conclusion have been reformatted to enhance clarity and logistical flow, adhering to HESS standards for concise scientific communication.

(2) **Introduction and Methodology:**
   a) The Introduction now provides a justification for the VIC model selection and an expanded description of multi-gauge calibration in nested catchment system and the cross-benefit of the proposed strategies.
   b) The Methods section was enriched with a LULC analysis of the UHRB, detailed VIC governing equations, and a detailed description of soil depth and RVIC parameter MPR-refinements.
   c) The EF-SPM framework and optimization algorithms were clarified, and distinct experimental aliases were introduced for better traceability.

(3) **Results and Discussion:**
   a) Statistical rigor was improved by incorporating pairwise t-tests to evaluate the significance of performance differences between cases.
   b) A new section (Sect. 4.3.2) was added to explicitly discuss parameter transferability under multi-gauge calibration.
   c) The discussion now includes a dedicated synthesis of the role of soil information in hydrological modeling.

(4) **Visualization:** All figures (notably Figs. 5–7) were revised for improved readability and updated with a professional, journal-standard color palette.

We believe the paper quality has significantly improved through this review process. We are also happy to address any comments that may further strengthen the paper quality. We are uploading our point-by-point response to the comments, an updated manuscript with red highlighting indicating changes, and a

clean updated manuscript without highlights.

In the point-by-point responses below, the original comments are displayed in **bold**. **EC-n/**, **RC1-n/**, and **RC2-n/** correspond to the Editor, Referee 1, and Referee 2 comments, respectively. The corresponding responses begin with **R/** and the revisions is highlighted in red, while important content are marked in **blue**. The quoted content is displayed in *italics*. All references cited in this response are listed at the end of the reply. Line numbers from the revised manuscript with marked-up changes are provided where relevant.

We thank you for your consideration,

Dengfeng Liu
Email: liudf@xaut.edu.cn

Responses to Editor

Thank you very much for your timely handling of our manuscript and for your careful consideration. We have provided detailed, point-by-point responses to all comments raised by you and the two anonymous reviewers, and have revised the manuscript accordingly. Below, we present our responses to your comments.

EC-1/ The editorial board has now received two reviewers' comments. Both reviewers acknowledge that your study addresses an important and timely question—the combined effects of spatially explicit parameterization and multi-gauge calibration in hydrological modelling. They also provide constructive feedback that will help strengthen the paper. After reading the manuscript myself, I concur with their overall assessments. Therefore, I'd like to invite you to submit a major revision. Please carefully address the comments below.

R/ We sincerely thank you for your recognition of our work and for providing us with the opportunity to revise and improve the manuscript. Your comments, together with those of the reviewers, are highly valuable for enhancing the quality and clarity of our study. We have carefully considered all the suggestions and have provided detailed, point-by-point responses below, along with corresponding revisions in the manuscript.

EC-2/ Abstract length: The abstract currently spans three paragraphs; please condense it to a single, concise paragraph that clearly states the objectives, methods, key results, and implications.

R/ Thank you for your helpful suggestion. In accordance with your recommendation, we have revised the abstract into a **single-paragraph format** and refined the wording to improve clarity and readability.

The revised abstract is provided below for your review.

Lines 17-46:

Abstract. Traditional hydrological modelling is facing transformative pressures from the rise of data-driven approaches and increasing demands for modelling realism. With improving data availability, spatially explicit parameterization and multi-gauge calibration offers a promising pathway to enhancing both the predictive capability and the realism of physically based distributed hydrological models. However, current understanding remains largely confined to their broad effects on aggregated simulated responses, while the underlying mechanisms and interactions through which these approaches benefit hydrological modelling remain poorly understood. To bridge this knowledge gap, this study develops an Experiment Framework to evaluate the effect of Spatially explicit Parameterization and Multi-gauge calibration, termed EF-SPM. Implemented through the Variable Infiltration Capacity (VIC) model combined with the multiscale parameter regionalization technique, the framework is applied to the Upper Han River Basin, a representative nested catchment, via intensive comparative calibration experiments. Results indicate that, compared to simpler configurations, considering both spatially explicit parameterization with multi-gauge calibration leads to consistent improvements in streamflow simulations across all sub-basins. Controlled experiments isolating individual effects further show that spatially explicit parameterization is particularly effective in improving simulations under moderate-flow to high-flow conditions (with an 18 % improvement in $\%BiasFHV_1$), yet at the cost of degraded performance during low-flow periods. On the other hand, multi-gauge calibration markedly enhances parameter identifiability by imposing stronger constraints on spatially shared parameters. This effectively mitigates information-gap-induced uncertainties, thereby enabling robust

parameter transfer to ungauged upstream sub-basins. Importantly, their combined application yields a clear cross-benefit in the multidimensional calibration objective space by substantially alleviating the trade-offs among gauge-specific objectives observed under uniform parameterization. This study integrates two promising directions in contemporary hydrological modelling, highlighting the importance of pursuing more expressive parameterization and stronger calibration constraints in parallel, rather than prioritizing one over the other. In doing so, it provides a steppingstone for advancing distributed hydrological modelling toward a modern Model–Data Infusion framework.

EC-3/ Justification of the VIC model: The introduction should include a clear rationale for choosing VIC over other well-established models (e.g., FLEX-Topo, HBV, Xinanjiang, Topmodel). Given that some of these models explicitly account for topographic controls or have a more flexible structure, please explain why VIC is particularly suited for the questions you address.

R/ This is a very important point that requires clarification, as it helps explain the rationale for our methodological choice and its appropriateness. To address this concern, we have added the following text to the Introduction for clarification.

Lines 99-111:

Although MPR has been implemented in a series of distributed hydrological models, the VIC model is adopted as the modelling framework in this study. This choice is motivated by the long-standing challenges in estimating spatially distributed parameters of VIC, which remain an active research topic. Recent studies have explored different strategies, including surrogate-based approaches (e.g., Sun et al. 2023) and MPR-based parameter estimation (e.g., Gou et al. 2021). In this context, the increasing use of spatially distributed parameters in VIC applications mirrors a broader shift towards spatially explicit parameterizations in modern hydrological modelling. In addition, recent developments in VIC, particularly the transition to version 5, enable all model parameters to be handled through a unified NetCDF-based I/O framework, thereby facilitating spatial parameter estimation (Hamman et al., 2018). To support the MPR-based spatial parameter estimation using distributed information (e.g., soil and vegetation properties), our team has developed an open-source Python-based deployment framework for VIC-5 (https://github.com/XudongZhengSteven/easy_vic_build). The framework integrates a range of transfer functions and offers an object-oriented, modular workflow for parameter estimation, thereby enabling the systematic comparison of spatially explicit and uniform parameterizations within a unified experimental setting and establishing the experimental basis for the present study.

EC-4/ Significance of the study area: The Upper Han River basin is the source of the South-North Water Transfer Project, a fact that greatly enhances the relevance of your work. Please highlight this in the introduction to better convey the practical importance of your study.

R/ Thank you for this constructive suggestion. As you pointed out, adding this clarification is essential for establishing the practical relevance of our study. Accordingly, we have added the following text to the Introduction.

Lines 150-156:

Here, we design an Experiment Framework to evaluate the effect of Spatially explicit Parameterization and Multi-gauge calibration, termed EF-SPM. This framework is implemented using the VIC model integrated with the MPR technique

and applied to the Upper Han River basin, a representative catchment comprising five sub-basins. As the source area of the South-to-North Water Transfer Project, the basin is of particular strategic importance, making hydrological simulation in this region highly relevant for water resources management and hydraulic engineering planning (Zhang et al., 2021).

EC-5/ Clarity of the two-step soil-layer reconciliation: As noted by Reviewer 1, the description of how soil data are reconciled with the VIC three-layer structure remains confusing. Please provide a clearer, step-by-step explanation. In addition, I invite you to reflect on the broader discussion about the role of soil in hydrological modelling (e.g., Gao et al., HESS, 2023). Moreover, please clarify how VIC represents groundwater discharge and baseflow—particularly because the current text (e.g., line 232) links the bottom layer to baseflow without describing the actual mechanism.

R/ Thank you for your detailed and constructive suggestions. In response to the three concerns raised above, we provide a point-by-point reply below.

(a) First, to further clarify the two-step soil reconciliation procedure and avoid ambiguity, we revised the notation by **replacing the potentially confusing symbol** d for soil layer thickness with h and updated Fig. 2 accordingly. We then **revised and expanded the description of this procedure** in Sect. 3.2, as shown below.

Lines 311-318:

In addition, the original soil dataset (i.e., SoilGrids1km) provides six soil layers, each characterized by distinct properties (e.g., soil texture), whereas VIC-3L adopts a three-layer soil structure. In order to feed original soil data into the transfer function, it is necessary to reconcile its soil-layer structure with the VIC vertical structure. As shown in Fig. 2, a two-step mechanism is applied here, involving aggregation and scaling. In the first step, the original soil data are aggregated into a standard three-layer structure defined by the layering numbers z_1 and z_2, with the corresponding layer thicknesses summed and the soil properties averaged within each standard layer. In the second step, the original thickness (h_0) is scaled by g_{10} to allow flexible adjustment of the final computational layer thickness.

Lines 349-353:

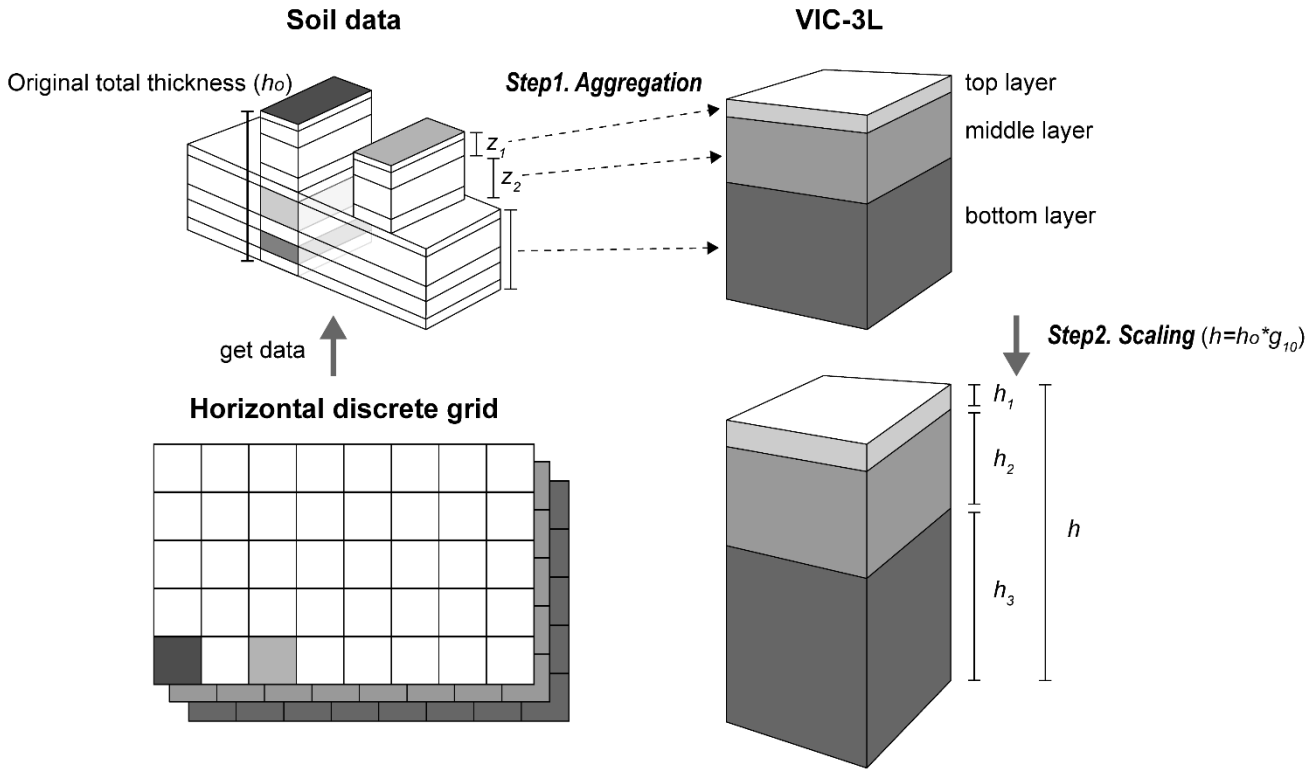

Figure 2. Schematic of the two-step mechanism for reconciling soil data with VIC three-layer (VIC-3L) vertical structure. In the aggregation step, the original soil layers are merged into three VIC soil layers based on the layering numbers z_1 and z_2. The scaling step then applies a scaling factor (g_{10}) to the original total thickness (h_o) to obtain the model-prepared layer thicknesses h_1, h_2, and h_3.

(b) Second, the role of soil in hydrological modelling is indeed a profound issue of central importance. Distributed models developed in the FH69 tradition (Freeze and Harlan., 1969) generally rely on Darcy-Richards equations to represent soil water movement process, and this dependence is hard-coded into the numerical representation of runoff generation, thereby inevitably embedding the assumption of a soil-centered model structure. In essence, this represents a **bottom-up modelling paradigm**, in which theories developed at the microscale, such as the field scale, are aggregated to represent hydrological processes at the macroscale (Blöschl and Sivapalan, 1995). This provides the foundation for many contemporary approaches to spatial parameter estimation based on soil information, including MPR.

However, such approaches may involve substantial uncertainty and several easily overlooked assumptions. First, the development of **transfer functions** is associated with **considerable subjective prior uncertainty,** because field measurements are often limited in sample size and insufficiently representative in either vertical depth or spatial extent. Second, **scale effects** arising from the extrapolation of microscale relationships to larger scales constitute an important source of uncertainty. In addition, some soil parameters may still **lack a priori information and can only be determined through calibration**, which may compromise model realism. A pertinent example in this study is soil depth: under spatially uniform parameterization, soil depth clearly cannot fully represent the differences in runoff generation processes among sub-basins in a physically realistic manner (Fig. 14).

On the other hand, **the role of ecosystems may be substantially underestimated** in the current hydrological modelling paradigm. Taking the VIC model used in this study as an example, vegetation information is incorporated into hydrological simulations only through simplified process representations, such as LULC, LAI, and NDVI. The root zone is treated even more simply, with its vertical distribution prescribed a priori using LULC-based lookup tables, which may lead to unrealistic simulations. Although this aspect is not explicitly evaluated in the present study, it is reasonable to expect that, as observational data become increasingly abundant, further **coupling ecosystem constraints into the modelling process** would yield substantial benefits and improve the physical realism of hydrological simulations (Gao et al., 2024). Feasible strategies include replacing static LAI with dynamic LAI and incorporating spatial estimates of root-zone properties (e.g., root zone depths) into the modelling framework. We likewise recognize such improvements as an important direction for the future development of hydrological modelling, given the importance of ecological process representation.

The present study is highly relevant to the discussion raised by Gao et al. (2023) in their HESS Opinion article. Both transfer-function-based MPR regionalization and spatially explicit parameterization largely reflect the current strong reliance of hydrological modelling on soil information, which may represent a problematic foundation of the prevailing modelling paradigm. Although this foundation may appear deeply entrenched, critically revising and improving it remains an important direction to which we are committed. Compared with previous approaches, the framework presented here reveals the practical benefits of spatially explicit parameterization and multi-gauge calibration, and thus directly contributes to this line of development.

Based on the above reflections and analysis, we have incorporated the relevant discussion to the end of Sect. 4.4.1, following the discussion of the spatial discretization of soil layer depth. This addition substantially enriches the scope and depth of the present study. The corresponding revised text is provided below for your review.

Lines 734-755:

A further point that merits discussion is that current spatial parameter estimation methods (e.g., MPR) rely heavily on soil information. This reliance arises because many distributed hydrological models follow the FH69 blueprint and adopt a bottom-up modelling paradigm in which the Darcy-Richards equations are embedded in the representation of runoff-generation processes, thereby providing the basis for estimating model parameters from soil information (Freeze and Harlan, 1969). However, such a widely adopted practice may involve substantial uncertainty and several assumptions that are not readily apparent. For example, the development of transfer functions is associated with considerable subjective prior uncertainty, given that many field measurements are often limited in sample size and by insufficient representativeness in terms of vertical depth or spatial extent. In addition, scale effects arising from the extrapolation of microscale relationships to larger scales constitute an important source of uncertainty. Furthermore, some soil parameters may still lack a priori information and can only be determined through calibration, which may compromise model realism. A pertinent example in this study is soil depth: under spatially uniform parameterization, soil depth clearly cannot fully represent the differences in runoff generation processes among sub-basins in a physically realistic manner (Fig. 14). On the other hand, the role of ecosystems may be substantially underestimated in current hydrological modelling. In the VIC model used here, vegetation is represented only through simplified descriptors such as LULC, LAI, and NDVI, while the root zone is treated even more simply through a priori vertical distributions prescribed by LULC-based lookup tables, which may compromise model realism. Increasing data availability creates new opportunities to more explicit incorporation of ecosystem constraints into hydrological modelling, for example through the integration of spatially

distributed root-zone information, with the potential to further improve model physical realism (Gao et al., 2024). The above discussion echoes the view advanced by Gao et al. (2023) that soils may overrated in hydrology, and perhaps even more so in hydrological modelling. Although the foundations of bottom-up modelling may appear deeply entrenched, their critical revision and improvement remain an important direction for continued development in the hydrological community. In this context, the present study contributes by highlighting the importance of the combined application of spatially explicit parameterization and multi-gauge calibration.

(c) To improve methodological transparency, we have **expanded Sect. 3.1 to further clarify runoff generation in the VIC model**, particularly the correspondence between soil layers and the different runoff components, as well as the conceptual interpretation of baseflow in VIC as a lumped representation of groundwater-related processes. The revised text is provided below for your review.

Lines 251-272:

In the VIC model, total runoff (Q_t) is partitioned into surface runoff (R) and baseflow (Q_b). Under the standard three-layer vertical soil structure (VIC-3L), surface runoff is generated from the thin top layer and the upper layer according to the variable infiltration capacity curve, as shown below (Liang et al., 1994):

$$R = \begin{cases} PE - (W_m - W_0), & PE + i_0 \geq i_m \\ PE - (W_m - W_0) + W_m \left(1 - \frac{PE + i_0}{i_m}\right)^{1+b}, & PE + i_0 < i_m \end{cases} \quad (3)$$

where PE denotes the effective precipitation; W_m represents the mean infiltration capacity of the grid cell; i_m is the maximum infiltration capacity at a point within the grid cell; W_0 is the initial soil moisture and i_0 is the corresponding initial infiltration capacity; and b is the shape parameter of the variable infiltration capacity curve. All of these quantities refer to the upper two soil layers. The VIC model does not explicitly represent soil drainage processes, such as interflow, or groundwater storage above the water table. Instead, these components are treated in a conceptual lumped manner and represented as baseflow from the bottom soil layer by a single term, Q_b, which is parameterized using the Arno baseflow scheme, as follows (Liang et al., 1994):

$$Q_b = \begin{cases} \frac{D_s D_m}{W_s \theta_b^s} \cdot \theta_b, & 0 \leq \theta_b < W_s \theta_b^s \\ \frac{D_s D_m}{W_s \theta_b^s} \cdot \theta_b + (D_m - \frac{D_s D_m}{W_s}) (\frac{\theta_b - W_s \theta_b^s}{\theta_b^s - W_s \theta_b^s})^c, & \theta_b \geq W_s \theta_b^s \end{cases} \quad (4)$$

where θ_b^s is the maximum soil moisture of the bottom layer; D_m is the maximum baseflow; D_s is the fraction of D_m at which nonlinear baseflow begins; W_s is the fraction of θ_b^s at which nonlinear baseflow begins; and c is the exponent of the nonlinear part of the Arno baseflow curve, which is usually set to 2. All of these quantities refer to the bottom soil layer. The Arno baseflow equation can be equivalently reformulated as a coupled linear reservoir and a nonlinear reservoir, and can thus be expressed in the following form, thereby alleviating parameter interactions (Nijssen et al., 2001):

$$Q_b = \begin{cases} D_1 \theta_b, & 0 \leq \theta_b < D_3 \\ D_1 \theta_b + D_2 (\theta_b - D_3)^{D_4}, & \theta_b \geq D_3 \end{cases} \quad (5)$$

where D_1 is the linear reservoir coefficient; D_2 is the nonlinear reservoir coefficient; D_3 denotes soil moisture at which baseflow transition occurs from linear to nonlinear; and D_4 has the same conceptual meaning as c. This formulation is also known as the Nijssen form of the Arno model, has been incorporated into the official VIC model and can be specified through the global parameter file.

Lines 306-309:

In the classic VIC-3L (i.e., VIC three-layer) configuration, the upper two soil layers respond rapidly to precipitation, generating surface flow according to the Xinanjiang formulation, whereas the bottom layer governs baseflow, resulting in a slower runoff response (Eqs. 3 and 5).

EC-6/ Model transferability: As a strong test of model realism, I encourage you to explore whether the new calibration approach improves transferability among nested catchments. For example, you could calibrate parameters on a subset of catchments and evaluate performance on the remaining ones with unchanged parameters. If feasible, such an analysis would add valuable insight.

R/ This is indeed an interesting issue worth further exploration and is closely related to the PUB (Prediction in Ungauged Basins) problem. In the context of a nested catchment, where sub-basins are **hydrological connected**, this issue can be further specified as **whether observations at downstream gauges can provide sufficient constraints on model parameters to allow their effective transfer to upstream sub-basin**.

We note that hydrological connectivity is the key mechanism through which synergistic constraints facilitate parameter transfer across sub-basins. In this context, parameter transfer across parallel catchments and the use of upstream observations to constrain downstream simulations are less informative, because both settings may still suffer from insufficient constraints. Specifically, parallel catchments lack an explicit flow-routing connection, meaning that their runoff hydrographs may be independent and therefore provide no mutual constraint; in the upstream–downstream case, however, part of the intermediate processes remains unobserved, leaving the downstream catchment under-constrained.

Furthermore, parameter transfer is only meaningful for parameters that are shared across sub-basins. Accordingly, the sub-basin-specific spatially explicit parameterization of soil layer depths considered in this study is not amenable to such transfer, because these parameters are independently defined for each sub-basin rather than shared among them. The remaining parameters, such as the global parameters in MPR scheme and the unit hydrograph parameters (t_p, μ, and m), are shared across sub-basins and are therefore transferable.

Based on the above considerations, we designed the following experiment with reference to Case 4. The VIC and RVIC parameters were estimated using the MPR-based spatially explicit parameterization, whereas the soil-layer depth parameters were kept spatially uniform (Table 3). In addition, **a leave-one-out multi-gauge strategy was adopted**, in which the streamflow KGE at the four gauges was used as the calibration objective, while the remaining single upstream gauge was excluded from calibration and used for parameter transfer and subsequent evaluation. To ensure robustness, this analysis was performed individually for two selected upstream sub-basins: Hanzhong and Lianghekou.

After two rounds of calibration, we conducted a comprehensive comparison between the Case 4 baseline and the transfer-based simulations, focusing on both performance metrics and hydrograph agreement. Our results demonstrate that **parameters calibrated using the four-gauge configuration can be effectively transferred to the "ungauged" upstream sub-basins without significant performance degradation**

(Table 11, Fig. 10, Figs. S6-8). This finding is consistently confirmed by both the Hanzhong and Lianghekou transfer cases, indicating that the transferability of parameters remains robust across different upstream locations within the study area. Overall, these results underscore the modelling gains inherent in multi-gauge calibration, which can effectively **mitigate information-gap-induced uncertainties in nested catchment systems**.

Furthermore, a deeper exploration of this issue could **yield valuable insights for regional hydrologic simulation and monitoring network optimization**. One can envision that as local observations are sequentially withheld, the global simulation accuracy may deteriorate, potentially reaching a tipping point where the absence of a particular gauge exerts a disproportionate impact. Such a pivotal gauge would be deemed more "informative" in providing critical constraints for regional hydrologic modelling. A systematic sensitivity analysis of **gauge importance** is beyond the scope of the present study, we have expanded our discussion on this point in the revised manuscript and reserved it for future investigation.

Based on the above analysis and discussion, an additional Sect. 4.3.2 has been added to the revised manuscript. For ease of your review, the added section is provided below.

Lines 600-643:

4.3.2 Transferability of parameters

Multi-gauge calibration enhances parameter identifiability and mitigates equifinality by imposing joint constraints on shared parameters. This mechanism is rooted in the intrinsic hydrological connectivity among sub-basins within a closed water balance system. Physically, the hydrograph at a downstream gauge integrates contributions from multiple upstream sub-basins; conversely. upstream streamflow dynamics can be constrained—albeit implicitly—by downstream observations through the routing process. Within this context, a challenge closely aligned with Prediction in Ungauged Basins (PUB) is parameter transferability. A critical question arises: when upstream observations are unavailable, can downstream gauges provide sufficiently informative constraints to facilitate the effective transfer of parameters to ungauged upstream sub-basins?

To answer this question, we established an experimental setup based on Case 4 (Table 3). Specifically, VIC and RVIC parameters were parameterized in a spatially explicit manner, while soil-layer depth parameters were maintained as spatially uniform, as only shared parameters across sub-basins are amenable to transfer. The calibration employed a leave-one-out multi-gauge strategy, optimizing streamflow KGE across four gauges while holding out the single upstream gauge for independent validation. To guarantee the robustness of our findings, parameter transferability was evaluated individually for the Hanzhong and Lianghekou sub-basins (Fig. 1).

Table 11 presents a comparison of ensemble mean KGE values for each sub-basin, contrasting the baseline Case 4 calibration with the transfer-based simulations using the five best-performing parameter sets. It is evident that, despite the absence of local upstream streamflow observations, the runoff simulation performance for the transfer targets exhibits only a marginal degradation compared to the baseline, remaining well within an acceptable range. This is further illustrated in Fig. 10, where the simulated hydrograph under the transfer scenario closely mirrors the Case 4 baseline—a consistency that becomes even more pronounced at the monthly time scale (Fig. S6). These results lead to the inference that, at least under the current model configuration, downstream gauges can provide sufficiently informative constraints to facilitate reliable runoff simulations in ungauged upstream sub-basins. This inference is further corroborated by the simulation results at the Lianghekou gauge, as shown in Figs. S7 and S8.

As a supplementary analysis, a paired t-test was conducted to compare the simulation performance at the Shiquan outlet between the Case 4 baseline and the transfer scenario, using the top 40 ranked optimization results (similar to the procedure in Table 8). The results reveal no statistically significant difference (t-statistics of 0.454 and 0.487, respectively), indicating that the current information deficiency in upstream sub-basins does not exert a significant impact on the simulations at the basin outlet. Overall, the aforementioned findings highlight the inherent advantages of multi-gauge calibration in nested basins for reducing information-gap-induced uncertainties. Furthermore, one can envision that as local observations are sequentially removed from the calibration process, the global simulation accuracy may deteriorate, potentially reaching a tipping point where the absence of a particular gauge exerts a disproportionate impact. Such a gauge would be deemed more informative in providing critical constraints for regional hydrologic modelling. A systematic investigation into this phenomenon would provide valuable insights for regional hydrologic simulation and monitoring network design, ensuring that resources are prioritized for the most strategically significant locations (Nasta et al., 2025). While such an exploration is beyond the scope of the present study, it remains a compelling direction for future research.

Table 11. Comparison of ensemble mean KGE between the baseline Case 4 calibration and the leave-one-out transfer simulations.

No.	Calibration period (2005–2014)					Validation period (2015–2018)				
	Hanzhong	Yangxian	Youshui	Lianghekou	Shiquan	Hanzhong	Yangxian	Youshui	Lianghekou	Shiquan
Case 4	0.418	0.657	0.533	0.715	0.787	0.521	0.637	0.446	0.730	0.689
Hanzhong	0.418	0.649	0.521	0.698	0.787	0.516	0.628	0.440	0.709	0.699
Lianghekou	0.415	0.646	0.525	0.700	0.787	0.520	0.622	0.430	0.694	0.695

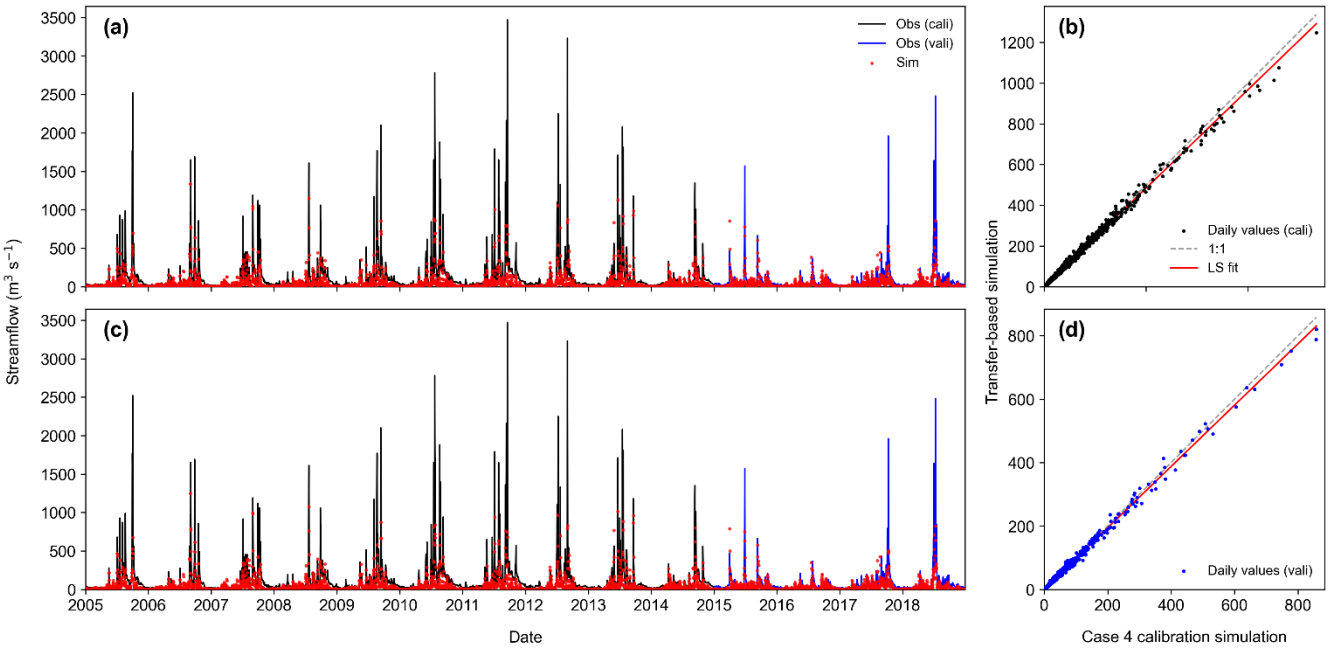

Figure 10. Leave-one-out evaluation of parameter transferability for the Hanzhong sub-basin. Panels (a) and (c) display the simulated daily hydrographs for the baseline Case 4 calibration and the transfer-based simulation, respectively. Panels (b) and (d) present the corresponding scatter plots comparing the Case 4 baseline and transfer-based simulation for the calibration and validation periods.

Supplementary:

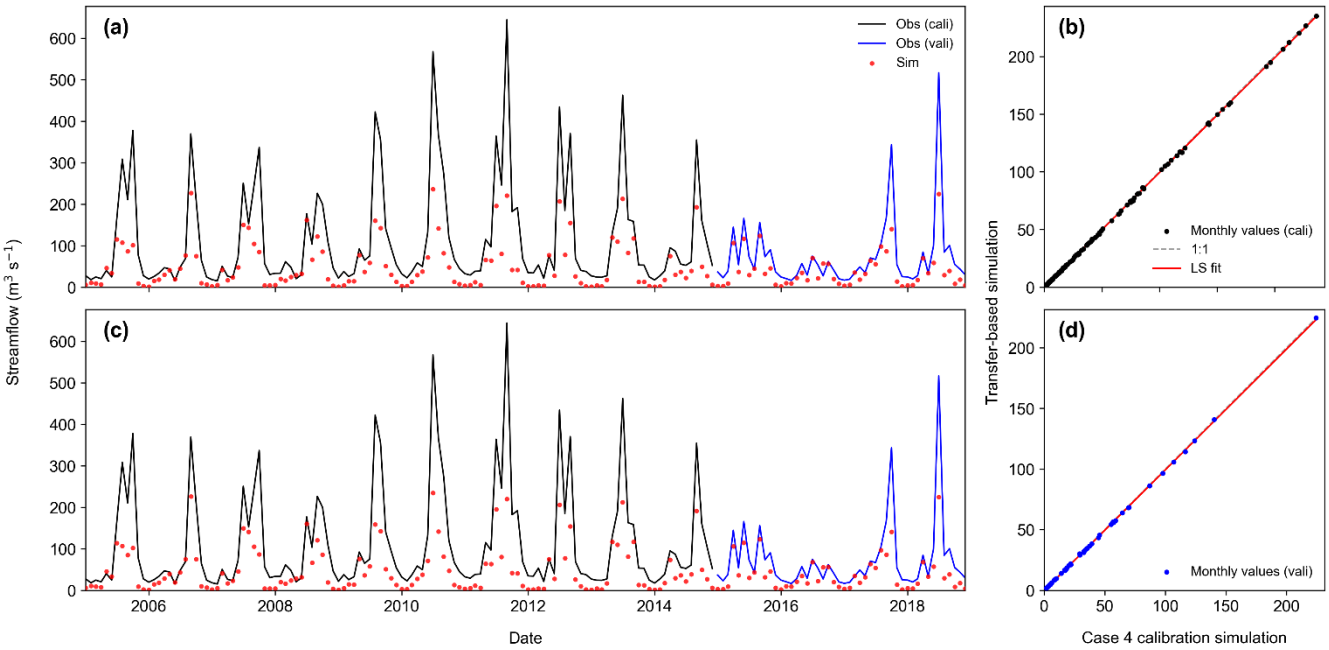

Figure S6. Leave-one-out evaluation of parameter transferability for the Hanzhong sub-basin. Panels (a) and (c) display the simulated monthly hydrographs for the baseline Case 4 calibration and the transfer-based simulation, respectively. Panels (b) and (d) present the corresponding scatter plots comparing the Case 4 baseline and transfer-based simulation for the calibration and validation periods.

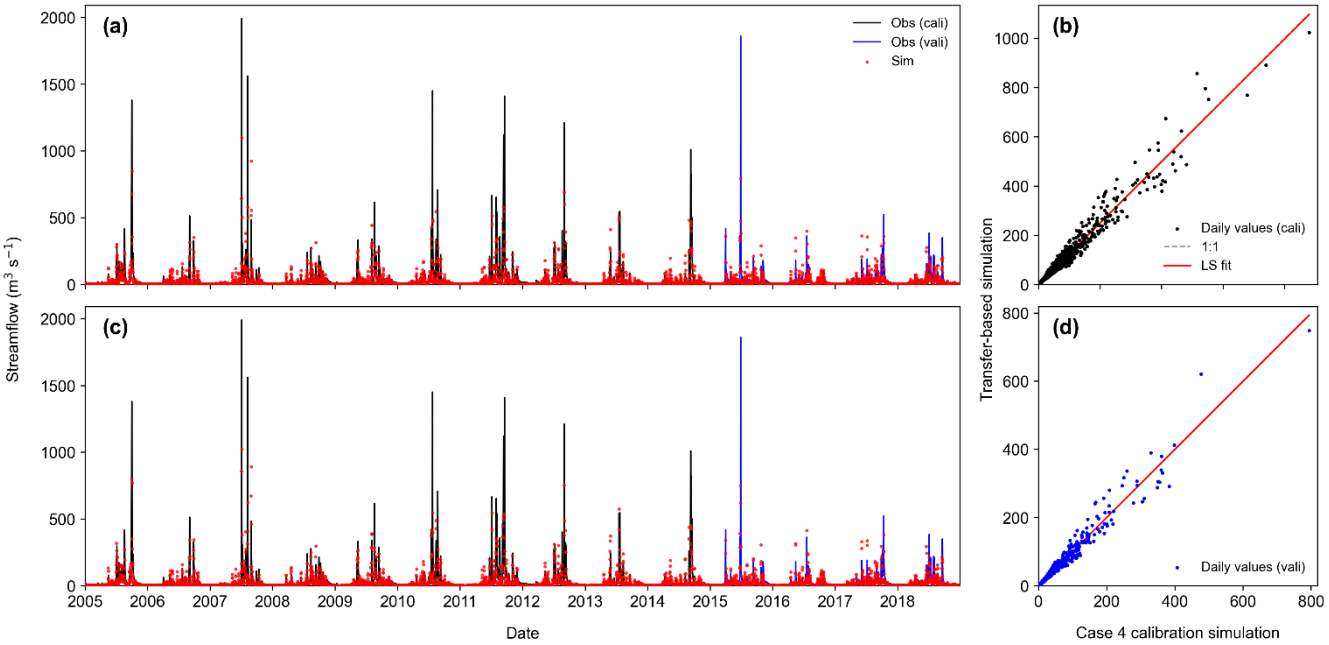

Figure S7. Leave-one-out evaluation of parameter transferability for the Lianghekou sub-basin, following the same visualization conventions as Fig. S6 but at a daily timescale.

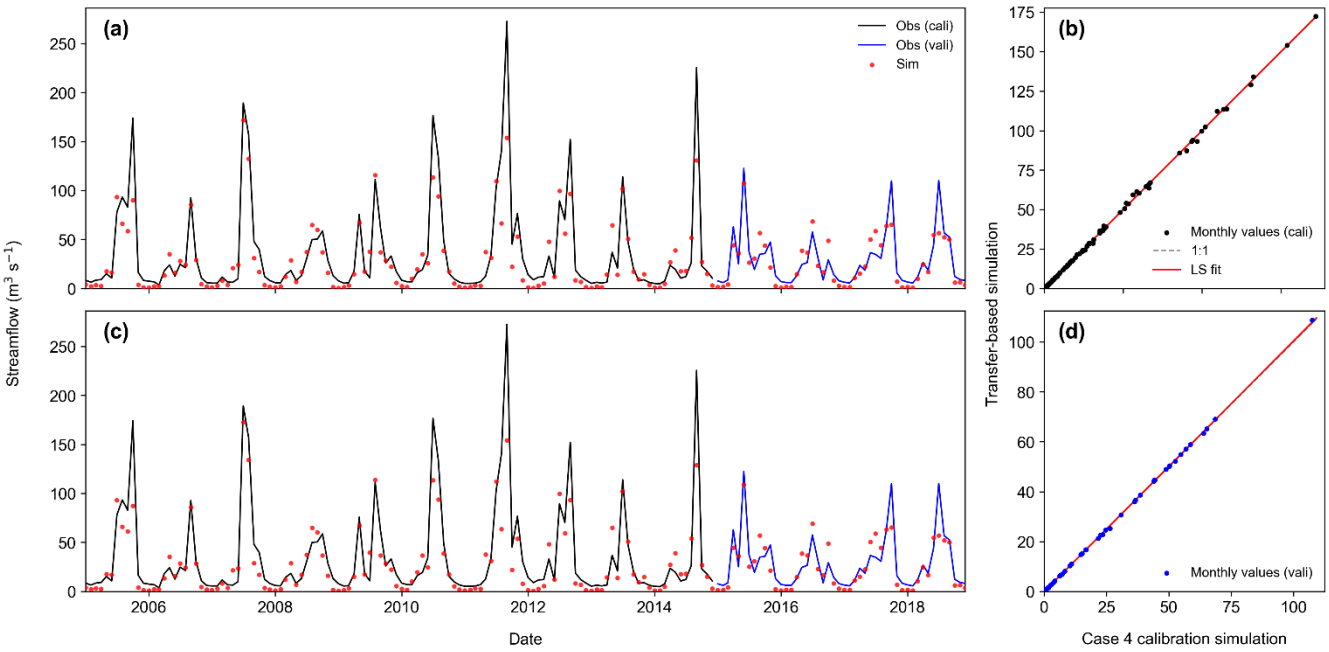

Figure S8. Leave-one-out evaluation of parameter transferability for the Lianghekou sub-basin, following the same visualization conventions as Fig. S6.

This new finding has also been incorporated into both the Abstract and the Conclusions section:

Lines 35-36:

This effectively mitigates information-gap-induced uncertainties, thereby enabling robust parameter transfer to ungauged upstream sub-basins.

Lines 847-849:

Furthermore, this strategy effectively reduces information-gap-induced uncertainties, thereby allowing for robust parameter transfer to ungauged upstream sub-basin.

EC-7/ The conclusion can be shortened and structured with bullet points to clearly highlight the mechanisms revealed by your study.

R/ Thank you for your suggestion. We have revised the conclusions section by streamlining the wording, removing less informative results, and summarizing the main findings in bullet points to improve its structure and clarity. The revised conclusions are provided below for your review.

Lines 816-867:

5 Conclusion

This study evaluated the benefits of spatially explicit parameterization and multi-gauge calibration for improving model realism and further disentangled their respective contributions, as well as their cross-benefits, through a dedicated experimental framework, termed EF-SPM. Built on the new-generation Variable Infiltration Capacity (VIC-5) model integrated with the multiscale parameter regionalization technique, the framework was applied to the representative nested Upper Han River Basin (UHRB) through intensive calibration experiments. Our key findings and conclusions are given in the following.

1. Both spatially explicit parameterization and multi-gauge strategy led to consistent gains in model performance across all sub-basins. Spatially explicit parameterization strategy mainly improved simulations under moderate-flow and high-flow conditions and substantially enhanced peak-flow representation, as indicated by 18% improvement in $\%BiasFHV_1$ at the main outlet; however, low-flow performance deteriorated. The multi-gauge calibration strategy offers a distinct advantage over its single-gauge counterpart by significantly enhancing parameter identifiability, demonstrated in the more clustered distribution of optimal and sampled candidate solutions within the parameter space. Furthermore, this strategy effectively reduces information-gap-induced uncertainties, thereby allowing for robust parameter transfer to ungauged upstream sub-basin.
2. The combined use of the two strategies yielded clear cross-benefits relative to their separate use. On the one hand, spatially explicit parameterization increases parameter dimensionality and thus poses a potential risk of reduced identifiability, whereas multi-gauge calibration strengthens observational constraints that help mitigate this issue. On the other hand, under spatially uniform parameterization, multi-gauge calibration appears to be subject to an optimization ceiling, manifested by pronounced trade-offs among gauge-specific objectives in the multidimensional objective space and by the formation of continuous, convex, arc-shaped Pareto fronts. By introducing spatially explicit parameterization, the feasible solution space is expanded, thereby raising the upper bound of achievable simulation accuracy.

Collectively, our findings suggest that enhancements in spatial representation and the strengthening of calibration constraints should be pursued in parallel during the model development. Focusing on either aspect in isolation is unlikely to produce genuinely reliable improvements and may result in right results for the wrong reasons. Looking ahead, continued advances in observational technologies and computational capacity are likely to move hydrological modelling towards a Model-Data Infusion framework, with the integration of increasingly informative data into the modelling process representing a shared direction for improving model realism.

Responses to RC1

Many thanks for taking the time and effort to review our paper. All comments from Referee #1 are addressed below with point-by-point responses.

RC1-0/ The manuscript presents a timely and highly relevant investigation into the combined effects of spatially explicit parameterization and multi-gauge calibration on hydrological modeling. The paper is well-written, logically structured, and provides a meaningful steppingstone for the advancement of modern Model-Data Infusion frameworks. However, there are several issues needed to be addressed before publication.

R/ Thanks for your positive feedback and recognition of our work. Your comments are valuable for revising and improving our paper. Below, we provide detailed responses to each of your concerns. The corresponding revisions will be incorporated into the subsequent revised manuscript. To maintain consistency with the original numbering, this comment is labeled as 0.

Specific comments:

RC1-1/ 1. In the parameter calibration, the authors did not consider three key VIC parameters that are commonly calibrated, namely Ds, Ws, and Dm. Although Gou et al. (2020) is cited in the manuscript, that study does not provide sensitivity analysis results for the specific basin investigated here. Including these parameters in the calibration process could potentially lead to different results. Therefore, the authors are encouraged to provide sensitivity analysis results for the study basin to justify the exclusion of these parameters. Otherwise, I believe that Ds, Ws, and Dm should be incorporated into the calibration.

R/ Thank you very much for your insightful comment and your expertise regrading the VIC model. As you correctly pointed out, D_s, W_s, D_m, b, and the three soil layer depths are among the most sensitive parameters in VIC and generally require calibration against basin-specific hydrological observations. This has been widely acknowledged in previous studies and in practical applications of the VIC model (Wen et al., 2012; Gou et al., 2020), and we fully agree with this point.

It should be clarified that there may be a **misunderstanding regarding parameter naming and the equivalence of the baseflow formulation**. Specifically, the free parameters adopted in this study include the variable infiltration curve parameter b, the **three baseflow parameters (D_1, D_2, D_3)**, **the three soil layer depths (d_1, d_2, d_3)**, the flow velocity v, and the diffusion coefficient D.

The VIC model employs the Arno baseflow formulation, which is generally expressed as follows:

$$Q_b = \begin{cases} \frac{D_s D_m}{W_s \theta_b^s} \cdot \theta_b, & 0 \le \theta_b < W_s \theta_b^s \\ \frac{D_s D_m}{W_s \theta_b^s} \cdot \theta_b + (D_m - \frac{D_s D_m}{W_s})(\frac{\theta_b - W_s \theta_b^s}{\theta_b^s - W_s \theta_b^s})^c, & \theta_b \ge W_s \theta_b^s \end{cases} \quad \text{(RC1-1)}$$

where θ_b is the soil moisture in the bottom layer (mm); θ_b^s denotes the maximum soil moisture in the bottom layer (mm); D_m represents the maximum baseflow (mm step^{-1}); D_s is the fraction of D_m at which nonlinear baseflow begins, with a range of 0-1; W_s denotes the fraction of θ_b^s at which nonlinear baseflow initiates, also ranging from 0 to 1; The parameter c is the exponent governing the nonlinear

portion of the Arno baseflow curve and is typically set to 2. Clearly, D_s, W_s, and D_m are the key parameters controlling the shape of the Arno baseflow curve and, consequently, the generation of baseflow; therefore, they require calibration.

In 2001, Nijssen et al. (2001) revisited the Arno model from the perspective of linear reservoirs and proposed an equivalent formulation:

$$Q_b = \begin{cases} D_1\theta_b, & 0 \le \theta_b < D_3 \\ D_1\theta_b + D_2(\theta_b - D_3)^{D_4}, & \theta_b \ge D_3 \end{cases} \quad \text{(RC1-2)}$$

where D_1 denotes the linear reservoir coefficient; D_2 is the nonlinear reservoir coefficient; D_3 represents the soil moisture (mm) at which nonlinear baseflow begins; D_4 is analogous to the parameter c. This formulation is also known as the **Nijssen form of the Arno model**. Its advantage lies in effectively **reducing the interactions among parameters, thereby facilitating calibration** (Mizukami et al., 2017). Currently, this formulation has been incorporated into the official VIC model and can be specified through the global parameter settings.

Therefore, Eq. (RC1-1) and Eq. (RC1-2) are equivalent, and **calibrating D_1, D_2, D_3 is effectively the same as calibrating D_s, D_m, W_s**. These two sets of parameters can be converted into each other (Mizukami et al., 2017).

We acknowledge that this misunderstanding may have been caused by the similar naming of D_1, D_2, D_3 and d_1, d_2, d_3, as well as the unclear description in Nijssen formulation. To enhance clarity, the revised manuscript now includes a **more detailed description of VIC runoff mechanisms** (covering both equivalent Arno model formulations), and **the soil layer depths are now denoted as h_1, h_2, h_3.**

The corresponding modifications are appended below for your review:

Lines 251-272:

In the VIC model, total runoff (Q_t) is partitioned into surface runoff (R) and baseflow (Q_b). Under the standard three-layer vertical soil structure (VIC-3L), surface runoff is generated from the thin top layer and the upper layer according to the variable infiltration capacity curve, as shown below (Liang et al., 1994):

$$R = \begin{cases} PE - (W_m - W_0), & PE + i_0 \ge i_m \\ PE - (W_m - W_0) + W_m\left(1 - \frac{PE + i_0}{i_m}\right)^{1+b}, & PE + i_0 < i_m \end{cases} \quad (3)$$

where PE denotes the effective precipitation; W_m represents the mean infiltration capacity of the grid cell; i_m is the maximum infiltration capacity at a point within the grid cell; W_0 is the initial soil moisture and i_0 is the corresponding initial infiltration capacity; and b is the shape parameter of the variable infiltration capacity curve. All of these quantities refer to the upper two soil layers. The VIC model does not explicitly represent soil drainage processes, such as interflow, or groundwater storage above the water table. Instead, these components are treated in a conceptual lumped manner and represented as baseflow from the bottom soil layer by a single term, Q_b, which is parameterized using the Arno baseflow scheme, as follows (Liang et al., 1994):

$$Q_b = \begin{cases} \frac{D_s D_m}{W_s \theta_b^s} \cdot \theta_b, & 0 \le \theta_b < W_s\theta_b^s \\ \frac{D_s D_m}{W_s \theta_b^s} \cdot \theta_b + (D_m - \frac{D_s D_m}{W_s})(\frac{\theta_b - W_s\theta_b^s}{\theta_b^s - W_s\theta_b^s})^c, & \theta_b \ge W_s\theta_b^s \end{cases} \quad (4)$$

where θ_b^s is the maximum soil moisture of the bottom layer; D_m is the maximum baseflow; D_s is the fraction of D_m at which nonlinear baseflow begins; W_s is the fraction of θ_b^s at which nonlinear baseflow begins; and c is the exponent of the nonlinear part of the Arno baseflow curve, which is usually set to 2. All of these quantities refer to the bottom soil layer. The Arno baseflow equation can be equivalently reformulated as a coupled linear reservoir and a nonlinear reservoir, and can thus be expressed in the following form, thereby alleviating parameter interactions (Nijssen et al., 2001):

$$Q_b = \begin{cases} D_1\theta_b, & 0 \le \theta_b < D_3 \\ D_1\theta_b + D_2(\theta_b - D_3)^{D_4}, & \theta_b \ge D_3 \end{cases} \quad (5)$$

where D_1 is the linear reservoir coefficient; D_2 is the nonlinear reservoir coefficient; D_3 denotes soil moisture at which baseflow transition occurs from linear to nonlinear; and D_4 has the same conceptual meaning as c. This formulation is also known as the Nijssen form of the Arno model, has been incorporated into the official VIC model and can be specified through the global parameter file.

Lines 297-301:

It is worth noting that VIC supports two equivalent baseflow parameterizations. Following Mizukami et al. (2017), we adopt the Nijssen formulation (i.e., D_1, D_2 and D_3) to avoid the parameter interactions present in the original Arno formulation (Eq. 5). Namely, calibration of D_1, D_2 and D_3 is equivalent to calibrating D_s, D_m and W_s, which have been identified as sensitive parameters in previous studies (Wen et al., 2012; Gou et al., 2021).

Lines 306-311:

In the classic VIC-3L (i.e., VIC three-layer) configuration, the upper two soil layers respond rapidly to precipitation, generating surface flow according to the Xinanjiang formulation, whereas the bottom layer governs baseflow, resulting in a slower runoff response (Eqs. 3 and 5). Different soil layer thicknesses produce significantly different runoff responses; consequently, these thicknesses (h_1, h_2, and h_3) are typically treated as free parameters during calibration to fine-tune the VIC behavior (Gou et al., 2020).

Lines 349-353:

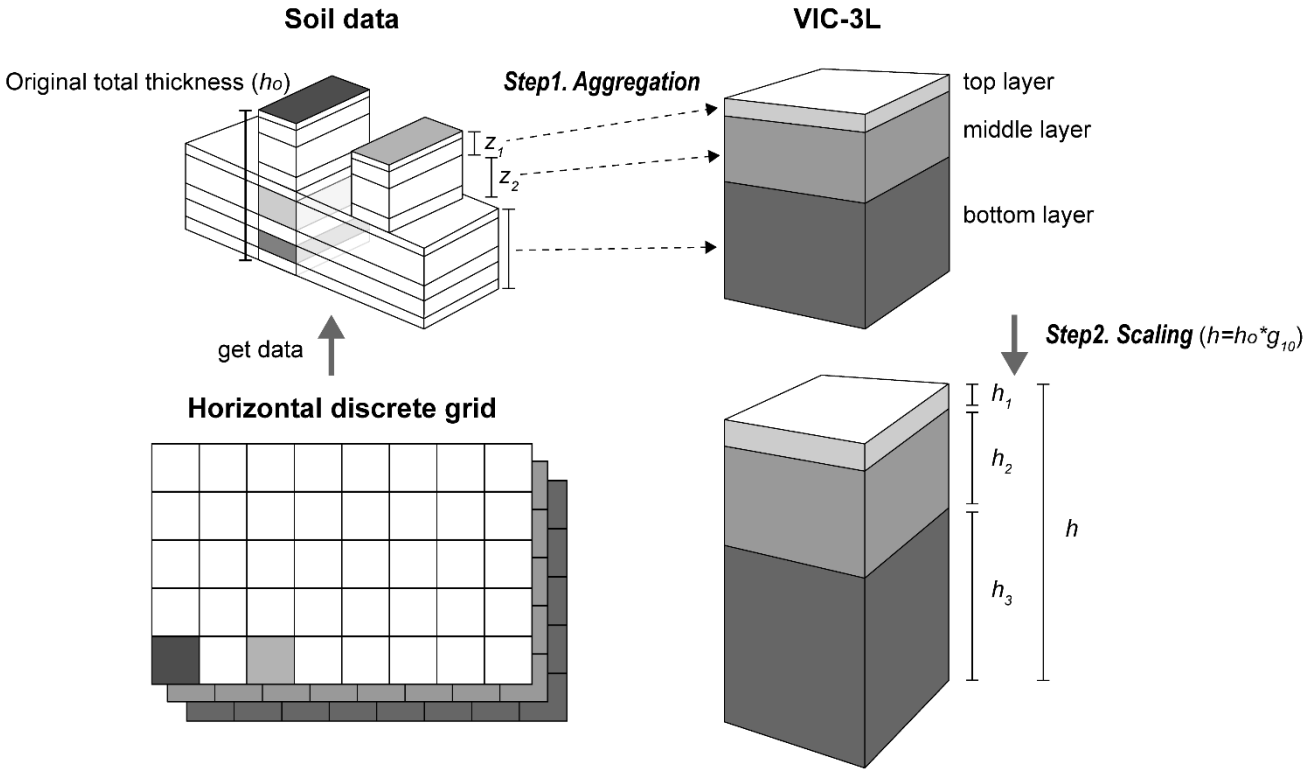

Figure 2. Schematic of the two-step mechanism for reconciling soil data with VIC three-layer (VIC-3L) vertical structure. In the aggregation step, the original soil layers are merged into three VIC soil layers based on the layering numbers z_1 and z_2. The scaling step then applies a scaling factor (g_{10}) to the original total thickness (h_o) to obtain the model-prepared layer thicknesses h_1, h_2, and h_3.

RC1-2/ 2. The use of MPR is indeed an effective approach for deriving distributed parameters; however, one of its key limitations lies in the uncertainty associated with the transfer functions. Previous studies have shown that different transfer functions can lead to substantially different calibration results. For example, Gou et al. (2021) adopted transfer functions for D1–D3 that differ from those used in this study. It remains unclear whether such differences could lead to different conclusions. The manuscript currently lacks analysis and discussion on this issue, which should be addressed to strengthen the robustness of the study.

Gou, Jiaojiao, et al. "CNRD v1. 0: a high-quality natural runoff dataset for hydrological and climate studies in China." Bulletin of the American Meteorological Society 102.5 (2021): E929-E947.

R/ This is a similar issue caused by the confusion noted in RC1-1, and we thank you for bringing it to our attention. In our parameter estimation using the MPR approach, we consulted a wide range of literature to find suitable transfer functions. Among these, **Gou et al. (2020), Gou et al. (2021), and Mizukami et al. (2017)** were particularly important, as they are pioneering works that provide essential transfer functions for applying the MPR approach within the VIC model.

However, we note that the **naming of parameters** and the **formulation of baseflow** module vary across

different studies, which we found to be a challenge during our practical implementation. For example, Gou et al. (2020, 2021) tend to use the traditional Arno baseflow formulation, representing the baseflow parameters with D_s, D_m, and W_s, and the three soil layer depths with $D_{1/2/3}$. In contrast, Mizukami et al. (2017) prefer the Nijssen version of the Arno baseflow formulation, using $D_{1/2/3}$ to denote the baseflow parameters equivalent to D_s, D_m, and W_s, while representing the three soil layer depths with $d_{1/2/3}$. **Integrating the formulations from these studies can be quite confusing and prone to errors**.

Since we adopted the Nijssen version of the Arno baseflow formulation, we followed the naming convention used by Mizukami et al. (2017) to maintain consistency with their work. Consequently, our notation appears different from that of Gou et al. (2021). Nevertheless, fundamentally, **calibration of $D_{1/2/3}$ is equivalent to that of D_s, D_m, and W_s**, as we discussed in our reply to RC1-1.

RC1-3/ 3. The manuscript currently lacks sufficient statistical validation of the calibration experiments. Without statistical evaluation, it is difficult to determine whether the reported improvements reflect meaningful advancements or merely small fluctuations. For instance, an KGE difference between 0.715 and 0.716 in Table 6 is not necessarily meaningful without significance testing. The authors are encouraged to incorporate appropriate statistical tests, such as the Wilcoxon signed-rank test or paired t-tests for comparisons. In addition, reporting the standard deviation (you only showed ensemble mean) across ensemble runs would substantially strengthen the credibility of the results.

R/ We sincerely appreciate your constructive suggestion. We fully agree that incorporating more explicit statistical tests is essential for strengthening the robustness and credibility of our conclusions. Following your recommendation, we have performed **pairwise paired t-tests across all cases**. The results are presented in Table 8, and a corresponding analysis has been added to Section 4.1 to further interpret these findings. For your convenience, the relevant revisions, including the updated text and table, are provided below.

Lines 466-471:

To further examine the statistical significance of performance differences across various cases, we conducted pairwise t-tests on the KGE at the Shiquan station (basin outlet), as shown in Table 8. It is evident that, at a significance level of 0.05, Case 4 performs significantly better than all other cases, while Case 8 performs significantly worse than all other cases. Moreover, the patterns identified in the previous analysis are further confirmed by this significance test, with the more complex configurations generally outperforming the simpler ones. It is worth noting that this comparison is conducted under a single-objective context, as Cases 6–8 only consider streamflow at the basin outlet as objective functions.

Lines 480-483:

Table 8. Pairwise t-test results of the ensemble performance metrics (KGE) for the top 40 ranked optimization results across all calibration cases. The table shows the t-statistics for the comparison between each pair of cases, with significant differences indicated by asterisks ($p < 0.05$).

	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8
Case 1	–	7.357*	6.820*	-11.652*	0.512	11.331*	2.965*	7.056*

Case 2	-7.357*	–	5.431*	-12.681*	0.189	11.040*	-0.471	6.920*
Case 3	-6.820*	-5.431*	–	-13.425*	-1.082	9.668*	-5.316*	6.343*
Case 4	11.652*	12.681*	13.425*	–	3.842*	13.970*	12.084*	8.459*
Case 5	-0.512	-0.189	1.082	-3.842*	–	8.108*	-0.234	6.308*
Case 6	-11.331*	-11.040*	-9.668*	-13.970*	-8.108*	–	-11.050*	1.568
Case 7	-2.965*	0.471	5.316*	-12.084*	0.234	11.050*	–	6.935*
Case 8	-7.056*	-6.920*	-6.343*	-8.459*	-6.308*	-1.568	-6.935*	–

On the other hand, reporting the standard deviation is generally advisable. However, in the context of this study, Table 7 only presents the ensemble mean of **top five optimal solutions**, resulting in a **very limited sample size**. Under such conditions, the calculation of standard deviation may introduce considerable uncertainty and potentially be misleading. For this reason, we did not include this metric. Nevertheless, our additional analysis indicates that, owing to the characteristics of the optimization algorithm and the relatively large number of calibration iterations performed, **the top five optimal solutions remain stable**. This point is important for clarifying the robustness of the results and for addressing your concern.

RC1-4/ 4. The description of the two-step mechanism for reconciling soil data with the VIC three-layer (VIC-3L) vertical structure is confusing. Since Table 2 already presents the transfer functions for D1–D3, it is unclear what additional role this two-step procedure plays. Because this component underpins the subsequent analysis, a more detailed and transparent explanation is essential.

R/ This issue is partly due to the naming inconsistency, as discussed in our responses to RC1-1 and RC1-2. D_1–D_3 refer to the parameters governing the Arno baseflow curve (Eq. (2)), not the soil layer depths d_1–d_3 associated with the two-step mechanism discussed here. The misunderstanding likely stems from the fact that **our manuscript follows a different convention than Gao et al. (2021), in which D1–D3 were used to represent soil layer thicknesses**.

On the other hand, unlike some previous VIC applications, the two-step mechanism adopted here accounts for the **dynamic, layer-specific assignment of soil attributes**. As a result, the soil layer delineation not only determines the layer thicknesses but also influences the corresponding soil attributes, such as soil texture.

Specifically, the original SoilGrids dataset provides six soil layers (0–5 cm, 5–15 cm, 15–30 cm, 30–60 cm, 60–100 cm, and 100–200 cm, hereafter SoilGrids-SL), each with distinct soil texture and bulk density attributes. However, the VIC model typically adopts a three-layer soil structure (hereafter, VIC-3L-SL). Therefore, the multiple soil attributes from SoilGrids-SL need to be assigned to the three layers of VIC-3L-SL. **To enable this process to be dynamically calibrated**, we implemented a two-step mechanism (Fig. 2). First, two layering number parameters (z_1 and z_2) determine the correspondence between SoilGrids-SL and VIC-3L-SL. Based on this mapping, the soil attributes from the relevant SoilGrids layers are aggregated to derive the soil attributes of the three VIC layers (i.e., averaged). Second, a scaling factor (g_{10}) is applied to uniformly adjust the total soil column thickness. This mechanism introduces three calibratable parameters (z_1, z_2, and g_{10}). **The first two parameters control the allocation of soil attributes and the relative thickness proportions of the three VIC soil layers, while the third parameter regulates the overall soil column thickness in VIC-3L**. This mechanism has been applied

in Mizukami et al. (2017), and has been shown to be important for overall model performance. Our results also demonstrate this effect, as discussed in Section 4.4.1 with respect to soil layer depth discretization.

We believe that this issue arose also from insufficient clarity in the original manuscript. Therefore, we have addressed it in two ways: first, by revising the parameter nomenclature (see our response to RC1-1); and second, by providing additional clarification of the two-step mechanism, which has been incorporated into the revised manuscript as follows.

Lines 309-318:

Different soil layer thicknesses produce significantly different runoff responses; consequently, these thicknesses (h_1, h_2, and h_3) are typically treated as free parameters during calibration to fine-tune the VIC behavior (Gou et al., 2020). In addition, the original soil dataset (i.e., SoilGrids1km) provides six soil layers, each characterized by distinct properties (e.g., soil texture), whereas VIC-3L adopts a three-layer soil structure. In order to feed original soil data into the transfer function, it is necessary to reconcile its soil-layer structure with the VIC vertical structure. As shown in Fig. 2, a two-step mechanism is applied here, involving aggregation and scaling. In the first step, the original soil data are aggregated into a standard three-layer structure defined by the layering numbers z_1 and z_2, with the corresponding layer thicknesses summed and the soil properties averaged within each standard layer. In the second step, the original thickness (h_0) is scaled by g_{10} to allow flexible adjustment of the final computational layer thickness.

RC1-5/ 5. The authors designed eight calibration experiments; however, the current numerical labeling makes it difficult for readers to remember the specific configurations during subsequent discussion. It is recommended that the authors adopt clearer and more descriptive naming conventions to distinguish the different experiments, which would improve readability and interpretability.

R/ You are right that the original experiment names could be confusing. We have improved their readability by **adding aliases**, as reflected in the updated Table 4 and Fig. 3. Corresponding references throughout the manuscript have also been updated to include these aliases for enhanced clarity.

Lines 393-397:

Table 4. Summary of experiments within the EF-SPM framework, showing the cases considered, the configurations compared, and their purposes.

Experiment	Cases	Defined purpose	Alias
Exp. 1-1	Case 1 and Case 5	Uniform vs. fully distributed parameterization (under multi-gauge calibration)	UD-MG
Exp. 1-2	Case 7 and Case 8	Uniform vs. fully distributed parameterization (under single-gauge calibration)	UD-SG
Exp. 2-1	Case 5 and Case 8	Multi-gauge vs. single-gauge calibration (under distributed parameterization)	MS-DP
Exp. 2-2	Case 1 and Case 7	Multi-gauge vs. single-gauge calibration (under uniform parameterization)	MS-UP
Exp. 3-1	Case 4 and Case 5	Uniform vs. spatially explicit soil layer depths parameterization (under multi-gauge calibration)	UD-MG: Depth
Exp. 3-2	Case 3, Case 2 and Case 4	Fixed vs. uniform vs. distributed RVIC parameterization (under multi-gauge calibration)	FUD-MG: RVIC
Exp. 4	Case 6 and Case 5	Simplest baseline setup vs. Full-Complexity setup	Baseline-Complex

*Alias abbreviation: UD = uniform vs. distributed parameterization; MS = multi-gauge vs. single-gauge calibration; FUD = fixed vs. uniform vs. distributed parameterization; MG = multi-gauge calibration; SG = single-gauge calibration; DP = distributed parameterization; Depth = soil layer depths; RVIC = RVIC-specific parameterization; Baseline-Complex = Simplest baseline setup vs. Full-Complexity setup.

Lines 403-405:

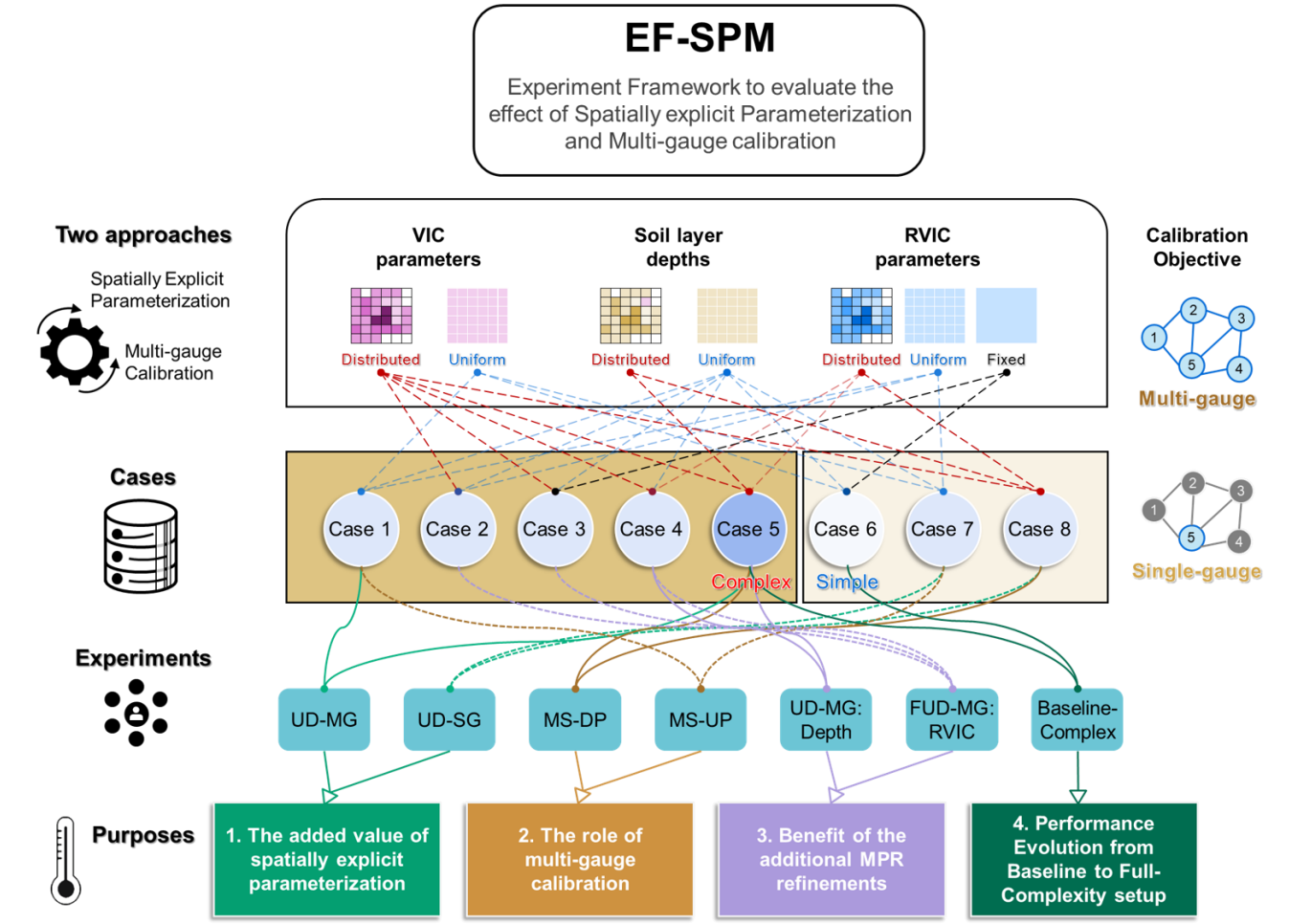

Figure 3. Schematic illustration of the Experiment Framework to evaluate the effect of Spatially explicit Parameterization and Multi-gauge calibration (EF-SPM).

RC1-6/ 6. When applying the NSGA-II algorithm for multi-objective optimization, the implementation details are not sufficiently described. Given that the study basin includes five gauging stations, does this imply that five separate objectives were defined for calibration? The authors should clearly specify how the objective functions were formulated and aggregated, and how the multi-objective framework was structured in practice.

R/ We thank you for pointing out the lack of clarity in our methodology. In response, we have added the following clarification in Section 3.4.

Lines 420-423:

Under the two optimization algorithms, CMA-ES optimizes the KGE at the basin outlet (Shiquan) as the single-objective function (KGE_{Q5}), whereas NSGA-II performs multi-objective optimization in the multi-dimensional Pareto space, considering the KGE of streamflow at the five subbasin gauges as separate objectives (KGE_{Q1}, KGE_{Q2}, KGE_{Q3}, KGE_{Q4}, KGE_{Q5}).

It should be noted that, under the NSGA-II framework, the five objectives **are treated as five**

independent dimensions in the Pareto space and are not aggregated. Further clarification of Fig. 10 will indeed help to better address this point and provide a more comprehensive understanding of the optimization process.

In Fig. 11 (below), each subplot represents one of the optimization objectives, corresponding to KGE_{Q1}, KGE_{Q2}, KGE_{Q3}, KGE_{Q4}, and KGE_{Q5} (i.e., the streamflow at five gauges). Within this five-dimensional Pareto space, NSGA-II explores the trade-offs between the conflicting objectives. The Pareto front, highlighted in red in the figure, represents a set of solutions where no other solution is strictly better across all objectives. In other words, each point on the Pareto front is non-dominated, meaning that it cannot be improved in any objective without worsening at least one other objective. However, it should be emphasized that there is no single solution that is superior across all objectives; instead, the Pareto front provides a range of equally optimal solutions, each reflecting a different trade-off between the objectives.

Therefore, **the multi-objective optimization algorithm of NSGA-II ultimately produces a set of solutions on the Pareto front through this process, without the need to aggregate the five objectives**. The final optimal solution is then selected from this set of solutions on the Pareto front by manually choosing the most balanced solution, considering the trade-offs between the objectives.

We hope that our response addresses your concerns and provides the clarity needed.

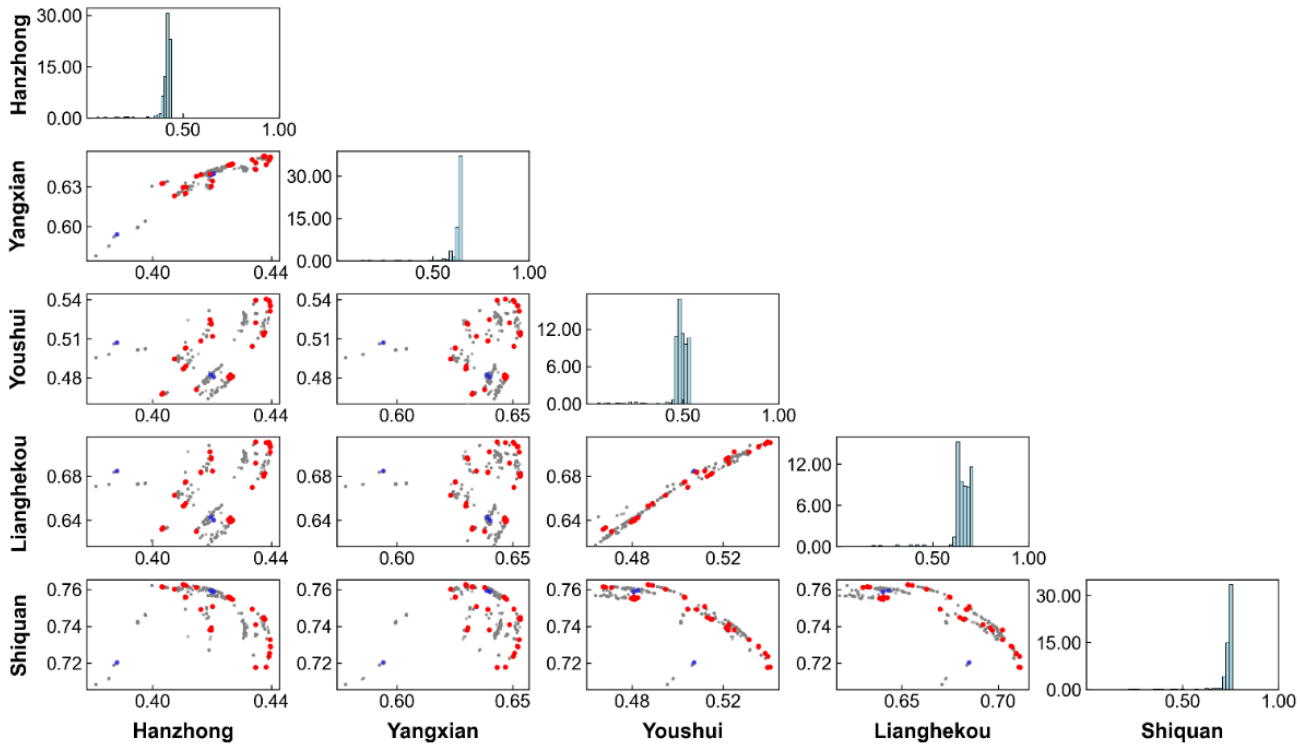

Figure 11. Objective space scatter plot matrix for Case 1. The matrix depicts the trade-offs between calibration objectives, where each axis corresponds to the Kling-Gupta Efficiency (KGE) of simulated streamflow for a sub-basin (gauge). Grey dots represent all candidate solutions from the calibration process. The initial and final Pareto fronts are highlighted in blue and red, respectively. Histograms along the diagonal show the distribution of the KGE for each individual objective.

RC1-7/ 7. The authors should include, in Table 5, the ranges for all parameters to be optimized. Providing the parameter bounds would improve transparency and allow readers to better assess

the robustness and reproducibility of the calibration procedure.

R/ Thank you for pointing out the insufficient methodological description in our manuscript. Following your suggestion, we have added Table 6 to provide a detailed description of the parameters subject to calibration and their feasible ranges, in order to improve the transparency of the methodology. The table and the added text are provided below for your convenience.

Lines 390-391:

The parameters to be calibrated in the EF-SPM experimental framework, along with their descriptions and feasible ranges, are summarized in Table 6.

Lines 400-401:

Table 6. Description and feasible ranges of the free parameters subject to calibration.

Category	Parameters	Unit	Related parameters and description	Feasible range	
				Lower	Upper
VIC parameters	g_1, g_2	–	g-parameters for Variable infiltration curve parameter (Distributed)	−2.0, 0.8	1.0, 1.2
	b	–	Variable infiltration curve parameter (Uniform)	0.001	0.5
	g_3	–	g-parameter for D_1 (Distributed)	1.75	3.5
	g_4	–	g-parameter for D_2 (Distributed)	1.75	3.5
	g_5	–	g-parameter for D_3 (Distributed)	0.001	2.0
	D_1	–	Linear reservoir coefficient (Uniform)	0.001	1.0
	D_2	–	Nonlinear reservoir coefficient (Uniform)	0.001	1.0
	D_3	–	Percentile of the bottom soil layer thickness (Uniform)	0.001	1.0
Soil layer depths	g_{10}	–	Scaling factor of total depth	0.1	4.0
	z_1, z_2	–	The layering number of the top two soil layers in the SoilGrids 6-layer soil profile	1, 3	2, 5
RVIC parameters	t_p	[h]	Time of peak occurrence	1.0	24.0
	μ	[h^{-1}]	Rising coefficient	2.0	10.0
	m	–	Recessing coefficient	0.5	6.0
	g_6, g_7, g_8	–	g-parameters for flow velocity (Distributed)	0.01, 0.1, 0.2	0.5, 0.3, 0.4
	g_9	–	g-parameter for flow diffusivity (Distributed)	0.01	0.5
	v	[m s^{-1}]	Flow velocity (Uniform)	0.01	3.0
	D	[m^2 s^{-1}]	Flow diffusivity (Uniform)	10.0	4000.0

RC1-8/ 8. Although the authors provided a zoomed-in view in Figure 5, the differences remain unclear. From my perspective, the four schemes perform almost identically in the high-flow segment.

R/ We appreciate your careful examination of Fig. 5. To better highlight the high-flow differences among the four schemes, we have updated Fig. 5 to include a **box-and-scatter representation**. This visualization makes the distribution of simulated high flows more apparent, showing that the spatially explicit parameterization schemes (Cases 5 and 8) tend to produce **higher peaks** compared to the uniform schemes (Cases 1 and 7). These differences are further supported by the signature metrics presented in Table 9 and 10. Additionally, **a logarithmic scale version of Fig. 5** has been included in the Supplementary Material (**Fig. S3**), which more clearly reveals the aforementioned phenomena.

We have added a clarifying statement in the manuscript to highlight these distinctions.

Line 502-505:

In general, spatially explicit parameterization (Cases 5 and 8) offer improved simulation of the high-flow segment, yielding higher peak values (as shown in the box-and-scatter representation Fig. 5); however, this comes at the expense of accurately capturing extreme low flows, namely, an overestimation of baseflow. This behaviour is also evident in the scatter comparison on a logarithmic scale (Fig. S3).

Supplementary:

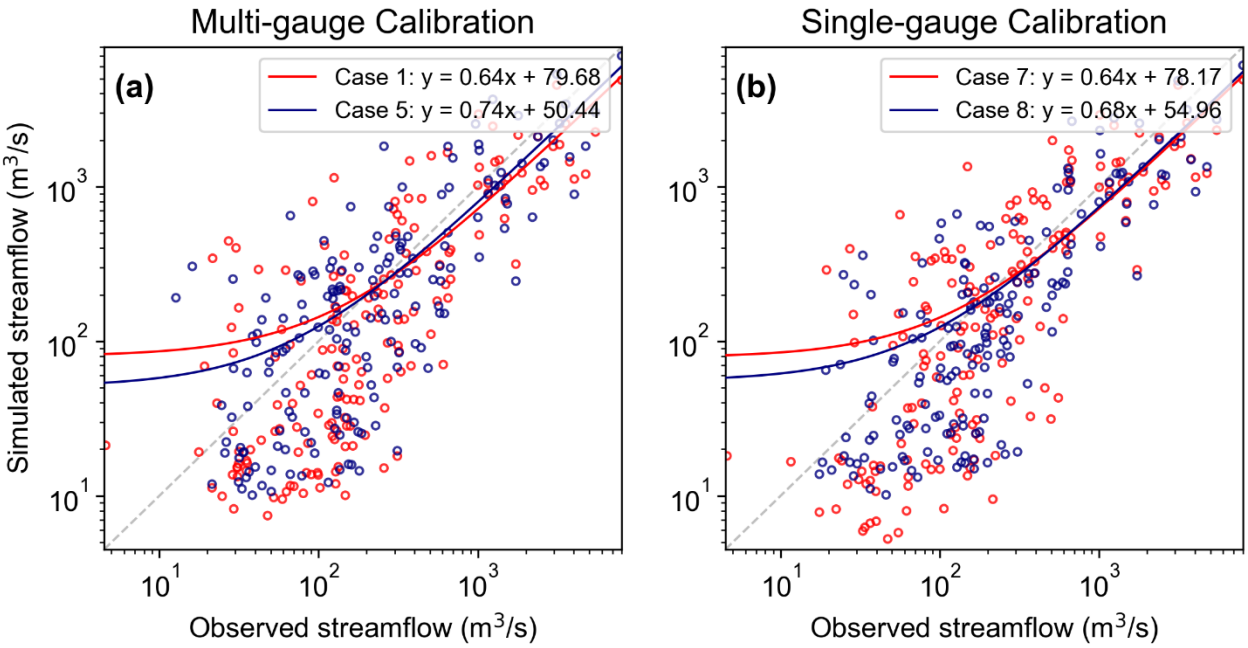

Figure S3. Scatterplots with least-squares regression lines comparing observed and simulated daily streamflow at the Shiquan station during the validation period, under different case configurations. The grey dashed line represents the 1:1 line. Flows below 1000 m^3 s^{-1} were randomly thinned for clarity, and both axes are shown on a logarithmic scale.

Lines 522-524:

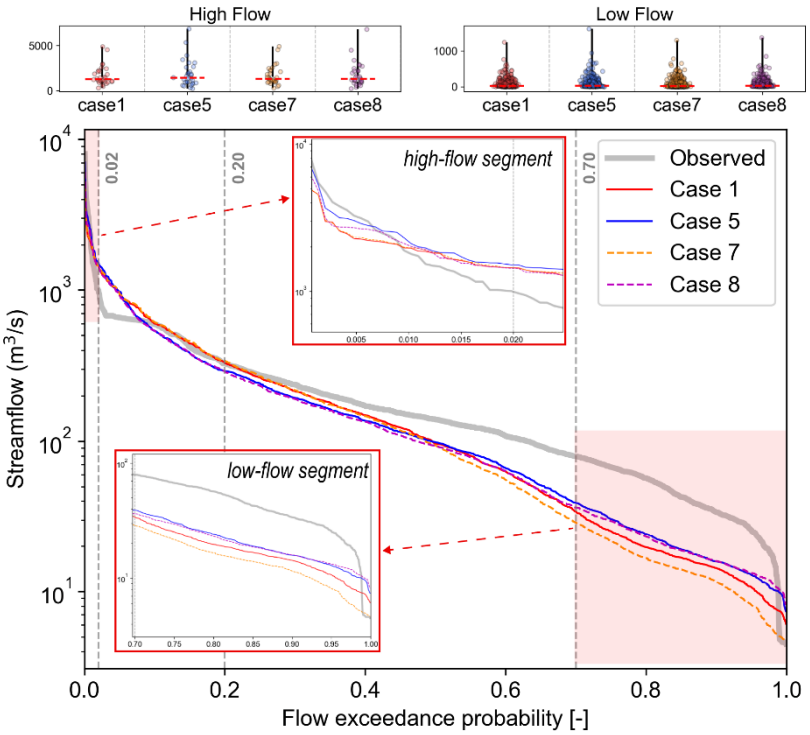

Figure 5. Comparison of observed and simulated flow duration curves for different case configurations at the Shiquan station during the validation period.

Responses to RC2

Many thanks for taking the time and effort to review our paper. All comments from Referee #2 are addressed below with point-by-point responses.

RC1-1/ Overall, I thought this was a well-designed study with a well written manuscript. The results demonstrated that spatially explicit parameterization and multi-gauge calibration increased overall accuracy of the VIC model simulated streamflow.

R/ Thank you for your positive and encouraging comments. Your comments are valuable for revising and improving our paper. Below, we provide detailed responses to each of your concerns. The corresponding revisions will be incorporated into the subsequent revised manuscript.

I have a few suggestions below:

RC1-2/ Introduction: The introduction needs a justification and explanation of the VIC model at the beginning or after line 70. The introduction, as it is currently, provides a specific, detailed explanation of spatial explicit parameterization without discussion, in detail, the implementation within VIC, despite the objectives detailing the application of MPR and VIC-refinements. I recommend at least one paragraph specifically justifying the use of VIC as the hydrologic model in this study.

R/ Your suggestion was very helpful in addressing the gaps in our introduction. Since the VIC model forms the foundation of this study, further clarification was indeed valuable. Accordingly, we have added additional explanation after line 70 (line numbers in unrevised manuscript). We also supplemented a paragraph to clarify the reasons for selecting the VIC model and the corresponding implementation approach in this study. The relevant revisions are provided below for your convenience.

Lines 74-79:

Of particular concern, the pathway poses a major challenge in estimating spatially explicit parameters and ensuring their transferability across scales, notably for distributed hydrological models such as the Variable Infiltration Capacity (VIC) model (Sun et al., 2023). Despite the widespread development and application of such models, this challenge remains unresolved, with many modelling studies—even recent ones—still rely on spatially uniform parameters derived from calibration (Yousefi Sohi et al., 2024; Shrestha et al., 2025).

Lines 99-111:

Although MPR has been implemented in a series of distributed hydrological models, the VIC model is adopted as the modelling framework in this study. This choice is motivated by the long-standing challenges in estimating spatially distributed parameters of VIC, which remain an active research topic. Recent studies have explored different strategies, including surrogate-based approaches (e.g., Sun et al. 2023) and MPR-based parameter estimation (e.g., Gou et al. 2021). In this context, the increasing use of spatially distributed parameters in VIC applications mirrors a broader shift towards spatially explicit parameterizations in modern hydrological modelling. In addition, recent developments in VIC, particularly the transition to version 5, enable all model parameters to be handled through a unified NetCDF-based I/O framework, thereby facilitating spatial parameter estimation (Hamman et al., 2018). To support the MPR-based spatial

parameter estimation using distributed information (e.g., soil and vegetation properties), our team has developed an open-source Python-based deployment framework for VIC-5 (https://github.com/XudongZhengSteven/easy_vic_build). The framework integrates a range of transfer functions and offers an object-oriented, modular workflow for parameter estimation, thereby enabling the systematic comparison of spatially explicit and uniform parameterizations within a unified experimental setting and establishing the experimental basis for the present study.

RC1-3/ I think the explanation of multi-gauge in nested catchment systems needs more explanation than the one sentence provided in line 105. I recommend expanding this to a full paragraph.

R/ Following your suggestion, we have expanded the explanation of multi-gauge calibration in nested catchments (originally in line 105) into a new paragraph. The revision is provided below for your convenience.

Lines 126-136:

In practice, streamflow is often used as a reliable and informative calibration target, as it integrates both surface and subsurface hydrological response over the entire basin. The presence of multi-gauge in nested catchment systems, measuring streamflow from different sub-basins, thus provides a suitable setting for multi-objective calibration of hydrological models (Liu et al., 2024). Such strategy offers two advantages. First, neighbouring sub-basins constitute comparable (often paired) units that support parameter estimation and comparative analyses (Vinogradov et al., 2011; Argentin et al., 2025). Second, the implied hydrological connections among sub-basins motivates joint calibration, which provides a broader, system-wide view and allows spatial relationships across the network to constrain shared parameters and improve identifiability. For example, parameter sets that reproduce outlet streamflow equally well can differ substantially at interior gauges, helping to discriminate among competing solutions. Quantifying the extent to which multi-gauge calibration improves model realism is therefore particularly relevant for models with spatially explicit parameters, where the nested structure reflects hierarchical hydrological responses.

RC1-4/ In objective 3 – Its unclear what you mean by cross-benefits between the two approaches. Can you expand on this by specifically defining the two approaches and generally discussing the expected cross-benefits?

R/ Thank you for pointing out this lack of clarity. We have revised the objectives section to explicitly define the two approaches (strategies) and have expanded description of Objective 3, highlighting the **potential sources of the expected cross-benefits**. The corresponding revisions are provided below for your review.

Lines 140-150:

Prompted by pressing issues in contemporary hydrological modelling, we examine two modelling strategies—spatially explicit parameterization and multi-gauge calibration—against a baseline of spatially uniform parameterization and single-gauge calibration. Specifically, this study aims to address three critical questions: (1) How spatially explicit parameterization, compared with spatially uniform parameterization, can enhance the realism of hydrological models while introducing additional equifinality; (2) To what extent multi-gauge calibration strategies in nested catchments can better constrain parameter identifiability; and (3) Whether cross-benefits exist between these two strategies and, if so, in what form. Notably, most previous studies have focus on individual strategies, potentially overlooking the cross-benefits that may emerge from their combination. We stress that relying on a single strategy alone may impose a practical limit

on model improvement; for example, simply increasing model complexity without complementary constraints can lead to under-constrained parameter estimation, a limitation that is expected to be alleviated through the introduction of multi-gauge calibration.

RC1-5/ Study Area: Can you provide a land use analysis on the Upper Han River Basin to visualize the amount of anthropogenic vs natural landscape? Please include an explanation of the presence or absence of impoundments, dams, or reservoirs in the watershed.

R/ We thank you for this valuable comment, which is very helpful for clarifying the characteristics of the study basin. In response, we have collected data from **CNLUCC (China multi-period land use remote sensing monitoring; Xu et al., 2018) and GDW (Global Dam Watch; Lehner et al., 2024) datasets** to conduct a land use analysis and spatial visualization, as shown in Fig. S1.

Our analysis shows that **natural land use dominates** the basin, accounting for approximately 75% of the area. Anthropogenic hydraulic structures are scarce, with only three reservoirs: Shimen and Nanshahe, located in the upper and mid-reaches, have maximum storage capacities of 110 and 43 MCM (million cubic meters), respectively, which are negligible for hydrological modelling. The only potential source of uncertainty is the Shiquan reservoir, with a maximum storage of 412 MCM, situated near the basin outlet. However, it **operates on a daily regulation scheme** and is primarily used for hydropower generation and flood control. As the model aims to simulate daily streamflow, **its impact on the simulated discharge is expected to be minimal**. This is further supported by **the large annual runoff at the station (~10.8 billion cubic meters), which substantially exceeds the reservoir storage capacity**. In light of these considerations, the reservoir is not explicitly represented in the model, and its influence is assumed to be negligible. It is worth noting that, compared with the UHRB, the Han River basin downstream of Shiquan contains many hydraulic structures along the mainstem, such as Xihe, Ankang, Dongwuyan, and Danjiangkou reservoirs, whereas anthropogenic impacts in the UHRB are comparatively minimal.

We have added the corresponding analysis and figure in the revised manuscript, attached below for your review.

Lines 169-182:

To illustrate, land use in the UHRB was analysed using the 2008 data from the China multi-period land use remote sensing monitoring dataset (CNLUCC; Xu et al., 2018) and the Global Dam Watch (GDW; Lehner et al., 2024) dataset, as shown in Fig. S1. The land use pattern in the basin is clearly dominated by forest, grassland, and cropland. Forested areas are mainly distributed along the northern and southern mountain slopes, whereas irrigated paddy fields are concentrated in the low-lying central plains along the Han River. Overall, natural land use types dominate the UHRB (~75%), while anthropogenic components account for only ~25%. Anthropogenic hydraulic structures are generally scarce: Shimen and Nanshahe reservoirs, located in the upper and mid-reaches, have maximum storage capacities of only 110 and 43 MCM (million cubic meters), respectively, which are negligible for hydrological modelling. The Shiquan Reservoir, with a maximum storage of 412 MCM, located upstream of the Shiquan hydrological station, represents the only potential source of uncertainty in the simulations. Nevertheless, it operates on a daily regulation scheme and is primarily used for hydropower generation and flood control. As the model aims to simulate daily streamflow, its operational effects are expected to be minimal. This is further supported by the large annual runoff at the station (~10.8 billion cubic meters), which far exceeds the reservoir storage capacity. In light of these considerations, the reservoir is not explicitly represented in the model, and its influence is assumed to be negligible.

Supplementary:

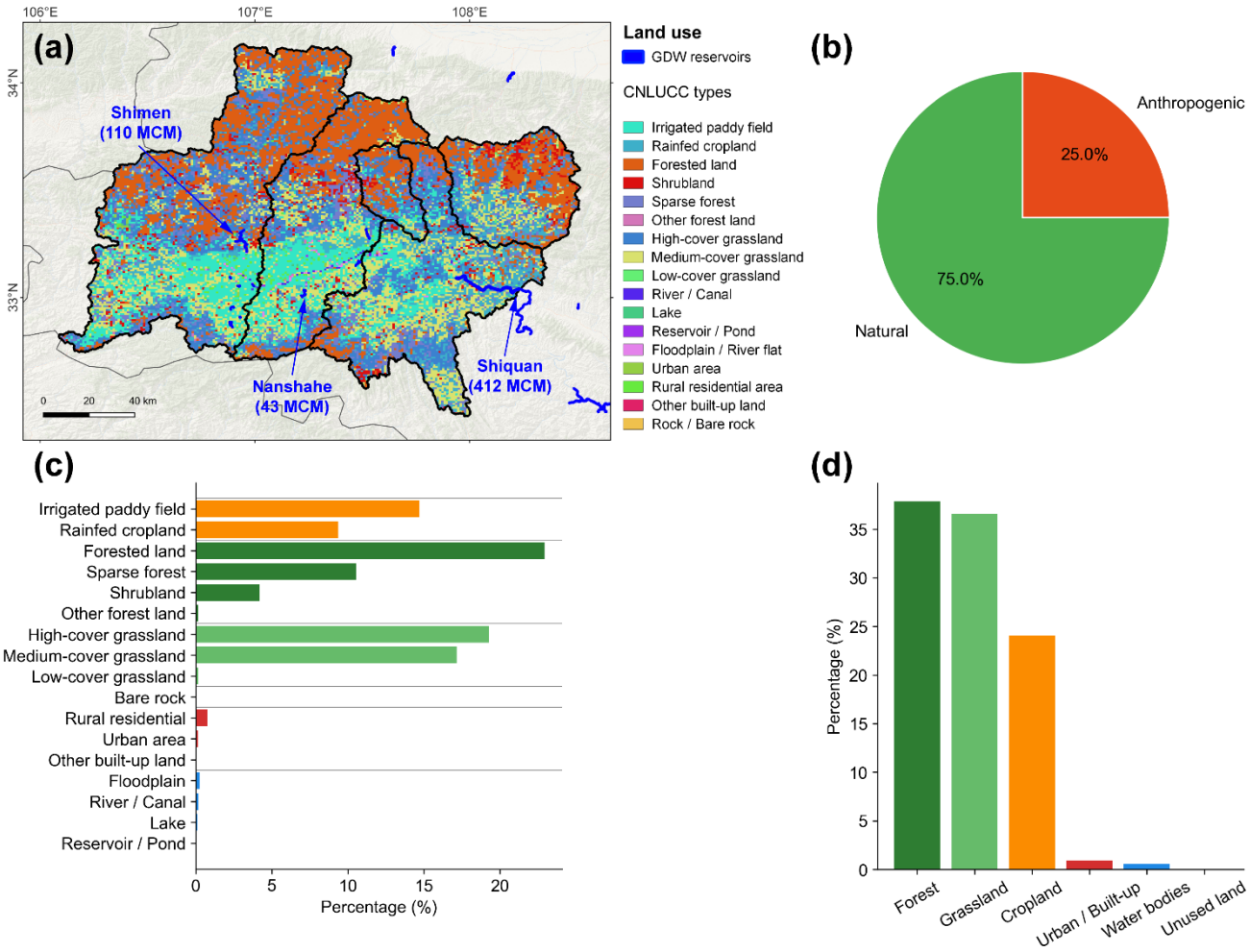

Figure S1. Land use analysis of the Upper Han River Basin based on the CNLUCC (Xu et al., 2018) and GDW (Lehner et al., 2024) datasets. (a) Spatial distribution of land use, with the three major dams (Shimen, Nanshahe, and Shiquan) highlighted in azure and having maximum storage capacities of 110, 43, and 412 MCM (million cubic meters), respectively. (b) Proportions of anthropogenic versus natural land use, where cropland and urban & built-up areas (based on the primary land use classification) are considered anthropogenic. (c) Secondary land use composition. (d) Primary land use composition.

RC1-6/ Methods: Figure 2: Could you add the additional steps of RVIC parameter refinement using MPR technique here?

R/ Thank you for your valuable suggestion. We recognize that the lack of a schematic illustration may reduce the transparency of the RVIC parameter refinement regionalization. To address this, we have included a schematic diagram to better clarify the methodology. To maintain the clarity and independence of the figures, this addition is presented as Fig. S2 in the Appendix rather than being incorporated into Figure 2. The newly added figure and the revised description are provided below for your review.

Lines 330-332:

Nonetheless, we use the MPR technique here to enhance the spatial representation of these two parameters by linking them to topography (Fig. S2).

Supplementary:

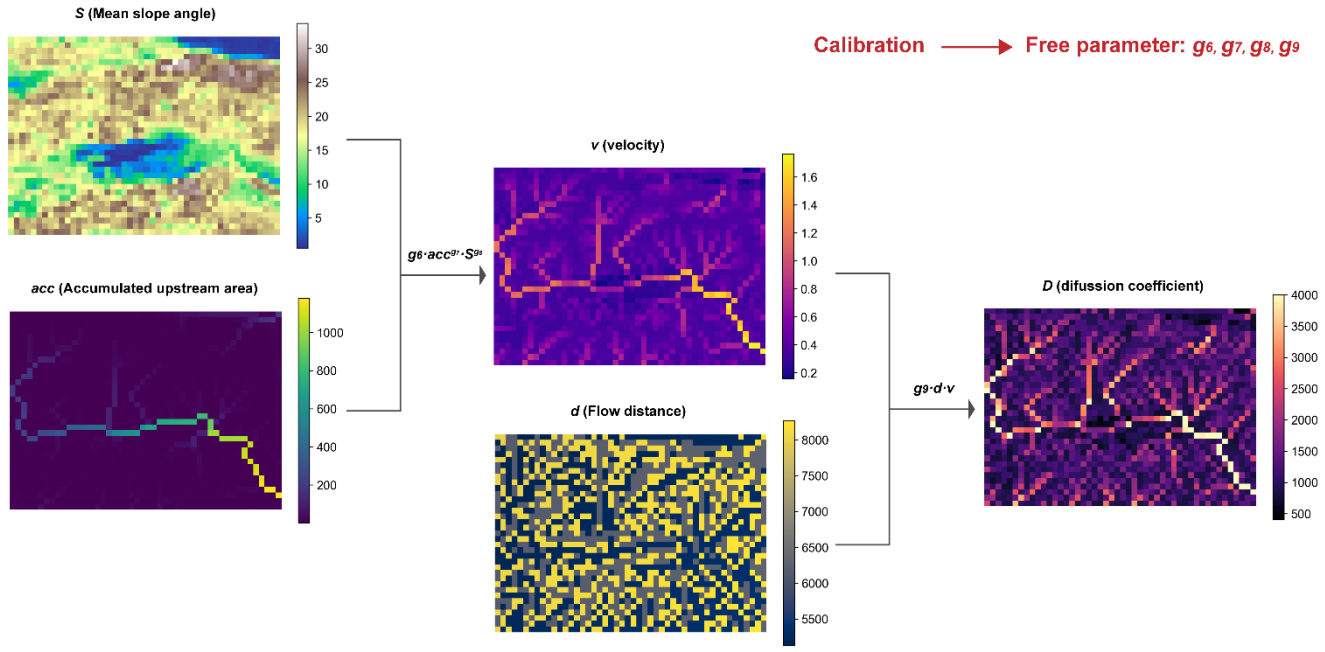

Figure S2. Refinement of the MPR-based regionalization of RVIC parameters. Spatially distributed velocity and diffusion coefficients are estimated using transfer functions (Table 2) driven by topographic attributes and calibratable g-parameters.

RC1-7/ Results: Figure 4: The regression plot between the observed and simulated streamflow makes it hard to visualize the patterns of the majority flow observations below 100m3/s, and appears to rely on a few of the larger flow metrics. I recommend a random thinning to effectively visualize the patterns in the majority of below 1000 m3/s.

R/ This is a very professional suggestion regarding the presentation. Following your advice, **flows below 1000 m3 s-1 were randomly thinned** to improve visual clarity. Furthermore, we have **added a scatter comparison in logarithmic space** in the Supplement, which clearly illustrates the simulated behavior discussed in the manuscript, showing that the distributed parameterization improves high-flow simulation while exacerbating overestimation at low flows. The figure color scheme has also been adjusted to comply with journal requirements. The revised captions and figures are provided below for your review.

Line 505:

This behaviour is also evident in the scatter comparison on a logarithmic scale (Fig. S3).

Lines 517-520:

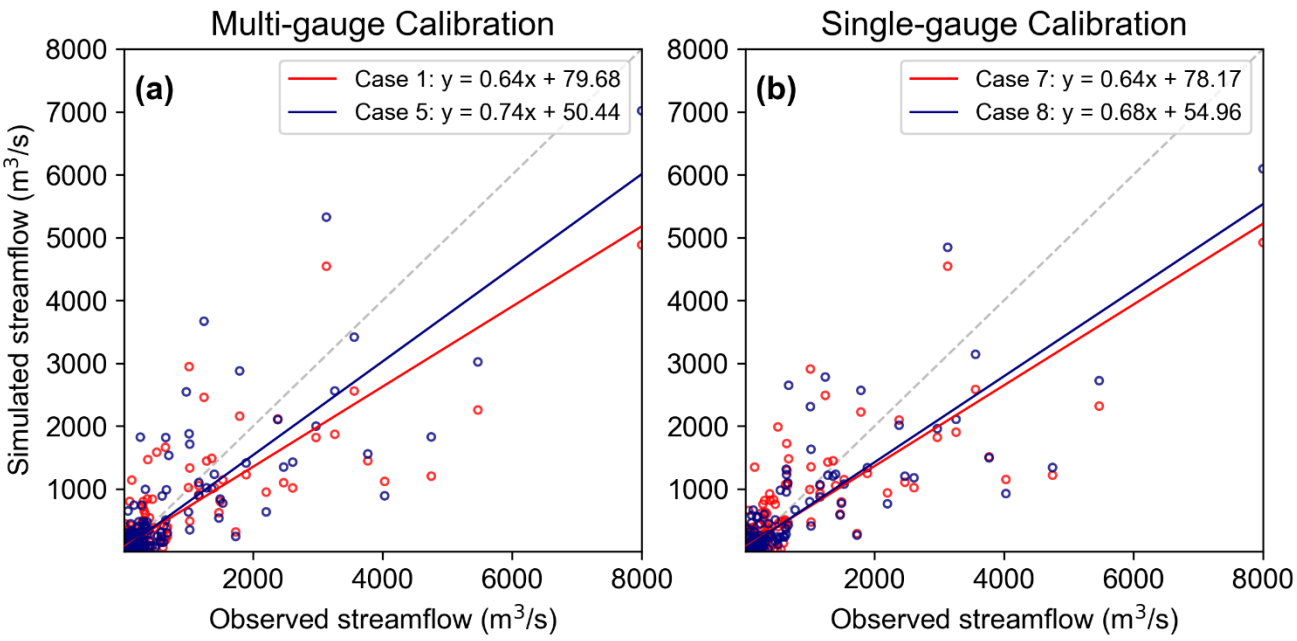

Figure 4. Scatterplots with least-squares regression lines comparing observed and simulated daily streamflow at the Shiquan station during the validation period, under different case configurations. The grey dashed line represents the 1:1 line. Flows below 1000 m^3 s^{-1} were randomly thinned for clarity.

Supplementary:

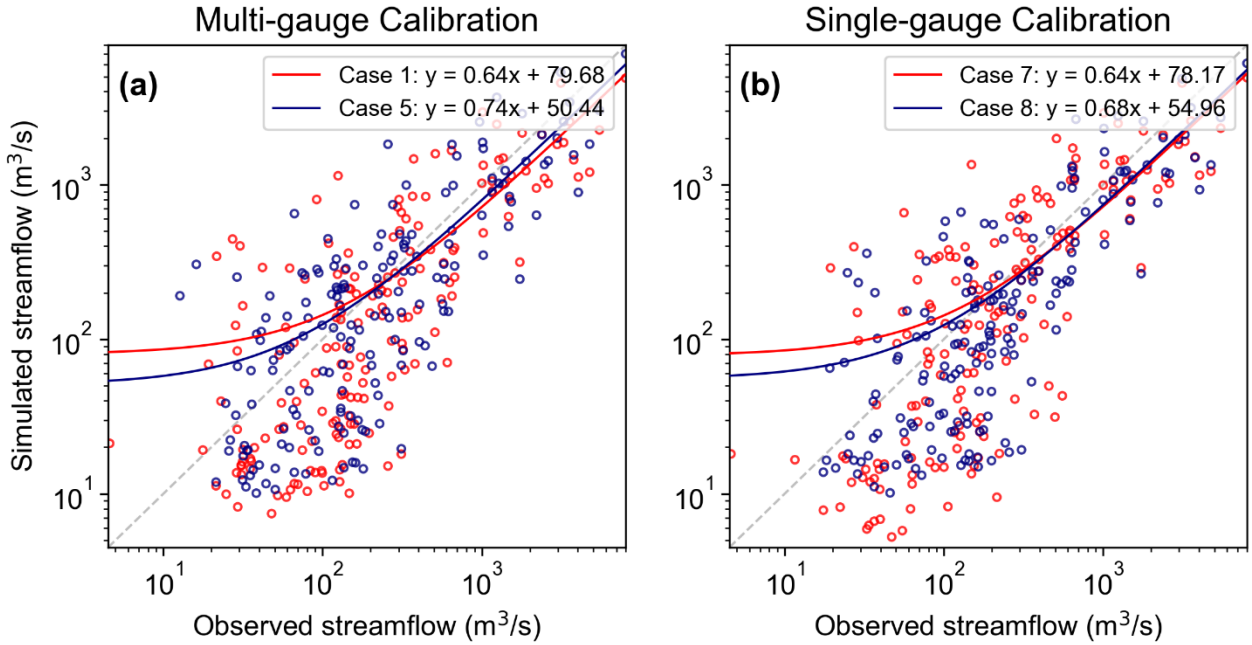

Figure S3. Scatterplots with least-squares regression lines comparing observed and simulated daily streamflow at the Shiquan station during the validation period, under different case configurations. The grey dashed line represents the 1:1 line. Flows below 1000 m^3 s^{-1} were randomly thinned for clarity, and both axes are shown on a logarithmic scale.

RC1-8/ Figure 6: I find the pie chart hard to visualize, and I would recommend converting to nested

barcharts.

R/ Thank you for your suggestion. We have replaced the pie chart with a nested bar chart that provides a clearer comparison. The revised figure and corresponding modifications are provided below for your review.

Lines 534-535:

This allows us to derive the relative contributions of each component, which are summarized as a nested barchart in Fig. 6.

Lines 542-544:

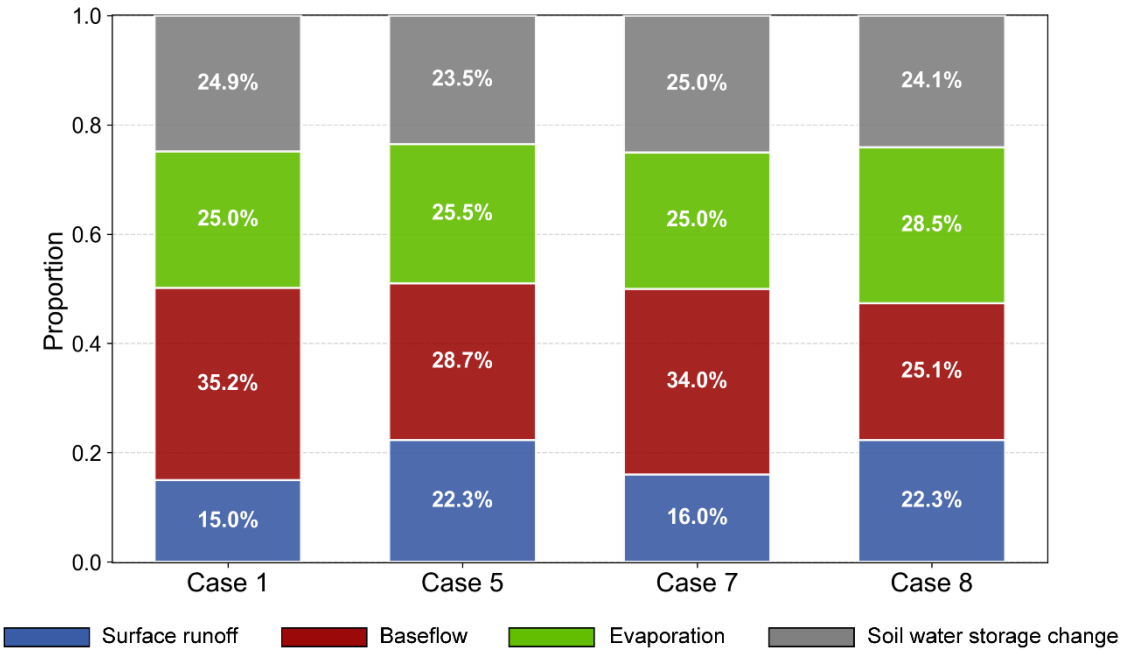

Figure 6. Relative contributions of four major water balance components for different case configurations during the validation period.

RC1-9/ Figure 7: I recommend adding the explanation of Case 1, 5, 7, and 8 in the caption so readers don't have to refer to table 4 in order to interpret the results.

R/ Thank you for your valuable comment. Following your suggestion, we have added additional descriptions of each case below Fig. 7 to improve readability and facilitate interpretation.

The revised Fig. 7 is provided below for your review.

Lines 564-566:

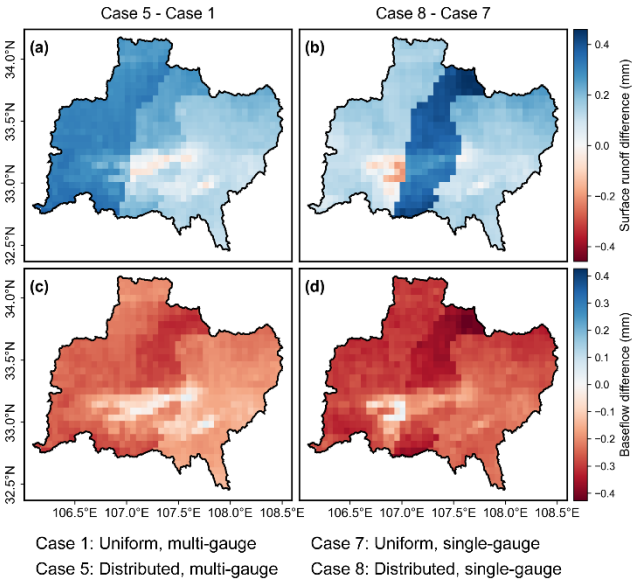

Figure 7. Spatial distribution of differences in (a, b) surface runoff and (c, d) baseflow volumes between the two parameterization schemes over the validation period. Differences are calculated as spatially explicit minus uniform parameterization.

Reference

Blöschl, G. and Sivapalan, M.: Scale issues in hydrological modelling: A review, Hydrol Process, 9, 251-290, 10.1002/hyp.3360090305, 1995.

Freeze, R. A. and Harlan, R. L.: Blueprint for a physically-based, digitally-simulated hydrologic response model, Journal of Hydrology, 9, 237-258, 10.1016/0022-1694(69)90020-1, 1969.

Gao, H., Fenicia, F., and Savenije, H.: HESS Opinions: Are soils overrated in hydrology?, Hydrology and Earth System Sciences, 27, 2607-2620, 10.5194/hess-27-2607-2023, 2023.

Gao, H., Hrachowitz, M., Wang-Erlandsson, L., Fenicia, F., Xi, Q., Xia, J., Shao, W., Sun, G., and Savenije, H.: Root zone in the Earth system, Hydrology and Earth System Sciences, 28, 4477-4499, 10.5194/hess-28-4477-2024, 2024.

Gou, J., Miao, C., Duan, Q., Tang, Q., di, Z., Liao, W., Wu, J., and Zhou, R.: Sensitivity analysis‐based automatic parameter calibration of the variable infiltration capacity (VIC) model for streamflow simulations over China, Water Resources Research, 56, 10.1029/2019WR025968, 2020.

Gou, J., Miao, C., Samaniego, L., Xiao, M., Wu, J., and Guo, X.: CNRD v1.0: A High-Quality Natural Runoff Dataset for Hydrological and Climate Studies in China, Bulletin of the American Meteorological Society, 102, 1-57, 10.1175/BAMS-D-20-0094.1, 2021.

Lehner, B., Beames, P., Mulligan, M., Zarfl, C., De Felice, L., van Soesbergen, A., Thieme, M., Garcia de Leaniz, C., Anand, M., Belletti, B., Brauman, K. A., Januchowski-Hartley, S. R., Lyon, K., Mandle, L., Mazany-Wright, N., Messager, M. L., Pavelsky, T., Pekel, J.-F., Wang, J., Wen, Q., Wishart, M., Xing, T., Yang, X., and Higgins, J.: The Global Dam Watch database of river barrier and reservoir information for large-scale applications, Scientific Data, 11, 1069, 10.1038/s41597-024-03752-9, 2024.

Mizukami, N., Clark, M. P., Newman, A. J., Wood, A. W., Gutmann, E. D., Nijssen, B., Rakovec, O., and Samaniego, L.: Towards seamless large-domain parameter estimation for hydrologic models, Water Resources Research, 53, 8020-8040, 10.1002/2017wr020401, 2017.

Nijssen, B., Lettenmaier, D., Lohmann, D., and Wood, E.: Predicting the Discharge of Global Rivers, Journal of Climate - J CLIMATE, 14, 3307-3323, 10.1175/1520-0442(2001)014<3307:PTDOGR>2.0.CO;2, 2001.

Wen, Z., Liang, X., and Yang, S.: A new multiscale routing framework and its evaluation for land surface modeling applications, Water Resources Research, 48, 8528, 10.1029/2011WR011337, 2012.

Xu, X., Liu, J., Zhang, S., Li, R., Yan, C., and Wu, S.: China multi-period land use remote sensing monitoring dataset (CNLUCC), Resource and Environmental Science Data Registration and Publication System, https://doi.org/10.12078/2018070201, 2018.

Zhang, J., Zhang, Y., Sun, G., Song, C., Dannenberg, M., Li, J., Liu, N., Zhang, K., Zhang, Q., and Hao, L.: Vegetation greening weakened the capacity of water supply to China's South-to-North Water Diversion

Project, Hydrology and Earth System Sciences, 25, 5623-5640, 10.5194/hess-25-5623-2021, 2021.

Finally, we would like to once again thank the Editor and all Reviewers for their thorough review of our paper. If there are any questions, suggestions, or discussions, please feel free to contact us.