

A Novel Framework for Robust Calibration and Evaluation of Framework for Dynamic Catchment Characteristics in Hydrological Models in Dynamic Catchments Modeling

5 Tian Lan^{1,2,3}, Jiajia Zhang¹, Wenqing Cheng⁴, Xiao Wang^{4*}, Hongbo Zhang^{1,2,3}, Xinghui Gong^{1,2,3}, Xue Xie⁵, Yongqin David Chen⁶, Chong-Yu Xu⁷

¹School of Water and Environment, Chang'an University, Xi'an 710054, China.

²Key Laboratory of Subsurface Hydrology and Ecological Effects in Arid Region of the Ministry of Education, Chang'an University, Xi'an 710054, China.

10 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.

⁴State Key Laboratory of Water Resources Engineering and Management, Wuhan University, Wuhan, China

⁵Center for Water Resources and Environment, and Guangdong Key Laboratory of Marine Civil Engineering, School of Civil Engineering, Sun Yat-sen University, Guangzhou 510275, China

⁶School of Humanities and Social Science, The Chinese University of Hong Kong, Shenzhen 518172, China.

15 ⁷Department of Geosciences, University of Oslo, P.O. Box 1047 Blindern, 0316 Oslo, Norway.

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

Abstract

Hydrological models often face challenges in accurately simulating hydrological processes within dynamic catchments due to simplifications of model structure. In a dynamic catchment processes due to structural deficiencies caused by oversimplifications. This results in compromised accuracy in where hydrological processes exhibit significant intra-annual or inter-annual variability, accurately capturing dynamic behaviours across different flow phases in seasonal catchments regimes is still challenging for models. To address these challenges, this study proposes a robust investigation calibration framework that incorporates dynamic catchment characteristics. Additionally, issues in dynamic catchments with a focus on two key aspects: the potential impacts influence of objective function configuration design on flow-phase-specific performance, and the limitations of sub-period calibration with dynamic parameters were investigated in this study. A pre-processing framework was developed to bridge models with catchment dynamics by clustering time series into sub periods with similar hydrological processes. Seven calibration experiments were conducted 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. The recommended calibration scheme effectively addressed challenges in dynamic parameter operations, significantly improving model performance across different flow phases and enhancing the simulation in dynamic catchments. In conclusion, incorporating dynamic parameters based on extracted catchment characteristics effectively mitigates structural deficiencies in hydrological models. This approach improves simulation accuracy across different flow phases, reduces uncertainty, and enhances the model's ability to capture dynamic hydrological processes in seasonal catchments. Our findings provide a practical solution for calibrating hydrological models in seasonal catchments, contributing to better understanding of the hydrological cycle. Among all calibration schemes, sub-period calibration with dynamic parameters exhibited the most reliable performance. Static parameter approaches often averaged catchment responses and poorly represented extreme flows, while the single-parameter dynamic scheme provided limited improvements. 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 strategy maintained accuracy through gradual parameter transitions despite higher variability in some catchments. This study demonstrates that sub-period calibration with dynamic catchment characteristics outperforms traditional static parameters by effectively capturing flow-phase variability and sustaining robust performance across dynamic catchments, offering a generalizable solution for simulating hydrological processes in dynamic catchments.

1 Introduction

Hydrological modelling is a crucial tool for water management, facilitating the management of hydrological challenges such as runoff prediction, disaster warning, and water resource management (Gosling et al., 2011; Shrestha et al., 2021; Grayson et al., 2010). However, understanding, modelling, and predicting hydrological processes with greater realism remains significant challenges in hydrological sciences (Clark et al., 2016). In many cases, when hydrological observational data are limited, process-driven hydrological models are more reliable than data-driven models for understanding catchment hydrological processes (Lakshmi and Sudheer, 2021; Bárdossy, 2007). Although more complex models, such as data-driven deep learning models, which possess several orders of magnitude more parameters (Kratzert et al., 2019a; Nearing et al., 2021), can model highly non-linear systems and often provide more accurate streamflow predictions compared to process-based models (Kratzert et al., 2018; Kratzert

55 et al., 2019a; Kratzert et al., 2019b; Nearing et al., 2021), the “black box” nature of purely data-driven models prevents a deeper understanding of hydrological processes (Karpatne et al., 2017). One of the most common flaws in the structure of process-driven models is their inherent oversimplification, leading to an incomplete representation of dynamic features driving the catchment (Pathiraja et al., 2016; Deng et al., 2016; Wang et al., 2022b). These simplifications, originating from a limited understanding of the physical mechanisms behind dynamic catchment characteristics (e.g., seasonal changes in climate and land cover), often lead to structural deficiency and compromised model accuracy (Xiong et al., 2019b; Pathiraja et al., 2016; Fowler et al., 2022). The oversimplification issue further manifests in the practice of using time-invariant model parameters (Zhou et al., 2022; Shamir et al., 2005), which focuses on the overall performance of the model, averaging the hydrological responses, resulting in sacrificing the simulation accuracy of high (low) flow to improve the simulation accuracy of low (high) flow (De Vos et al., 2010; Lin et al., 2010). To address this issue, it is imperative to re-examine historical hydrological and meteorological data, extract catchment dynamics and address structural deficiencies in the models.

Hydrological models are a crucial tool for water management, facilitating the management of hydrological challenges such as runoff projection, disaster warning, and water resource management (Shao et al., 2023; Shrestha et al., 2021; Razavi et al., 2025). These models conceptualize hydrological 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 (Pathiraja et al., 2016; Deng et al., 2016; Wang et al., 2022b; Wen et al., 2021). A dynamic catchment is defined as one in 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 periods (Longyang and Zeng, 2023; Yoshida et al., 2022). Understanding, modelling, and predicting dynamic hydrological processes with greater realism remain significant challenges in hydrological sciences (Clark et al., 2016).

A key challenge in modelling catchments with significant 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 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 regimes (such as extreme high flow, high flow, middle flow, low flow, and extreme low flow).

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. 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~~regimes and observational periods. In critical runoff events like floods and droughts, this static approach may fail to capture the dynamic ~~nature~~catchment characteristic of hydrological processes, underscoring the need for more flexible calibration methods (~~Pfannerstill et al., 2014; Clark et al., 2021~~)(~~Martel et al., 2025; Clark et al., 2021~~). Hence, ~~researchers have investigated~~ various calibration techniques ~~incorporating have been developed to incorporate~~ dynamic catchment characteristics. One method involves revising the objective function based on selected evaluation criteria to improve model performance (~~Thyer et al., 2009; Ji et al., 2023~~)(~~Araya et al., 2023; Ji et al., 2023~~). Calibrations using multi-objective optimization algorithms better highlight different flow ~~phases~~regimes, but face potential challenges such as increased computational complexity, sensitivity to parameter settings, and slower convergence with more objective functions (~~Yapo et al., 1998; Shafii and De Smedt, 2009~~)(~~Song et al., 2023; Shafii and De Smedt, 2009~~). Alternative approaches, like multi-weighted objective functions, can improve the simulation accuracy of specific time and flow ~~phases~~regimes. While these methods enhanced different flow ~~phases~~regimes and water balance, they may not effectively address structural deficiencies and cannot fundamentally enhance the model's overall performance (~~Kollat et al., 2012; Fowler et al., 2018; Wagener et al., 2003~~)(~~Kollat et al., 2012; Fowler et al., 2018; Anderson and Radić, 2022~~).

Another strategy involves ~~using the use of~~ dynamic parameters in hydrological models. ~~Implementing 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 ~~can address~~addresses structural limitations ~~in model structure of models~~ and ~~fundamentally improve~~improves predictive performance, ~~not only for across the full range of hydrological processes, rather than being restricted to~~ specific flow ~~phases~~regimes or periods ~~but across the entire spectrum of hydrological processes~~ (Zhang and Liu, 2021; Krapu and Borsuk, 2022)(~~Zhang and Liu, 2021; Krapu and Borsuk, 2022~~). Recent studies have significantly advanced hydrological simulations by integrating the dynamic catchment characteristics ~~of catchments~~. Clustering based on catchment characteristics, such as precipitation, evapotranspiration, and soil moisture, facilitates the ~~segmentation of dynamic hydrological processes into distinct sub-periods~~ (Choi and Beven, 2007; De Vos et al., 2010; Lakshmi and Sudheer, 2021). ~~Wei et al. (2021) clustering of dynamic hydrological processes into distinct sub-periods~~ (Acuña Espinoza et al., 2024; De Vos et al., 2010; 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 (~~Bronstert, 2004; Hundecha and Bárdossy, 2004~~)(~~Zhang and Liu (2021)~~Wi and Steinschneider, 2022; Hundecha and Bárdossy, 2004). 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.

~~This study aims to investigate the challenges encountered during calibration operations in hydrological models, particularly in catchments with seasonal dynamics, and propose a robust calibration scheme to address these issues. The research focuses on two primary areas: the impact of objective function settings and the complications arising from sub-period calibration with dynamic parameters. To address these concerns, a series of experiments are conducted. First, experiments 1 to 3 examine whether adjusting objective functions can enhance model performance across different flow phases, with a traditional objective function serving as~~

the control (Kollat et al., 2012; Fowler et al., 2018; Wagener et al., 2003). Second, experiments 4 to 6 explore the effects of parameter correlation (Wagener and Kollat, 2007; Höge et al., 2018; Bárdossy and Singh, 2008), the dimensionality disaster on global optimization (Orth et al., 2015; Daggupati et al., 2015; Xie et al., 2021), and the uncertainty introduced by abrupt parameter shifts (Kim and Han, 2017). The performance of these experiments is evaluated using multi-performance metrics, with model state variables and fluxes providing insights into the model's internal behaviours. Ultimately, experiment 7 provides a recommended optimal scheme for modelling seasonal catchments, offering a balanced solution to the identified calibration challenges. All experiments are verified using the MOPEX dataset, and the recommended calibration scheme is demonstrated through four case studies, illustrating its improvements in addressing the identified calibration challenges.

2 Dynamic catchment characteristics

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 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 through multiple metrics and internal diagnostics across 219 MOPEX catchments.

2 Study area

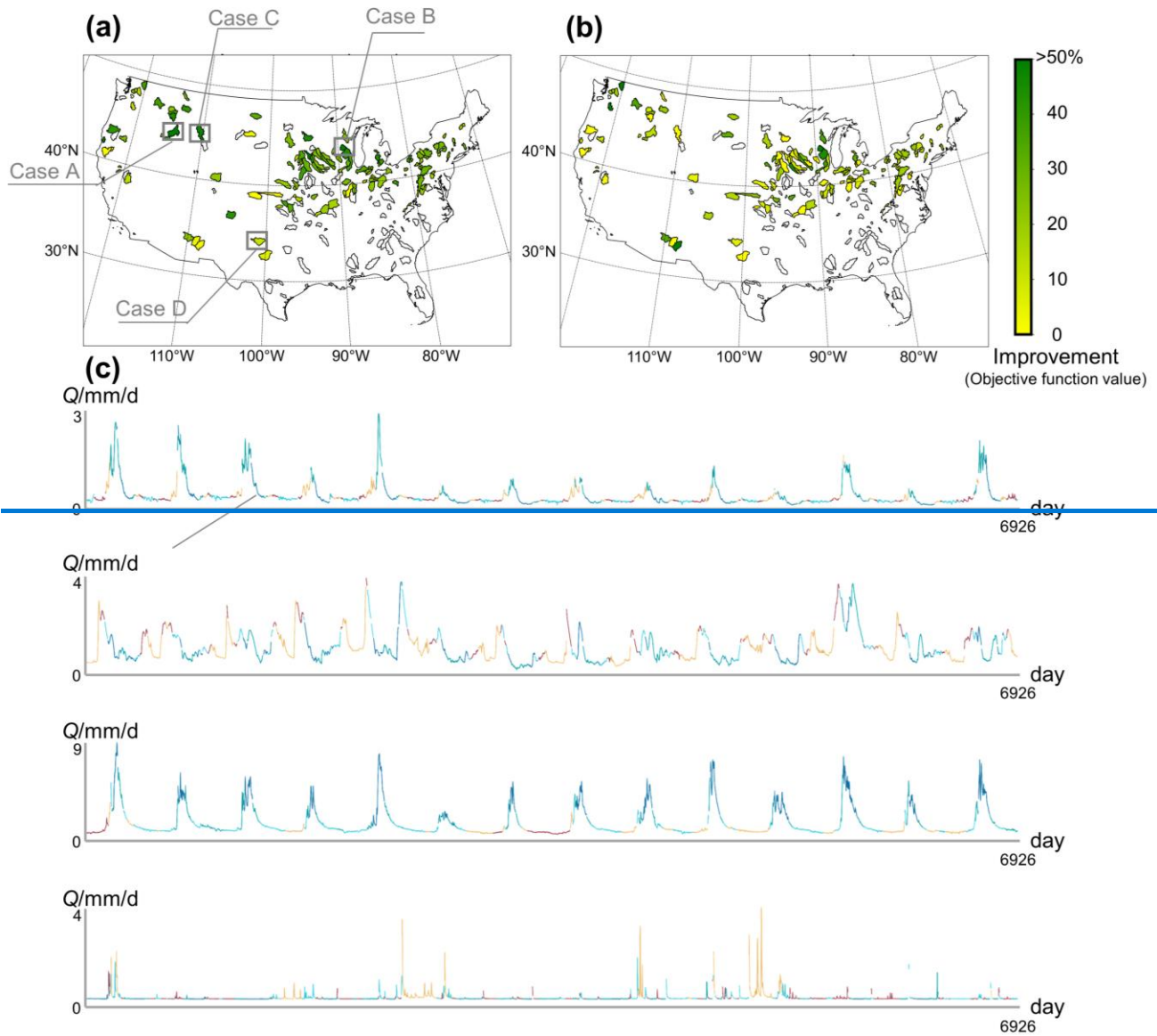
The Model Parameter Estimation Experiment (MOPEX) is an international project aimed at developing enhanced techniques for the ~~the~~ priori estimation of parameters in hydrologic models and land surface parameterization schemes of weather and climate models (Duan et al., 2006) (Duan et al., 2006). A comprehensive MOPEX database has been developed that contains historical hydrometeorological data and land-surface characteristics data for numerous hydrologic ~~basins~~ catchments in the United States (US) and other countries. This study utilizes the dataset from 219 ~~basins~~ catchments spatially distributed across the contiguous US (Fig. 1(a)). Rigorous screening criteria were applied to ensure the acquisition of high-quality data. The screening process involved three key considerations: (1) no missing or ~~abnormal~~ 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 ~~basins~~ catchments, including diverse meteorological and underlying surface conditions. The dataset for selected ~~basins~~ 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 (refer to Supporting Information S2.1) (McCabe et al., 2015). The Normalized Difference Vegetation Index (NDVI) was used as one of the land-surface indices to represent the vegetation coverage of the basins, which had a spatial resolution of 8 km and a temporal resolution of half-monthly intervals (Tucker et al., 2010). Based on these criteria, a total of 219 catchments were selected (Fig. 1(a)), 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 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 hydrologic and meteorological characteristics, making them ideal for testing various model structures under diverse conditions (Duan et al., 2006). Additionally, four catchments—Case A (N13302500), Case B (N04073500), Case C (N06192500), and Case D (N08085500)—are analysed in more detail as case studies (Duan et al., 2006).

2.2 Catchment dynamics

A systematic approach for Extracting Seasonal Dynamic Catchment Characteristics (EDCC) has been developed. This method utilizes machine learning techniques to extract seasonal dynamic information from the catchment, and finally divide time series data into periods with similar seasonal dynamics for further analysis. This process includes three main steps: ① Data Sampling: A 15-day moving window is used to sample hydro-meteorological data, capturing early-stage catchment characteristics. A seasonal characteristic index system is developed, including climatic indices (e.g., precipitation, temperature, evaporation) and land-surface indices (e.g., NDVI, antecedent soil moisture). ② Data Cleaning: A Seasonality Index (SI) is calculated to analyse temporal patterns and identify significant seasonality. The Maximal Information Coefficient (MIC) is used to select indices significantly correlated with streamflow, and Principal Components Analysis (PCA) is applied to eliminate redundant information. ③ Data Clustering: The Fuzzy C-Means (FCM) algorithm is used to cluster data based on climatic and land-surface indices.

The sub-period division results for the four study cases are illustrated in the hydrograph presented in Fig. 1(c), where identical colours denote sub-periods with similar catchment dynamics. A detailed description of the EDCC approach is provided in Supporting Information S2.



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 analysed 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

<u>Runoff ratio</u>	<u>0.72</u>	<u>0.50</u>	<u>0.02</u>	<u>0.40</u>	<u>0.42</u>
<u>Aridity index</u>	<u>2.60</u>	<u>1.01</u>	<u>0.38</u>	<u>1.19</u>	<u>1.23</u>
<u>Forest cover (%)</u>	<u>71.96</u>	<u>36.95</u>	<u>16.76</u>	<u>31.31</u>	<u>57.36</u>
<u>Land use</u>	<u>Evergreen Forest, Pasture/Hay</u>	<u>Evergreen Forest, Shrub/Scrub</u>	<u>Evergreen Forest, Grassland/Herbaceous</u>	<u>Deciduous Forest, Cultivated Crops</u>	<u>Deciduous Forest, Pasture/Hay</u>

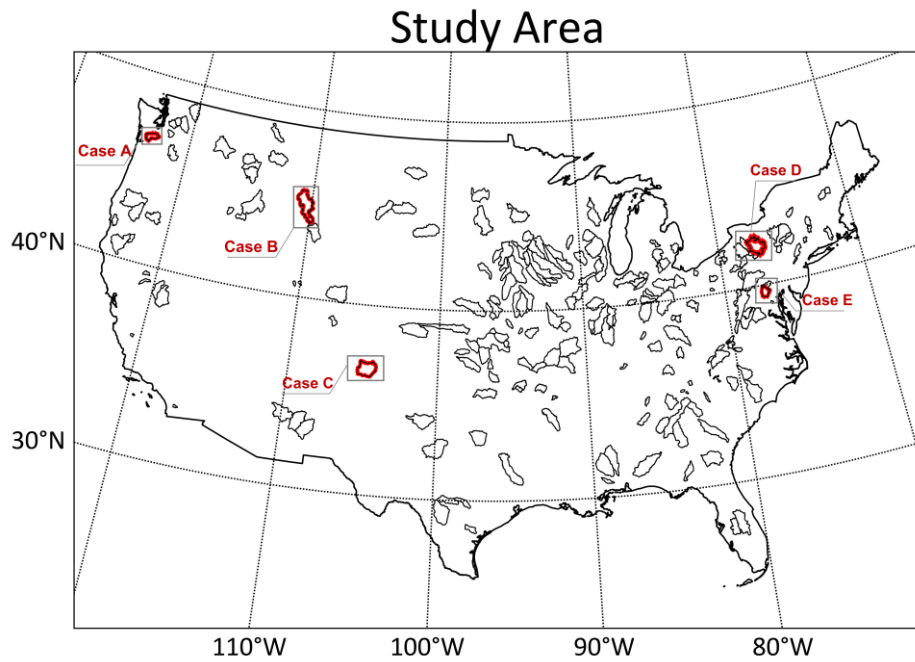


Figure 1. (a), Location map of the catchment area used in this study, where cases A, B, C, D, and E correspond to basins N13302500, N04073500, N06192500, and N08085500, respectively. The shading indicates the improvement in the objective function during the calibration period for the recommended scheme, as detailed in Section 4.1. (b), Improvement during the validation period, with shading reflecting results from Section 4.1. (c), Visualization of clustering results on the hydrograph for the respective study cases. catchments 12027500, 6192500, 7211500, 1643000, and 1531000 (from west to east) are highlighted with red outlines for reference.

3 Methods

3.1 Calibration

Seven calibration experiments are designed and compared to explore potential issues in model calibration for seasonal catchments, and a recommended scheme is proposed based on the results (see Fig. 2). These experiments assess challenges associated with calibration dynamic catchments. Experiments 1 to 3 compare the conventional calibration experiment, the Pareto-based multi-objective calibration experiment, and the multi weighted objective functions calibration experiment. These experiments are employed to investigate issues related to time constant parameters, the trade off between objective functions, and the emphasis across different flow phases. Additionally, three sub period calibration experiments (experiments 4-6) are utilized to explore problems brought by dynamic parameters, including correlation between parameters, the dimensional disaster of parameters, the abrupt shift of parameters and the transition of state variables and fluxes between sub periods.

3.1.1 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 regimes, 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 model

For illustrative purposes, the HYMOD (Hydrological MODEL) model (Moore, 2009)(Moore, 1985) is utilized in this study, since. The HYMOD model is a simple conceptual rainfall-runoff model with a simple structure (five parameters), low input requirements, and clear empirical physical meanings (interpretations). It has been successfully used in streamflow prediction across America and many other regions (Vrugt et al., 2003; Wagener et al., 2001). To enhance model performance in snowy areas, the Degree-day model is applied in this study to account for the snow melt (Supporting Information S1.6) (Wang et al., 2022a).

The structure of the HYMOD model is shown in Fig. 2). The HYMOD consists of 2. Precipitation (P) and potential evapotranspiration (PET) drive a probability-distributed soil-moisture accounting module with store characterized by a maximum capacity (H_{uz}) and a shape parameter (B). Actual evaporation (AE) is limited by potential evapotranspiration and soil water availability. The remaining rainfall infiltrates to replenish the soil-moisture storage (XH_{uz}). When XH_{uz} reaches its maximum capacity (H_{uz}), the surplus is released as excess rainfall (saturation-excess runoff, OV). This excess rainfall is then partitioned by α into inputs to the quick-flow and the slow-flow pathways. Quick flow is routed through a cascade of three adjustable parameters: H_{uz} , B , and α , and two flow routing modules with two adjustable parameters: linear reservoirs (states X_{q1} – X_{q3}) governed by K_q and, producing outflow Q_q , while the slow flow is routed through a single linear reservoir governed by K_s . The definition of the model, producing outflow Q_s . The simulated discharge (Q_{sim}) is computed as the sum of Q_q and Q_s (Wang et al., 2022a). Detailed information on the HYMOD model parameters, state variables, and fluxes are presented is provided in the Supporting Information (Table S1). To transform snowfall2.

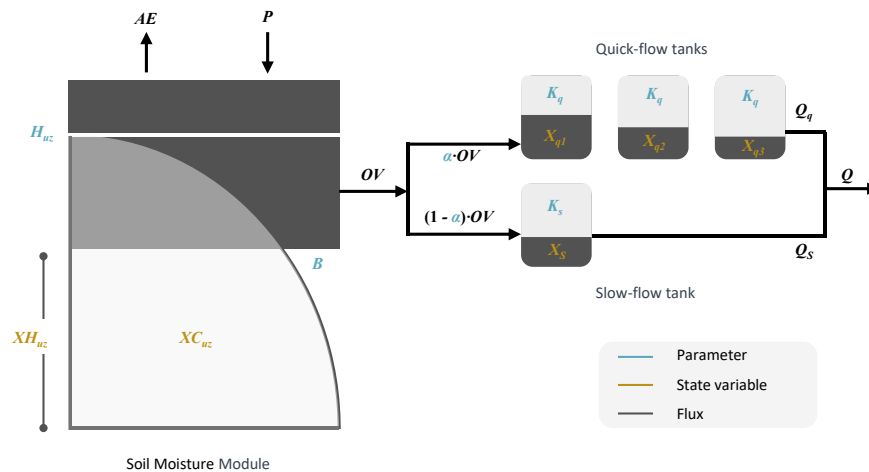


Figure 2. Schematic diagram of the HYMOD structure and principles (Vrugt et al., 2003; Wagener et al., 2001).

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 mm	Scaled distribution function shape

α	Parameter	0–0.99 mm	Quick or slow split
K_q	Parameter	0.5–0.99 mm	Quick-flow routing tanks' rate
K_s	Parameter	0–0.5 mm	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 effective multiple sub-periods characterized by relatively homogeneous hydrological conditions, allowing dynamic parameters to better reflect temporal variations in catchment behaviour across different phases (Liu et al., 2024; Pan et al., 2022; Zhang and Liu, 2021; Zhang et al., 2015). 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 via screening and dimensionality reduction; and (3) applying unsupervised clustering to cluster the time series into hydrologically coherent sub-periods for further model integration.

Describing catchment dynamics: To characterize the temporal dynamics of catchment behavior, 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, the Degree-day model, a widely employed method in hydrology and climatology for estimating snowmelt in snowy basins or glacierized areas (P), temperature (T), and potential evapotranspiration (PE), along with several 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, where the optimal window length is determined via a time-windowed Bayesian inference framework based on temperature data is applied in this study-predictive log-score (PLS) performance (Hsueh et al., 2024). This method is designed to preserve long-term trend signals, suppress short-term high-frequency noise, and enhance the stability and robustness of dynamic catchment characteristic extraction.

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. First, a threshold-based screening is applied to identify variables with significant seasonality, thereby retaining only the relevant subsystems and forming an initial pool of candidate indicators (see Supporting Information S2.6) (Wang et al., 2022a). 1 for detailed criteria). Subsequently, the Maximal Information Coefficient (MIC) is used to quantify both linear and nonlinear associations between candidate indicators and streamflow, ensuring that selected indicators are hydrologically relevant. To further address multicollinearity and reduce dimensionality, Principal Component Analysis (PCA) is conducted, and the first two principal components are retained for

clustering. This multi-step filtering and reduction process ensures robust extraction of dynamic catchment characteristics and establishes a solid foundation for sub-period clustering based on hydrological similarity.

Throughout the experiments, the Shuffled Complex Evolution algorithm (SCE-UA) is employed to search for the globally optimal parameter set (see Supporting Information S2.5) (Duan et al., 1993). All parameters not mentioned are set to their default values. Considering both high flow and low flow, the objective function is set as $OF = 1 - 0.5 \cdot (NSE + LNSE)$, where NSE is sensitive to peaks and discharge dynamics, while the LNSE emphasizes low flows due to the logarithm of discharge (Nash and Sutcliffe, 1970). The HYMOD model was configured for basins over a 19-year period from 1982 to 2000, with the first year serving as warm-up period, the following 13 years serving as the calibration period and the last 5 years as the validation period. It should be noted that this framework is applicable to all models, and the use of the HYMOD model is merely for the purpose of clearly demonstrating the dynamic changes in parameters and model operations.

3.1.2 Calibration experiments

Experiment 1, as a control scheme, is employed to compare and investigate potential issues arising from time-invariant parameters. The global optimization procedure with time-constant parameters focuses on the overall performance of the hydrological model. However, it may average the hydrological responses and fail to ensure accurate simulation across all time intervals or flow phases. Fundamentally, the use of time-constant parameters may result in model simulations that compromise accuracy during high-flow phases while improving accuracy for low-flow phases. This implies a trade-off where the pursuit of enhanced accuracy in low-flow simulations comes at the expense of accuracy in high-flow simulations, and vice versa (Xiong et al., 2019a; Deng et al., 2016; Wang et al., 2017).

Experiment 2 investigates solving complex optimization problems with multi-objective optimization. The Non-dominated Sorting Genetic Algorithm II (NSGA-II) is selected for achieving global parameter optimization. In hydrological models, NSGA-II typically configures a pair of opposing objective functions, such as performance metrics for high and low flows (e.g., NSE and Log-NSE). However, the dynamic characteristics of the extracted catchment often exhibit greater complexity, which may not be fully captured by the objective functions used in NSGA-II. It is important to note that, despite the continuous advancements and refinements in Pareto-based multi-objective optimization algorithms, challenges such as heightened computational complexity, sensitivity to parameter settings, and potential slower convergence persist, especially with an increasing number of objective functions. Furthermore, while a diverse set of optimal parameter sets along the Pareto front improves low-flow simulation accuracy at the expense of high-flow simulation accuracy, it does not inherently address model structural deficiencies or enhance the overall performance of hydrological models (Kollat et al., 2012; Fowler et al., 2018; Wagener et al., 2003).

Experiment 3 is dedicated to investigating the impact of objective function configuration on model performance. The objective function is designed to encompass diverse flow phases ensuring simulation accuracy across various conditions. This is crucial for risk management of flood and drought events. To address this, a multi-weighted objective function is proposed, integrating both the RMSE and the FDC. RMSE is advantageous due to its sensitivity to prediction errors, while FDC serves as a hydrological tool graphically illustrating the proportion of time various flow rates are equalled or exceeded. The simultaneous incorporation of FDC and RMSE, specifically calculated the RMSE_Q95, RMSE_Q70, RMSE_Qmid, RMSE_Q20, and RMSE_Q5 based on no exceedance probability, addresses the simulation accuracy of extremely high, high, moderate, low, and extremely low flows, respectively (Pfannerstill et al., 2014). The weights of each part of the objective function are derived from the results of experiment 1, where the weights were assigned using a combination of AHP, PP, and CRITIC methods to evaluate the results of each part of

the objective function. For detailed information, please refer to Supporting Information S2.7. This approach allows decision-makers to assign weights to different flow phases based on practical considerations while ensuring an objective and scientific evaluation.

Experiment 4 is a sub-period calibration scheme specifically designed to investigate the impact of complex correlations among parameters. Parameter with highest sensitivity is enabled to change over time to bridge the gap between the extracted dynamic information from EDCC and the hydrological model. When considering the dynamization of individual parameters based on the extracted information, the correlations among parameters may lead to dynamic changes in one parameter being counteracted by adjustments in other fixed parameters, which is also called “compensation” (Wagener and Kollat, 2007). As a result, the extracted dynamic catchment characteristics may not be effectively reflected in the hydrological model by making one parameter dynamic, resulting in limited improvements to the model's accuracy. In this regard, this experiment designated the parameter with the highest sensitivity as the dynamic parameter.

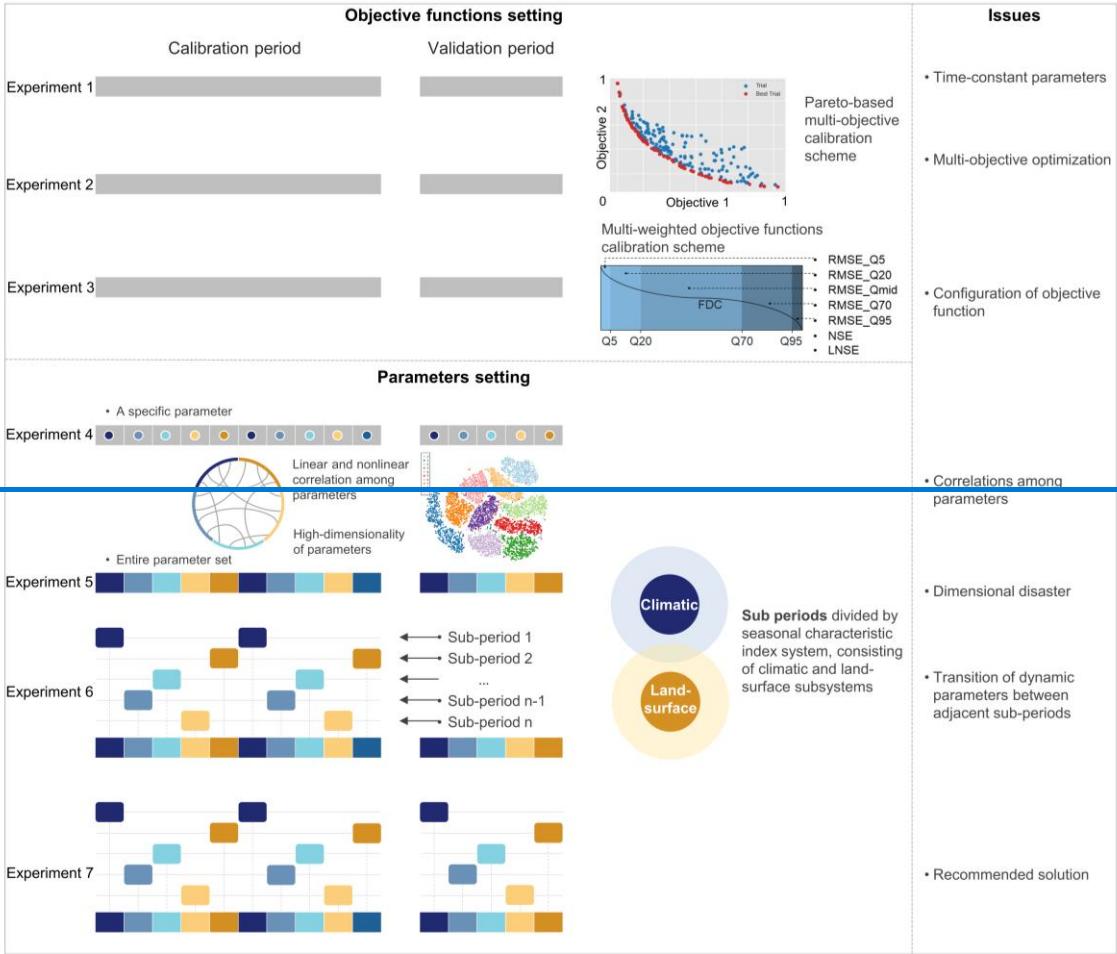
Variance Based Sensitivity Analysis (VBSA) is employed to perform global sensitivity analysis and to quantify the individual contributions of each parameter to the overall variability in the model output (Pianosi et al., 2015). During the calibration period, the dynamic parameter and other fixed parameters are simultaneously optimized. For example, in the HYMOD model with five parameters, if H_{ie} (maximum height of the soil moisture accounting tank) shows the highest sensitivity and there are five sub-periods, a total of nine parameters are optimized in each model run— H_{ie} and the four fixed parameters across the five sub-periods. The transition of state variables and fluxes between two consecutive sub-periods is achieved by inheriting the last values of the former period as the initial values of the next period. In the validation period, the model is run with the specified dynamic parameter and other fixed parameters. The transitions of parameters, state variables, and fluxes between two consecutive sub-periods are handled similarly to the calibration period.

Experiment 5 is dedicated to investigating the potential challenges arising from the dimensional disaster of parameters. The determination of the number of sub-periods is influenced by both the dynamic characteristics of the catchment and the unsupervised clustering of hydrological processes. As the number of sub-periods increases, the number of parameters in sub-period calibration grows multiplicatively. For example, in the study case area with n sub-periods (n is the number of the sub-periods), the number of parameters involved in the HYMOD model calibration is $5 \times n$. The simultaneous operation of high-dimensional parameters poses the risk of triggering the curse of dimensionality for hydrological model parameters, possibly resulting in the instability of the optimization algorithm and hydrological model. To address the issues associated with high-dimensional parameters in sub-period calibration, this study examines the simultaneous optimization with an increasing multiplicative number of parameters, with the transition of state variables and fluxes between adjacent sub-periods following the same principles as experiment 4.

Experiment 6 investigates the potential issues arising from the abrupt shift of dynamic parameters between sub-periods. During the calibration period, the model runs over the whole period, with the optimal performance in each sub-period serving as the objective function for each run. For instance, in a basin divided into five distinct sub-periods, the model runs n iterations (n is the number of the sub-periods), optimizing parameters for the specific objective function corresponding to each sub-period during each run. Subsequently, the simulated flow data from each sub-period are then merged and compared with the observed flow. In the validation period, the transitions of parameters, state variables, and fluxes between two consecutive sub-periods are handled the same as in experiment 5. The model runs one time using the dynamic parameter sets and the parameter set is switched between two consecutive sub-periods. As a result, the transition of the state variables and fluxes between two consecutive sub-periods is abrupt and achieved by inheriting the last values of the former period as the initial values of the next period. Experiment 6 is

designed with the purpose of mitigating the effects of the correlations and high dimensions of the parameters are excluded, and the influences caused by the parameter transitions are elucidated (Kim and Han, 2016).

Experiment 7 is a recommend solution aimed at addressing the aforementioned potential issues. In the calibration period, the model run follows the same procedure as that of the calibration period of experiment 6. In the validation period, the simulated flow data from each sub-period are merged and compared with the observed flow. Specifically, the model runs n times, combining the simulated flow data in the sub-periods.



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 conducted in the principal component space, allowing structural patterns in the catchment dynamics to be captured effectively. This data-driven clustering approach reveals the temporal heterogeneity of hydrological processes and provides a robust basis for integrating dynamic parameters into hydrological models.

In this study, 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 validation 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. 3).

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 was 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 validation. All other model parameters are fixed at their default values. Unless specified otherwise, model calibration is guided by the following objective function:

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

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 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 3), representing different flow regimes. Weights are derived from Experiment 1 using AHP, PP, and CRITIC methods (refer to Supporting Information S1.7).

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 separate values for each sub-period. This increases the number of parameters proportionally to the number of sub-periods, creating a high-dimensional calibration space. State and flux continuity between sub-periods follows the same inheritance mechanism as 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 introduces smooth parameter transitions during validation. This parallel calibration strategy aims to maintain continuity in parameter evolution while preserving water balance within each sub-period.

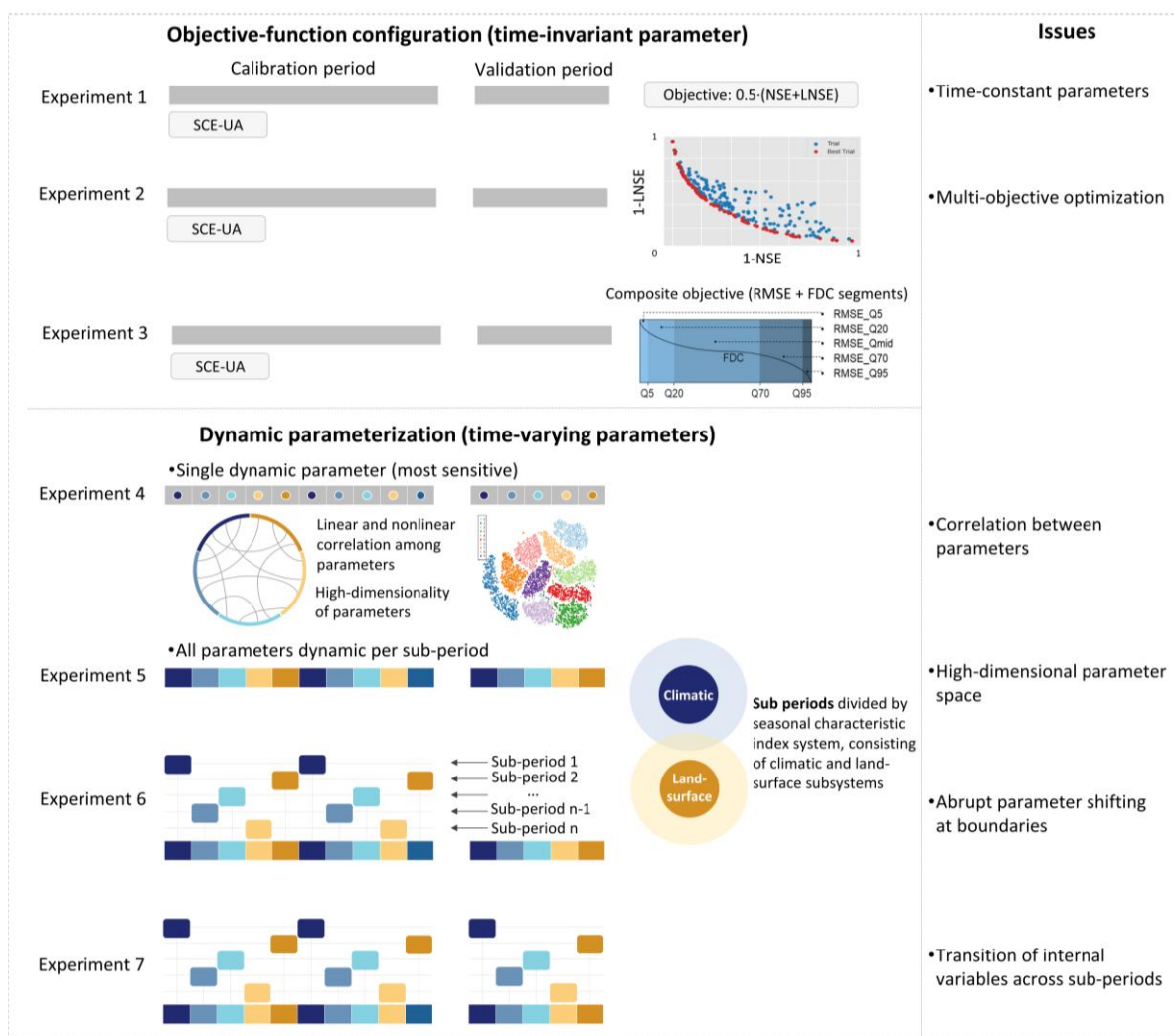


Figure 23. Schematic illustration of the seven calibration experiments. The colour bands represent state variables and fluxes, which are continuously transferred within the same period. In experiments 1, 2, and 3, the parameters are time-invariant, but the experiments differ in their objective function configurations. Conversely, in experiments 4, 5, and 6 maintain a consistent objective function, but vary the parameters across different experiments. In experiment 4, the dynamic of only the specific parameter is 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 validation 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 validation period, the simulated flow data from each separate sub-period are combined and compared with the observed flow.

3.2 Evaluation

3.4 Model evaluation

3.2.1 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; Van Werkhoven et al., 2009).

Werkhoven et al., 2009). However, a limitation exists with many common performance metrics: ~~they~~They only focus on overall or specific segments of the discharge series, neglecting other parts that may have the greatest practical impact. Hence, for diagnostic analysis, discharge segments of the flow duration curve (FDC) are used to identify discharge levels where model performance is poor (Pfannerstill et al., 2014; Laaha and Blöschl, 2006; Smakhtin, 2001). (Pfannerstill et al., 2014; Laaha and Blöschl, 2006; Smakhtin, 2001). In this study, performance across different dimensions of streamflow is assessed using the criteria defined in Table 4.3, which provides a comprehensive evaluation of model performance.

Table 4.3. Description of performance metrics.

Metric	Formula	Description
NSE	$NSE = 1 - \frac{\sum_{i=1}^n (Q_{obs,i} - Q_{sim,i})^2}{\sum_{i=1}^n (Q_{obs,i} - \bar{Q}_{obs})^2}$	Sensitive to peaks and discharge dynamiedynamics
LNSE	$LNSE = 1 - \frac{\sum_{i=1}^n (\log Q_{obs,i} - \log Q_{sim,i})^2}{\sum_{i=1}^n (\log Q_{obs,i} - \log \bar{Q}_{obs})^2}$	Emphasizing low flows with the log of discharge
RMSE_Q5	$RMSE_{Q5} = \sqrt{\frac{1}{n_{Q5}} \sum_{i \in I_{Q5}} (Q_{obs,i} - Q_{sim,i})^2}$	RMSE in FDC Q5 very-high-segment volume
RMSE_Q20	$RMSE_{Q20} = \sqrt{\frac{1}{n_{Q20}} \sum_{i \in I_{Q5 < Q < Q20}} (Q_{obs,i} - Q_{sim,i})^2}$	RMSE in FDC between Q5 and Q20 high-segment volume
RMSE_Qmid	$RMSE_{Qmid} = \sqrt{\frac{1}{n_{Qmid}} \sum_{i \in I_{Q20 < Q < Q70}} (Q_{obs,i} - Q_{sim,i})^2}$	RMSE in FDC between Q20 and Q70 mid-segment volume
RMSE_Q70	$RMSE_{Q70} = \sqrt{\frac{1}{n_{Q70}} \sum_{i \in I_{Q70 < Q < Q95}} (Q_{obs,i} - Q_{sim,i})^2}$	RMSE in FDC between Q70 and Q95 low-segment volume
RMSE_Q95	$RMSE_{Q95} = \sqrt{\frac{1}{n_{Q95}} \sum_{i \in I_{Q < Q95}} (Q_{obs,i} - Q_{sim,i})^2}$	RMSE in FDC Q95 very-low-segment volume
RMSE	$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Q_{obs,i} - Q_{sim,i})^2}$	RMSE sensitive to flood peaks
MSE	$MSE = \frac{1}{n} \sum_{i=1}^n (Q_{obs,i} - Q_{sim,i})^2$	MSE is sensitive to high flow
MSEL	$MSEL = \frac{1}{n} \sum_{i=1}^n (\log Q_{obs,i} - \log Q_{sim,i})^2$	MSEL is sensitive to low flow
MAE	$MAE = \frac{1}{n} \sum_{i=1}^n Q_{obs,i} - Q_{sim,i} $	MAE is measuring the overall discharge

Notes: Where $Q_{obs,i}$ and $Q_{sim,i}$ represent the observed and simulated streamflow at time step i , and \bar{Q}_{obs} is the mean observed flow. n is the total number of time steps. For log-transformed metrics (e.g., LNSE and MSEL), \log denotes the natural logarithm. Note that the flow duration curve (FDC) is usually split into different segments to describe different flow characteristics of a catchment (Gupta et al., 2009b; Cheng et al., 2012; Pfannerstill et al., 2014). (Gupta et al., 2009; Cheng et al., 2012; Pfannerstill et al., 2014). The RMSE with quadratic character is usually used to evaluate poor model performance due to the strong sensitivity to extreme positive and negative error values.

3.24.2 State variables and fluxes

The evaluation of state variables and fluxes links sub-period calibration and dynamic parameterization with 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 space, is

visualized in graphs and categorized by the operation of different sub-periods. This visualization helps illustrate issues with 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 were 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 up to filling the entire plausible 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 Defined sub-periods based on catchment dynamics

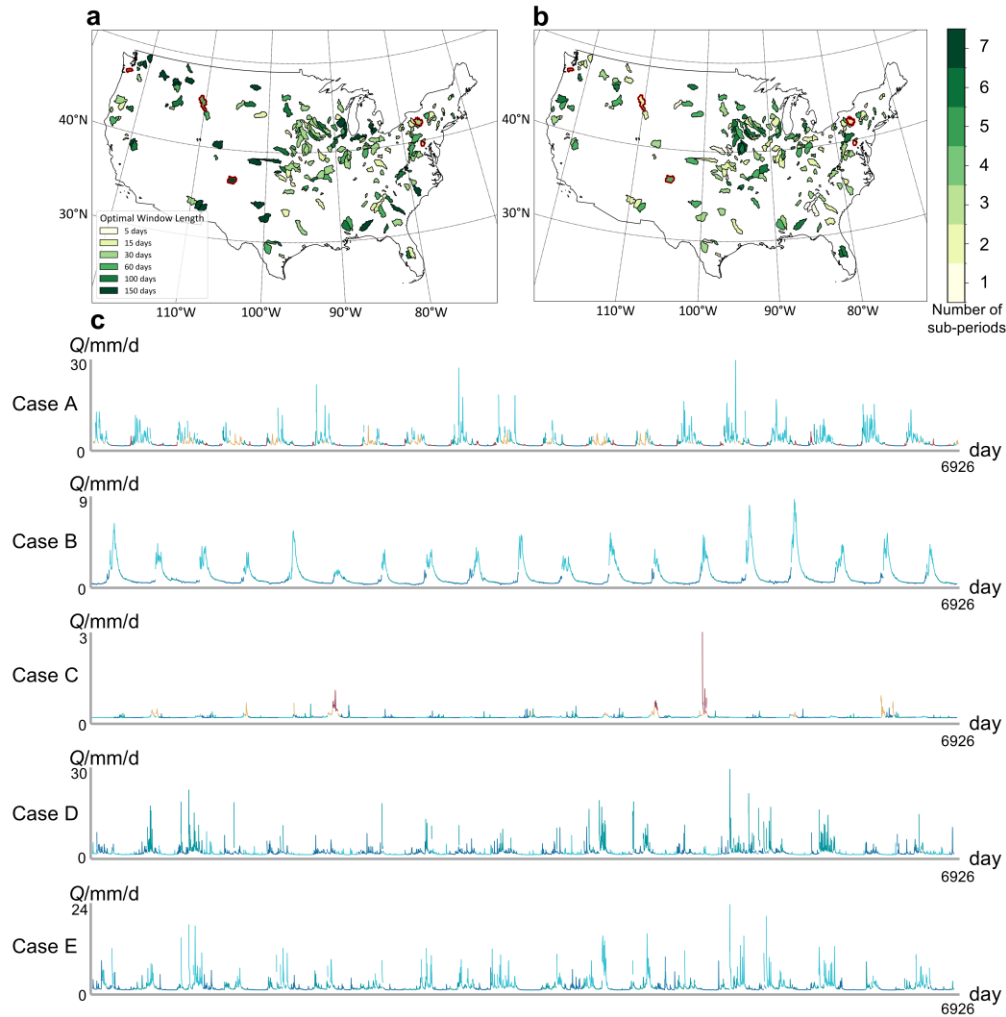
To support the implementation of sub-period calibration, periods were identified for all 219 catchments based on variations in dynamic catchment characteristics. The results indicate that dynamic catchment patterns are widespread across the study area, with 219 catchments exhibiting significant variation in at least one hydrometeorological variable (precipitation, temperature, potential evapotranspiration, NDVI, or runoff). Spatially, precipitation seasonality is more significant in the central and western regions; potential evapotranspiration seasonality is widespread, especially in northern areas; runoff seasonality is most evident in the central and northeastern regions; and vegetation seasonality is also common, with only a few high-latitude catchments lacking significant dynamic variation.

To extract relevant information and cluster the time series into distinct periods, a data-driven method was applied. First, the optimal sampling window for each catchment was determined using a Bayesian inference approach, with values ranging from 5 to 150 days (mean = 59.45 days). The Maximal Information Coefficient (MIC) was then used to filter out indicators with weak correlation to runoff. Principal Component Analysis (PCA) was applied to reduce dimensionality, and the first two components explained, on average, 83.5% of the total variance. Based on the reduced feature space, Fuzzy C-Means (FCM) clustering was used to group time steps, with an average of 4.2 periods identified per catchment.

To illustrate the applicability of this approach under diverse hydro-climatic conditions, five representative catchments were selected, covering a range of climate zones and dominant hydrological drivers. These catchments were also used in the subsequent modelling experiments. As shown in Fig. 4a and Fig. 4b, their optimal window lengths ranged from 30 to 150 days, with 12 to 31 indicators retained after screening. In all five cases, the number of identified periods ranged from 3 to 5. When compared with hydrographs, the identified periods aligned well with key hydrological processes, such as rising and recession limbs (Fig. 4c). 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 clusterings still captured major dynamic catchment characteristics. These period clusterings provide a physically interpretable structure that supports the dynamic parameterization and modelling experiments introduced in the following sections. Considering the performance of the seven modelling experiments

460

across both calibration and validation periods, Experiments 5 and 7 are considered the recommended experiments for capturing dynamic catchment characteristics. Experiment 5, with multi-parameter dynamic calibration, achieves high predictive accuracy across 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.



465

Figure 4. **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 respective study cases.

4.2 Model performance

470

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. 5, the NSE and LNSE values during both calibration and validation periods reveal differences in the ability of different 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: $OF = 0.27 \cdot RMSE_{Q5} + 0.16 \cdot RMSE_{Q20} + 0.08 \cdot RMSE_{Qmid} + 0.24 \cdot$

475 $RMSE_{Q70} + 0.25 \cdot RMSE_{Q95}$, 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
480 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. More significant improvements emerged with multiple dynamic parameters.
Experiment 5 achieved median NSE and LNSE values of approximately 0.7–0.8 in both calibration and validation 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
485 performed well during the calibration period; however, its abrupt parameter switching led to a significant decline of LNSE and
increased dispersion in the validation period. Experiment 7 addressed these shortcomings by applying a gradual parameter-
switching strategy during the validation period. As shown in Fig. 5, 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,
490 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.

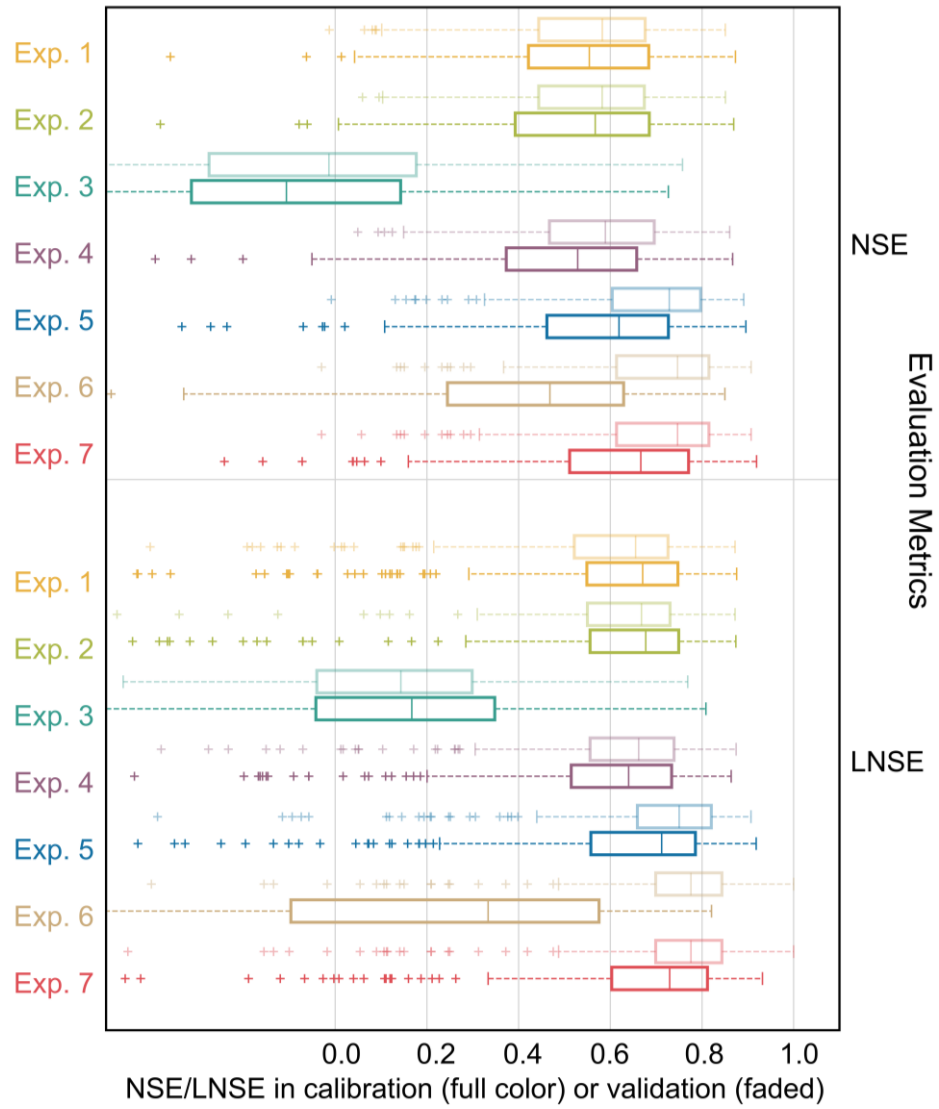
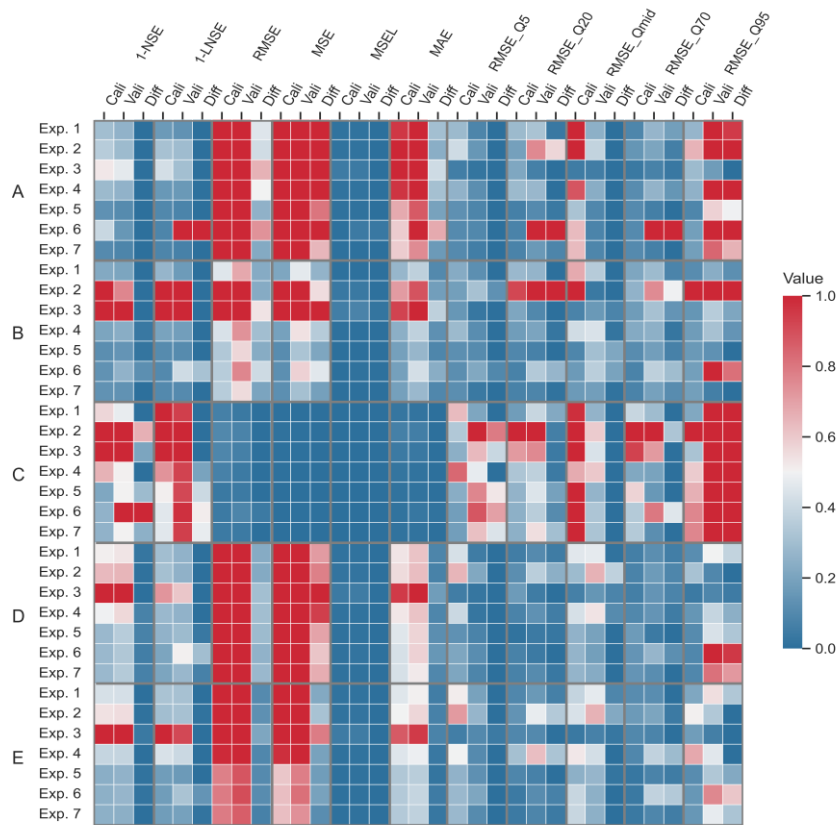


Figure 5. Performance of seven calibration experiments on the MOPEX dataset across 219 catchments. Boxplot color denotes different experiments. The whiskers extend a maximum of 1.5 times the interquartile range. Values beyond the whiskers are marked as outliers and are denoted as +.

To further examine how the different experiments under various hydrological conditions, a detailed assessment of five representative catchments is conducted with diverse dynamic patterns and baseline model performance. Fig. 6 presents the model performance of the seven experiments in four study cases. In case A, experimental study cases, Experiment 1 exhibited poor demonstrated low simulation accuracy and weak limited parameter transferability in diverse flow phases, particularly in extremely low and high flow phases. Experiment 2 demonstrates improved performance in its targeted flow phases. However, this advantage comes at the expense of sacrificing accuracy in other flow regimes. For instance, in experiment 2-1, when focusing on the NSE, case A experienced a decrease in NSE from 0.40 to 0.34 during the calibration period, and from 0.54 to 0.29 in the validation period. However, the LNSE increased from 0.61 to 2.02 during the calibration period, and from 0.44 to 1.27 in the validation period. When focusing on balancing NSE and LNSE, the parameter sets on the Pareto optimal frontier are similar in performance to those obtained in experiment 1, with NSE and LNSE during calibration being 0.48 and 0.64, respectively, and during validation being 0.48 and 0.44. The results in experiment 2 across different flow regimes, particularly under extremely low-flow and high-flow conditions. The results in Experiments 2 and 3 show limitations on both objective functions compared to the recommended scheme (experiment 7), which demonstrates superior results: NSE and LNSE during calibration are 0.19 and 0.28,

respectively, and Experiments 5 and 7. Adjusting the weights between NSE and LNSE improved accuracy for mid-phase flows but failed to account for other flow phases. For instance, in case A, considering NSE, the metric increased from 0.48 (Experiment 1) to 0.62 (Experiment 2) during the calibration period, and from 0.50 to 0.64 during the validation period. However, both RMSE_Q5 and RMSE_Q95 increased. Relative to Experiment 1, the RMSE_Q95 exhibited a significant deterioration during both the calibration and validation periods in Experiment 2. Despite prioritizing high and low flows through a weighted objective function ($OF = 0.27 \cdot RMSE_Q5 + 0.16 \cdot RMSE_Q20 + 0.08 \cdot RMSE_Qmid + 0.24 \cdot RMSE_Q70 + 0.25 \cdot RMSE_Q95$), experiment, Experiment 3 underperforms compared to experiment Experiment 1. While the objective function emphasizes these targeted phases, adjusting its weights unexpectedly failed to improve performance in the target flow phase and even worsened the model's performance in other evaluation metrics, indicating that this scheme exhibits instability in its performance across different flow phases. Experiment 4 demonstrates For instance, in case A, the NSE decreased from 0.48 to -0.74 in the calibration period, and from 0.50 to -0.27 in the validation period, compared with Experiment 1. In case C, the performance decline was more significant, with NSE values during both the calibration and validation periods approaching zero. Experiment 4 exhibited only marginal improvements compared to experiment over Experiment 1 across most metrics. Experiment 5 exhibits the poorest overall model performance during both calibration and validation periods. In some basins, experiment 5 led to operational errors, resulting in invalid results. Experiments 6 and 7, utilizing In contrast, Experiments 6 and 7, which employed the same calibration procedures, achieved the best strong overall performance during the calibration period, particularly in reproducing high flow flows and flood peak responses peaks. However, in during the validation period, experiment Experiment 6 displayed showed inconsistent performance, while excelling in certain metrics, aspects such as high-flow, but showed simulation, it experienced significant deterioration in others (NSE, LNSE, and RMSE_Q5). Similar to experiment 5, it may yield invalid values due to the failure of the hydrological model triggered by the calibration scheme e.g., NSE, MAE, and RMSE_Q95). The extent of performance decline in Experiment 6 varied among catchments: in case D, RMSE_Q95 increased by only 0.61 mm/d compared to the calibration period, whereas in case C, the deterioration was most severe, with RMSE_Q95 increasing by 17.64 mm/d. This significant decline can be attributed to extremely dry conditions, where runoff volumes approached zero (less than 0.01 mm), making small deviations translate into disproportionately large relative errors. Notably, experiment 7 across all study cases, Experiments 5 and 7 consistently maintained excellent performance in during the validation period, closely mirroring their calibration results and surpassing outperforming other experiments in nearly all metrics. Furthermore Moreover, analysis of parameter transferability revealed the smallest minimal differences between calibration and validation periods for experiment Experiments 5 and 7, while experiment 6 exhibited the largest discrepancies. Hence, experiment Experiments 5 and 7 demonstrates demonstrate the superior performer performance across all evaluation metrics, exhibiting improvements in simulations across various flow phases. In cases B, C, and D, the evaluation results are similar to those derived from case A.

To further expand the scope of testing to 130 basins in the MOPEX dataset, which exhibit significant seasonal dynamics (refer to Fig. 1), experiment 7 outperformed control experiment 1, demonstrating significant improvement in simulation performance. The evaluation metrics enhanced from 0.43 to 0.34 during the calibration period and from 0.52 to 0.44 during the validation period. In summary, experiments 2 and 3 exhibit performances comparable to experiment 1. Experiment 4 displays slightly superior results. Experiments 5 and 6 manifest poor overall performance, raising the possibility of generating erroneous outputs. Experiment 7 emerges as the best performer across all evaluation metrics for both calibration and validation periods, demonstrating consistent improvements in hydrological simulations across diverse flow phases.



4.2

Figure 6. The model performance of seven experiments in five study cases was assessed using multiple evaluation metrics. Lower values reflect superior performance.

4.3 State variables and fluxes

The state variables and fluxes reflect the internal operation of the hydrological model (definitions are provided in Table S1 in the Supporting Information). The assessment results of state variables and fluxes through seven calibration experiments for case study A are illustrated in Fig. 37 and Fig. 48 (results of cases B, C, D, and E are shown in S5S3 of Supporting Information). 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 brevity. A slight improvement is shown in Experiment 4, contrary to expectations, did not improve compared with the model's performance-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. Notably, the state variable X_q and flux Q_q in experiment 4, the display is abnormally flat compared to the control experiment 1, indicating a wrong response of the rapid runoff module to input variations. Experiment 5 exhibits invalid values in both state variables and flux time series, due to abrupt changes at the transitions between sub-periods. Experiment 6 in the validation period exhibits similarities to experiment 5.

Despite these setbacks, experiment 7 introduced significant improvements, marking a turning point in the simulation results. In experiment 7, flux responses differ from those in experiment 1 as follows: (1) Sub-periods 1 and 2 in case A, identified by the chosen seasonal characteristics, are the coldest with the least runoff. Although significant one-day maximum precipitation events ($RX1day$) occur, the quick runoff component exhibits minimal response during sub-periods 1 and 2. In contrast, the slow runoff process, Q_s , appears more influential, likely attributed to freeze-thaw cycles and snowmelt. For the subsequent sub-periods, Q_s , Q_q , and overland flow OV all react to precipitation events. While Q_q shows a slight decline, and the contribution of saturated runoff

OV increases. These changes improve the shortcomings of underestimating low flows and overestimating high flows in the runoff simulation process (Guo et al., 2018; Höge et al., 2018; Pande and Moayeri, 2018; Wang et al., 2018). (2) Experiment 7 provides a more accurate simulation for diverse flow phases than experiment 1. (3) Anomalies in state variables occur in experiment 7 during sub-period 5, likely due to H_{uz} constraints ($H_{uz} = 43$), maintaining XH_{uz} and XC_{uz} at low levels. This may arise from the optimization algorithm failing to converge, resulting in a local optimum 'trap' during this sub-period. However, experiment 7's simulation of flow in sub-period 5 is better than that of experiment 1, further demonstrating the robustness of experiment 7 and the reliability of dynamic parameter sets compared to an individual dynamic parameter. Similar results were observed in cases B, C, and D (see Fig. S5-S12). In sum, experiments 1 and 3 show negligible differences in state variables and flux time series. Experiment 4 fails to improve simulation results and triggers abnormal responses within certain model components. Experiment 5 exhibits invalid values, and experiment 6 encounters similar issues at sub-period boundaries during the validation period. Experiment 7 shows improved simulation performance, particularly in addressing the underestimation of high flows and overestimation of low flows, as evidenced by the model's internal variables.

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 subperiods. This 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. 7 and Fig. 8, 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. Overall, Experiments 1, 2, and 3 exhibit negligible differences in state variables and flux series, although Experiment 3 produces a decline in low-flow accuracy. Experiment 4 shows marginal improvements compared with time-invariant parameterization; however, it indicates that a single dynamic parameter is insufficient to capture overall dynamic catchment characteristics. Experiment 6 applies abrupt parameter switching across sub-periods, which disrupts water continuity. In contrast, Experiments 5 and 7 display significant improvements in simulation performance, particularly by mitigating the underestimation of high flows and the overestimation of low flows, as evidenced by the behaviour of internal model variables.

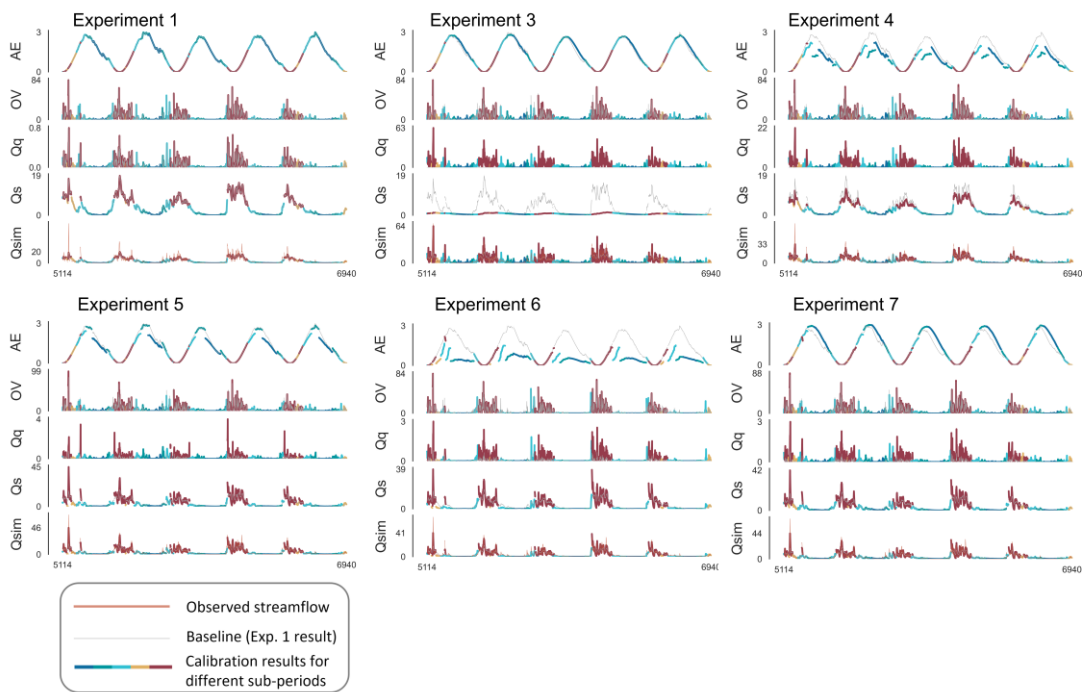


Figure 7. Fluxes simulation results of experiments during the representative validation period for case A. The figure shows the flux simulation results from Experiments 1 to 7, with different colours representing different sub-periods.

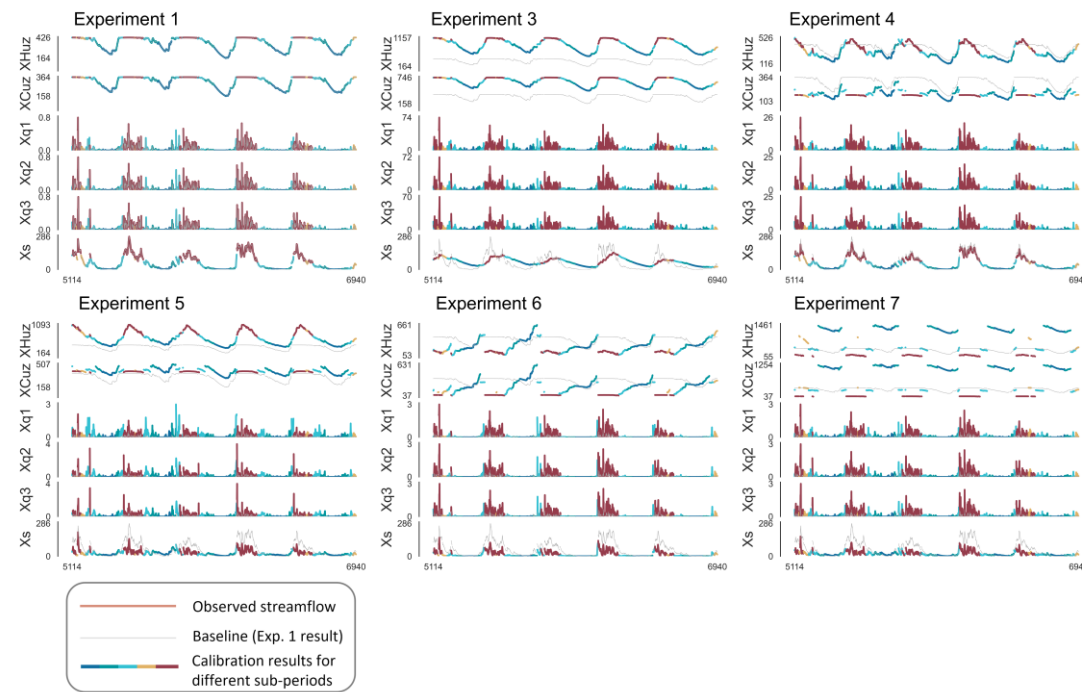


Figure 8. State variables simulation results of experiments during the representative validation period for case A. The figure shows the state variable simulation results from Experiments 1 to 7, with different colours representing different sub-periods.

The comparative analysis of ~~experiment~~Experiments 1, 5, and ~~experiment~~7 further illustrates the performance improvements-~~in performance~~ introduced by ~~experiment~~Experiments 5 and 7. Fig. 5(e) and Fig. 5(d) ~~illustrate~~7 illustrates the flux mapping of various sub-periods in the study case A, comparing Experiments 1, 5, and 7. Flux-mapping figures for the other study cases are detailed in the ~~control experiment 1 and the recommended scheme (experiment 7)~~Supporting Information (Fig. S13-S16). Each scatter point in the figures represents a parameter set generated during the SCE-UA algorithm optimization process. The colour

and relative position of each scatter point on the axes illustrate the variation in runoff components for sub-periods under specific parameter sets, as well as the corresponding objective function value. To facilitate comparison, the results of ~~experiment~~Experiment 1 are also presented by the same sub-periods as ~~experiment~~Experiments 5 and 7.

Notably, the differences in optimization performance between ~~experiments~~Experiments 1 and 7 reveal key insights into model behaviour. ~~Both experiment~~Across all study cases, both Experiment 1 and ~~experiment~~Experiment 7 show the poorest results in sub-periods 1 and 2, with the largest (worst) objective function values. In the remaining three sub-periods, the objective function values ~~were~~are significantly better. Compared to ~~the traditional scheme, the recommended scheme~~Experiment 1, Experiment 7 consistently identified more optimal parameter sets with smaller objective function values within the same period. ~~For example, in Fig. 9b (Experiment 7), most of the dark blue scatter points for sub-period 5 cluster around a vertical axis value of approximately 0.25, whereas in Experiment 1, scatter points for the same sub-period are more widely distributed near 0.5.~~ Shifting the focus to flux components, the spatial distribution of scatter points in the flux maps reveals varied runoff components and internal model behaviour for each sub-period. ~~Notably~~In Experiment 7, clusters of scatter points of the same color appear more compact, while in ~~the traditional scheme, they are more dispersed along both vertical and horizontal axes. This pattern indicates that,~~ despite similar objective function values, ~~the recommended scheme~~Experiment 7 possesses a narrower range of optimal equifinality parameters during the parameter evolution process, reducing the model's internal fluxes equifinality and uncertainty. ~~From sub-periods 1 to 4~~Furthermore, Fig. 9b shows that in Experiment 7, the color bars along the vertical axis are shorter and more evenly distributed, demonstrating that from sub-periods 1 to 5, the SCE-UA algorithm more rapidly converges to near-optimal solutions, showing a narrower range of variability in the optimization process. ~~In sum, the improvements observed in experiment 7 not only highlight the importance of refining dynamic parameters but also underscore the model's capacity to simulate complex hydrological processes across different sub-periods.~~

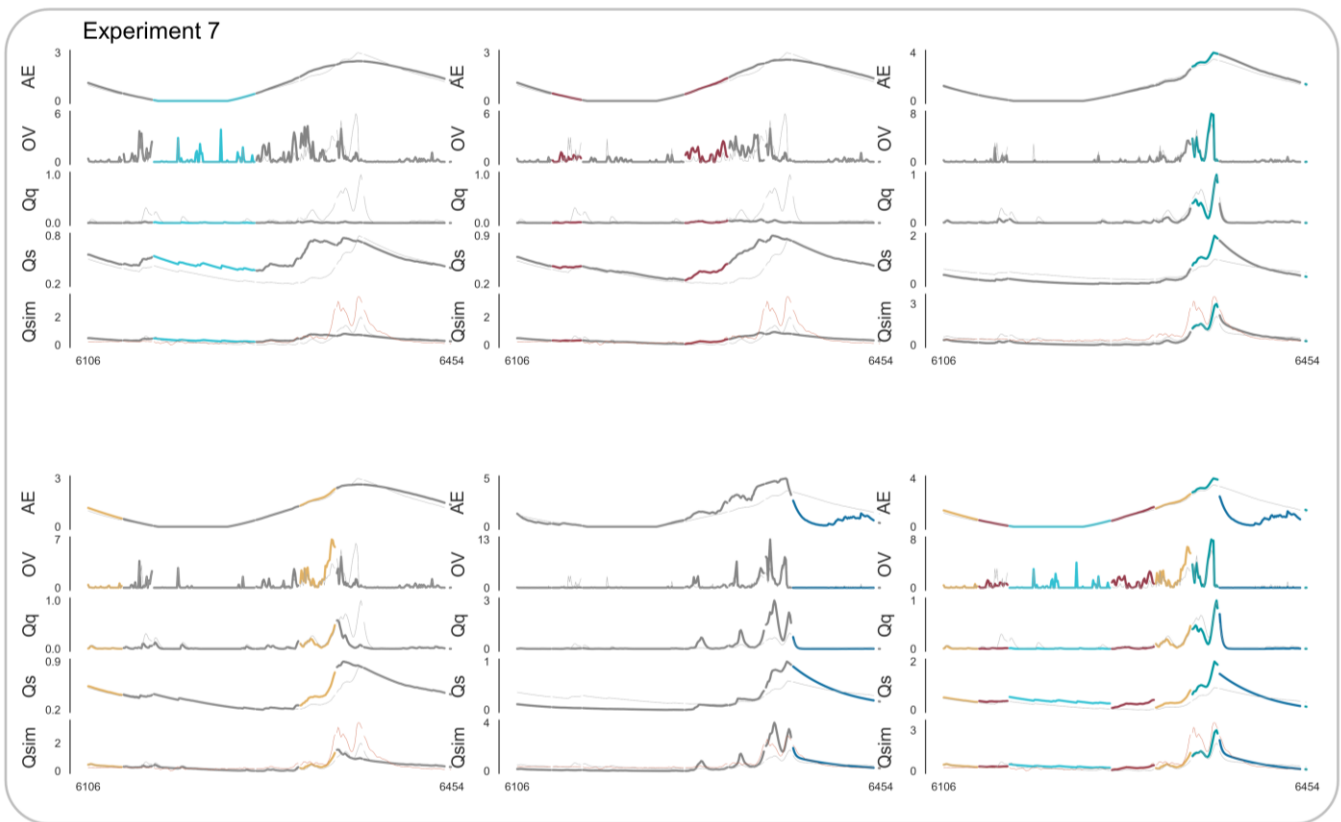
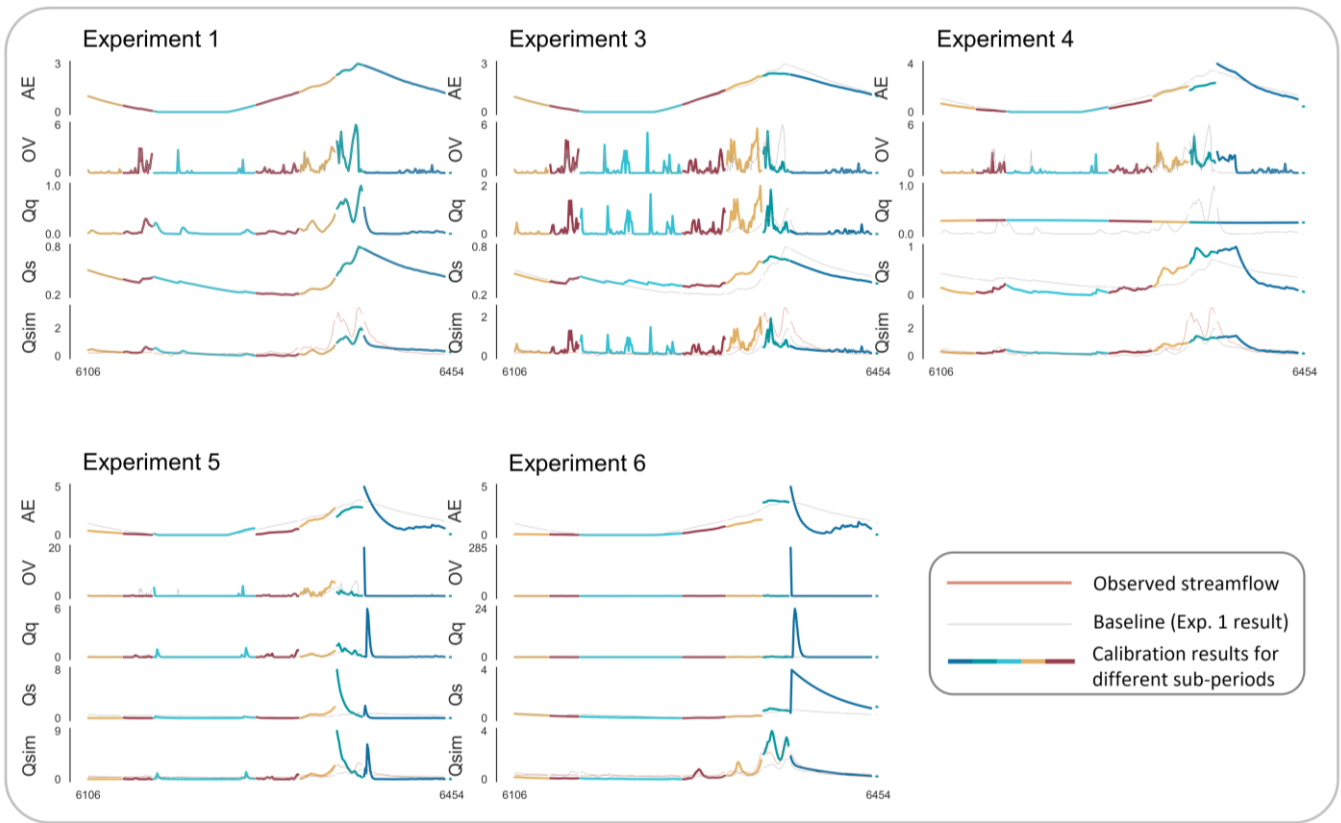


Figure 3. Fluxes simulation results of experiments during the representative validation period for the case A. The figure shows the flux simulation results from Experiments 1 to 7, with different colours representing different sub periods. In Experiment 7, five separate calibrations were performed for five sub-periods, and the results were then aggregated to obtain the final simulation.

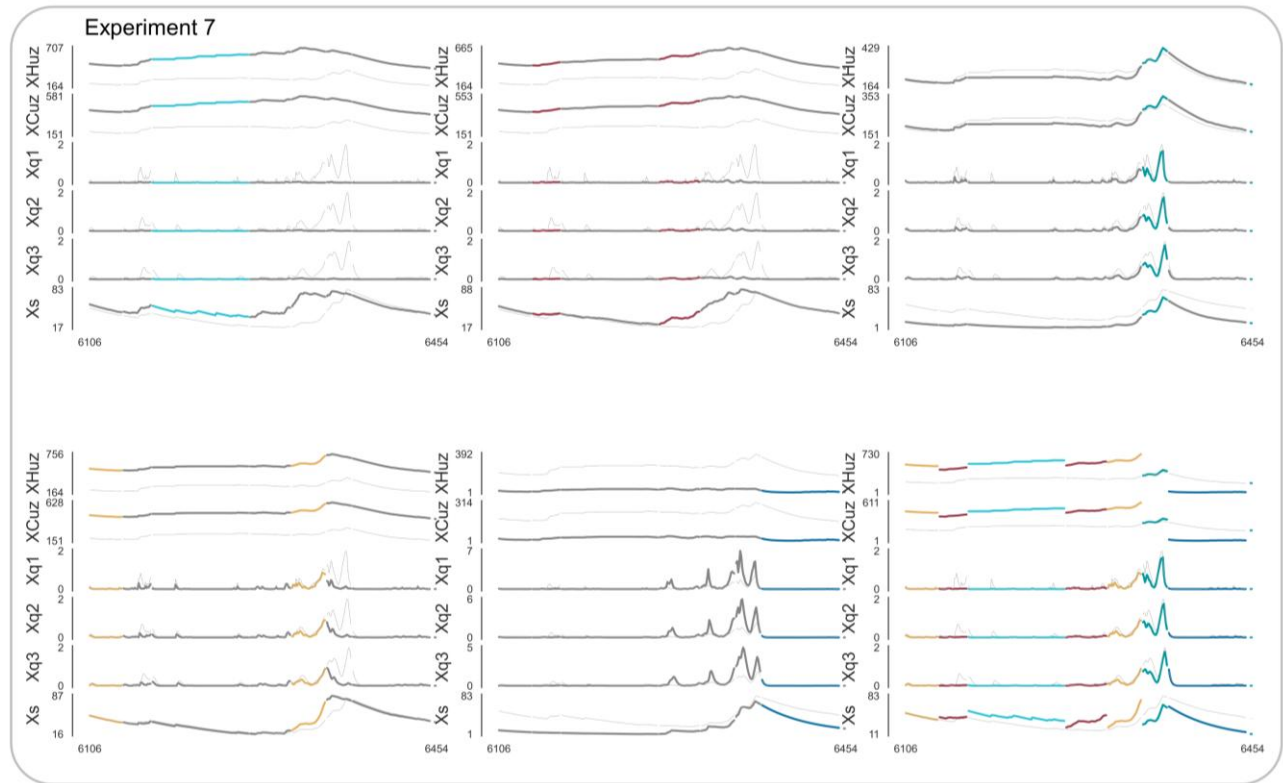
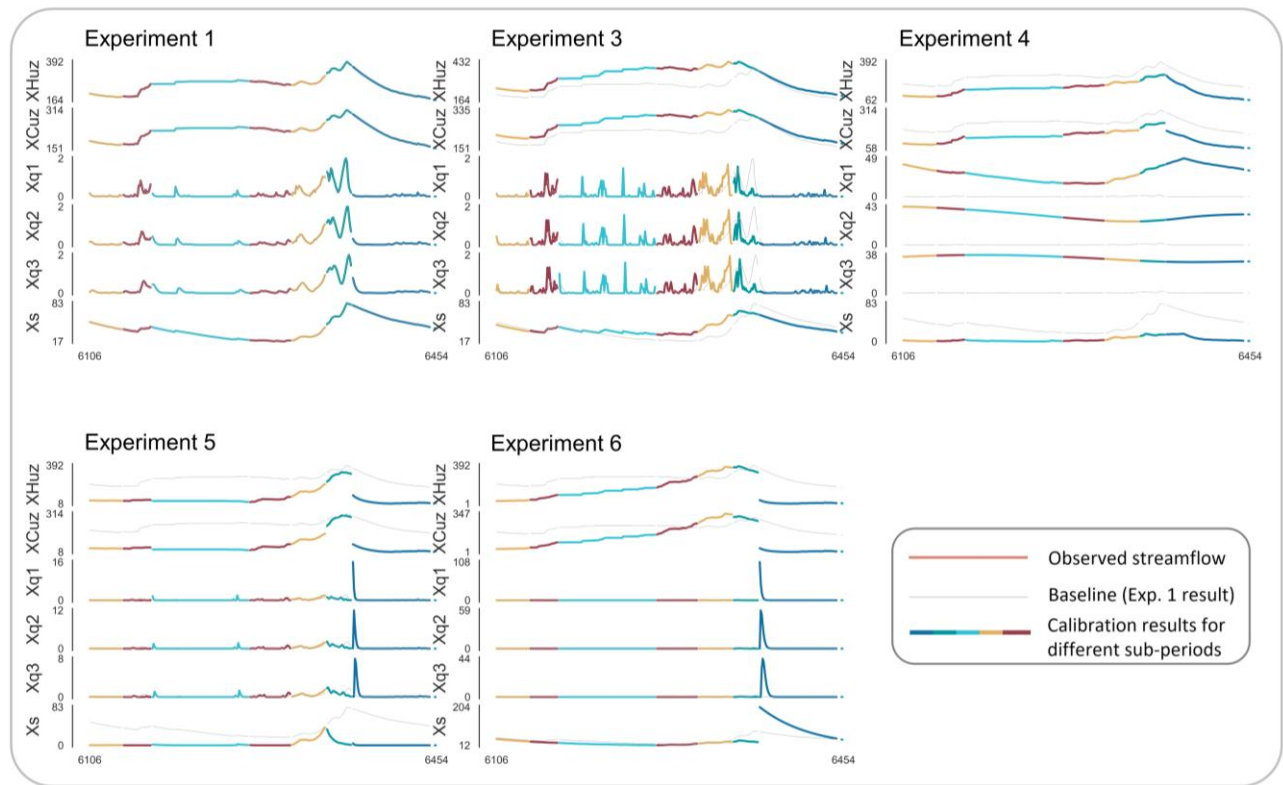


Figure 4. State variables simulation results of experiments during the representative validation period for the case A. The figure shows the state variable simulation results from Experiments 1 to 7, with different colours representing different sub-periods. In Experiment 7, five separate calibrations were performed for five sub-periods, and the results were then aggregated to obtain the final simulation.

Further comparison with Experiment 5 (Fig. 9c) shows that parameter sets within each sub-period were tightly clustered in the vertical direction, indicating consistently high performance within individual sub-periods. However, these clusters were widely dispersed along the horizontal axis. For instance, the cluster for Sub-period 2 (dark red) is concentrated at higher Q_s values (approximately 0.9), whereas the cluster for Sub-period 4 (orange) is concentrated at much lower values (approximately 0.5). Such horizontal separation suggests that different runoff generation mechanisms (fluxes) are adopted across sub-periods to achieve high performance, which may compromise the physical consistency of the overall simulated discharge (Q_{sim}). This inconsistency is particularly evident in case E, where runoff generation mechanisms across sub-periods appeared nearly independent, while the separation is less significant in case B. In contrast, scatter clusters in Experiment 7 (Fig. 9b) are more tightly aligned along the horizontal axis, indicating the adoption of more consistent and physically reasonable runoff strategies across sub-periods. Nevertheless, Experiment 7 poses a potential risk of discontinuities in internal state variables at sub-period boundaries, a phenomenon that was particularly evident in case D (Fig. S15). In summary, the improvements observed in Experiments 5 and 7 underscore both the importance of refining dynamic parameters and the model's ability to simulate complex hydrological processes across sub-periods. However, Experiment 5 may compromise physical consistency in runoff generation processes, while Experiment 7 faces the challenge of ensuring smooth transitions of state variables across boundaries.

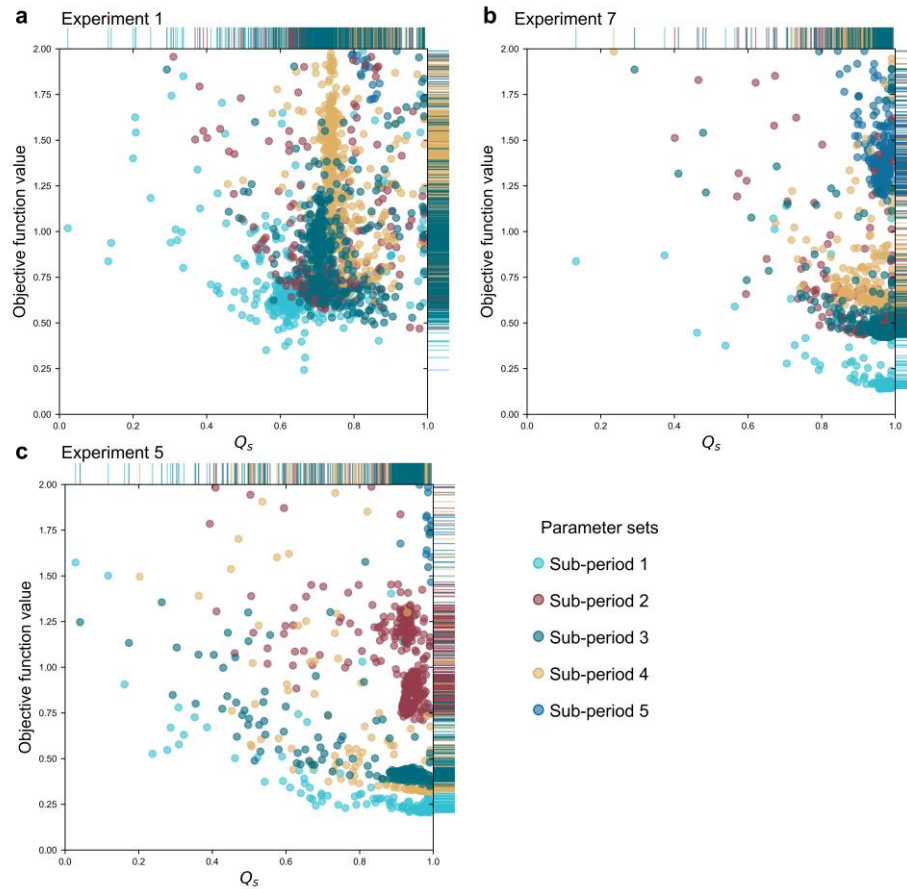


Figure 9. a, Flux mapping for case A in the conventional scheme, b, Experiment 7, and c, Experiment 5, where the horizontal axis represents the proportion of Q_s in the runoff.

4.3.4 Parameters

The dynamic parameter sets, optimized by various calibration experiments across ~~four~~five case studies, are depicted in Fig. ~~5(b)~~10. Experiments 1, 2, and 3 utilized a time-invariant parameter set, adjusted through the objective function to reflect the catchment's average characteristics. Experiment 4 allowed the parameter H_{uz} , which exhibits the highest sensitivity, to vary across different sub-periods while maintaining other parameters constant. However, the dynamics of H_{uz} in response to catchment characteristic changes across sub-periods did not significantly improve the model's performance within the ~~four~~five case studies. In Experiment 5 ~~allowed for~~, all parameters ~~to~~ vary across sub-periods, ~~but~~as illustrated in Fig. 10. The greater colour variation of parameters compared with Experiment 4 indicates a stronger response to catchment dynamics; however, no consistent variation pattern emerged in response to catchment characteristics. In ~~experiments~~Experiments 6 and ~~the recommended experiment~~ 7, certain parameters, such as K_s , exhibited minimal correlation with sub-period characteristics within catchments. As indicated in Fig. 10, the colour variation of bubbles in the K_s column is limited. In some cases, however, deeper bubble colours appear during sub-periods with concentrated precipitation or higher antecedent soil moisture, indicating the K_s value is highest during sub-periods with abundant and concentrated precipitation, higher temperatures, and higher antecedent runoff (soil moisture), and lowest during relatively cold and dry sub-~~period~~periods. However, this correlation ~~was~~is not significant. Furthermore, K_q ~~did~~does not exhibit a clear pattern of variation, due to the poor response of α , which indirectly changed the model structure and bypassed the quick flow module. This phenomenon reflects the uncertainty inherent in model parameters and structure. In sum, compared to time-invariant schemes, ~~the recommended scheme exhibits~~Experiments 5 and 7 exhibit superior performance in identifying key parameters and their responses to catchment dynamics. The dynamic ~~nature~~characteristic of parameters emphasizes the importance of calibration across sub-periods. Although the dynamic parameter set enhanced the model's response capability to ~~seasonal~~catchment dynamics, the overall response of the entire dynamic parameter set to ~~seasonal dynamic characteristics~~catchment dynamics remains relatively poor. The reasons for the improved simulation performance of the dynamic parameter set will be explored in the discussion section.

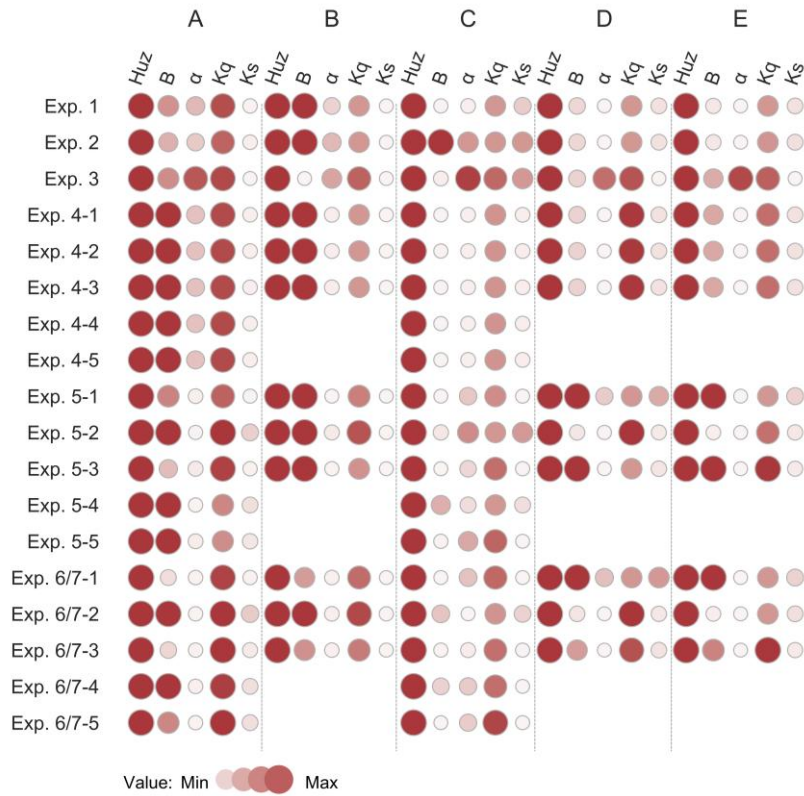


Figure 5. (a). Model performance of the seven experiments in four study cases. (b). 10. Assessment of dynamic parameter sets across various calibration experiments. (c)/(d). Flux mapping for case A in both the conventional (e) The parameter boundaries shown in the figure are H_{uz} (0-1500), B (0-2), α (0-1), K_q (0.5-1), and recommended (d) schemes, where the horizontal axis represents the proportion of Q_s in the runoff. K_s (0-0.5).

5 Discussion

5.1 Why dynamic parameter sets improve simulation performance

Despite the significant improvement in the simulation performance of hydrological models based on [seasonal-dynamic characteristicscatchment dynamics](#), the response of discretized dynamic parameters (even highly sensitive ones) to these [seasonalcatchment](#) dynamics is not satisfactory. However, a dynamic parameter set can collectively carry the extracted information of [seasonal-dynamic featurescatchment 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. 6(a)11a and Fig. 6(e)11c 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). MIC values above 0.35 among most parameters suggest that the dynamics of individual parameters may be affected by others ([Bárdossy, 2007](#)). This explains [unimproved model performance when altering individual parameters during different sub-periods in experiment 4](#). The analysis results of parameter sensitivity based on ([Bárdossy, 2007](#)). 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 seasonal dynamic of the 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., 2019a; Vaze et al., 2010; Zhou et al., 2022). Bárdossy (2007) (Xiong et al., 2019; Vaze et al., 2010; 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. 5(e) and Fig. 5(d), 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; Gupta et al., 2009a; Santos et al., 2018). A rigid interpretation of objective functions can lead to misinterpretations; for instance, in Fig. 6(b), 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 (Clark et al., 2015).

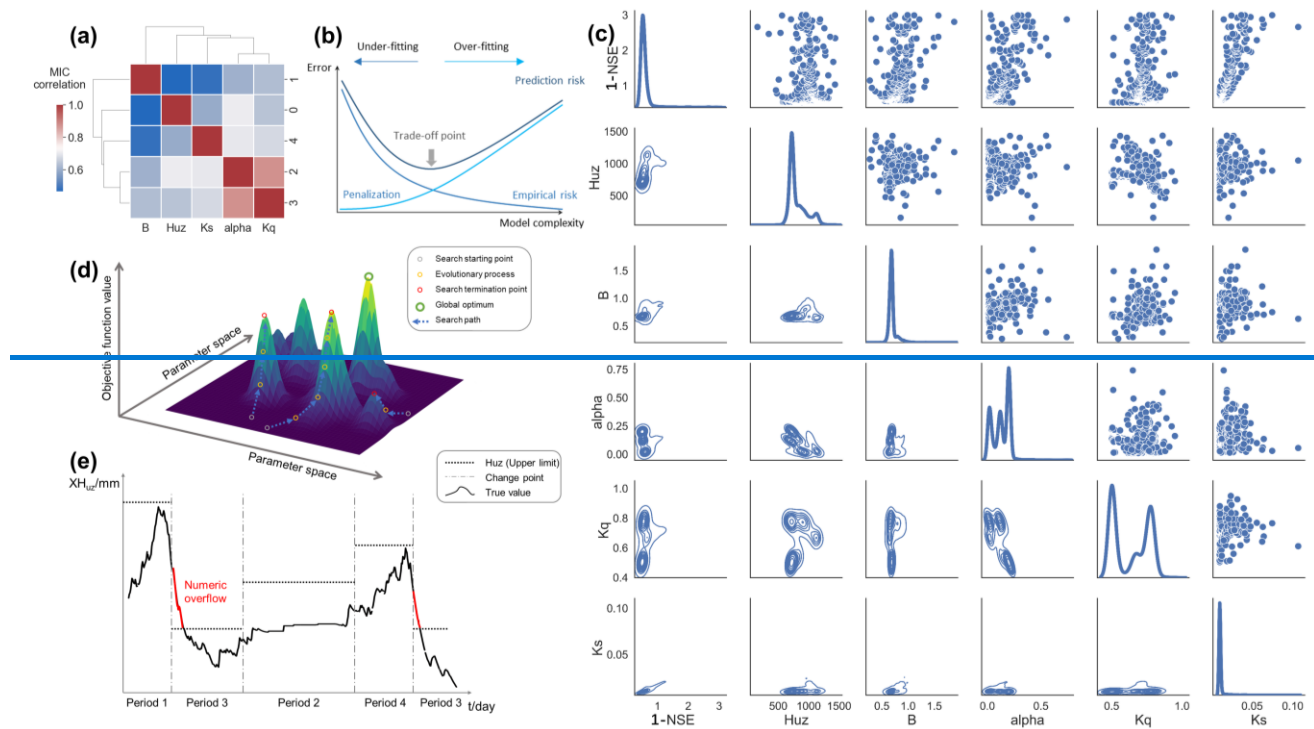
5.1.3 Evolution process of parameters

While the causes of abnormal 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 (Weise, 2009; Maier et al., 2014; Kallel et al., 1998). 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. 6(d)). 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 (Weise,

[2009](#)); [\(Weise, 2009\)](#). 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. [\(Kim and Han, 2016; Beven and Binley, 2006; Sivakumar, 2004; Laloy and Vrugt, 2012\)](#); [\(Kim and Han, 2016; Beven and Binley, 2006; Sivakumar, 2004; Laloy and Vrugt, 2012\)](#). 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. [6\(e\)](#)–[11c](#)). 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 [causingmaking](#) the optimization algorithm incapable of producing valid results.



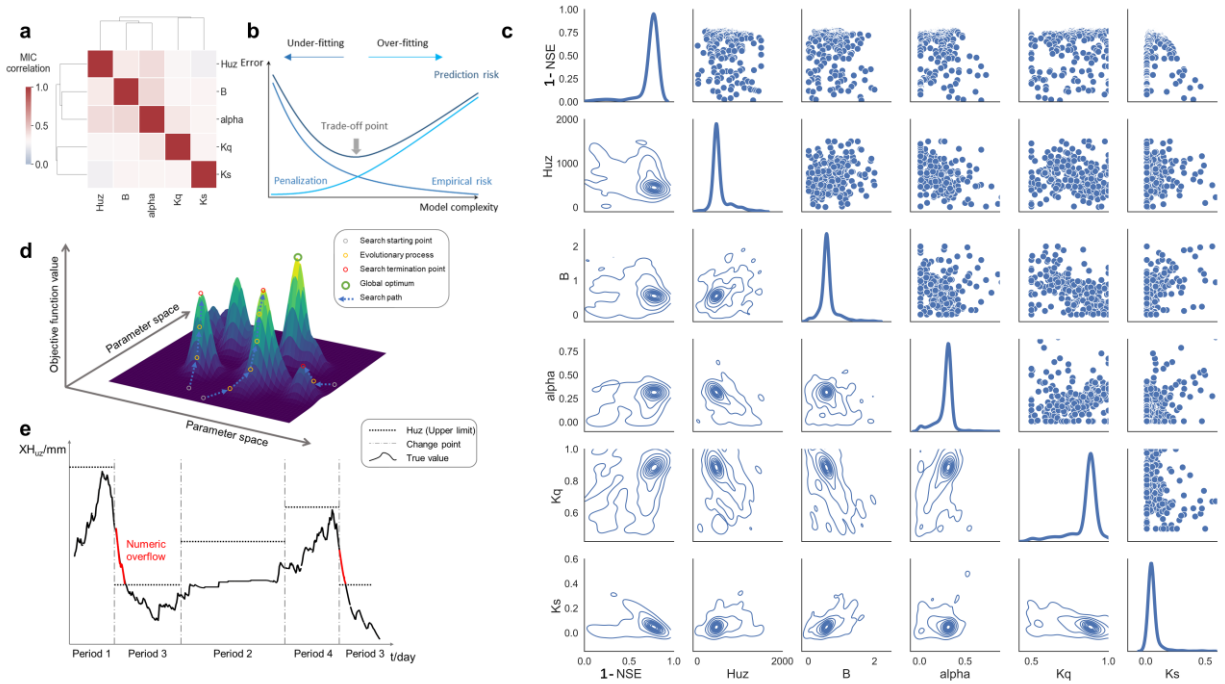


Figure 6-(11. a) Linear or nonlinear correlations between parameters based on MICs in case A, with red indicating the strongest correlation among parameters. **(b)** Conceptual diagram illustrating the trade-off between empirical fitting to data and the penalization of model complexity, and its impact 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 3.3) is employed to extract the dynamic catchment characteristics of hydrological processes, enabling model parameters to adjust across hydrological periods. This approach improved simulation accuracy and robustness in dynamic catchments, demonstrating the necessity and effectiveness of incorporating dynamic parameters into conceptual hydrological models. However, a critical question arises: To what extent do dynamic parameter variations represent the true dynamic variability of catchment properties, and to what extent do they compensate for structural deficiencies of the model itself? 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 (five in the HYMOD model). 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. 12a, imposing the constraint leads to posterior distributions that are more concentrated within each sub-period, with reduced dispersion, reflecting greater stability. Parameter transferability between calibration and validation periods also improved, as illustrated in Fig. 12b, with smaller declines in model performance across periods. However, these gains in parameter stability are accompanied by significant reductions in NSE and LNSE, rendering performance inferior to unconstrained sub-period calibration. This trade-off highlights the compensatory role of dynamic parameters in addressing structural limitations of fixed model formulations. When the capacity of parameters to compensate is constrained, the observed performance decline reflects underlying structural inadequacies in representing key hydrological processes.

The demand for dynamic parameters is often symptomatic of structural insufficiency. A structurally adequate model should maintain stable parameters that represent physical catchment properties. When the model formulation fails to capture essential processes, the “optimal” parameters must vary dynamically to compensate for these omissions. Evidence from the GLUE framework has shown that posterior parameter distributions can diverge almost completely between wet and dry seasons, implying that sub-period calibration with distinct parameter sets effectively corrects structural errors and improves accuracy (Blasone et al., 2008). Moreover, concepts from Data-Based Mechanistic (DBM) modeling and state-dependent parameter (SDP) approaches suggest that time- or state-dependent gains—such as nonlinear filters linked to soil moisture or runoff—can be identified from data. These gains compensate for missing nonlinearities in effective rainfall, often exhibiting dynamic catchment patterns over longer periods (McIntyre et al., 2011). On this basis, the proposed sub-period calibration framework is positioned as a practical means of using parameters as proxy variables to alleviate structural deficiencies, thereby enhancing streamflow simulation accuracy in dynamic catchments.

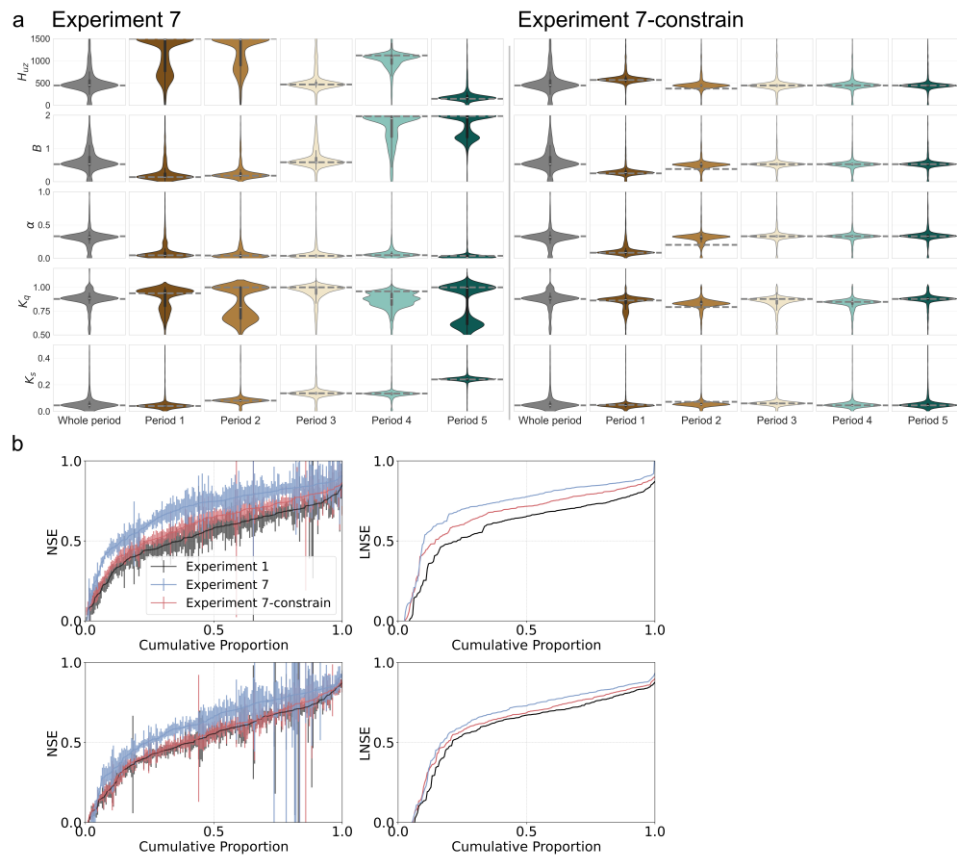


Figure 12. a. Distributions of the optimal parameter spaces across sub-periods under different climatic and land-surface conditions for case A. Each violin depicts one parameter space, with parameter values on the y-axis; the violin width reflects the probability density of the parameter

values. Parameter bounds are: $H_{0\pm}$ (0-1500), B (0-2), α (0-1), K_g (0.5-1), and K_s (0-0.5). Results for all study cases are provided in S5 of the Supporting Information. **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 a lack of an incomplete understanding of catchment hydrological processes, the structure of traditional conceptual hydrological models often fails to represent catchment dynamics, leading to inaccurate simulation of different flow regimes. To address model simulations, here we aimed to improve simulation in dynamic catchment characteristics, it is essential to develop a methodology to effectively calibrate the re-examine the time-varying information in historical hydrological models, addressing meteorological data and consider the variation 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 imposed by model structure. The Extracting Seasonal Dynamic Catchment Characteristics (EDCC) approach was taken to cluster the dynamic processes of the catchment into diverse sub-periods, linking hydro-meteorological data period calibration with models through objective function and dynamic parameters. The identified clusters were calibrated using seven contrasting calibration. Seven experiments applied to the MOPEX dataset and 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 from the perspectives of using multiple metrics and internal model processes. A recommended calibration scheme was ultimately proposed 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.
- Based on the EDCC approach, which clusters sub-periods according to catchment dynamics, the recommended scheme achieved better simulation results. Sub-period calibration enables conceptual hydrological models to overcome structural deficiencies and simulate the dynamic processes of the catchment.
- 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 [are](#) provided.

Author contributions

Tian Lan devised the modelling concept. Tian Lan and Xiao Wang wrote the code, and prepared the original draft manuscript. Hongbo Zhang, Xinghui Gong, Xue Xie, Yongqin David Chen, and Chong-Yu Xu provided supervision, and reviewed/edited the manuscript. [Jiajia Zhang and Wenqing Cheng contributed to supplementary calculations and textual revisions during the manuscript revision.](#)

Code availability

~~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://zenodo.org/records/11058234> (last access: 23 November 2024).~~

~~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.

Acknowledgments

~~This study is financially supported by the the National Natural Science Foundation of China (NSFC) (Grant No. 52209006, 52379003, and 42071055), China Postdoctoral Science Foundation (Grant No. 2021M700018), and Research Council of Norway FRINATEK Project 274340.~~

~~Data Availability Statement: Data and code reported in this paper can be downloaded from Lan (2025) at <https://doi.org/10.5281/zenodo.16676391>. The MOPEX dataset is available from Duan et al. (2006). The Sensitivity Analysis For Everyone (SAFE) toolbox is available at <https://safetoolbox.github.io/> (Pianosi et al., 2015).~~

References

~~Bárdossy, A.: Calibration of hydrological model parameters for ungauged catchments, Hydrology and Earth System Sciences, 11, 703–710, 10.5194/hess-11-703-2007, 2007.~~

855 Bárdossy, A. and Singh, S. K.: Robust estimation of hydrological model parameters, *Hydrol Earth Syst Sc*, 12, 1273–1283, 10.5194/hess-12-1273-2008, 2008.

Beven, K.: Prophecy, reality and uncertainty in distributed hydrological modelling, *Adv Water Resour*, 16, 41–51, 10.1016/0309-1708(93)90028-e, 1993.

Beven, K. and Binley, A.: The future of distributed models: Model calibration and uncertainty prediction, *Hydrological Processes*, 6, 279–298, 10.1002/hyp.3360060305, 2006.

860 Bronstert, A.: Rainfall-runoff modelling for assessing impacts of climate and land-use change, *Hydrological Processes*, 18, 567–570, 10.1002/hyp.5500, 2004.

Cheng, L., Yaeger, M., Viglione, A., Coopersmith, E., Ye, S., and Sivapalan, M.: Exploring the physical controls of regional patterns of flow duration curves – Part 1: Insights from statistical analyses, *Hydrol. Earth Syst. Sci.*, 16, 4435–4446, doi.org/10.5194/hess-16-4435-2012, 2012.

865 Choi, H. T. and Beven, K.: Multi-period and multi-criteria model conditioning to reduce prediction uncertainty in an application of TOPMODEL within the GLUE framework, *J Hydrol*, 332, 316–336, 10.1016/j.jhydrol.2006.07.012, 2007.

Clark, M. P., Schaefli, B., Schymanski, S. J., Samaniego, L., Luce, C. H., Jackson, B. M., Freer, J. E., Arnold, J. R., Moore, R. D., Istanbuluoğlu, E., and Ceola, S.: Improving the theoretical underpinnings of process-based hydrologic models, *Water Resour Res*, 52, 2350–2365, 10.1002/2015wr017910, 2016.

870 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, 10.1029/2020wr029001, 2021.

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 Resour Res*, 51, 5929–5956, 10.1002/2015wr017096, 2015.

875 Dagupati, 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, 10.1002/hyp.10536, 2015.

de Vos, N. J., Rientjes, T. H. M., and Gupta, H. V.: Diagnostic evaluation of conceptual rainfall-runoff models using temporal clustering, *Hydrological Processes*, 24, 2840–2850, 10.1002/hyp.7698, 2010.

880 Deng, C., Liu, P., Guo, S., Li, Z., and Wang, D.: Identification of hydrological model parameter variation using ensemble Kalman filter, *Hydrol Earth Syst Sc*, 20, 4949–4961, 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, *J Hydrol*, 320, 3–17, 10.1016/j.jhydrol.2005.07.031, 2006.

885 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, Doi 10.1007/Bf00939380, 1993.

Fowler, K., Coxon, G., Freer, J., Peel, M., Wagener, T., Western, A., Woods, R., and Zhang, L.: Simulating runoff under changing climatic conditions: A framework for model improvement, *Water Resour Res*, 54, 9812–9832, 10.1029/2018wr023989, 2018.

890 Fowler, K., Peel, M., Sait, 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, *Hydrol Earth Syst Sc*, 26, 6073–6120, 10.5194/hess-26-6073-2022, 2022.

895 Gomez, J.: Stochastic global optimization algorithms: A systematic formal approach, *Inform Sciences*, 472, 53–76, 10.1016/j.ins.2018.09.021, 2019.

Gosling, S. N., Taylor, R. G., Arnell, N. W., and Todd, M. C.: A comparative analysis of projected impacts of climate change on river runoff from global and catchment scale hydrological models, *Hydrology and Earth System Sciences*, 15, 279–294, 10.5194/hess-15-279-2011, 2011.

900 Grayson, R. B., Moore, I. D., and McMahon, T. A.: Physically based hydrologic modeling: 1. A terrain-based model for investigative purposes, *Water Resour Res*, 28, 2639–2658, 10.1029/92wr01258, 2010.

Guo, D., Johnson, F., and Marshall, L.: Assessing the Potential Robustness of Conceptual Rainfall-Runoff Models Under a Changing Climate, *Water Resour Res*, 54, 5030–5049, 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, 10.1016/j.jhydrol.2009.08.003, 2009a.

905 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, doi.org/10.1016/j.jhydrol.2009.08.003, 2009b.

Höge, M., Wöhling, T., and Nowak, W.: A Primer for Model Selection: The Decisive Role of Model Complexity, *Water Resources Research*, 54, 1688–1715, 10.1002/2017wr021902, 2018.

910 Hundecha, Y. and Bárdossy, A.: Modeling of the effect of land use changes on the runoff generation of a river basin through parameter regionalization of a watershed model, *J Hydrol*, 292, 281–295, 10.1016/j.jhydrol.2004.01.002, 2004.

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, *J Hydrol*, 621, 129666, 10.1016/j.jhydrol.2023.129666, 2023.

Kallel, F., Ophir, J., Magee, K., and Krouskop, T.: Elastographic Imaging of Low-Contrast Elastic Modulus Distributions in Tissue, *Ultrasound in Medicine & Biology*, 24, 409–425, 10.1016/s0301-5629(97)00287-1, 1998.

915 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, 10.1109/TKDE.2017.2720168, 2017.

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, doi.org/10.1029/2018WR023750, 2019.

Kim, K. B. and Han, D.: Exploration of sub-annual calibration schemes of hydrological models, *Hydrology Research*, 48, 1014–1031, 10.2166/nh.2016.296 %J Hydrology Research, 2016.

Kim, K. B. and Han, D.: Exploration of sub-annual calibration schemes of hydrological models, *Hydrology Research*, 48, 1014–1031, 10.2166/nh.2016.296, 2017.

Kollat, J. B., Reed, P. M., and Wagener, T.: When are multiobjective calibration trade-offs in hydrologic models meaningful?, *Water Resources Research*, 48, 10.1029/2011wr011534, 2012.

Krapu, C. and Borsuk, M.: A differentiable hydrology approach for modeling with time-varying parameters, *Water Resour Res*, 58, e2021WR031377, 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, *Hydrol. Earth Syst. Sci.*, 22, 6005–6022, 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, doi.org/10.1029/2019WR026065, 2019a.

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, *Hydrol. Earth Syst. Sci.*, 23, 5089–5110, 10.5194/hess-23-5089-2019, 2019b.

Laaha, G. and Blöschl, G.: Seasonality indices for regionalizing low flows, *Hydrological Processes*, 20, 3851–3878, 10.1002/hyp.6161, 2006.

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, 10.1016/j.envsoft.2021.104981, 2021.

Laloy, E. and Vrugt, J. A.: High-dimensional posterior exploration of hydrologic models using multiple-try DREAM(ZS) and high-performance computing, *Water Resour Res*, 48, 10.1029/2011wr010608, 2012.

Lin, K., Zhang, Q., and Chen, X.: An evaluation of impacts of DEM resolution and parameter correlation on TOPMODEL modeling uncertainty, *J Hydrol*, 394, 370–383, 10.1016/j.jhydrol.2010.09.012, 2010.

Maier, H. R., Kapelan, Z., Kasprzyk, J., Kollat, J., Matott, L. S., Cunha, M. C., Dandy, G. C., Gibbs, M. S., Keedwell, E., Marelli, A., Ostfeld, A., Savic, D., Solomatine, D. P., Vrugt, J. A., Zecchin, A. C., Minsker, B. S., Barbour, E. J., Kuczera, G., Pasha, F., Castelletti, A., Giuliani, M., and Reed, P. M.: Evolutionary algorithms and other metaheuristics in water resources: Current status, research challenges and future directions, *Environ Modell Softw*, 62, 271–299, 10.1016/j.envsoft.2014.09.013, 2014.

McCabe, G. J., Hay, L. E., Boek, A., Markstrom, S. L., and Atkinson, R. D.: Inter-annual and spatial variability of Hamon potential evapotranspiration model coefficients, *J Hydrol*, 521, 389–394, 10.1016/j.jhydrol.2014.12.006, 2015.

Moore, R. J.: The probability distributed principle and runoff production at point and basin scales, *Hydrological Sciences Journal*, 30, 273–297, 10.1080/02626668509490989, 2009.

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, 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, 10.1029/2020wr028091, 2021.

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, *J Hydrol*, 523, 147–159, 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, *J Hydrol*, 563, 657–668, 10.1016/j.jhydrol.2018.06.035, 2018.

Pande, S. and Moayeri, M.: Hydrological interpretation of a statistical measure of basin complexity, *Water Resour Res*, 54, 7403–7416, 10.1029/2018wr022675, 2018.

Pathiraja, S., Marshall, L., Sharma, A., and Moradkhani, H.: Hydrologic modeling in dynamic catchments: A data assimilation approach, *Water Resources Research*, 52, 3350–3372, 10.1002/2015wr017192, 2016.

Pfannerstill, M., Guse, B., and Fohrer, N.: Smart low flow signature metrics for an improved overall performance evaluation of hydrological models, *J Hydrol*, 510, 447–458, 10.1016/j.jhydrol.2013.12.044, 2014.

Pianosi, F., Sarrazin, F., and Wagener, T.: A Matlab toolbox for Global Sensitivity Analysis, *Environ Modell Softw*, 70, 80–85, 10.1016/j.envsoft.2015.04.009, 2015.

Santos, L., Thirel, G., and Perrin, C.: Technical note: Pitfalls in using log-transformed flows within the KGE criterion, *Hydrology and Earth System Sciences*, 22, 4583–4591, 10.5194/hess-22-4583-2018, 2018.

Schoups, G., van de Giesen, N. C., and Savenije, H. H. G.: Model complexity control for hydrologic prediction, *Water Resources Research*, 44, 10.1029/2008wr006836, 2008.

Shafii, M. and De Smedt, F.: Multi-objective calibration of a distributed hydrological model (WetSpa) using a genetic algorithm, *Hydrology and Earth System Sciences*, 13, 2137–2149, 10.5194/hess-13-2137-2009, 2009.

Shamir, E., Imam, B., Gupta, H. V., and Sorooshian, S.: Application of temporal streamflow descriptors in hydrologic model parameter estimation, *Water Resour Res*, 41, 10.1029/2004wr003409, 2005.

Shrestha, S., Bae, D. H., Holc, P., Ghimire, S., and Pokhrel, Y.: Future hydrology and hydrological extremes under climate change in Asian river basins, *Scientific Reports*, 11, 17089, 10.1038/s41598-021-96656-2, 2021.

Sivakumar, B.: Dominant processes concept in hydrology: moving forward, *Hydrological Processes*, 18, 2349–2353, 10.1002/hyp.5606, 2004.

Smakhtin, V. U.: Low flow hydrology: a review, *Journal of Hydrology*, 240, 147–186, Doi 10.1016/S0022-1694(00)00340-1, 2001.

Thyer, M., Renard, B., Kavetski, D., Kuczera, G., Franks, S. W., and Srikanthan, S.: Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: A case study using Bayesian total error analysis, *Water Resour Res*, 45, 10.1029/2008wr006825, 2009.

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, 10.1080/01431160500168686, 2010.

- van Werkhoven, K., Wagener, T., Reed, P., and Tang, Y.: Sensitivity guided reduction of parametric dimensionality for multi-objective calibration of watershed models, *Advances in Water Resources*, 32, 1154–1169, [10.1016/j.advwatres.2009.03.002](https://doi.org/10.1016/j.advwatres.2009.03.002), 2009.
- Vaze, J., Post, D. A., Chiew, F. H. S., Perraud, J. M., Viney, N. R., and Teng, J.: Climate non-stationarity—Validity of calibrated rainfall-runoff models for use in climate change studies, *J Hydrol*, 394, 447–457, [10.1016/j.jhydrol.2010.09.018](https://doi.org/10.1016/j.jhydrol.2010.09.018), 2010.
- Wagener, T. and Kollat, J.: Numerical and visual evaluation of hydrological and environmental models using the Monte Carlo analysis toolbox, *Environ Modell Softw*, 22, 1021–1033, [10.1016/j.envsoft.2006.06.017](https://doi.org/10.1016/j.envsoft.2006.06.017), 2007.
- Wagener, T., McIntyre, N., Lees, M. J., Wheater, H. S., and Gupta, H. V.: Towards reduced uncertainty in conceptual rainfall-runoff modelling: Dynamic identifiability analysis, *Hydrological Processes*, 17, 455–476, [10.1002/hyp.1135](https://doi.org/10.1002/hyp.1135), 2003.
- 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 Resour Res*, 54, 2129–2151, [10.1002/2018wr022546](https://doi.org/10.1002/2018wr022546), 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, [10.1016/j.jhydrol.2017.03.027](https://doi.org/10.1016/j.jhydrol.2017.03.027), 2017.
- Wang, Y., Wang, J., Xie, J., and Lu, H.: Improvements in the degree-day model, incorporating forest influence, and taking China's Tianshan Mountains as an example, *J Hydrol Reg Stud*, 44, [10.1016/j.ejrh.2022.101215](https://doi.org/10.1016/j.ejrh.2022.101215), 2022a.
- Wang, Z., Yang, Y., Zhang, C., Guo, H., and Hou, Y.: Historical and future Palmer Drought Severity Index with improved hydrological modeling, *J Hydrol*, 610, 127941, [10.1016/j.jhydrol.2022.127941](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, [10.1007/s00477-020-01896-9](https://doi.org/10.1007/s00477-020-01896-9), 2021.
- Weise, T.: Global optimization algorithms theory and application, Self-Published Thomas Weise, 361, 153, 2009.
- 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, *J Hydrol*, 598, 125772, [10.1016/j.jhydrol.2020.125772](https://doi.org/10.1016/j.jhydrol.2020.125772), 2021.
- 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, *J Hydrol*, 568, 758–768, [10.1016/j.jhydrol.2018.11.038](https://doi.org/10.1016/j.jhydrol.2018.11.038), 2019a.
- Xiong, M. S., Liu, P., Cheng, L., Deng, C., Gui, Z. L., Zhang, X. J., and Liu, Y. H.: 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, [10.1016/j.jhydrol.2018.11.038](https://doi.org/10.1016/j.jhydrol.2018.11.038), 2019b.
- Yapo, P. O., Gupta, H. V., and Sorooshian, S.: Multi-objective global optimization for hydrologic models, *J Hydrol*, 204, 83–97, [10.1016/s0022-1694\(97\)00107-8](https://doi.org/10.1016/s0022-1694(97)00107-8), 1998.
- 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, doi.org/10.1029/2007wr006716, 2008.
- Zhang, X. and Liu, P.: A time-varying parameter estimation approach using split sample calibration based on dynamic programming, *Hydrol Earth Syst Sc*, 25, 711–733, [10.5194/hess-25-711-2021](https://doi.org/10.5194/hess-25-711-2021), 2021.
- 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, 127305, 2022.
- 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.
- 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.
- 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, *Hydrol Earth Syst Sc*, 12, 1273–1283, <https://doi.org/10.5194/hess-12-1273-2008>, 2008.
- Cheng, L., Yaeger, M., Viglione, A., Coopersmith, E., Ye, S., and Sivapalan, M.: Exploring the physical controls of regional patterns of flow duration curves – Part 1: Insights from statistical analyses, *Hydrol. Earth Syst. Sci.*, 16, 4435–4446, <https://doi.org/10.5194/hess-16-4435-2012>, 2012.
- Clark, M. P., Schaefli, B., Schymanski, S. J., Samaniego, L., Luce, C. H., Jackson, B. M., Freer, J. E., Arnold, J. R., Moore, R. D., Istanbuloglu, E., and Ceola, S.: Improving the theoretical underpinnings of process-based hydrologic models, *Water Resour Res*, 52, 2350–2365, <https://doi.org/10.1002/2015WR017910>, 2016.
- 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), <https://doi.org/10.1029/2020WR029001>, 2021.
- 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 Resour Res*, 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.
- Deng, C., Liu, P., Guo, S., Li, Z., and Wang, D.: Identification of hydrological model parameter variation using ensemble Kalman filter, *Hydrol Earth Syst Sc*, 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, *J Hydrol*, 320, 3-17, <https://doi.org/10.1016/j.jhydrol.2005.07.031>, 2006.
- 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.
- Fowler, K., Coxon, G., Freer, J., Peel, M., Wagener, T., Western, A., Woods, R., and Zhang, L.: Simulating runoff under changing climatic conditions: A framework for model improvement, *Water Resour Res*, 54, 9812-9832, <https://doi.org/10.1029/2018WR023989>, 2018.
- 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, *Hydrol Earth Syst Sc*, 26, 6073-6120, <https://doi.org/10.5194/hess-26-6073-2022>, 2022.
- 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.
- Guo, D., Johnson, F., and Marshall, L.: Assessing the Potential Robustness of Conceptual Rainfall-Runoff Models Under a Changing Climate, *Water Resour Res*, 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.
- 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.
- Hundecha, Y. and Bárdossy, A.: Modeling of the effect of land use changes on the runoff generation of a river basin through parameter regionalization of a watershed model, *J Hydrol*, 292, 281-295, <https://doi.org/10.1016/j.jhydrol.2004.01.002>, 2004.
- 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, *J Hydrol*, 621, 129666, <https://doi.org/10.1016/j.jhydrol.2023.129666>, 2023.
- Kallel, F., Ophir, J., Magee, K., and Krouskop, T.: Elastographic Imaging of Low-Contrast Elastic Modulus Distributions in Tissue, *Ultrasound in Medicine & Biology*, 24, 409-425, [https://doi.org/10.1016/S0301-5629\(97\)00287-1](https://doi.org/10.1016/S0301-5629(97)00287-1), 1998.
- 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.
- 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.
- Kim, K. B. and Han, D.: Exploration of sub-annual calibration schemes of hydrological models, *Hydrology Research*, 48, 1014-1031, <https://doi.org/10.2166/nh.2016.296>, 2017.
- Kollat, J. B., Reed, P. M., and Wagener, T.: When are multiobjective calibration trade-offs in hydrologic models meaningful?, *Water Resources Research*, 48, <https://doi.org/10.1029/2011WR011534>, 2012.
- Krapu, C. and Borsuk, M.: A differentiable hydrology approach for modeling with time-varying parameters, *Water Resour Res*, 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, *Hydrol. Earth Syst. Sci.*, 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.
- 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, *Hydrol. Earth Syst. Sci.*, 23, 5089-5110, <https://doi.org/10.5194/hess-23-5089-2019>, 2019b.
- Laaha, G. and Blöschl, G.: Seasonality indices for regionalizing low flows, *Hydrological Processes*, 20, 3851-3878, <https://doi.org/10.1002/hyp.6161>, 2006.
- 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.
- Laloy, E. and Vrugt, J. A.: High-dimensional posterior exploration of hydrologic models using multiple-try DREAM(ZS) and high-performance computing, *Water Resour Res*, 48, <https://doi.org/10.1029/2011WR010608>, 2012.
- 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.
- Maier, H. R., Kapelan, Z., Kasprzyk, J., Kollat, J., Matott, L. S., Cunha, M. C., Dandy, G. C., Gibbs, M. S., Keedwell, E., Marchi, A., Ostfeld, A., Savic, D., Solomatine, D. P., Vrugt, J. A., Zecchin, A. C., Minsker, B. S., Barbour, E. J., Kuczera, G., Pasha, F., Castelletti, A., Giuliani,

- M., and Reed, P. M.: Evolutionary algorithms and other metaheuristics in water resources: Current status, research challenges and future directions, *Environ Modell Softw*, 62, 271-299, <https://doi.org/10.1016/j.envsoft.2014.09.013>, 2014.
- 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, *J Hydrol*, 521, 389-394, <https://doi.org/10.1016/j.jhydrol.2014.12.006>, 2015.
- 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.
- 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.
- 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, *J Hydrol*, 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, *J Hydrol*, 563, 657-668, <https://doi.org/10.1016/j.jhydrol.2018.06.035>, 2018.
- Pande, S. and Moayeri, M.: Hydrological interpretation of a statistical measure of basin complexity, *Water Resour Res*, 54, 7403-7416, <https://doi.org/10.1029/2018WR022675>, 2018.
- Pathiraja, S., Marshall, L., Sharma, A., and Moradkhani, H.: Hydrologic modeling in dynamic catchments: A data assimilation approach, *Water Resources Research*, 52, 3350-3372, <https://doi.org/10.1002/2015WR017192>, 2016.
- Pianosi, F., Sarrazin, F., and Wagener, T.: A Matlab toolbox for Global Sensitivity Analysis, *Environ Modell Softw*, 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.: 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.
- Santos, L., Thirel, G., and Perrin, C.: Technical note: Pitfalls in using log-transformed flows within the KGE criterion, *Hydrology and Earth System Sciences*, 22, 4583-4591, <https://doi.org/10.5194/hess-22-4583-2018>, 2018.
- 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.
- Shafii, M. and De Smedt, F.: Multi-objective calibration of a distributed hydrological model (WetSpa) using a genetic algorithm, *Hydrology and Earth System Sciences*, 13, 2137-2149, <https://doi.org/10.5194/hess-13-2137-2009>, 2009.
- Shamir, E., Imam, B., Gupta, H. V., and Sorooshian, S.: Application of temporal streamflow descriptors in hydrologic model parameter estimation, *Water Resour Res*, 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.
- 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.
- Sivakumar, B.: Dominant processes concept in hydrology: moving forward, *Hydrological Processes*, 18, 2349-2353, <https://doi.org/10.1002/hyp.5606>, 2004.
- Smakhtin, V. U.: Low flow hydrology: a review, *Journal of Hydrology*, 240, 147-186, [https://doi.org/10.1016/S0022-1694\(00\)00340-1](https://doi.org/10.1016/S0022-1694(00)00340-1), 2001.
- 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.
- 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.
- van Werkhoven, K., Wagener, T., Reed, P., and Tang, Y.: Sensitivity-guided reduction of parametric dimensionality for multi-objective calibration of watershed models, *Advances in Water Resources*, 32, 1154-1169, <https://doi.org/10.1016/j.advwatres.2009.03.002>, 2009.
- Vaze, J., Post, D. A., Chiew, F. H. S., Perraud, J. M., Viney, N. R., and Teng, J.: Climate non-stationarity – Validity of calibrated rainfall-runoff models for use in climate change studies, *J Hydrol*, 394, 447-457, <https://doi.org/10.1016/j.jhydrol.2010.09.018>, 2010.
- Wagener, T. and Kollat, J.: Numerical and visual evaluation of hydrological and environmental models using the Monte Carlo analysis toolbox, *Environ Modell Softw*, 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 Resour Res*, 54, 2129-2151, <https://doi.org/10.1002/2018WR022546>, 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.

- Wang, Y., Wang, J., Xie, J., and Lu, H.: Improvements in the degree-day model, incorporating forest influence, and taking China's Tianshan Mountains as an example, *J Hydrol-Reg Stud*, 44, <https://doi.org/10.1016/j.ejrh.2022.101215>, 2022a.
- Wang, Z., Yang, Y., Zhang, C., Guo, H., and Hou, Y.: Historical and future Palmer Drought Severity Index with improved hydrological modeling, *J Hydrol*, 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.
- Weise, T.: Global optimization algorithms-theory and application, Self-Published Thomas Weise, 361, 153, <http://www.it-weise.de/>, 2009.
- 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, *J Hydrol*, 598, 125772, <https://doi.org/10.1016/j.jhydrol.2020.125772>, 2021.
- 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, *J Hydrol*, 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.
- 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, *Hydrol Earth Syst Sc*, 25, 711-733, <https://doi.org/10.5194/hess-25-711-2021>, 2021.
- 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.