

A Robust Calibration and Evaluation Framework for Dynamic Catchment Characteristics in Hydrological Modeling

Tian Lan^{1,2,3,4}, Jiajia Zhang¹, Wenqing Cheng⁵, Xiao Wang^{5*}, Hongbo Zhang¹, Xinghui Gong¹, Xue Xie⁶, Yongqin David Chen^{7,8}, Chong-Yu Xu^{9,10}

- 5 ¹School of Water and Environment, Chang'an University, Xi'an 710054, China.
²State Key Laboratory of Hydrosience and Engineering, Tsinghua University, Beijing, China.
³Key Laboratory of Eco-Hydrology and Water Security in Arid and Semi-Arid Regions of Ministry of Water Resources, Chang'an University, Xi'an 710054, China.
10 ⁴Key Laboratory of Water Security Guarantee in Guangdong-Hong Kong-Macao Greater Bay Area of Ministry of Water Resources, Guangzhou, China.
⁵State Key Laboratory of Water Resources Engineering and Management, Wuhan University, Wuhan, China.
⁶College of Water Conservancy and Civil Engineering, Shandong Agricultural University, Tai'an 271018, China.
⁷School of Humanities and Social Science, The Chinese University of Hong Kong, Shenzhen, China.
⁸Department of Geography and Resource Management, The Chinese University of Hong Kong, Shatin, China.
15 ⁹Yellow River Research Institute, North China University of Water Resources and Electric Power, Zhengzhou 450046, China.
¹⁰Department of Geosciences, University of Oslo, P.O. Box 1047 Blindern, N-0316 Oslo, Norway.

Correspondence to: Xiao Wang (xiao_wang@whu.edu.cn)

Abstract

20 Hydrological models often face challenges in accurately simulating hydrological processes within dynamic catchments due to
simplifications of model structure. In a dynamic catchment where hydrological processes exhibit significant intra-annual or inter-
annual variability, accurately capturing dynamic behaviours across different flow regimes is still challenging for models. To
address these challenges, this study investigates calibration issues in dynamic catchments with a focus on two key aspects: the
influence of objective function design on flow-phase-specific performance, and the limitations of sub-period calibration with
25 dynamic parameters. Seven calibration experiments were designed to explore issues related to time-invariant parameters, objective
function configurations, parameter correlations, dimensionality in global optimization, and abrupt parameter shifts. The
experiments were conducted using the MOPEX dataset, which includes 219 basins across the United States, and were evaluated
based on performance metrics, as well as state variables and fluxes. Among all calibration schemes, sub-period calibration with
dynamic parameters exhibited the most reliable performance. Static parameter approaches often averaged catchment responses and
30 poorly represented extreme flows, whereas enabling temporal variability to only a subset of parameters yielded limited
improvement. In contrast, multi-parameter dynamic schemes significantly improved NSE and LNSE values and enhanced
parameter transferability across flow phases, where the high-dimensional calibration strategy balanced dynamic adaptability with
physical consistency, while the parallel calibration maintained accuracy through gradual parameter transitions despite higher
variability in some catchments. This study demonstrates that sub-period calibration with dynamic catchment characteristics
35 outperforms traditional static parameters by effectively capturing flow-regime variability and sustaining robust performance under
changing catchment conditions, offering a generalizable solution for simulating hydrological processes in dynamic catchments.

1 Introduction

Hydrological models serve as essential tools in water management, supporting tasks such as runoff projection, disaster warning,
and water-resource planning (Shao et al., 2023; Shrestha et al., 2021; Razavi et al., 2025). These models conceptualize hydrological
40 processes with physically based parameters and state variables, enabling transparent simulations and process-informed diagnostic
analysis of the catchment. However, limited understanding of the mechanisms underlying seasonal climate patterns, vegetation
dynamics, and water storage variability has led existing model structures to rely on simplified representations of hydrological
processes and steady-state assumptions (i.e., time-invariant parameters). Such assumptions only partially capture the dynamic
catchment characteristics (Deng et al., 2016; Wang et al., 2022b; Wen et al., 2021). A dynamic catchment is defined as one in
45 which hydrological processes exhibit significant intra-annual or inter-annual variability, making their simulation particularly
challenging. Dynamic catchment characteristics denote the time-varying states of a catchment that describe the temporal evolution
of hydrological processes, such as precipitation seasonality and changes in vegetation cover under significant human disturbances.
As a result, models tend to capture only the “average” behaviour of catchments, often at the cost of reduced accuracy in high- or
low-flow phases (Longyang and Zeng, 2023; Yoshida et al., 2022). Understanding, modelling, and predicting dynamic
50 hydrological processes with greater realism remain significant challenges in hydrological sciences (Bouaziz et al., 2022).

A key challenge in modelling catchments with dynamic variability lies in how to adapt the model to accurately reflect time-varying
hydrological responses. Calibration aims to adjust model parameters using local observations, thereby tailoring a general model
structure to the hydrological responses of a specific catchment. This process typically involves the definition of objective functions
and the systematic exploration of parameter space. The mathematical form of the objective function determines which aspects of
55 model performance are emphasized, such as the accuracy of peak flows or the representation of overall water balance (Gupta et
al., 2009; Fauer et al., 2021). In catchments with strong dynamics, however, the calibrated parameter sets may reflect not only the

actual catchment behaviour but also implicit structural limitations and assumptions about boundary conditions. Consequently, the calibrated parameters often reflect trade-offs shaped by the objective function and model structure, leading to an averaged performance across flow phases (such as extreme high flow, high flow, middle flow, low flow, and extreme low flow).

60 One common strategy to improve model performance under structural limitations is to refine the configuration of the objective function to better emphasize key hydrological processes. Such a challenge is not unique to dynamic catchments, but represents a general issue in hydrological model calibration. Traditional calibration of hydrological models typically employs global evaluation metrics and time-constant parameters, focusing on the model's overall performance. However, this approach might average hydrological responses and fail to ensure accurate simulations across various flow phases and observational periods. In critical runoff events like floods and droughts, this static approach may fail to capture the dynamic catchment characteristics of hydrological processes, underscoring the need for more flexible calibration methods (Martel et al., 2025; Clark et al., 2021). In dynamic catchments, such limitations become more evident as hydrological responses vary across seasons or years. To alleviate these limitations, various calibration strategies have been proposed to incorporate dynamic catchment characteristics. One method involves revising the objective function based on selected evaluation criteria to improve model performance (Araya et al., 2023; Ji et al., 2023). Calibrations using multi-objective optimization algorithms better highlight different flow phases, but face potential challenges such as increased computational complexity, sensitivity to parameter settings, and slower convergence with more objective functions (Song et al., 2023; Carletti et al., 2022). Alternative approaches, like multi-weighted objective functions, can improve the simulation accuracy of specific time and flow phases. While these methods enhanced different flow phases and water balance, they may not effectively address structural deficiencies and cannot fundamentally enhance the model's overall performance (Lin et al., 2024; Anderson and Radić, 2022).

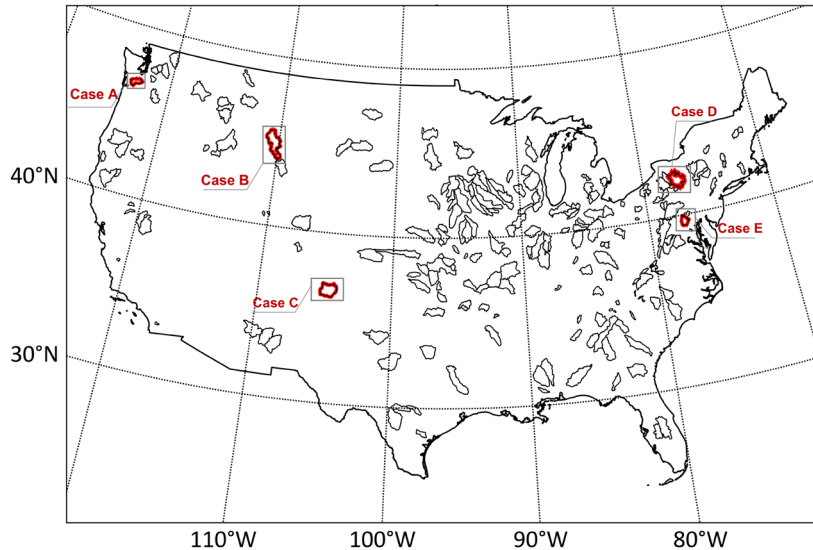
Another strategy involves the use of dynamic parameters in hydrological models. A dynamic parameter is defined as a model parameter that varies across sub-periods rather than remaining fixed over the entire simulation period. Sub-periods are segments of the simulation period characterized by relatively homogeneous hydrological conditions, which are typically identified through clustering of the time series. The implementation of dynamic parameters addresses structural limitations of models and improves predictive performance across the full range of hydrological processes, rather than being restricted to specific flow regimes or periods (Zhang and Liu, 2021; Krapu and Borsuk, 2022). Recent studies have significantly advanced hydrological simulations by integrating the dynamic catchment characteristics. Clustering based on catchment characteristics, such as precipitation, evapotranspiration, and soil moisture, facilitates the clustering of dynamic hydrological processes into distinct sub-periods (Acuña Espinoza et al., 2024; Lakshmi and Sudheer, 2021). Wei et al. (2021) further broadened this perspective by highlighting the hydrological processes that arise from the interplay of various factors, including meteorological conditions, surface characteristics, and anthropogenic interference. This interaction among water balance components, such as soil, vegetation, and topography, exhibits temporal variability, which ideally should be captured by process-driven hydrologic simulation models. These changes need to be taken into account through model parameters (Wi and Steinschneider, 2022; Reichert et al., 2021). Zhang and Liu (2021) suggested that temporal variations in parameters reflect the evolving environment. However, some fundamental problems still need to be addressed before applying the dynamic parameters. Sub-period calibration with dynamic parameters involves the hydrological model structure, global optimization, physical mechanisms of dynamic catchment characteristics, as well as complex relationships between the parameters, state variables, and fluxes.

To address the model deficiencies and improve simulation across all flow regimes, it is imperative to re-examine the time-varying information in historical hydrological and meteorological data, extract dynamic catchment characteristics, and address the variation

95 in calibration. This study investigates calibration challenges in dynamic catchments and proposes a structured framework to address
two major issues: the influence of objective function design on flow-phase-specific performance, and the limitations of sub-period
calibration with dynamic parameters. Seven experiments are developed to systematically evaluate these aspects. Experiments 1–3
focus on the effects of time-invariant parameters and various objective function configurations. Experiments 4–7 explore issues in
dynamic parameter calibration, such as parameter correlation, dimensionality, and state transitions. Model performance is assessed
100 through multiple metrics and internal diagnostics across 219 MOPEX catchments.

2 Study area

The Model Parameter Estimation Experiment (MOPEX) is an international project aimed at developing enhanced techniques for
a priori estimation of parameters in hydrologic models and land surface parameterization schemes of weather and climate models
(Duan et al., 2006). A comprehensive MOPEX database has been developed that contains historical hydrometeorological data and
105 land-surface characteristics data for numerous hydrological catchments in the United States (US) and other countries. This study
utilizes the dataset from 219 catchments spatially distributed across the contiguous US (Fig. 1a). Rigorous screening criteria were
applied to ensure the acquisition of high-quality data. The screening process involved three key considerations: (1) no missing or
non-physical data throughout the study period; (2) minimal interference from anthropogenic influences in both temporal and spatial
dimensions; and (3) a large spatial distribution scale of the selected catchments, including diverse meteorological and underlying
110 surface conditions. The dataset for selected catchments includes the hydrometeorological forcing data, land-surface data, and
streamflow data, covering the period from 1983 to 2000. Hydrometeorological data includes daily precipitation data (P),
temperature data (T), and streamflow (Q) provided by the MOPEX dataset, as well as potential evaporation data (PE) calculated
by the Hamon model (McCabe et al., 2015). The Normalized Difference Vegetation Index (NDVI) was used as one of the land-
surface indicators to represent the vegetation coverage of the catchments, which had a spatial resolution of 8 km and a temporal
115 resolution of half-monthly intervals (Tucker et al., 2010). Based on these criteria, a total of 219 catchments were selected (Fig. 1a),
spanning a wide range of hydrological and meteorological characteristics, making them ideal for testing various model structures
under diverse conditions (Duan et al., 2006).



120 **Figure 1.** Location map of the catchment area used in this study, where cases A, B, C, D, and E correspond to catchments 12027500, 6192500,
7211500, 1643000, and 1531000 (from west to east) are highlighted with red outlines for reference.

In addition to the large-sample analysis of the MOPEX dataset, five representative catchments, Case A (12027500), Case B (6192500), Case C (7211500), Case D (1643000), Case E (1531000), are analyzed in more detail as case studies. These catchments encompass a variety of Köppen climate classifications and different dominant dynamic catchment characteristics, facilitating comparison of calibration strategies and evaluation of their robustness under diverse hydroclimatic conditions. Their locations and characteristics are listed in Table 1 and will be analyzed in depth in the subsequent sections.

Table 1. Summary of catchment characteristics for study cases.

ID	12027500	6192500	7211500	1643000	1531000
Location	122.99°W	110.40°W	104.76°W	77.25°W	77.24°W
Area (km ²)	895	3551	2850	817	2056
Climate	Csb	Dfc	Bsk	Cfa	Dfb
Mean <i>P</i> (mm)	1548.78	735.71	491.70	1068.49	870.53
Mean <i>PE</i> (mm)	596.53	731.59	1279.88	897.63	711.06
Mean <i>Q</i> (mm)	1110.19	369.79	10.08	430.15	366.76
Mean elevation (m)	253.06	2441.28	2262.91	191.80	492.25
Mean slope (°)	12.16	15.26	9.44	4.99	8.25
Runoff ratio	0.72	0.50	0.02	0.40	0.42
Aridity index	2.60	1.01	0.38	1.19	1.23
Forest cover (%)	71.96	36.95	16.76	31.31	57.36
Land use	Evergreen Forest, Pasture/Hay	Evergreen Forest, Shrub/Scrub	Evergreen Forest, Grassland/Herbaceous	Deciduous Forest, Cultivated Crops	Deciduous Forest, Pasture/Hay

3 Methods

Hydrological processes within catchments commonly exhibit significant annual and inter-annual variability. However, conventional hydrological models often fail to capture these temporal dynamics due to structural simplifications, resulting in averaged responses and reduced simulation accuracy. To address these limitations and improve model performance across different flow phases, this study investigates two strategies: (1) refining the configuration of objective functions during calibration to enhance sensitivity to temporal variations; and (2) integrating dynamic catchment characteristics into the modelling framework through dynamic parameterization, while systematically investigating the associated calibration challenges. This study aims to evaluate the effectiveness and limitations of various calibration strategies under dynamic catchment conditions and to develop a robust calibration framework for catchments with temporal dynamics.

3.1 Hydrological models

The investigated strategies involve only parameter configuration and calibration procedures, without requiring structural modifications to the hydrological model. Such strategies are applicable to lumped, semi-distributed, and fully distributed models, including both conceptual and physically based types. To evaluate the applicability of different calibration strategies under dynamic catchment conditions, the simple conceptual hydrological model HYMOD (Hydrological MODel) (Moore, 1985) is employed. HYMOD is characterized by a parsimonious structure with five parameters, low input requirements, and empirically interpretable physical meaning. It has been widely applied in streamflow prediction across America and other regions (Vrugt et al., 2003; Wagener et al., 2001). To improve performance in snow-affected regions, a degree-day model is incorporated to represent snowmelt processes (Supporting Information S1.6) (Wang et al., 2022a). HYMOD serves as a representative model for conceptual validation and mechanism analysis, rather than constraining the applicability of the proposed framework. To further assess cross-model generalizability, additional comparisons between conventional calibration and dynamic calibration are conducted using the

semi-distributed TOPMODEL and the GR4J model, as presented in the Supporting Information S4. The structure and configuration of the HYMOD model are presented in Supporting Information S1.7, with detailed information on model parameters, state variables, and fluxes provided in Table 2.

150 **Table 2.** HYMOD model parameters, state variables, and fluxes (Vrugt et al., 2003; Wagener et al., 2001).

Label	Property	Range	Description
H_{uz}	Parameter	10–1500 mm	Maximum height of the soil moisture accounting tank
B	Parameter	0–1.99	Scaled distribution function shape
α	Parameter	0–0.99	Quick or slow split
K_q	Parameter	0.5–0.99	Quick-flow routing tanks' rate
K_s	Parameter	0–0.5	Slow-flow routing tank's rate
XH_{uz}	State variable	mm	Upper-zone soil moisture tank state height
XC_{uz}	State variable	mm	Upper-zone soil moisture tank state contents
X_q	State variable	mm	Quick-flow tank state contents
X_s	State variable	mm	Slow-flow tank state contents
AE	Flux	mm d ⁻¹	Actual evapotranspiration flux
OV	Flux	mm d ⁻¹	Excess rainfall flux
Q_q	Flux	mm d ⁻¹	Quick-flow flux
Q_s	Flux	mm d ⁻¹	Slow-flow flux
Q_{sim}	Flux	mm d ⁻¹	Total simulated streamflow flux

3.2 Clustering hydrological processes

Sub-period calibration provides a practical means of linking dynamic catchment characteristics with hydrological models. In sub-period calibration, the simulation period is clustered into multiple sub-periods characterized by relatively homogeneous hydrological conditions, allowing dynamic parameters to better reflect temporal variations in catchment behaviour across different streamflow regimes (Zhang and Liu, 2021). In this study, the clustering of sub-periods is guided by temporal variations in key hydrometeorological and land-surface variables. The methodological framework consists of three key steps: (1) constructing a dynamic catchment characteristic index system to describe catchment states; (2) extracting dynamic catchment characteristics through screening and dimensionality reduction; and (3) applying unsupervised clustering to cluster the time series into sub-periods with similar hydrological processes for subsequent sub-period calibration.

160 **Describing dynamic catchment characteristics:** To characterize the temporal dynamics of catchment behaviour, a dynamic catchment characteristic index system comprising a climatic subsystem and a land-surface subsystem is constructed to represent the time-varying states of the catchment. The climatic subsystem includes core hydrometeorological variables such as precipitation (P), temperature (T), and potential evapotranspiration (PE), along with corresponding extreme climatic indicators. The land-surface subsystem reflects evolving surface conditions through indicators such as antecedent runoff, runoff coefficient, and the normalized difference vegetation index (NDVI). All indicators are sampled using a moving window approach, with the optimal window length determined through a time-windowed Bayesian inference framework based on predictive log-score (PLS) performance (Hsueh et al., 2024). The framework is designed to preserve long-term trend signals, suppress short-term high-frequency noise, and improve the stability and robustness of dynamic catchment characteristic extraction.

170 **Extracting dynamic catchment characteristics:** Not all indicators exhibit significant dynamic catchment variability; therefore, filtering irrelevant or redundant variables is essential to retain meaningful catchment dynamics. A threshold-based screening is applied to identify variables exhibiting significant seasonality, retaining only relevant subsystems and forming an initial pool of candidate indicators (see Supporting Information S2.1 for detailed criteria). The Maximal Information Coefficient (MIC) is then employed to quantify linear and nonlinear associations between candidate indicators and streamflow, ensuring hydrological

relevance. To mitigate multicollinearity and reduce dimensionality, Principal Component Analysis (PCA) is performed, with the first two principal components retained for clustering. This multi-step filtering and reduction procedure ensures robust extraction of dynamic catchment characteristics and provides a solid basis for sub-period clustering according to hydrological similarity.

Clustering hydrological processes: Based on the extracted dynamic catchment characteristics, the time series is clustered into distinct sub-periods using the unsupervised Fuzzy C-Means (FCM) clustering algorithm. The optimal number of clusters is determined through a combination of clustering validity indicators, including the Partition Coefficient (SC), Separation Index (S), and Xie–Beni (XB) index, which collectively assess clustering compactness and separation. In addition, the elbow method is employed as a supplementary diagnostic to identify the inflection point beyond which further increases in cluster number yield diminishing returns. Clustering is performed in the principal component space, enabling effective capture of structural patterns in catchment dynamics. The resulting sub-periods provide a robust foundation for integrating dynamic parameters into hydrological models.

In addition, the sub-period clustering is developed exclusively using data from the calibration period. To independently evaluate the generalization capability and robustness of the model under unseen conditions, no model training or parameter adjustment is performed during the evaluation period.

3.3 Calibration experiments

To systematically evaluate how calibration strategies capture catchment dynamics and improve the simulation of diverse flow regimes, a diagnostic framework comprising seven calibration strategies is developed. These experiments sequentially address key challenges in representing time-varying hydrological behaviour, with a focus on objective function design and time-varying parameterization (Fig. 2).

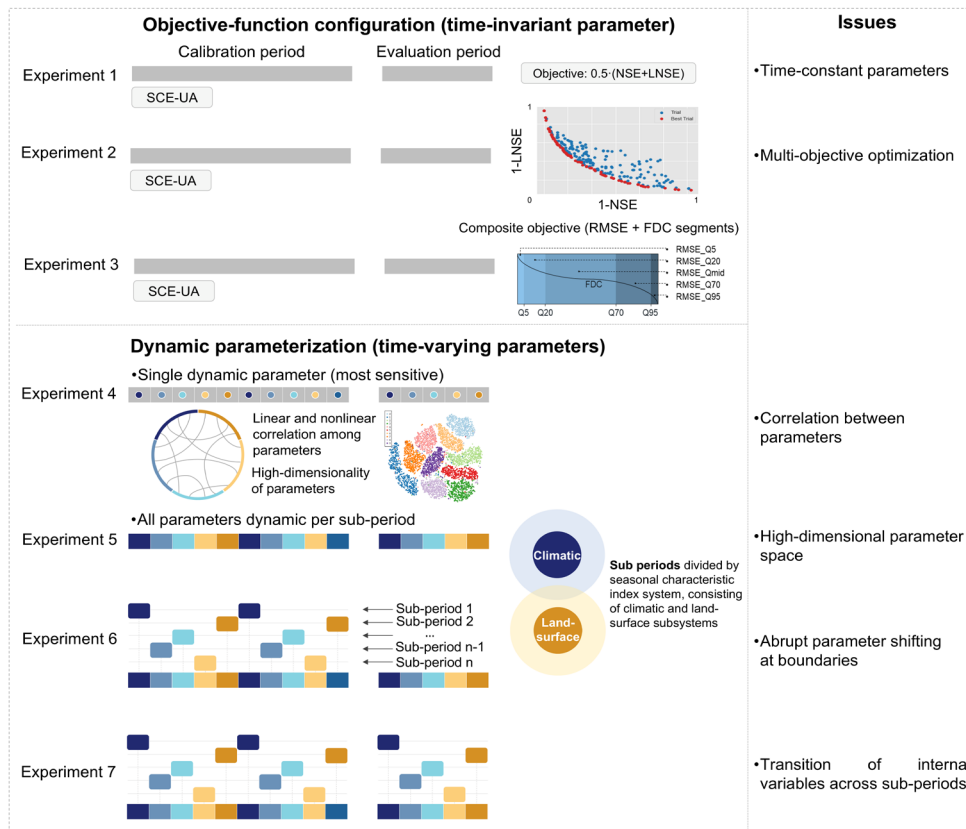
Experiments 1–3 use time-invariant parameters and focus on the design and weighting of objective functions. Experiment 1 establishes a baseline with standard global calibration. Experiment 2 applies a multi-objective approach to explore trade-offs between high and low flows. Experiment 3 designs a composite objective function to enhance simulation performance across a range of flow conditions. Experiments 4–7 incorporate time-varying parameters to better represent temporal catchment variability and examine related calibration challenges. Experiment 4 allows only the most sensitive parameter to vary, assessing partial dynamization and parameter compensation. Experiment 5 makes all parameters dynamic, raising issues of parameter dimensionality. Experiment 6 investigates the effects of abrupt parameter shifts on model continuity. Experiment 7 introduces smooth parameter transitions to reduce instability while preserving responsiveness to catchment dynamics.

Throughout the experiments, the Shuffled Complex Evolution algorithm (SCE-UA) is employed to search for the globally optimal parameter set (Duan et al., 1993). The HYMOD model is configured for catchments over 19 years from 1982 to 2000, with 1982 as the warm-up year, 1983–1995 for calibration, and 1996–2000 for evaluation. All other model parameters are held at their default values. Unless specified otherwise, model calibration is guided by the following objective function:

$$OF = 0.5 \cdot NSE + 0.5 \cdot LNSE \quad (1)$$

where the Nash–Sutcliffe Efficiency (NSE) evaluates overall model performance with greater sensitivity to high flows, and log-transformed Nash–Sutcliffe Efficiency (LNSE) emphasizes model performance for low flows. **Experiment 1** uses time-invariant parameters calibrated over the entire period without sub-period clustering. It serves as a baseline for assessing standard global calibration. **Experiment 2** approximates a multi-objective calibration by combining NSE and LNSE into a weighted objective: $w \times NSE + (1-w) \times LNSE$. The weight w varies from 0 to 1 (step = 0.05), forming a series of single-objective optimizations using

210 SCE-UA with time-invariant parameters. This setup explores trade-offs between flow regimes without changing the optimization algorithm. **Experiment 3** adopts a composite objective function to improve simulation across flow regimes. It integrates RMSE with flow duration curve (FDC)-based metrics (RMSE_Q95, Q70, Qmid, Q20, Q5, as listed in Table S1), representing different flow phases. Weights are derived from Experiment 1 using AHP, PP, and CRITIC methods (refer to Supporting Information S1.8). **Experiment 4** introduces time-varying parameters by allowing only the most sensitive parameter to vary across sub-periods, while all others remain fixed. State variables and fluxes are passed between sub-periods through an inheritance approach. **Experiment 5** extends the dynamic calibration to all parameters, with distinct values assigned to each sub-period. As a result, the number of parameters increases in proportion to the number of sub-periods, generating a high-dimensional calibration space. State and flux continuity between sub-periods is maintained through the same inheritance mechanism used in Experiment 4. **Experiment 6** investigates the impact of abrupt parameter transitions across sub-periods. Parameters are optimized independently for each sub-period. During model runs, parameter sets switch discretely between sub-periods, while state variables and fluxes are inherited to maintain continuity. **Experiment 7** adopts the same calibration structure as Experiment 6 but incorporates smooth parameter transitions during evaluation. The parallel calibration strategy is designed to preserve continuity in parameter evolution while maintaining water balance within each sub-period.



225 **Figure 2.** Schematic illustration of the seven calibration experiments. The colour bands represent state variables and fluxes, which are continuously transferred within the same period. In Experiments 1, 2, and 3, the parameters are time-invariant, but the experiments differ in their objective function configurations. Conversely, experiments 4, 5, and 6 maintain a consistent objective function, but vary the parameters across different experiments. In Experiment 4, the dynamics of only the specific parameter are operated, and the other fixed parameters are optimized simultaneously. In Experiment 5, the parameter set is dynamized. The parameter sets in different sub-periods are optimized simultaneously. In Experiment 6, the data from the individual sub-periods are used for minimizing the objective function, while the model is run for the whole period. In the evaluation period, the parameter set between two consecutive sub-periods is updated accordingly. In Experiment 7, the calibration is the same as in Experiment 6. In the evaluation period, the simulated flow data from each separate sub-period are combined and compared with the observed flow.

3.4 Model evaluation

3.4.1 Multi-criteria evaluation

Model simulations are typically evaluated using performance metrics, which can be divided into statistical and signature metrics (Pfannerstill et al., 2014; Yilmaz et al., 2008; Clark et al., 2021). However, a limitation exists with common performance metrics: They only focus on overall or specific segments of the streamflow series, neglecting other parts that may have the greatest practical impact. Hence, for diagnostic analysis, streamflow segments of the flow duration curve (FDC) are used to identify flow phases where model performance is poor (Pfannerstill et al., 2014; Schwemmler et al., 2021). Performance across multiple streamflow segments is assessed through the criteria defined in Table S1, providing a comprehensive evaluation of model performance.

3.4.2 State variables and fluxes

The evaluation of state variables and fluxes links sub-period calibration and dynamic parameterization to internal model continuity and responsiveness, helping to diagnose performance differences across experiments. The internal behaviour of the hydrological model, involving the time series of state variables and fluxes that constitute subspaces within the model state space, is visualized in graphs and categorized by the operation of different sub-periods. Such visualization facilitates the identification of issues in calibration experiments. For instance, unreasonable values exceeding operational boundaries often signal errors in model operation triggered by abrupt parameter shifts. Similarly, unresponsive values may indicate either operational errors or unique catchment characteristics. Furthermore, a flux map is developed and applied to evaluate the equifinality or uncertainty of internal model behaviour by plotting different components of model fluxes (Khatami et al., 2019). The flux map is a ternary or binary plot where each dimension represents a model runoff flux, and each model run is projected as a single point based on the proportions of its equifinal runoff fluxes to the total simulated Q . To HYMOD model, the components with Q_{q1} , Q_{q2} , and Q_s are defined, which represents the runoff component of the output of quick-release reservoirs of linear routing component (OV_1), the output of quick-release reservoirs of nonlinear routing component (OV_2) and the output of slow-release reservoir (Q_s). The point cloud pattern from ternary or binary plots can vary from very constrained to filling the entire feasible flux space, which represents the different dominant components of runoff. Thus, the point cloud on the flux maps is an expression of the model uncertainty; filling a larger space on the flux map indicates higher degrees of model uncertainty.

4 Results

4.1 Clustered sub-periods based on catchment dynamics

To support sub-period calibration, periods are identified for all 219 catchments based on variations in dynamic catchment characteristics. The results show that dynamic patterns are widespread across the study area, with all catchments exhibiting significant variation in at least one hydrometeorological variable, including precipitation, temperature, potential evapotranspiration, NDVI, or runoff. Spatially, precipitation seasonality is more significant in the central and western regions, potential evapotranspiration seasonality is widespread, particularly in the north, runoff seasonality is strongest in the central and northeastern regions, and vegetation seasonality is also common, except in a few high-latitude catchments.

A data-driven method is applied to extract relevant information and partition the time series into distinct periods. The optimal sampling window identified by Bayesian inference ranges from 5 to 150 days (mean = 59.45 days). After MIC-based screening and PCA-based dimensionality reduction, the first two principal components explain 83.5% of the variance on average. FCM clustering in the reduced feature space identifies an average of 4.2 periods per catchment. To demonstrate applicability under diverse hydroclimatic conditions, five representative catchments are selected for subsequent modelling experiments. As shown in

Fig. 3b, the optimal window lengths range from 30 to 150 days, with 12–31 indicators retained and 3–5 periods identified across the five cases. When compared with hydrographs, the identified periods aligned well with key hydrological processes, such as rising and recession limbs (Fig. 3c). In catchments with strong dynamic signals (e.g., Case A and Case B), the identified periods showed stable interannual patterns, while in catchments with greater variability (e.g., Case D and Case E), the clustering still captured major dynamic catchment characteristics.

275

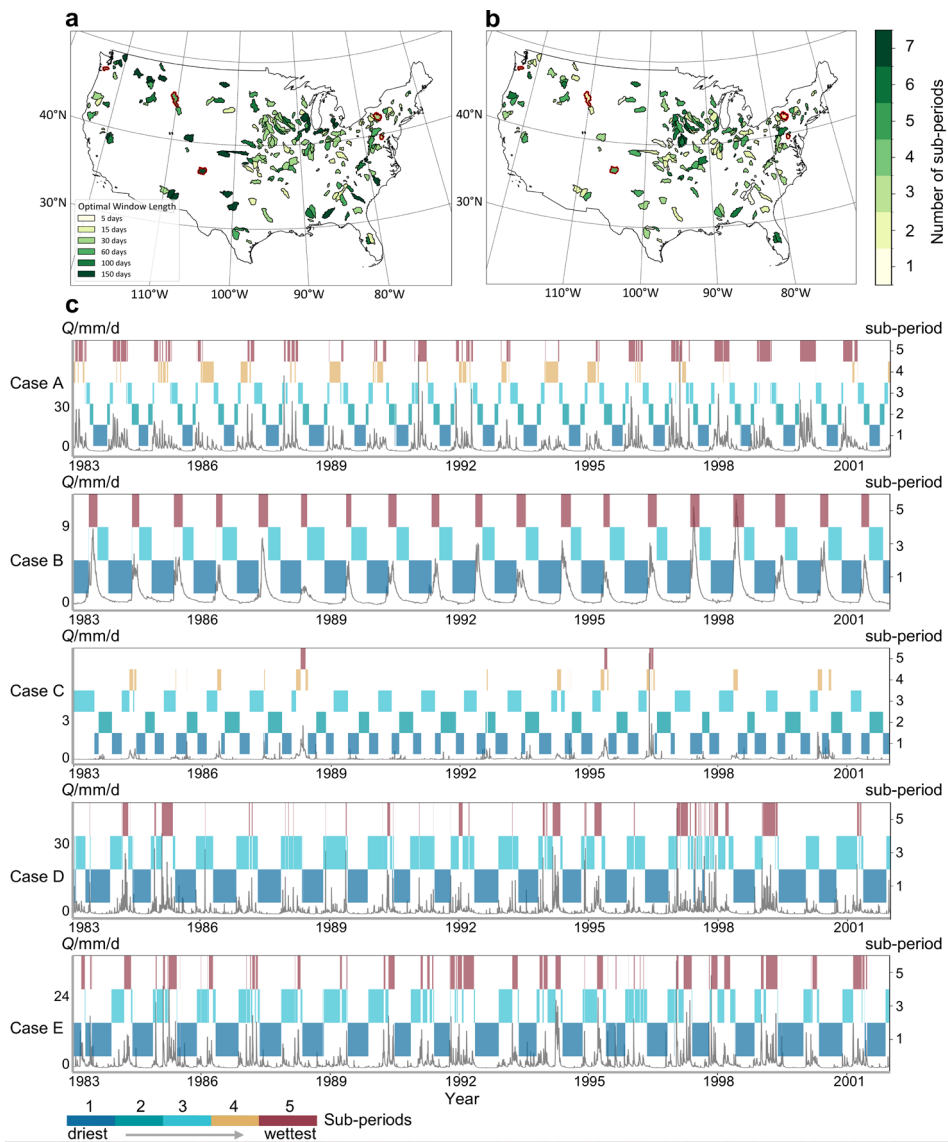


Figure 3. **a**, Optimal window lengths of catchment area used in this study for the sub-period clustering. **b**, Number of subperiods reflecting results from Section 3.2. **c**, Visualization of clustering results on the hydrograph for the study cases, where colors indicate reordered sub-periods ranked from driest to wettest according to mean streamflow, with a consistent color mapping across the entire time series.

280

Analysis of the statistical properties of each sub-period (Supporting Information S2.3) indicates that the identified sub-periods are closely associated with underlying physical and climatic drivers. Across the representative cases, the sub-periods consistently correspond to distinct hydrological states shaped by variations in evaporative demand, antecedent storage, runoff response, and the relative contributions of quick flow and baseflow, rather than arbitrary temporal clustering. These physically interpretable sub-periods provide a robust basis for subsequent dynamic parameterization and modelling experiments. In addition, two diagnostic

285

comparative analyses are presented in the Supporting Information S5 to further evaluate whether the observed performance improvements arise from improved representation of catchment dynamics rather than increased parameter dimensionality.

Considering the performance of the seven modelling experiments across both calibration and evaluation periods, Experiments 5 and 7 are considered the recommended experiments for capturing dynamic catchment characteristics (Fig. 4). Experiment 5, with multi-parameter dynamic calibration, achieves high predictive accuracy across diverse flow regimes, although it may slightly compromise physical consistency in runoff generation. Experiment 7, incorporating smooth parameter transitions, maintains comparable accuracy while promoting more consistent and physically reasonable runoff strategies across sub-periods, thus offering a balanced approach between model performance and hydrological interpretability. Detailed analysis of the results will be presented in the following sections.

4.2 Model performance

To compare seven experiments in dynamic catchments and to identify potential limitations in model calibration, the evaluation is conducted across 219 catchments characterized by hydrological variability. As shown in Fig. 4, the NSE and LNSE values during both calibration and evaluation periods reveal differences in the ability of diverse calibration schemes to capture high- and low-flow conditions. The median NSE reached only 0.4–0.5 in Experiments 1 and 2, and although the LNSE approached 0.7, negative values are frequently observed. It is suggested that global optimization or simple weighted objective functions often lead to an averaging of catchment responses, thereby limiting accuracy for both high- and low-flow conditions. Experiment 3 employed an objective function defined as: $OF = 0.27 \cdot RMSE_{Q5} + 0.16 \cdot RMSE_{Q20} + 0.08 \cdot RMSE_{Qmid} + 0.24 \cdot RMSE_{Q70} + 0.25 \cdot RMSE_{Q95}$, the weighting scheme explicitly accounted for extremely high (Q95), high (Q70), medium (Qmid), low (Q20), and extremely low (Q5) flows. Despite this design, both NSE and LNSE declined relative to Experiment 1. The decrease may be attributed to excessive parameter adjustments aimed at fitting a limited number of extreme events, which reduced the predictive accuracy of the overall streamflow process. When single dynamic parameters are introduced in Experiment 4, median NSE and LNSE increased to approximately 0.55 and 0.8, respectively, with narrower interquartile ranges. These outcomes indicate that dynamic parameters enhanced the ability of the hydrological model to capture temporal variability, although structural errors persisted, as reflected in local outliers. Experiment 5 achieved median NSE and LNSE values of approximately 0.7–0.8 in both calibration and evaluation periods. Although high-dimensional optimization increased computational demand and LNSE variability in some basins, overall performance represented a balanced trade-off between dynamic adaptability and physical consistency. Experiment 6 also performed well during the calibration period; however, its abrupt parameter switching led to a decline of LNSE and increased dispersion in the evaluation period. Experiment 7 addressed these shortcomings by applying a gradual parameter-switching strategy during the evaluation period. As shown in Fig. 4, the boxplots are more compact and shifted toward higher values, indicating that stable and consistent performance was achieved across most basins. However, compared with Experiment 5, Experiment 7 displayed a greater number of outliers, particularly in LNSE, where they tended to cluster at lower values, suggesting higher variability in model performance across catchments. The overall accuracy remained comparable to that of Experiment 5. In summary, compared with static calibration schemes (Experiments 1–3), single dynamic parameter calibration (Experiment 4) improved simulative accuracy, while multi dynamic parameter calibration produced further gains. Among all experiments, Experiments 5 and 7 demonstrated the most robust and accurate performance. Representative catchments further confirm the basin-scale conclusions. Across all five cases, Experiments 5 and 7 consistently maintain superior performance during evaluation, with improved robustness across flow phases. Detailed multi-metric results for the five representative cases are provided in the Supporting Information S3.

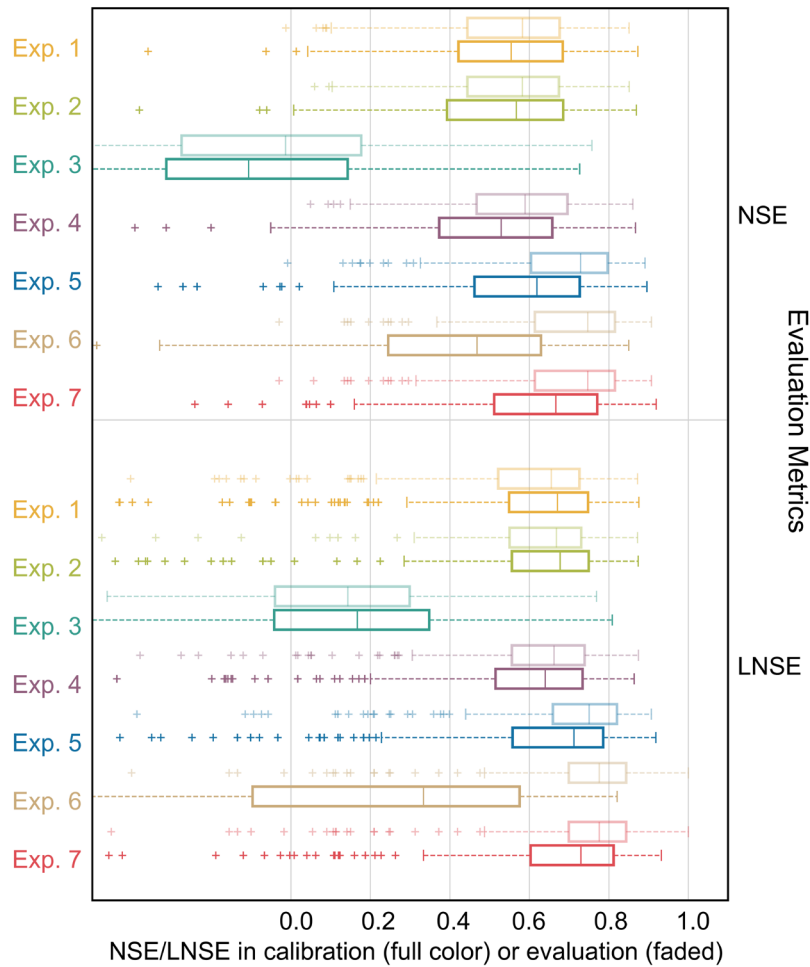


Figure 4. Performance of seven calibration experiments on the MOPEX dataset across 219 catchments. Boxplot colour denotes different experiments. The whiskers extend a maximum of 1.5 times the interquartile range. Values beyond the whiskers are marked as outliers and are denoted as +.

4.3 State variables and fluxes

The state variables and fluxes reflect the internal operation of the hydrological model. The assessment results of state variables and fluxes through seven calibration experiments for case study A are illustrated in Fig. 5 and Fig. 6 (results of cases B, C, D, and E are shown in Supporting Information S6). Experiments 1, 2, and 3 exhibited only minimal differences in both state variables and flux time series, with only the results of Experiments 1 and 3 shown for clarity. A slight improvement is shown in Experiment 4 compared with the time-invariant parameter schemes; however, small mismatches remain during flow recessions and peak timings. This indicates that the dynamic adjustment of a single parameter is insufficient to represent the full range of catchment dynamics. In Experiment 6, abrupt parameter switching is applied across sub-periods. The state variable X_q and flux Q_q in Experiment 6, exhibit step changes or even discontinuities at the switching boundaries, with large deviations during low-flow sub-periods. The phenomenon is particularly evident in cases B and D. These results indicate that abrupt switching disrupts water balance continuity, thereby reducing performance in low-flow simulations. Despite these setbacks, Experiments 5 and 7 introduced significant improvements across all study cases. In Experiment 5, multi-parameter dynamic calibration is applied while continuity of state variables and fluxes is maintained. As shown in Fig. 5 and Fig. 6, in case A, the flux variables Q_q and Q_s transition smoothly across sub-periods without visible discontinuities, the state variables XH_{uz} and XC_{uz} also connect consistently across sub-periods, indicating that multi-parameter dynamic calibration captures the catchment dynamics of soil moisture and storage processes. However, Experiment 5 shows limitations in maintaining the consistency of simulated discharge (Q_{sim}). For example, in case B,

the baseline extent of Q_{sim} exhibited slight drift, reflected in systematic differences in response intensity to similar rainfall events across adjacent sub-periods. The fluxes and state variables in Experiment 7 exhibit results similar to those in Experiment 5. However, when sub-period simulations are concatenated, slight inconsistencies occasionally emerge at the sub-period boundaries, with flood peaks being slightly overestimated or baseflows being underestimated. A comparative flux-mapping analysis of Experiments 1, 5, and 7 is provided in the Supporting Information S6 to further examine differences in runoff component consistency and parameter equifinality across sub-periods.

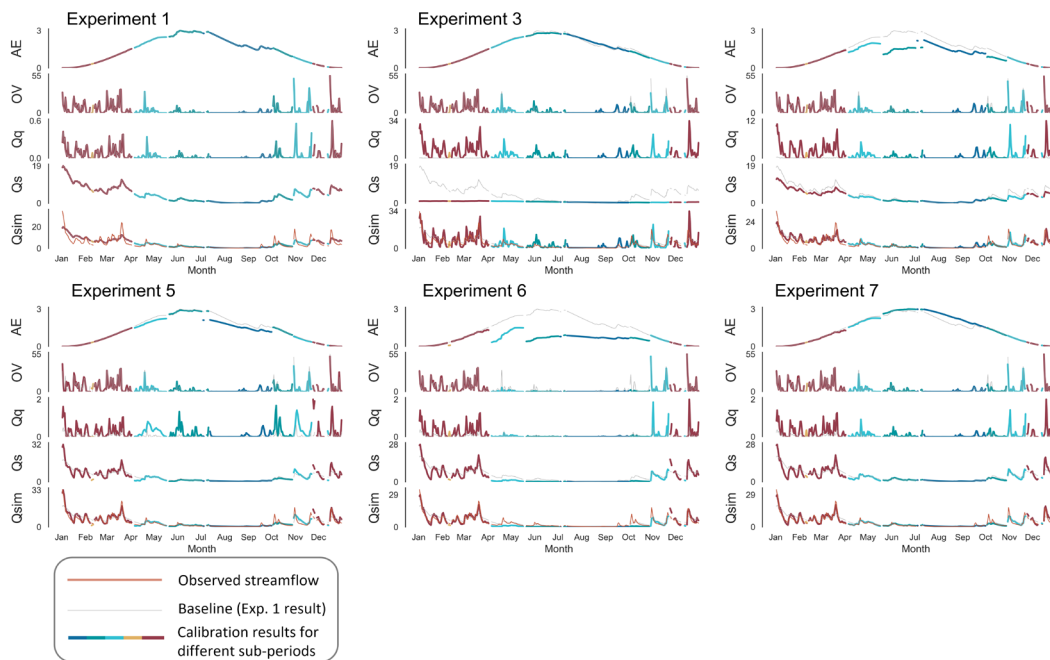


Figure 5. Flux simulation results for Experiments 1–7 during 1998, a representative year from the middle of the evaluation period for Case A. The figure shows the flux simulation results from Experiments 1 to 7, with different colours representing different sub-periods.

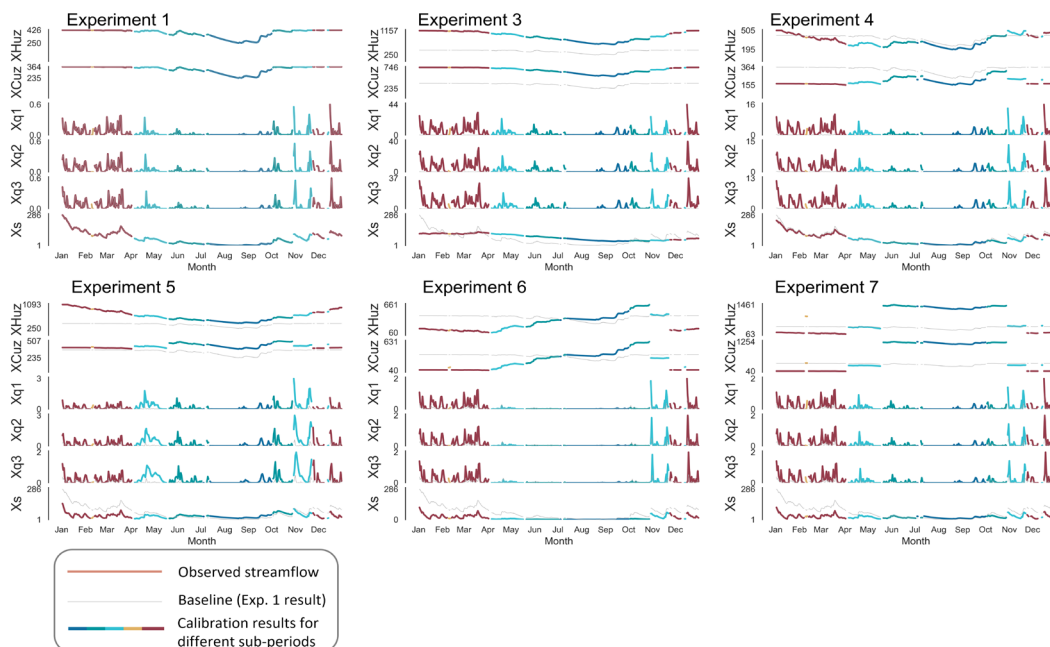


Figure 6. State variables simulation results for Experiments 1–7 during 1998, a representative year from the middle of the evaluation period for Case A. The figure shows the state variable simulation results from Experiments 1 to 7, with different colours representing different sub-periods.

5.1 Why dynamic parameter sets improve simulation performance

Despite the significant improvement in the simulation performance of hydrological models based on catchment dynamics, analyses of dynamic parameter behaviour (presented in Supporting Information S7) indicate that the response of discretized dynamic parameters (even highly sensitive ones) to these catchment dynamics is not satisfactory. However, a dynamic parameter set can collectively carry the extracted information of dynamic catchment characteristics, compensating for model structural deficiencies and improving model performance. Therefore, this study further explores the potential reasons from three aspects: the correlations between parameters, equifinality in the hydrological model, and the evolution process of parameters.

5.1.1 Complex correlation between parameters

Fig. 7a and Fig. 7c demonstrate that there are both significantly linear and nonlinear correlations among the parameters of the hydrological model in the study case A (results of other cases are shown in Supporting Information S8). MIC values above 0.35 among most parameters suggest that the dynamics of individual parameters may be affected by others (Gillespie et al. 2021). This explains the unimproved model performance when altering individual parameters during different sub-periods in Experiment 4. The analysis results of parameter sensitivity based on the scatter plot method also confirmed the influence of the correlation between parameters. In the recommended scheme (Experiment 7), parameters like K_s (the slow-flow routing tank's rate) exhibit a weak responsive relationship to the dynamic catchment, validating the significance of clustering sub-periods based on catchment dynamics. Due to the complex linear or nonlinear correlations between parameters, the variation of individual parameters can be compensated for by changes or adjustments in other parameters, leading to no significant changes in the simulation performance of the model (Xiong et al., 2019; Gou et al., 2020; Zhou et al., 2022). Bárdossy (2007) suggested that parameters within a hydrological model parameter group should not be considered individually but rather treated as a whole.

5.1.2 Equifinality in the hydrological model

The parameter sets derived by the SCE-UA algorithm for flux mapping encounter inherent limitations (Beven, 1993; Padiyedath Gopalan et al., 2018). This arises due to the algorithm's inherent directionality in the optimization process, which potentially overlooks certain parameter sets capable of producing equifinality results. Analysis of parameter sensitivity through flux mapping and scatterplot methodology reveals a distinctive feature towards the end of the search path: A tail-like pattern in the scatterplot in Fig. 7a and Fig. 7b, indicating a series of parameter sets with equifinality identified by the optimization algorithm. These scatter points represent parameter sets producing similar results, though originating from distinctly different physical processes. Hence, it may fail to infer that model runs exhibiting higher performance values consistently correspond to more realistic scenarios. The evaluation of model performance, particularly when quantified in a scalar manner, emerges as a weak, unreliable, and unrealistic approach for model assessment. The representation of model processes cannot be sufficiently measured by a solitary performance metric or a limited range of values (Khatami et al., 2019; Knoben et al., 2020). A rigid interpretation of objective functions can lead to misinterpretations; for instance, in Fig. 7b, model runs with marginally lower NSE values might offer more realistic underlying processes compared to those with better NSE values (Gomez, 2019). It is vital to acknowledge that high model performance does not inherently equal realism and may be influenced by numerical artifacts arising from various sources of uncertainty. Moreover, our constrained understanding of catchment processes, involving runoff generation mechanisms and complex runoff events, makes it challenging to determine the likelihood of specific parameter sets occurring in reality (Troin et al., 2021).

5.1.3 Evolution process of parameters

While the causes of non-physical dynamic parameter values are complex, they might be partially attributed to the failure of global optimization algorithms to converge and find approximated global optimal solutions during the evolutionary process. Hydrological model parameter response surfaces exhibit a range of complex characteristics, including high non-linearity, multi-modality, non-convexity, irregularity, discontinuity, noise, roughness, and non-differentiability (Bian et al., 2024; Herrera et al., 2021). To better describe the evolutionary process of the parameters, a fitness landscape is used, where the vertical axis represents the objective function values and the horizontal axis represents the parameter space (Fig. 7d). The evolutionary process is the process of searching for a global optimum. During this process, deceptive gradients of the objective function values can mislead the optimizer away from the global optimum; the increase in the number of local optima also makes the search path for the global optimum more complex and challenging (Bian et al., 2024). Terminating at a local optimum can prevent the optimized parameters from accurately responding to environmental changes.

5.2 Problems caused by parameter abrupt shifts

Abrupt parameter shifts disrupt the assumption of long-term water balance in traditional hydrological models, potentially leading to invalid values for state variables in adjacent sub-periods (Myers et al., 2021; Fowler et al., 2022). For instance, during the transition of soil maximum storage height (H_{uz}), the H_{uz} value for the next sub-period might be lower than the former actual state variable value (XH_{uz}). Similarly, numerical overflow errors might lead to model crashes and the generation of invalid results (Fig. 7e). These errors could also propagate through various modules of the model, such as the high-speed runoff module and slow-speed runoff module, disrupting the proper functioning of other parts of the model and making the optimization algorithm incapable of producing valid results.

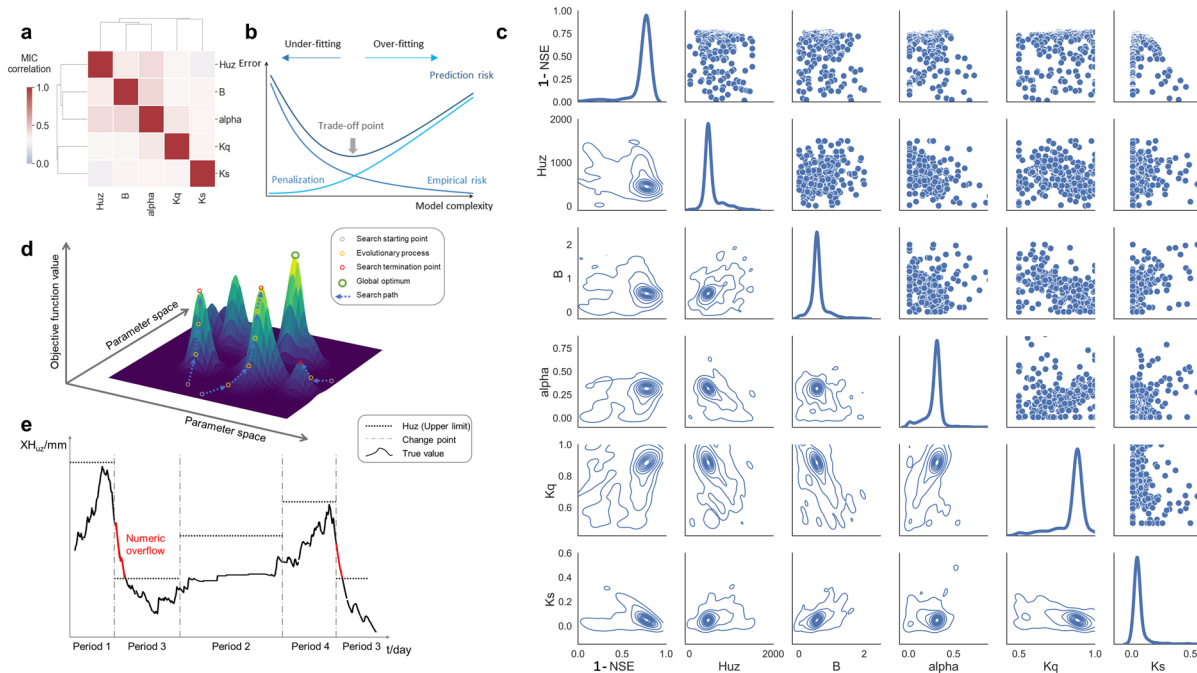


Figure 7. **a**, Linear or nonlinear correlations between parameters based on MICs in case A, with red indicating the strongest correlation among parameters. **b**, Conceptual diagram illustrating the trade-off between empirical fitting to data and the penalization of model complexity, and its impact on prediction error (Schoups et al., 2008). **c**, Parameter sensitivity analysis for case A through scatter plots. **d**, Three-dimensional fitness landscape showing the objective function values on the vertical axis, parameter space on the horizontal axis, and various evolutionary paths that elements can follow within the parameter space, indicated by arrows. **e**, Conceptual diagram of errors resulting from abrupt parameter shifts.

5.3 Parameter response to catchment dynamics

In this study, the sub-period clustering method (Section 3.2) is used to extract temporally varying catchment information, allowing model parameters to adjust across distinct hydrological periods. The resulting improvements in simulation accuracy and robustness highlight the value of incorporating time-varying parameters into hydrological models (Refsgaard et al., 2021). A key question, however, is whether these parameter variations reflect genuine changes in catchment behaviour or primarily compensate for structural limitations of the model itself (Thornton et al., 2022). To address this problem, a diagnostic experiment is designed. Building on the sub-period calibration framework (Experiment 7), a soft constraint based on globally optimal parameters is introduced, integrating prior information on overall catchment behaviour into sub-period parameter estimation. This design balances the flexibility of dynamic parameter adjustment with the need to preserve physical consistency. The diagnostic objective function is defined as:

$$OF=1-(0.5*NSE+0.5*LNSE) + \text{Penalty} \quad (2)$$

where the penalty term quantifies the deviation of the sub-period parameter set $\hat{\theta}_i$ from the globally optimal parameter set θ_i . The penalty is formulated as the mean of the absolute relative errors: $\text{Penalty} = \frac{1}{N} \times \sum \left| \frac{\hat{\theta}_i - \theta_i}{\theta_i} \right|$, where i denotes the parameter index, and N is the total number of parameters (HYMOD). This setting allows assessment of how model responses change when parameter variability is constrained within a more stable and physically consistent range.

As shown in Fig. 8, the imposed constraint produces more concentrated posterior parameter distributions and improves parameter transferability between calibration and evaluation periods. However, these gains are accompanied by clear reductions in NSE and LNSE relative to unconstrained dynamic calibration, indicating that part of the performance improvement under dynamic calibration arises from compensation for structural deficiencies in the fixed model formulation. The results suggest that the need for dynamic parameters is closely linked to the incomplete representation of key hydrological processes. Within this context, the proposed dynamic calibration framework provides a practical strategy for using dynamic parameters as proxy variables to partially account for unresolved structural limitations and improve streamflow simulation in dynamic catchments (Beven, 2019).

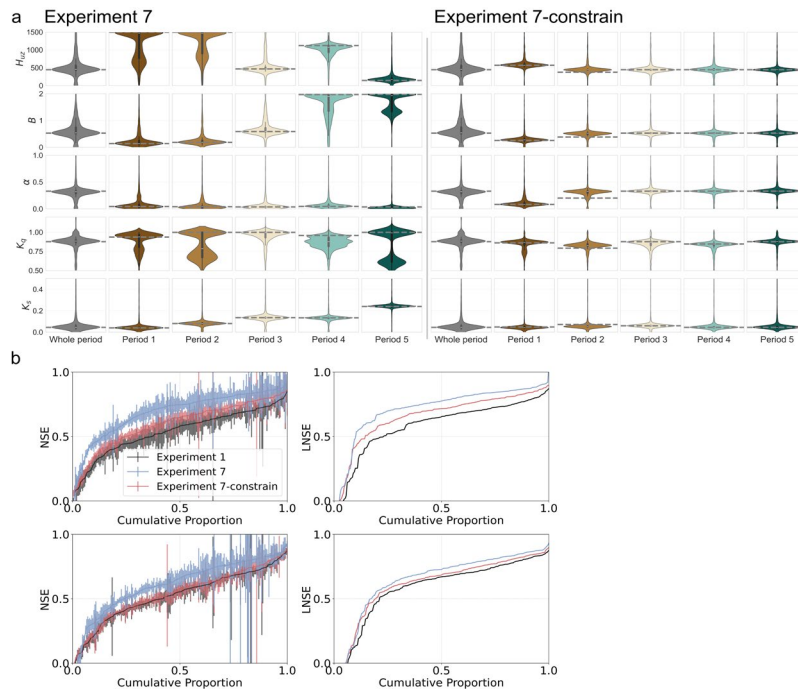


Figure 8. a, Distributions of the optimal parameter spaces across sub-periods under different climatic and land-surface conditions for case A. Each violin depicts one parameter space, with parameter values on the y-axis; the violin width reflects the probability density of the parameter

values. Parameter bounds are: H_{lc} (0-1500), B (0-2), α (0-1), K_q (0.5-1), and K_r (0-0.5). Results for all study cases are provided in Supporting Information S9. **b**, Cumulative distribution functions (CDFs) of NSE and LNSE comparing the experiment built on Experiment 7 with added parameter constraints (blue), Experiment 7 (red), and Experiment 1 (black); higher values indicate better performance. Upper panels show calibration; lower panels show verification. Shaded bands denote 90% bootstrap confidence limits to indicate sampling uncertainty.

6 Conclusions

Due to limitations in observational data and an incomplete understanding of catchment hydrological processes, traditional conceptual hydrological models often fail to represent catchment dynamic characteristics, leading to generalized simulation of different flow regimes. To address model deficiencies and improve simulation in dynamic catchment characteristics, it is essential to re-examine the time-varying information in historical hydrological and meteorological data and consider the variation in calibration. The study investigates calibration challenges in dynamic catchments and proposes a structured framework to address two major issues: The influence of objective function design on flow-phase-specific performance and the limitations of sub-period calibration with dynamic parameters. Seven experiments were conducted to systematically evaluate these aspects. Experiments 1–3 focused on the effects of time-invariant parameters and different objective function configurations, while Experiments 4–7 explored challenges in dynamic parameter calibration, including parameter correlation, high dimensionality, and state transitions. Model performance was comprehensively assessed using multiple metrics and internal diagnostics across 219 MOPEX catchments. The following specific conclusions could be drawn from this study:

- Adjusting the configuration of the objective function can enhance the simulation of emphasized flow phases, but at the cost of sacrificing simulation performance for other flow phases, making it difficult to improve overall model performance.
- Due to issues of model structural deficiencies, correlation among parameters, dimensional disaster in optimization, and the transition of dynamic parameters between adjacent sub-periods, improving model performance through individual parameters alone is not feasible. Model parameters should be considered as a group of parameters.
- Among all calibration experiments, Experiments 5 and 7 effectively addressed the challenges associated with dynamic parameter operations and flow-phase-specific performance, balancing dynamic adaptability and physical consistency. These calibration strategies are thus recommended for application in dynamic catchments, where capturing temporal variability and maintaining model reliability are critical.

The calibration and evaluation framework proposed in this study not only addresses defects caused by the simplification of model structure for hydrological models but also enhances model simulation accuracy across different flow phases and effectively reduces model uncertainty. The evaluation framework comprehensively assesses the performance of hydrological models through multi-criteria evaluation and reveals sources of uncertainty in model internal operation from the perspectives of state variables and fluxes. Despite the positive results of this study, developing more realistic models will aid in our understanding of hydrological processes and improve hydrological forecasting.

Supplement link

Supporting Information is provided.

T.L. conceived the modelling framework. T.L. and X.W. developed the code and prepared the original draft manuscript. H.Z., X.G., X.X., Y.D.C., and C.-Y.X. provided supervision and reviewed/edited the manuscript. J.Z. and W.C. contributed to supplementary calculations and textual revisions during the manuscript revision.

Code availability

480 The MOPEX dataset is available at Duan et al. (2006). The Sensitivity Analysis For Everyone (SAFE) toolbox is available at <https://safetoolbox.github.io/> (last access: 23 November 2024) (Pianosi et al., 2015). Model set-up configurations have been reported in <https://doi.org/10.5281/zenodo.16676391> (last access: 26 September 2025).

Competing interests

The authors declare that they have no conflict of interest.

485 Acknowledgments

This study is financially supported by the National Natural Science Foundation of China(NSFC) (Grant 52209006, and 42071055); Open Research Fund Program of the State Key Laboratory of Hydrosience and Engineering (sklhse-KF-2025-A-03); Scientific Research Program Funded by Education Department of Shaanxi Provincial Government (Program No.25JE005); Open Research Fund of Key Laboratory of Water Security Guarantee inGuangdong-Hong Kong-Macao Greater Bay Area of Ministry of
490 WaterResources (Grant number [2026]KJ015); Key Laboratory of Hydrometeorological Disaster Mechanism and Warning of Ministry of Water Resources (HYMED2024010); and Research Council of Norway (FRINATEK, project no. 274310).

References

- Acuña Espinoza, E., Loritz, R., Álvarez Chaves, M., Bäuerle, N., and Ehret, U.: To bucket or not to bucket? Analyzing the performance and interpretability of hybrid hydrological models with dynamic parameterization, *Hydrology and Earth System Sciences*, 28, 2705-2719, <https://doi.org/10.5194/hess-28-2705-2024>, 2024.
- 495 Anderson, S. and Radić, V.: Evaluation and interpretation of convolutional long short-term memory networks for regional hydrological modelling, *Hydrology and Earth System Sciences*, 26, 795-825, <https://doi.org/10.5194/hess-26-795-2022>, 2022.
- Araya, D., Mendoza, P. A., Muñoz-Castro, E., and McPhee, J.: Towards robust seasonal streamflow forecasts in mountainous catchments: impact of calibration metric selection in hydrological modeling, *Hydrology and Earth System Sciences*, 27, 4385-4408, <https://doi.org/10.5194/hess-27-4385-2023>, 2023.
- 500 Bárdossy, A.: Calibration of hydrological model parameters for ungauged catchments, *Hydrology and Earth System Sciences*, 11, 703-710, <https://doi.org/10.5194/hess-11-703-2007>, 2007.
- Bárdossy, A. and Singh, S. K.: Robust estimation of hydrological model parameters, *Hydrology and Earth System Sciences*, 12, 1273-1283, <https://doi.org/10.5194/hess-12-1273-2008>, 2008.
- 505 Beven, K.: How to make advances in hydrological modelling. *Hydrology Research*, 50(6), 1481-1494, <https://doi.org/10.2166/nh.2019.134>, 2019.
- Beven, K.: Prophecy, reality and uncertainty in distributed hydrological modelling. *Advances in Water Resources*, 16(1), 41-51, [https://doi.org/10.1016/0309-1708\(93\)90028-E](https://doi.org/10.1016/0309-1708(93)90028-E), 1993.
- Bian, K., & Priyadarshi, R.: Machine learning optimization techniques: a survey, classification, challenges, and future research issues. *Archives of Computational Methods in Engineering*, 31(7), 4209-4233, <https://doi.org/10.1007/s11831-024-10110-w>, 2024.
- 510 Bouaziz, L. J., Aalbers, E. E., Weerts, A. H., Hegnauer, M., Buiteveld, H., Lammersen, R., ... & Hrachowitz, M.: Ecosystem adaptation to climate change: the sensitivity of hydrological predictions to time-dynamic model parameters. *Hydrology and Earth System Sciences*, 26(5), 1295-1318, <https://doi.org/10.5194/hess-26-1295-2022>, 2022.
- Carletti, F., Michel, A., Casale, F., Burri, A., Bocchiola, D., Bavay, M., & Lehning, M.: A comparison of hydrological models with different level of complexity in Alpine regions in the context of climate change. *Hydrology and Earth System Sciences*, 26(13), 3447-3475, <https://doi.org/10.5194/hess-26-3447-2022>, 2022.
- 515 Clark, M. P., Vogel, R. M., Lamontagne, J. R., Mizukami, N., Knoben, W. J. M., Tang, G., Gharari, S., Freer, J. E., Whitfield, P. H., Shook, K. R., and Papalexiou, S. M.: The Abuse of Popular Performance Metrics in Hydrologic Modeling, *Water Resources Research*, 57, e2020WR029001, <https://doi.org/10.1029/2020WR029001>, 2021.

- 520 Clark, M. P., Fan, Y., Lawrence, D. M., Adam, J. C., Bolster, D., Gochis, D. J., Hooper, R. P., Kumar, M., Leung, L. R., Mackay, D. S., Maxwell, R. M., Shen, C., Swenson, S. C., and Zeng, X.: Improving the representation of hydrologic processes in Earth System Models, *Water Resources Research*, 51, 5929-5956, <https://doi.org/10.1002/2015WR017096>, 2015.
- Daggupati, P., Yen, H., White, M. J., Srinivasan, R., Arnold, J. G., Keitzer, C. S., and Sowa, S. P.: Impact of model development, calibration and validation decisions on hydrological simulations in West Lake Erie Basin, *Hydrological Processes*, 29, 5307-5320, <https://doi.org/10.1002/hyp.10536>, 2015.
- 525 Deng, C., Liu, P., Guo, S., Li, Z., and Wang, D.: Identification of hydrological model parameter variation using ensemble Kalman filter, *Hydrology and Earth System Sciences*, 20, 4949-4961, <https://doi.org/10.5194/hess-20-4949-2016>, 2016.
- Duan, Q., Schaake, J., Andréassian, V., Franks, S., Goteti, G., Gupta, H. V., Gusev, Y. M., Habets, F., Hall, A., Hay, L., Hogue, T., Huang, M., Leavesley, G., Liang, X., Nasonova, O. N., Noilhan, J., Oudin, L., Sorooshian, S., Wagener, T., and Wood, E. F.: Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops, *Journal of Hydrology*, 320, 3-17, <https://doi.org/10.1016/j.jhydrol.2005.07.031>, 2006.
- 530 Duan, Q. Y., Gupta, V. K., and Sorooshian, S.: Shuffled Complex Evolution Approach for Effective and Efficient Global Minimization, *Journal of Optimization Theory and Applications*, 76, 501-521, <https://doi.org/10.1007/BF00939380>, 1993.
- Fauer, F. S., Ulrich, J., Jurado, O. E., and Rust, H. W.: Flexible and consistent quantile estimation for intensity–duration–frequency curves, *Hydrology and Earth System Sciences*, 25, 6479-6494, <https://doi.org/10.5194/hess-25-6479-2021>, 2021.
- 535 Fowler, K., Peel, M., Saft, M., Nathan, R., Horne, A., Wilby, R., ... & Peterson, T.: Hydrological shifts threaten water resources. *Water Resources Research*, 58(8), e2021WR031210, <https://doi.org/10.1029/2021WR031210>, 2022.
- Fowler, K., Peel, M., Saft, M., Peterson, T. J., Western, A., Band, L., Petheram, C., Dharmadi, S., Tan, K. S., Zhang, L., Lane, P., Kiem, A., Marshall, L., Griebel, A., Medlyn, B. E., Ryu, D., Bonotto, G., Wasko, C., Ukkola, A., Stephens, C., Frost, A., Weligamage, H. G., Saco, P., Zheng, H. X., Chiew, F., Daly, E., Walker, G., Vervoort, R. W., Hughes, J., Trotter, L., Neal, B., Cartwright, I., and Nathan, R.: Explaining changes in rainfall-runoff relationships during and after Australia's Millennium Drought: a community perspective, *Hydrology and Earth System Sciences*, 26, 6073-6120, <https://doi.org/10.5194/hess-26-6073-2022>, 2022.
- 540 Gillespie, L. M., Hättenschwiler, S., Milcu, A., Wambsganss, J., Shiha, A., & Fromin, N.: Tree species mixing affects soil microbial functioning indirectly via root and litter traits and soil parameters in European forests. *Functional Ecology*, 35(10), 2190-2204, <https://doi.org/10.1111/1365-2435.13877>, 2021.
- 545 Gomez, J.: Stochastic global optimization algorithms: A systematic formal approach, *Inform Sciences*, 472, 53-76, <https://doi.org/10.1016/j.ins.2018.09.021>, 2019.
- Gou, J., Miao, C., Duan, Q., Tang, Q., Di, Z., Liao, W., ... & Zhou, R.: Sensitivity analysis - based automatic parameter calibration of the VIC model for streamflow simulations over China. *Water Resources Research*, 56(1), e2019WR025968, <https://doi.org/10.1029/2019WR025968>, 2020.
- 550 Guo, D., Johnson, F., and Marshall, L.: Assessing the potential robustness of conceptual rainfall-runoff models under a changing climate, *Water Resources Research*, 54, 5030-5049, <https://doi.org/10.1029/2018WR022636>, 2018.
- 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, *Journal of Hydrology*, 377, 80-91, <https://doi.org/10.1016/j.jhydrol.2009.08.003>, 2009.
- 555 Herrera, Paulo A., Miguel Angel Marazuola, and Thilo Hofmann.: Parameter estimation and uncertainty analysis in hydrological modeling. *Wiley Interdisciplinary Reviews: Water*, 9.1: e1569, <https://doi.org/10.1002/wat2.1569>, 2021.
- Höge, M., Wöhling, T., and Nowak, W.: A primer for model selection: The decisive role of model complexity, *Water Resources Research*, 54, 1688-1715, <https://doi.org/10.1002/2017WR021902>, 2018.
- Hsueh, H. F., Guthke, A., Wöhling, T., & Nowak, W.: Optimized predictive coverage by averaging time - windowed Bayesian distributions. *Water Resources Research*, 60(5), e2022WR033280, <https://doi.org/10.1029/2022WR033280>, 2024.
- 560 Ji, H. K., Mirzaei, M., Lai, S. H., Dehghani, A., and Dehghani, A.: The robustness of conceptual rainfall-runoff modelling under climate variability – A review, *Journal of Hydrology*, 621, 129666, <https://doi.org/10.1016/j.jhydrol.2023.129666>, 2023.
- Karpatne, A., Atluri, G., Faghmous, J. H., Steinbach, M., Banerjee, A., Ganguly, A., Shekhar, S., Samatova, N., and Kumar, V.: Theory-guided data science: A new paradigm for scientific discovery from data, *IEEE Transactions on Knowledge and Data Engineering*, 29, 2318-2331, <https://doi.org/10.1109/TKDE.2017.2720168>, 2017.
- 565 Khatami, S., Peel, M. C., Peterson, T. J., and Western, A. W.: Equifinality and flux mapping: A new approach to model evaluation and process Representation Under Uncertainty, *Water Resources Research*, 55, 8922-8941, <https://doi.org/10.1029/2018WR023750>, 2019.
- Knoben, W. J., Freer, J. E., Peel, M. C., Fowler, K. J. A., & Woods, R. A.: A brief analysis of conceptual model structure uncertainty using 36 models and 559 catchments. *Water Resources Research*, 56(9), e2019WR025975, <https://doi.org/10.1029/2019WR025975>, 2020.
- 570 Krapu, C. and Borsuk, M.: A differentiable hydrology approach for modeling with time-varying parameters, *Water Resources Research*, 58, e2021WR031377, <https://doi.org/10.1029/2021WR031377>, 2022.
- Kratzert, F., Klotz, D., Brenner, C., Schulz, K., and Herrnegger, M.: Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks, *Hydrology and Earth System Sciences*, 22, 6005-6022, <https://doi.org/10.5194/hess-22-6005-2018>, 2018.
- Kratzert, F., Klotz, D., Herrnegger, M., Sampson, A. K., Hochreiter, S., and Nearing, G. S.: Toward improved predictions in ungauged basins: Exploiting the power of machine learning, *Water Resources Research*, 55, 11344-11354, <https://doi.org/10.1029/2019WR026065>, 2019a.
- 575 Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., and Nearing, G.: Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets, *Hydrology and Earth System Sciences*, 23, 5089-5110, <https://doi.org/10.5194/hess-23-5089-2019>, 2019b.
- Lakshmi, G. and Sudheer, K. P.: Parameterization in hydrological models through clustering of the simulation time period and multi-objective optimization based calibration, *Environ Modell Softw*, 138, <https://doi.org/10.1016/j.envsoft.2021.104981>, 2021.
- 580 Lin, Y., Wang, D., Zhu, J., Sun, W., Shen, C., & Shangguan, W.: Development of objective function-based ensemble model for streamflow forecasts. *Journal of Hydrology*, 632, 130861, <https://doi.org/10.1016/j.jhydrol.2024.130861>, 2024.
- Longyang, Q. and Zeng, R.: A hierarchical temporal scale framework for data - driven reservoir release modeling, *Water Resources Research*, 59, <https://doi.org/10.1029/2022WR033922>, 2023.

- 585 Martel, J.-L., Brissette, F., Arsenault, R., Turcotte, R., Castañeda-Gonzalez, M., Armstrong, W., Mailhot, E., Pelletier-Dumont, J., Rondeau-Genesse, G., and Caron, L.-P.: Assessing the adequacy of traditional hydrological models for climate change impact studies: a case for long short-term memory (LSTM) neural networks, *Hydrology and Earth System Sciences*, 29, 2811-2836, <https://doi.org/10.5194/hess-29-2811-2025>, 2025.
- McCabe, G. J., Hay, L. E., Bock, A., Markstrom, S. L., and Atkinson, R. D.: Inter-annual and spatial variability of Hamon potential evapotranspiration model coefficients, *Journal of Hydrology*, 521, 389-394, <https://doi.org/10.1016/j.jhydrol.2014.12.006>, 2015.
- 590 Myers, D. T., Ficklin, D. L., Robeson, S. M., Neupane, R. P., Botero - Acosta, A., & Avellaneda, P. M.: Choosing an arbitrary calibration period for hydrologic models: How much does it influence water balance simulations?. *Hydrological Processes*, 35(2), e14045, <https://doi.org/10.1002/hyp.14045>, 2021.
- Moore, R. J.: The probability-distributed principle and runoff production at point and basin scales, *Hydrological Sciences Journal*, 30, 273-297, <https://doi.org/10.1080/02626668509490989>, 2009.
- 595 Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I — A discussion of principles, *Journal of Hydrology*, 10, 282-290, [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6), 1970.
- Nearing, G. S., Kratzert, F., Sampson, A. K., Pelissier, C. S., Klotz, D., Frame, J. M., Prieto, C., and Gupta, H. V.: What Role Does Hydrological Science Play in the Age of Machine Learning?, *Water Resources Research*, 57, e2020WR028091, <https://doi.org/10.1029/2020WR028091>, 2021.
- 600 Orth, R., Staudinger, M., Seneviratne, S. I., Seibert, J., and Zappa, M.: Does model performance improve with complexity? A case study with three hydrological models, *Journal of Hydrology*, 523, 147-159, <https://doi.org/10.1016/j.jhydrol.2015.01.044>, 2015.
- Padiyedath Gopalan, S., Kawamura, A., Takasaki, T., Amaguchi, H., and Azhikodan, G.: An effective storage function model for an urban watershed in terms of hydrograph reproducibility and Akaike information criterion, *Journal of Hydrology*, 563, 657-668, <https://doi.org/10.1016/j.jhydrol.2018.06.035>, 2018.
- 605 Pande, S. and Moayeri, M.: Hydrological interpretation of a statistical measure of basin complexity, *Water Resources Research*, 54, 7403-7416, <https://doi.org/10.1029/2018WR022675>, 2018.
- Pianosi, F., Sarrazin, F., and Wagener, T.: A Matlab toolbox for lobar sensitivity analysis, *Environmental Modelling & Software*, 70, 80-85, <https://doi.org/10.1016/j.envsoft.2015.04.009>, 2015.
- Razavi, S., Duffy, A., Eamen, L., Jakeman, A. J., Jardine, T. D., Wheeler, H., Hunt, R. J., Maier, H. R., Abdelhamed, M. S., and Ghoreishi, M.: 610 Convergent and transdisciplinary integration: On the future of integrated modeling of human - water systems, *Water Resources Research*, 61, <https://doi.org/10.1029/2024WR038088>, 2025.
- Refsgaard, J. C., Stisen, S., & Koch, J.: Hydrological process knowledge in catchment modelling—Lessons and perspectives from 60 years development. *Hydrological Processes*, 36(1), e14463, <https://doi.org/10.1002/hyp.14463>, 2021.
- Reichert, P., Ammann, L., & Fenicia, F.: Potential and challenges of investigating intrinsic uncertainty of hydrological models with stochastic, 615 time - dependent parameters. *Water Resources Research*, 57(3), e2020WR028400, <https://doi.org/10.1029/2020WR028400>, 2021.
- Schoups, G., van de Giesen, N. C., and Savenije, H. H. G.: Model complexity control for hydrologic prediction, *Water Resources Research*, 44, <https://doi.org/10.1029/2008WR006836>, 2008.
- Schwemmler, R., Demand, D., & Weiler, M.: Diagnostic efficiency—specific evaluation of model performance. *Hydrology and Earth System Sciences*, 25(4), 2187-2198, <https://doi.org/10.5194/hess-25-2187-2021>, 2021.
- 620 Shamir, E., Imam, B., Gupta, H. V., and Sorooshian, S.: Application of temporal streamflow descriptors in hydrologic model parameter estimation, *Water Resources Research*, 41, <https://doi.org/10.1029/2004WR003409>, 2005.
- Shao, M., Fernando, N., Zhu, J., Zhao, G., Kao, S. C., Zhao, B., Roberts, E., and Gao, H.: Estimating future surface water availability through an integrated climate - hydrology - management modeling framework at a basin scale under CMIP6 scenarios, *Water Resources Research*, 59, <https://doi.org/10.1029/2022WR034099>, 2023.
- 625 Shrestha, S., Bae, D.-H., Hok, P., Ghimire, S., and Pokhrel, Y.: Future hydrology and hydrological extremes under climate change in Asian river basins, *Scientific Reports*, 11, 17089, <https://doi.org/10.1038/s41598-021-96656-2>, 2021.
- Song, Y., Knoben, W. J., Clark, M. P., Feng, D., Lawson, K. E., and Shen, C.: When ancient numerical demons meet physics-informed machine learning: adjoint-based gradients for implicit differentiable modeling, *Hydrology and Earth System Sciences Discussions*, 2023, 1-35, <https://doi.org/10.5194/hess-28-3051-2024>, 2023.
- 630 Thornton, J. M., Therrien, R., Mariéthoz, G., Linde, N., & Brunner, P.: Simulating fully - integrated hydrological dynamics in complex alpine headwaters: potential and challenges. *Water Resources Research*, 58(4), e2020WR029390, <https://doi.org/10.1029/2020WR029390>, 2022.
- Troin, M., Arsenault, R., Wood, A. W., Brissette, F., & Martel, J. L.: Generating ensemble streamflow forecasts: A review of methods and approaches over the past 40 years, *Water Resources Research*, 2020WR028392, <https://doi.org/10.1029/2020WR028392>, 2020.
- 635 Tucker, C. J., Pinzon, J. E., Brown, M. E., Slayback, D. A., Pak, E. W., Mahoney, R., Vermote, E. F., and El Saleous, N.: An extended AVHRR 8-km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data, *International Journal of Remote Sensing*, 26, 4485-4498, <https://doi.org/10.1080/01431160500168686>, 2010.
- Vrugt, J. A., Gupta, H. V., Bastidas, L. A., Bouten, W., & Sorooshian, S.: Effective and efficient algorithm for multiobjective optimization of hydrologic models. *Water resources research*, 39(8), <https://doi.org/10.1029/2002WR001746>, 2003.
- 640 Wagener, T., Boyle, D. P., Lees, M. J., Wheeler, H. S., Gupta, H. V., & Sorooshian, S.: A framework for development and application of hydrological models. *Hydrology and Earth System Sciences*, 5(1), 13-26, <https://doi.org/10.5194/hess-5-13-2001>, 2001.
- Wagener, T. and Kollat, J.: Numerical and visual evaluation of hydrological and environmental models using the Monte Carlo analysis toolbox, *Environmental Modelling & Software*, 22, 1021-1033, <https://doi.org/10.1016/j.envsoft.2006.06.017>, 2007.
- Wang, S., Ancell, B. C., Huang, G. H., and Baetz, B. W.: Improving robustness of hydrologic ensemble predictions through probabilistic pre- and post-processing in sequential data assimilation, *Water Resources Research*, 54, 2129-2151, <https://doi.org/10.1002/2018WR022546>, 645 2018.
- Wang, S., Huang, G. H., Baetz, B. W., and Ancell, B. C.: Towards robust quantification and reduction of uncertainty in hydrologic predictions: Integration of particle Markov chain Monte Carlo and factorial polynomial chaos expansion, *Journal of Hydrology*, 548, 484-497, <https://doi.org/10.1016/j.jhydrol.2017.03.027>, 2017.

- 650 Wang, Z., Yang, Y., Zhang, C., Guo, H., and Hou, Y.: Historical and future Palmer Drought Severity Index with improved hydrological modeling, *Journal of Hydrology*, 610, 127941, <https://doi.org/10.1016/j.jhydrol.2022.127941>, 2022b.
- Wei, X. T., Huang, S. Z., Huang, Q., Leng, G. Y., Wang, H., He, L., Zhao, J., and Liu, D.: Identification of the interactions and feedbacks among watershed water-energy balance dynamics, hydro-meteorological factors, and underlying surface characteristics, *Stochastic Environmental Research and Risk Assessment*, 35, 69-81, <https://doi.org/10.1007/s00477-020-01896-9>, 2021.
- 655 Wen, H., Brantley, S. L., Davis, K. J., Duncan, J. M., and Li, L.: The limits of homogenization: What hydrological dynamics can a simple model represent at the catchment scale?, *Water Resources Research*, 57, <https://doi.org/10.1029/2020WR029528>, 2021.
- Wi, S. and Steinschneider, S.: Assessing the physical realism of deep learning hydrologic model projections under climate change, *Water Resources Research*, 58, <https://doi.org/10.1029/2022WR032123>, 2022.
- Xie, K., Liu, P., Zhang, J., Wang, G., Zhang, X., and Zhou, L.: Identification of spatially distributed parameters of hydrological models using the dimension-adaptive key grid calibration strategy, *Journal of Hydrology*, 598, 125772, <https://doi.org/10.1016/j.jhydrol.2020.125772>, 2021.
- 660 Xiong, M., Liu, P., Cheng, L., Deng, C., Gui, Z., Zhang, X., and Liu, Y.: Identifying time-varying hydrological model parameters to improve simulation efficiency by the ensemble Kalman filter: A joint assimilation of streamflow and actual evapotranspiration, *Journal of Hydrology*, 568, 758-768, <https://doi.org/10.1016/j.jhydrol.2018.11.038>, 2019.
- Yilmaz, K. K., Gupta, H. V., and Wagener, T.: A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model, *Water Resources Research*, 44, <https://doi.org/10.1029/2007WR006716>, 2008.
- 665 Yoshida, T., Hanasaki, N., Nishina, K., Boulange, J., Okada, M., and Troch, P.: Inference of parameters for a global hydrological model: Identifiability and predictive uncertainties of climate - based parameters, *Water Resources Research*, 58, <https://doi.org/10.1029/2021WR030660>, 2022.
- Zhang, X. and Liu, P.: A time-varying parameter estimation approach using split-sample calibration based on dynamic programming, *Hydrology and Earth System Sciences*, 25, 711-733, <https://doi.org/10.5194/hess-25-711-2021>, 2021.
- 670 Zhou, L., Liu, P., Gui, Z., Zhang, X., Liu, W., Cheng, L., and Xia, J.: Diagnosing structural deficiencies of a hydrological model by time-varying parameters, *Journal of Hydrology*, 605, <https://doi.org/10.1016/j.jhydrol.2021.127305>, 2022.