Supplement of Hydrol. Earth Syst. Sci., 30, 2455–2471, 2026
https://doi.org/10.5194/hess-30-2455-2026-supplement

Hydrology and Earth System Sciences
Open Access

Supplement of

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

Tian Lan et al.

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

Contents of this file

Introduction

This supporting information includes nine sections that support the analysis. The *S1 Assistive techniques for the proposed framework* introduce the techniques used in this study. The *S2 Seasonal dynamic catchment characteristics* sections are used to support the *4.1 Clustered sub-periods based on catchment dynamics* section in the main manuscript. The *S3 Model performance in study cases* is used to support the *4.3 Model performance* section in the main manuscript. The *S4 Applicability analysis of the dynamic calibration framework* is used to support the *3.1 Hydrological models* section in the main manuscript. The *S5 Evaluation of performance gains beyond parameter dimensionality* is used to support the *4.1 Clustered sub-periods based on catchment dynamics* section in the main manuscript. The *S6 State variables and fluxes assessment in the study cases* section is used to support the *4.2 State variables and fluxes* section in the main manuscript. The *S7 Parameters* is used to support the *5.1 Why dynamic parameter sets improve simulation performance* in the main manuscript. The *S8 Correlation between model parameters* is used to support the *5.1.1 Complex correlation between parameters* section in the main manuscript. The *S9 Violin plots* is used to support the *5.3 Parameter response to catchment dynamics* in the main manuscript.

S1 Assistive techniques for the proposed framework

S1.1 Hamon potential evapotranspiration equation

To fill in missing data and ensure data consistency, this study employs the Hamon potential evapotranspiration equation to calculate potential evapotranspiration. The Hamon-derived evaporation equation is as simple as possible. The equation uses only two main input parameters, including the temperature and sunshine hours. The developed equation applies to both humid and dry climatic conditions (McCabe et al., 2015). According to this method, the evaporation may be calculated as follows (Hamon, 1961; Morton, 1971):

$$E = B_1(D)^{B_2} 10^{\left(\frac{B_3 T_{\mathrm{mean}}}{T_{\mathrm{mean}}+273}\right)} \tag{1}$$

where D is the maximum sunshine duration ratio, and T_{mean} is the mean air temperature. The value of D may be estimated using the equation given below. B_1, B_2, and B_3 are constants with typical values of 0.63, 2.0, and 7.5, respectively.

$$D = \frac{1}{90} \arccos\left(-\tan(\varnothing) \,.\, \tan 23.45^{\circ} \sin\left(\frac{J-80}{365}\right) 360^{\circ}\right) \tag{2}$$

In this equation, "ϕ" is the latitude and J represents the Julian day (Ghumman et al., 2021).

S1.2 Maximal information coefficient (MIC)

The MIC, as proposed by Reshef et al. (2011), is a measurement approach that doesn't rely on the distributional assumptions of datasets. This method captures extensive mutual information between variables, whether they exhibit functional or non-functional relationships. In the case of functional relationships, the MIC algorithm provides a score similar to the coefficient of determination (R^2) of the datasets. We analyze the datasets using the SG-MIC algorithm (utilizing simulated annealing and genetic techniques) developed for optimal MIC calculation. Convergence of the SG-MIC is established based on Markov theory (Zhang et al., 2014).

S1.3 Principal component analysis (PCA)

PCA is a multivariate statistical procedure that reduces data redundancy and unveils embedded patterns. This algorithm employs orthogonal transformations to convert a set of potentially correlated variables into linearly uncorrelated vectors with lower dimensionality. These resulting vectors, referred to as Principal Components (PCs), are orthogonal because the eigenvectors of the covariance matrix are symmetric. The first PC captures the maximum possible variance (Frey and Pimentel, 1978; Wright et al., 2009; Fan et al., 2017; Wold et al., 1987).

S1.4 Fuzzy C-means clustering (FCM)

The FCM clustering approach is a widely used probabilistic-type clustering method originally proposed by Dunn (1973) and further improved by Bezdek et al. (1984). The FCM algorithm can be summarized as follows: First, initialize the membership function matrix μ_{ij} based on the selected value of m. Second, compute the cluster center z_j and the Euclidean distance d_{ij}. Finally, update the membership function μ_{ij} once the iteration has converged. The termination criterion for the FCM algorithm is a low relative change in the cluster center values. It's important to note that the FCM algorithm is sensitive to initial conditions (Hathaway and Bezdek, 2001).

S1.5 Shuffled Complex Evolution approach (SCE-UA)

The shuffled complex evolution approach (SCE-UA), as an effective global optimization method, is a commonly used algorithm because it is open source and was the first algorithm aimed specifically at calibrating hydrological models (Khakbaz and Kazeminezhad, 2012; Eckhardt and Arnold, 2001; Duan et al., 1994; Sorooshian et al., 2010). The technical details of the SCE-UA can be shown in the flowchart (see Fig. S1) (Duan et al., 1994). In the SCE-UA, the upper limit of the objective function evaluation is set to 10,000 times. All other settings of the SCE-UA technique are the default.

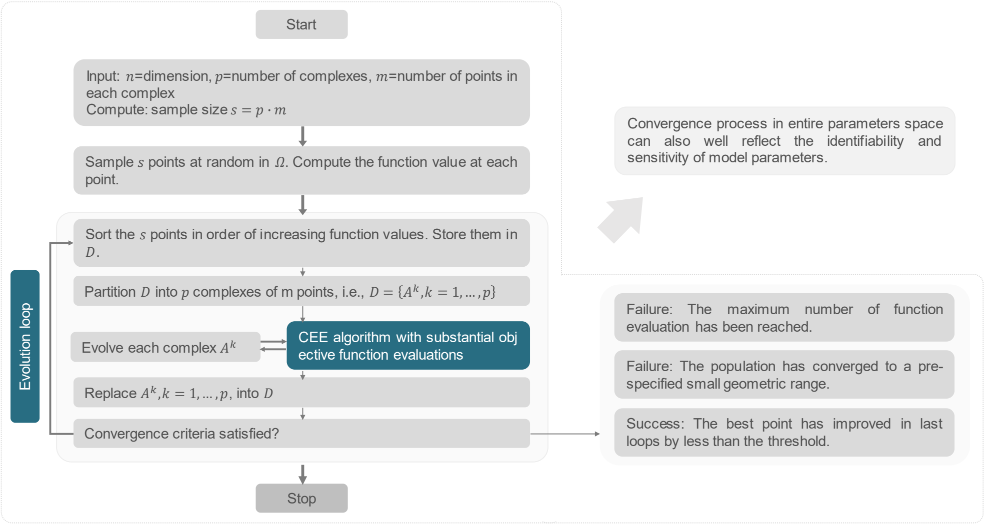

Figure S1. The flowchart of the SCE-UA algorithm (Duan et al., 2010; Duan et al., 1993; Duan et al., 1994).

S1.6 HYMOD model

The structure of the HYMOD model is shown in Fig. S2. Precipitation (P) and potential evapotranspiration (PET) drive a probability-distributed soil-moisture 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 recharge 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 linear reservoirs (states X_{q1}–X_{q3}) governed by K_q, producing outflow Q_q, while the slow flow is routed through a single linear reservoir governed by K_s, producing outflow Q_s. The simulated discharge (Q_{sim}) is computed as the sum of Q_q and Q_s (Wang et al., 2022a).

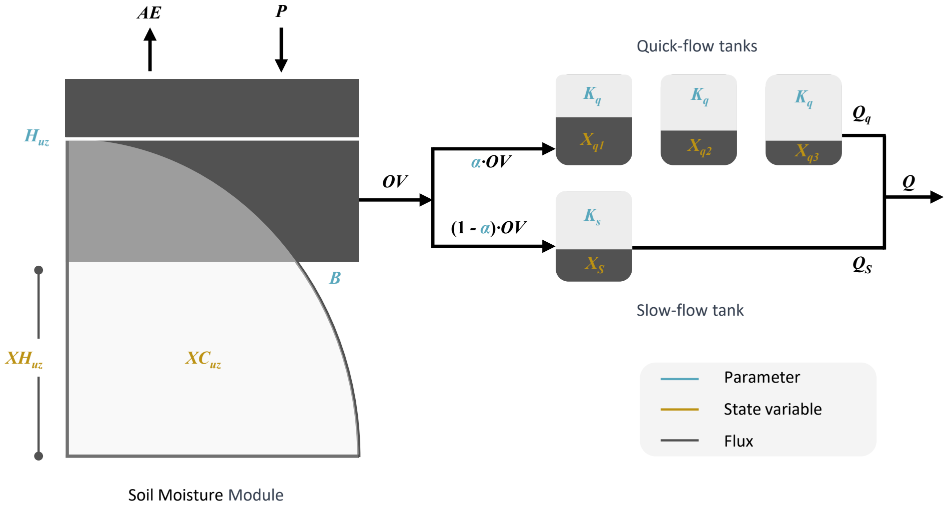

Figure S2. Schematic diagram of the HYMOD structure and principles (Vrugt et al., 2003; Wagener et al., 2001).

S1.7 Degree Day Model

The Degree Day Model is a widely used method for estimating the melting of ice and snow. The Degree Day Model is based on a positive linear relationship between glaciers and snowmelt and temperature, especially the positive degree days on the surface of ice and snow. It assumes that when the daily average temperature exceeds the critical temperature for melting, a certain amount of melting occurs. Key parameters include the temperature threshold and the degree day factor (Hock, 2003; Wang et al., 2022).

$$M = \begin{cases} \mathrm{DDF} \cdot \left(T - T_t\right) & T > T_t \\ 0 & T \le T_t \end{cases} \tag{3}$$

In the formula: is the amount of melt (mm·d^{-1}); is the degree day factor (mm·d^{-1}·°C^{-1}); is the daily average air temperature (°C); is the critical temperature for melting (°C).

S1.8 AHP, PP, and CRITIC

The composite objective function in Experiment 3 incorporated three distinct weighting methods to assign metric weights, thereby exploring the sensitivity of calibration outcomes to different weighting strategies.

The **Analytic Hierarchy Process (AHP)** is a widely used subjective weighting method that structures complex decision problems into a hierarchy of objectives, criteria, and alternatives. Through pairwise comparisons, it derives a set of prioritized weights valued for its clarity, simplicity, and reliability in multi-criteria decision-making.

The **Projection Pursuit (PP)** method is an objective weighting technique that projects high-dimensional data into a lower-dimensional space. By optimizing a projection index function, it identifies the direction that best reveals the underlying structure of the data, making it effective for data mining and optimization.

The **CRiteria Importance Through Intercriteria Correlation (CRITIC)** method (Diakoulaki et al., 1995) is an objective weighting approach that determines weights based on contrast intensity (represented by standard deviation) and conflict (quantified by inter-criteria correlation). The information content H_j for each criterion is calculated as $H_j = S_j \sum_{i=1}^{p}\left(1 - \lvert r_{ij} \rvert\right)$, where a higher H_j indicates greater influence, leading to a higher weight $W_j = \frac{H_j}{\sum_{j=1}^{p} H_j}$.

S1.9 Multi-criteria evaluation

Table S1. Description of performance metrics.

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

Where $Q_{obs,i}$ and $Q_{sim,i}$ represent the observed and simulated streamflow at time step i, and $\overline{Q}_{obs}$ 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 FDC is usually split into various segments to describe different flow characteristics of a catchment (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.

S2 Seasonal dynamic catchment characteristics

S2.1 Threshold-based screening of seasonal indicators

The selection of candidate indicators for sub-period clustering follows a hierarchical, variable-specific screening protocol designed to retain only variables that exhibit hydrologically meaningful seasonality. For hydrological fluxes (precipitation, potential evapotranspiration, and streamflow), seasonality is assessed by the Seasonality Index (SI); a variable is considered seasonally structured when the multi-year average SI>0.6, indicating a significant intra-annual concentration of fluxes. Temperature is treated as a state variable and evaluated by climatic regime: catchments in Köppen tropical (A) and polar (E) zones are regarded as non-seasonal due to minimal intra-annual temperature variation, while other climate classes are retained. Vegetation seasonality is assessed using the intra-annual NDVI range (ΔNDVI), and catchments with ΔNDVI>0.3 are treated as having substantial seasonal phenology. After this seasonality filter, indicators are further screened for streamflow relevance using the Maximal Information Coefficient (MIC), retaining indicators with MIC>0.35 to capture both linear and non-linear dependencies with discharge; remaining multicollinearity is addressed by principal component analysis (PCA) before clustering. These combined, conservative thresholds ensure the candidate pool for each catchment contains only indicators with robust, process-relevant seasonal signals.

S2.2 MIC values among clustering indicators and streamflow

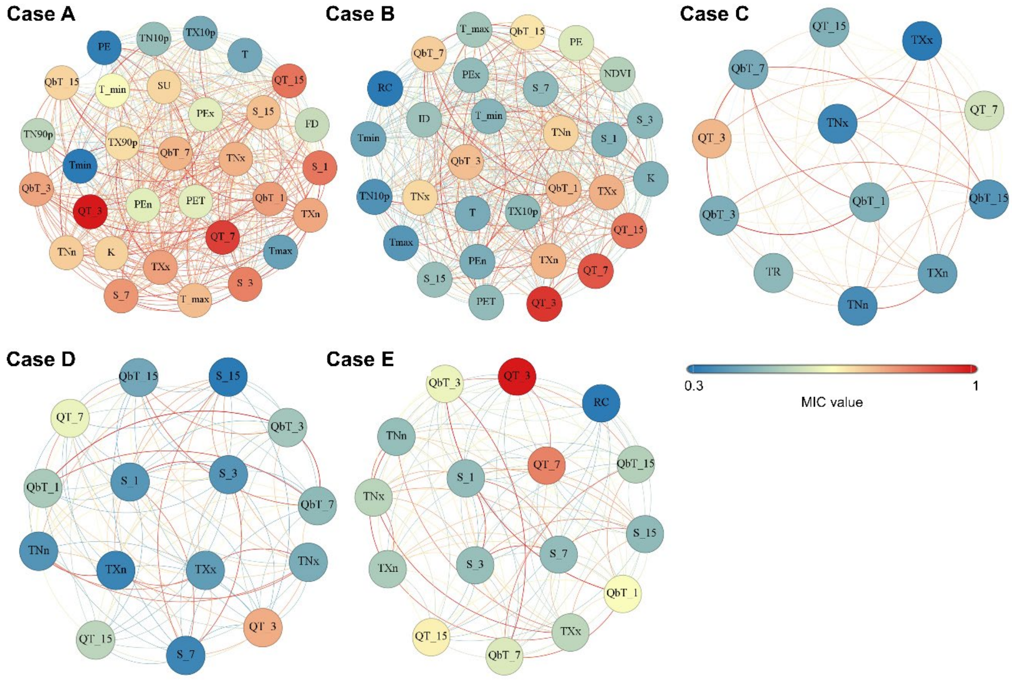

Figure S3. The interconnected network of the nonlinear relationships (MIC values) among the candidate clustering indicators and the streamflow in the cases A, B, C, D, and E. The nodes correspond to the MIC values between all the candidate inputs and the streamflow. The color of the nodes is proportional to the MIC values. The edges correspond to the MIC values occurring for any two variables (minimum is red; maximum is blue). MIC = maximal information coefficient.

S2.3 Characteristic values in diverse sub-periods in study cases

In cases A and B, where seasonal signals are more evident, sub-periods correspond to phases characterized by high evaporative demand, low-temperature conditions, and enhanced runoff responses linked to vegetation recovery and increasing antecedent moisture, reflecting distinct seasonal cycles and wet–dry transitions. In cases D and E, where larger variations in hydrological states are reflected, sub-period boundaries align more closely with variations in antecedent storage, runoff coefficients, and the relative contributions of quick flow and baseflow, indicating transitions from low connectivity and weak response states to more humid conditions with higher runoff efficiency. In case C, despite the retention of fewer dynamic features, runoff responses across sub-periods exhibit a continuous gradient from low to high response states, suggesting that the clustering captures meaningful hydrological variability rather than arbitrary segmentation. The identified sub-periods, therefore, provide a physically interpretable structure that supports subsequent dynamic parameterization and modelling experiments.

Table S2. The selected seasonal characteristic indicators in diverse sub-periods in case A.

Index	Period 1	Period 2	Period 3	Period 4	Period 5
PE	2.57	2.29	1.49	0.39	0.55
PE_T	175.03	135.31	73.91	16.42	25.63
PE_x	3.21	2.88	2.00	0.72	0.96
PE_n	2.44	1.53	0.53	0.02	0.08
T	16.58	12.56	8.57	3.94	6.10
T_{max}	23.47	18.16	13.42	8.13	9.99
T_{min}	9.69	6.97	3.72	-0.25	2.20
$\bar{T}_{\mathrm{max}}$	23.41	18.62	13.58	7.71	9.52
$\bar{T}_{\mathrm{min}}$	9.85	6.54	3.64	0.23	1.97
FD	0.02	1.42	7.98	27.19	16.07
SU	21.00	6.73	0.88	0.00	0.02
TX_{x}	32.84	29.17	23.62	14.24	16.25
TN_{x}	14.06	12.00	9.53	7.09	8.32
TX_{n}	16.09	10.93	6.52	0.15	2.90
TN_{n}	4.60	0.11	-2.93	-9.35	-5.89
TN10p	0.00	0.39	4.15	20.84	11.03
TX10p	0.00	0.03	1.91	22.69	12.30
TN90p	18.29	4.02	0.65	0.00	0.06
TX90p	17.49	5.11	0.52	0.00	0.01
K	997.90	755.02	516.85	257.27	347.99
$\overline{S_{t\text{-}1}}$	111.31	164.07	220.90	242.42	245.34
$\overline{S_{t\text{-}3}}$	334.85	491.22	662.06	727.00	735.88
$\overline{S_{t\text{-}7}}$	786.24	1143.16	1541.27	1696.14	1712.96
$\overline{S_{t\text{-}15}}$	1710.36	2446.45	3279.77	3632.05	3643.51
$Q_{t\text{-}3}$	1.00	2.78	8.56	9.35	26.80
$Q_{t\text{-}7}$	2.35	6.35	20.22	22.74	61.44
$Q_{t\text{-}15}$	5.18	13.61	45.82	54.96	122.36
$Q_{bt\text{-}1}$	0.25	0.48	1.16	1.56	2.50
$Q_{bt\text{-}3}$	0.74	1.45	3.46	4.68	7.47
$Q_{bt\text{-}7}$	1.75	3.45	8.10	10.98	17.20
$Q_{bt\text{-}15}$	3.83	7.75	17.64	23.83	35.44

Table S3. The selected seasonal characteristic indicators in diverse sub-periods in case B.

Index	Period 1	Period 2	Period 3
PE	3.15	0.6	4.1
PE_T	229.96	32.5	188.59
PE_x	4.43	1.3	4.11
PE_n	2.95	0.06	1.96
T	8.77	-4.66	9.7
T_{max}	17.15	0.92	17.87
T_{min}	0.4	-10.24	1.52
$\bar{T}_{\mathrm{max}}$	19.64	1.13	12.91
$\bar{T}_{\mathrm{min}}$	2.17	-9.99	-2.24
ID	0.17	27.02	1.19
TX_{x}	27.49	11.6	23.35
TN_{x}	7.88	-1.58	4.41
TX_{n}	7.66	-10.51	1.04
TN_{n}	-5.42	-18.51	-11.52

TN10p	0.02	11.15	0.51
TX10p	0.02	11.43	0.04
K	658.23	50.73	350.12
NDVI	0.56	0.25	0.56
C	0.84	0.22	0.82
$\overline{S_{t\text{-}1}}$	104.37	114.89	178.78
$\overline{S_{t\text{-}3}}$	313.85	343.58	537.74
$\overline{S_{t\text{-}7}}$	736.49	796.77	1259.4
$\overline{S_{t\text{-}15}}$	1601.47	1688.03	2704.69
$Q_{t\text{-}3}$	2.2	1.26	8.25
$Q_{t\text{-}7}$	5.27	2.89	19.15
$Q_{t\text{-}15}$	12.03	6.07	40.16
$Q_{bt\text{-}1}$	0.59	0.31	1.3
$Q_{bt\text{-}3}$	1.79	0.94	3.88
$Q_{bt\text{-}7}$	4.26	2.19	8.93
$Q_{bt\text{-}15}$	9.57	4.68	18.39

Table S4. The selected seasonal characteristic indicators in diverse sub-periods in case C.

Index	Period 1	Period 2	Period 3	Period 4	Period 5
TR	0	0	0	0	0
TX_{x}	22.41	29.81	31.71	23.64	25.43
TN_{x}	3.66	11.2	12.47	5.23	6.44
TX_{n}	-6.69	-2.26	8.63	-5.12	-4.71
TN_{n}	-15.05	-14.33	-6.41	-14.47	-14.8
$Q_{t\text{-}3}$	0.01	0.01	0.02	0.14	0.42
$Q_{t\text{-}7}$	0.04	0.04	0.05	0.41	1.26
$Q_{t\text{-}15}$	0.09	0.11	0.12	0.96	2.88
$Q_{bt\text{-}1}$	0.19	0.24	0.27	2.03	5.83
$Q_{bt\text{-}3}$	0.02	0.02	0.01	0.11	0.33
$Q_{bt\text{-}7}$	0.04	0.04	0.03	0.26	0.77
$Q_{bt\text{-}15}$	0.09	0.09	0.07	0.55	1.58

Table S5. The selected seasonal characteristic indicators in diverse sub-periods in case D.

Index	Period 1	Period 2	Period 3
TX_{x}	31.29	20.29	19.79
TN_{x}	18.74	7.91	7.91
TX_{n}	17.76	1.05	1.49
TN_{n}	5.19	-9.24	-8.91
$\overline{S_{t\text{-}1}}$	132.36	189.75	217.26
$\overline{S_{t\text{-}3}}$	396.94	569.25	651.26
$\overline{S_{t\text{-}7}}$	926.7	1327.99	1511.82
$\overline{S_{t\text{-}15}}$	1990.64	2840.62	3210.22
$Q_{t\text{-}3}$	1.27	3.59	14.34
$Q_{t\text{-}7}$	3.01	8.7	31.63
$Q_{t\text{-}15}$	6.85	19.23	62.45
$Q_{bt\text{-}1}$	0.21	0.51	1.18
$Q_{bt\text{-}3}$	0.63	1.52	3.54
$Q_{bt\text{-}7}$	1.49	3.53	8.15
$Q_{bt\text{-}15}$	3.3	7.56	16.86

Table S6. The selected seasonal characteristic indicators in diverse sub-periods in case E.

Index	Period 1	Period 2	Period 3
TX_x	28.44	15.15	20.28
TN_x	14.87	3.45	6.65
TX_n	14.35	-4.45	0.19
TN_n	1.15	-13.13	-10.15
C	0.2	0.6	0.92
$\overline{S_{t\text{-}1}}$	140.95	191.09	200.82
$\overline{S_{t\text{-}3}}$	422.65	573.26	602.37
$\overline{S_{t\text{-}7}}$	986.18	1338.57	1400.37
$\overline{S_{t\text{-}15}}$	2116.76	2869.92	2973.89
$Q_{t\text{-}3}$	1.14	2.4	9.38
$Q_{t\text{-}7}$	2.7	5.75	21.46
$Q_{t\text{-}15}$	6.08	13.28	43.11
$Q_{bt\text{-}1}$	0.16	0.35	0.82
$Q_{bt\text{-}3}$	0.48	1.06	2.46
$Q_{bt\text{-}7}$	1.14	2.47	5.7
$Q_{bt\text{-}15}$	2.55	5.33	11.84

S3 Model performance in five study cases

To further examine how the different experiments perform under various hydrological conditions, a detailed assessment of five representative catchments is conducted with diverse dynamic patterns and baseline model performance. Fig. S4 presents the model performance of the seven experiments in five study cases. In all study cases, Experiment 1 demonstrated low simulation accuracy and limited parameter transferability across different flow phases, particularly under extremely low-flow and high-flow conditions. The results in Experiments 2 and 3 show limitations on both objective functions compared to 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 evaluation period. However, both RMSE_Q5 and RMSE_Q95 increased. Relative to Experiment 1, the RMSE_Q95 exhibited a diminished performance during both the calibration and evaluation periods in Experiment 2. Despite prioritizing high and low flows through a weighted objective function, Experiment 3 underperforms compared to 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. 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 evaluation period, compared with Experiment 1. In case C, the performance decline was more significant, with NSE values during both the calibration and evaluation periods approaching zero. Experiment 4 exhibited only marginal improvements over Experiment 1 across most metrics. In contrast, Experiments 6 and 7, which employed the same calibration procedures, achieved strong overall performance during the calibration period, particularly in reproducing high flows and flood peaks. However, during the evaluation period, Experiment 6 showed inconsistent performance—while excelling in certain aspects such as high-flow simulation, it experienced significant degradation in others (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. The decline can be attributed to extremely dry conditions, where runoff volumes approached zero (less than 0.01 mm), making small deviations translate into large relative errors. Notably, across all study cases, Experiments 5 and 7 consistently maintained excellent performance during the evaluation period, closely mirroring their calibration results and outperforming other experiments in nearly all metrics. Moreover, analysis of parameter transferability revealed minimal differences between calibration and evaluation periods for Experiments 5 and 7. Hence, Experiments 5 and 7 demonstrate the superior performance across all evaluation metrics, exhibiting improvements in simulations across various flow phases.

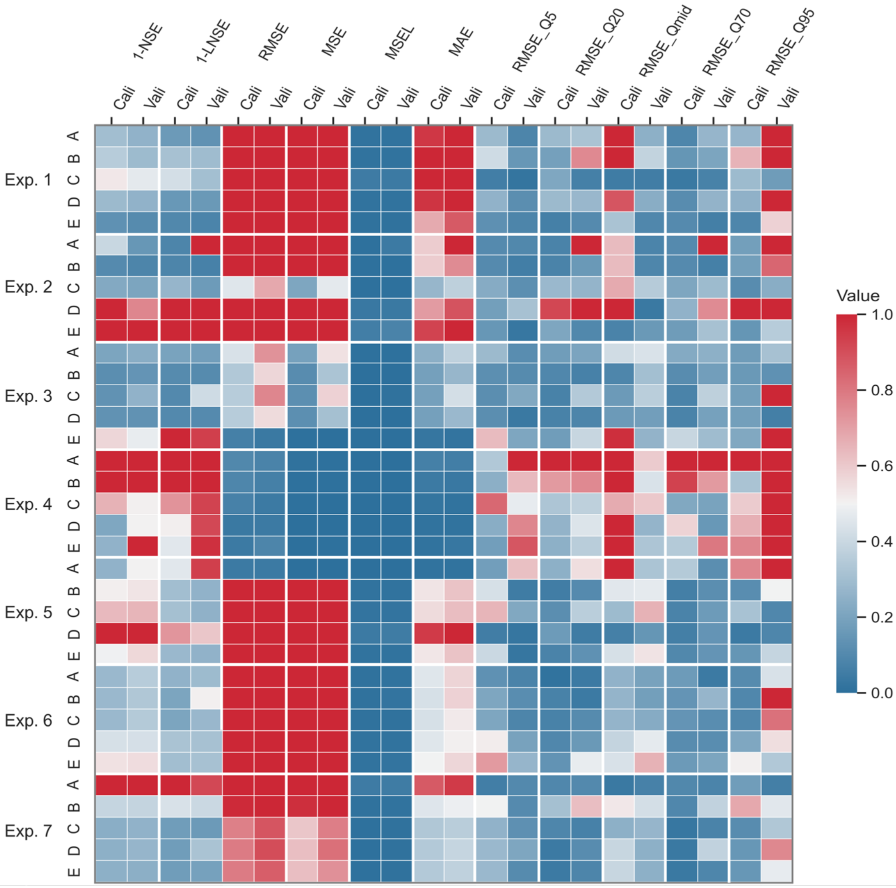

Figure S4. The model performance of seven experiments in five study cases was assessed using multiple evaluation metrics. Lower values reflect superior performance.

S4 Applicability analysis of the dynamic calibration framework

To further evaluate the generalizability of the proposed dynamic calibration framework under more complex model structures, the semi-distributed TOPMODEL is adopted to replace HYMOD for a comparative analysis of Experiment 1 and Experiment 7. TOPMODEL (TOPOgraphy based hydrological MODEL), originally proposed by Beven and Kirkby (1979) and further developed in later studies (Fuentes-Andino et al., 2017), is a terrain-based semi-distributed hydrological model. Parameter ranges are defined according to basin characteristics and previous studies in the same region (Fig. S5 and Table S7).

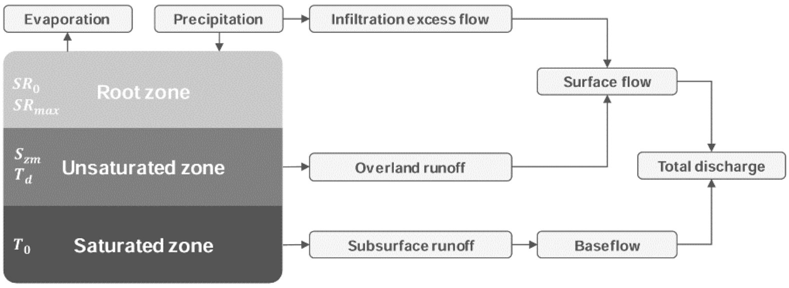

Figure S5. The schematic diagram of the TOPMODEL.

Table S7. Model parameters and their value ranges for the TOPMODEL.

Parameter	Range	Description
T_0 (m^2/day)	0.01-10	Saturated hydraulic conductivity
T_d (m/day)	10^{-4}-5	Unsaturated zone time delay per unit deficit
SR_0 (m)	10^{-5}-1	Initial root zone soil moisture
$SR_{\max}$ (m)	10^{-5}-1	Maximum soil root zone storage deficit
S_{zm} (m)	10^{-3}-1	Maximum moisture deficit

Given the higher computational cost of semi-distributed models and the fact that the objective here is to verify framework effectiveness under a more complex model structure rather than reassess large-sample spatial generalizability, three representative sub-basins within the Hanjiang River basin, namely Hanzhong, Mumahe, and Xunhe, are selected as study cases. Using the proposed method for defining sub-periods based on catchment dynamics, the time series is clustered into four sub-annual periods characterized by distinct hydrological behaviour (Fig. S6). These periods correspond to a dry period, rainfall period I, rainfall period II (the wettest phase), and rainfall period III. Precipitation magnitude and variability are lowest during the dry period and highest during rainfall period II. Although rainfall periods I and III exhibit comparable climatic conditions, rainfall period III is associated with higher antecedent soil moisture.

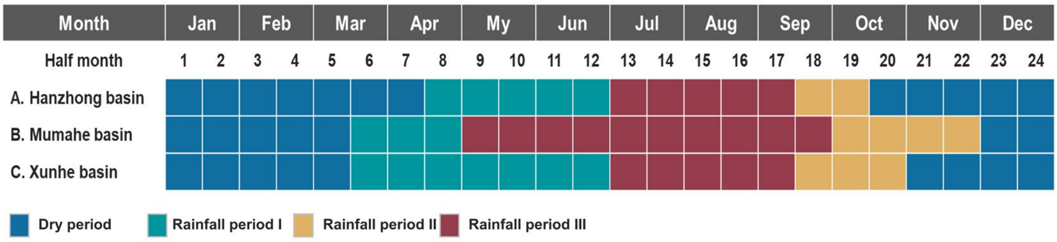

Figure S6. The heat map of the clustering results. The historical data are clustered into four different subannual periods for calibration, including the dry period, rainfall period I, rainfall period II (wet period), and rainfall period III.

Taking the Hanzhong River basin as an example, the precipitation data and streamflow data in the four subannual periods are used to further verify the clustering results in the calibration period (1980–1987) and evaluation period (1988–1990), as shown in Fig. S7. In the dry period, both the total amount and the variance values of all the precipitation series reach the minimums. In contrast, the total amount and the variance values of the precipitation series in the wet period reach the maximum, and the frequency of heavy rain is the highest. In Fig. S7, similar precipitation patterns in the same subperiod of different calendar years are found, which verifies the clustering results in clustering process I (Fig. S7a). In the two normal subannual periods with similar climatic patterns, the streamflow volume is quite different. The streamflow volume is higher in rainfall period III than in rainfall period I (Fig. S7b). The results demonstrate that the runoff yield is higher in the rainfall III period due to the higher antecedent soil moisture content, implying the requirement for clustering based on land-surface indices.

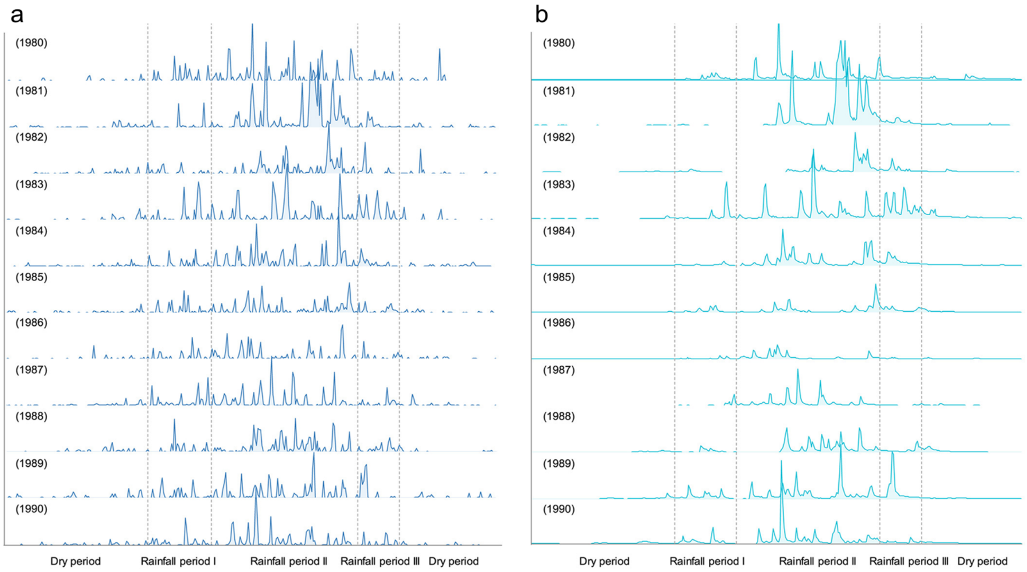

Figure S7. The precipitation data (0–0.1 m/day; a) and the streamflow data (0–8000 m^3/s; b) are divided into four subannual periods in the calibration period (1980–1987) and the evaluation period (1988–1990) in the Hanzhong basin.

Model performance is assessed by comparing Experiment 7 and Experiment 1 within TOPMODEL, with evaluation metrics summarized in Table S8. Improvements are quantified by differences between the two experiments across both calibration and evaluation periods. During the calibration period, overall performance is enhanced, with NSE increasing by 4.4% on average and LNSE increasing by 47.3%. Metrics based on the five flow duration curve (5FDC) components indicate more significant improvements in low- and mid-flow conditions, as reflected by larger reductions in RMSE_Q20 and RMSE_mid. These results indicate that dynamic calibration improves simulation accuracy across multiple flow regimes, particularly under low- and intermediate-flow conditions. Consistent improvements are observed during the evaluation period, where NSE and LNSE increase by 9.8% and 75.7%, respectively. Reductions in RMSE_Q5 and RMSE_Q20 further demonstrate enhanced performance under low-flow conditions. The consistent gains across both calibration and evaluation periods indicate that the effectiveness of dynamic calibration is not restricted to simple lumped models, but remains evident in more complex semi-distributed hydrological models.

Table S8. Multimetric performance comparison of Experiment 1 and Experiment 7 using TOPMODEL.

		NSE	LNSE	RMSE_Q5	RMSE_Q20	RMSE_mid	RMSE_Q70	RMSE_Q95
				Calibration				
Hanzhong	Exp.1	0.9	0.657	0.651	0.288	0.239	0.104	0.139
	Exp.7	0.933	0.822	0.791	0.194	0.054	0.102	0.143
Mumahe	Exp.1	0.863	0.44	0.745	0.987	0.586	0.236	0.068
	Exp.7	0.899	0.87	0.476	0.114	0.065	0.106	0.068
Xunhe	Exp.1	0.891	0.669	0.465	0.308	0.158	0.167	0.11
	Exp.7	0.938	0.797	0.199	0.069	0.138	0.064	0.05
				Evaluation				
Hanzhong	Exp.1	0.806	0.522	0.309	0.198	0.297	0.143	0.089
	Exp.7	0.908	0.791	0.23	0.172	0.166	0.145	0.108
Mumahe	Exp.1	0.813	0.344	0.92	1.071	0.555	0.177	0.103
	Exp.7	0.89	0.835	0.146	0.058	0.121	0.063	0.099
Xunhe	Exp.1	0.795	0.57	0.642	0.584	0.335	0.082	0.214
	Exp.7	0.853	0.757	0.368	0.486	0.312	0.133	0.137

As an additional consistency check, the conceptual lumped model GR4J is further incorporated for comparative evaluation using the 219 MOPEX basins. GR4J is a widely used daily rainfall–runoff model driven by precipitation and potential evapotranspiration (Fig. S8 and Table S9), with four parameters representing production store capacity (x_1), groundwater exchange (x_2), routing store capacity (x_3), and unit hydrograph time base (x_4) (Perrin et al., 2003). Similar improvements in NSE and LNSE are observed during both calibration and evaluation under Experiment 7 relative to the static-parameter scheme (Experiment 1), further supporting the broad applicability of the proposed calibration framework across models of varying complexity (Fig. S9).

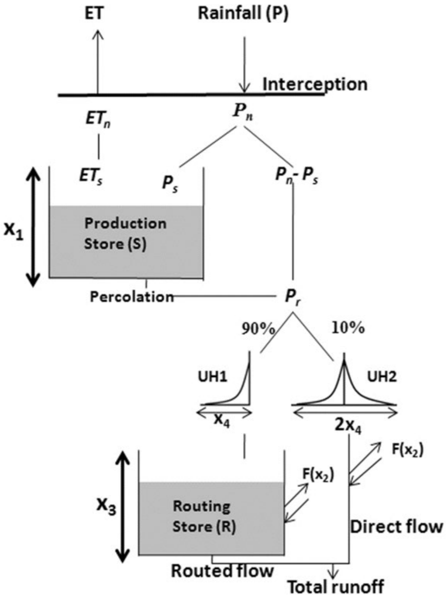

Figure S8. The schematic diagram of the GR4J model (Kunnath-Poovakka & Eldho, 2019).

Table S9. Model parameters and their value ranges for the GR4J model.

Parameter	Range	Description
x_1 (mm)	100-1200	Maximum capacity of the production store
x_2 (mm)	-5–3	Groundwater exchange coefficient
x_3 (mm)	20–300	One day ahead maximum capacity of the routing store
x_4 (day)	1.1–2.9	Time base of unit hydrograph UH1

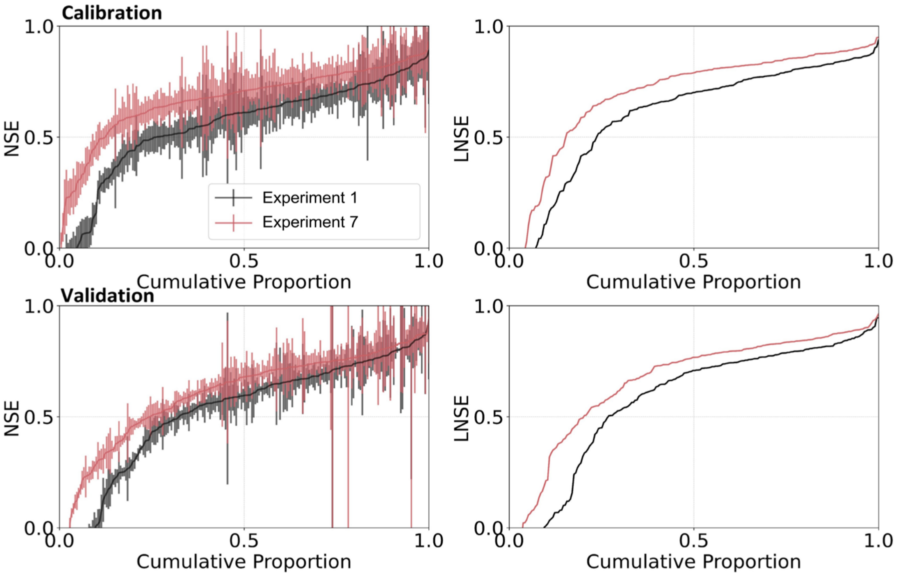

Figure S9. Improvements in NSE and LNSE metrics for the GR4J model using the dynamic calibration during calibration and evaluation periods.

S5 Evaluation of performance gains beyond parameter dimensionality

S5.1 SWAT model-based comparison under a more complex model structure

To further verify that the performance improvements achieved by the proposed calibration framework do not arise solely from increased parameter flexibility in a simple lumped model, the distributed hydrological model SWAT, characterized by a more complex structure, a larger parameter set, and a more complete process representation, is introduced as a reference model for comparative evaluation under the same Hanjiang River Basin and identical periods. The Soil and Water Assessment Tool (SWAT), originally developed by Arnold et al. (2012), is a semi-distributed process-based hydrological model operating at the catchment scale. Its core framework is built upon Hydrological Response Units (HRUs) and a modular coupling mechanism to simulate key catchment hydrological processes and daily streamflow dynamics. Model calibration is performed using the SUFI-2 algorithm implemented in the SWAT-CUP platform to obtain an approximate global optimal parameter set. A common parameter range is adopted for the Hanzhong and Xunhe River basins (Table S10), whereas an independent parameter range is defined for the Mumahe River basin to account for snow processes and channel storage effects (Table S11). On this basis, conventionally calibrated SWAT under the static scheme (Experiment 1) is compared with TOPMODEL and HYMOD under the proposed calibration framework (Experiment 7), allowing the contribution of dynamic information extraction to performance improvement to be further evaluated under a more complex model setting.

Table S10. Selected SWAT parameters and value ranges for the Hanzhong and Xunhe River basins.

Parameter	Description	Range
v__ALPHA_BF.gw	Baseflow recession constant (day^{-1})	[0.01, 1]
v__GW_DELAY.gw	Groundwater delay time (day)	[0, 500]
v__GW_REVAP.gw	Groundwater “revap” coefficient	[0.02, 0.2]
v__REVAPMN.gw	Threshold water depth in the shallow aquifer for revap (mm)	[0, 500]
v__ESCO.hru	Soil evaporation compensation factor	[0, 1]
v__CH_N2.rte	Manning’s roughness coefficient for the main channel	[-0.01, 0.3]
v__CH_K2.rte	Effective hydraulic conductivity in the main channel alluvium (mm h^{-1})	[0.01, 500]
r__CN2.mgt	SCS runoff curve number under moisture condition II	[-0.5, 0.5]
r__SOL_K().sol	Saturated hydraulic conductivity of the soil layer (mm h^{-1})	[-0.5, 0.5]
r__SOL_AWC().sol	Available water capacity of the soil layer (mm H_2O mm^{-1} soil)	[-0.5, 0.5]
r__SOL_BD.sol	Moist bulk density of the soil layer (Mg m^{-3})	[-0.5, 0.5]
r__RCHRG_DP.gw	Deep aquifer percolation fraction	[0, 1]
r__GWQMN.gw	Threshold water depth in the shallow aquifer for baseflow (mm)	[0, 1]

Table S11. Selected SWAT parameters and value ranges for the Mumahe River basin.

Parameter	Description	Range
v__ALPHA_BF.gw	Baseflow recession constant (day^{-1})	[0.01, 1]
v__ALPHA_BNK.rte	Baseflow alpha factor for bank storage	[0, 1]
v__GW_DELAY.gw	Groundwater delay time (day)	[0, 500]
v__GW_REVAP.gw	Groundwater “revap” coefficient	[0.02, 0.2]
v__ESCO.hru	Soil evaporation compensation factor	[0, 1]
v__CH_N2.rte	Manning’s roughness coefficient for the main channel	[-0.01, 0.3]
v__CH_K2.rte	Effective hydraulic conductivity in the main channel alluvium (mm h^{-1})	[0.01, 500]
r__CN2.mgt	SCS runoff curve number under moisture condition II	[-0.5, 0.5]
r__SOL_K().sol	Saturated hydraulic conductivity of the soil layer (mm h^{-1})	[-0.5, 0.5]
r__SOL_AWC().sol	Available water capacity of the soil layer (mm H_2O mm^{-1} soil)	[-0.5, 0.5]

r__SOL_BD.sol	Moist bulk density of the soil layer (Mg m⁻³)	[-0.5, 0.5]
r__GWQMN.gw	Threshold water depth in the shallow aquifer for baseflow (mm)	[0, 1]
v__SFTMP.bsn	Snowfall temperature threshold (°C)	[-20, 20]

As shown in Table S12, TOPMODEL-Experiment 7 exhibits the most stable and best overall performance across the three catchments during both calibration and evaluation, with NSE and LNSE consistently higher than those of SWAT-Experiment 1. Although HYMOD-Experiment 7 is slightly less effective than TOPMODEL in controlling high-flow errors in some catchments, it still generally outperforms SWAT, with the advantages being particularly clear in LNSE and low-flow related metrics. A clearer performance decline is also observed for SWAT during evaluation. For example, in the Hanzhong River basin during evaluation, RMSE_Q95 decreases from 1.805 mm/d in SWAT to 0.108 mm/d in TOPMODEL-Experiment 7 and 0.195 mm/d in HYMOD-Experiment 7, corresponding to reductions of 94.0% and 89.2%, respectively. These results suggest that a more complex model structure does not inherently guarantee better generalization performance. Instead, effective identification of dynamic hydrological states across sub-periods and improved representation of low-flow control processes may play an important role in enhancing model performance. In summary, even when SWAT, with its larger parameter set and more complex structure, is used as the reference model, the proposed calibration framework still consistently achieves superior performance. Its core advantage is more likely to arise from the effective identification of dynamic hydrological states and dominant runoff generation and routing processes across sub-periods, rather than from higher parameter dimensionality alone.

Table S12. Multimetric performance comparison of SWAT-Experiment 1, TOPMODEL-Experiment 7, and HYMOD-Experiment 7 during calibration and evaluation.

Catchment	NSE	LNSE	RMSE_Q5	RMSE_Q20	RMSE_mid	RMSE_Q70	RMSE_Q95
Calibration period							
SWAT-Experiment 1							
Hanzhong	0.745	0.275	0.326	0.176	0.259	0.808	1.19
Mumahe	0.682	0.245	0.346	0.117	0.296	1.113	1.78
Xunhe	0.778	0.754	0.097	0.223	0.19	0.037	0.228
TOPMODEL-Experiment 7							
Hanzhong	0.933	0.822	0.791	0.194	0.054	0.102	0.143
Mumahe	0.899	0.87	0.476	0.114	0.065	0.106	0.068
Xunhe	0.938	0.797	0.199	0.069	0.138	0.064	0.05
HYMOD-Experiment 7							
Hanzhong	0.832	0.802	0.272	0.078	0.047	0.125	0.353
Mumahe	0.676	0.738	0.213	0.032	0.091	0.288	0.532
Xunhe	0.748	0.817	0.165	0.153	0.157	0.166	0.724
Evaluation period							
SWAT-Experiment 1							
Hanzhong	0.626	0.113	0.402	0.168	0.235	0.964	1.805
Mumahe	0.59	0.086	0.396	0.137	0.234	1.163	1.825
Xunhe	0.6	0.723	0.135	0.132	0.103	0.257	0.264
TOPMODEL-Experiment 7							
Hanzhong	0.908	0.791	0.23	0.172	0.166	0.145	0.108
Mumahe	0.89	0.835	0.146	0.058	0.121	0.063	0.099
Xunhe	0.853	0.757	0.368	0.486	0.312	0.133	0.137
HYMOD-Experiment 7							
Hanzhong	0.706	0.728	0.302	0.069	0.097	0.134	0.195
Mumahe	0.655	0.675	0.338	0.056	0.165	0.524	0.717
Xunhe	0.747	0.742	0.186	0.071	0.076	0.092	0.292

S5.2 Diagnostic comparison with a fixed-season control scheme

To evaluate whether the observed performance improvements arise from improved representation of catchment dynamics rather than from increased degrees of freedom, a fixed seasonal scheme (Experiment 7-Seasons) is introduced for comparison with Experiment 7, in which sub-periods are adaptively identified based on dynamic catchment characteristics. In the fixed seasonal scheme, twelve calendar months are grouped into four clusters using the Fuzzy C-Means algorithm, based on standardized monthly hydrometeorological variables, including precipitation, potential evapotranspiration, temperature, NDVI, and streamflow. The calibration procedure is kept consistent with that of Experiment 7 (Section 3.3). Cumulative distribution functions of NSE and LNSE during calibration and evaluation (Fig. S10a) show consistently higher performance under Experiment 7. During evaluation, performance remains stable under Experiment 7, whereas a clear decline is observed under the fixed seasonal scheme (Fig. S10b). These results indicate that increasing parameter dimensionality alone does not improve model generalization; performance is primarily governed by the representation of temporal variability. When parameter discretization is aligned with hydrologically coherent temporal patterns, more physically meaningful information is extracted from observational data. Regime-specific performance is further evaluated using RMSE across flow quantiles from Q_5 to Q_{95} (Fig. S10c). Lower RMSE values are consistently obtained under Experiment 7 across all flow conditions in both calibration and evaluation periods, with the largest improvements observed under extreme flow regimes, particularly for high flows (Q_5) and low flows (Q_{95}).

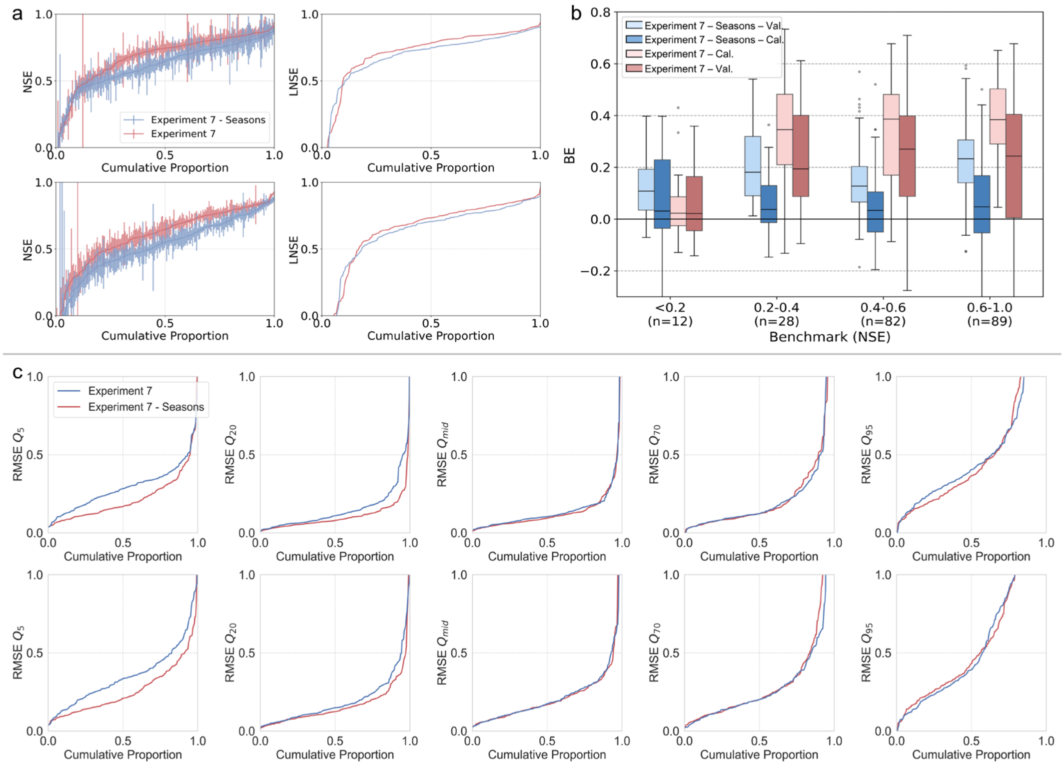

Figure S10. a, CDFs of NSE and LNSE for Experiment 7 with the fixed-season sub-period (blue) and dynamic definition (red). **b,** Benchmark Efficiency (BE) grouped by NSE ranges of Experiment 1. BE quantifies the relative improvement of a given scheme over a predefined baseline, with higher values indicating greater enhancement in model performance. Darker boxes denote calibration results, whereas lighter boxes denote evaluation results. **c,** RMSE distributions across flow phases for the same experiments. Lower values indicate improved performance. The upper panels correspond to the calibration period, and the lower panels to the evaluation period, with flow phases arranged from high to low (left to right).

S6 State variables and fluxes assessment in study cases

The results of state variables and fluxes in cases A, B, C, D, and E in the entire study period are shown in Fig. S11 to Fig. S20.

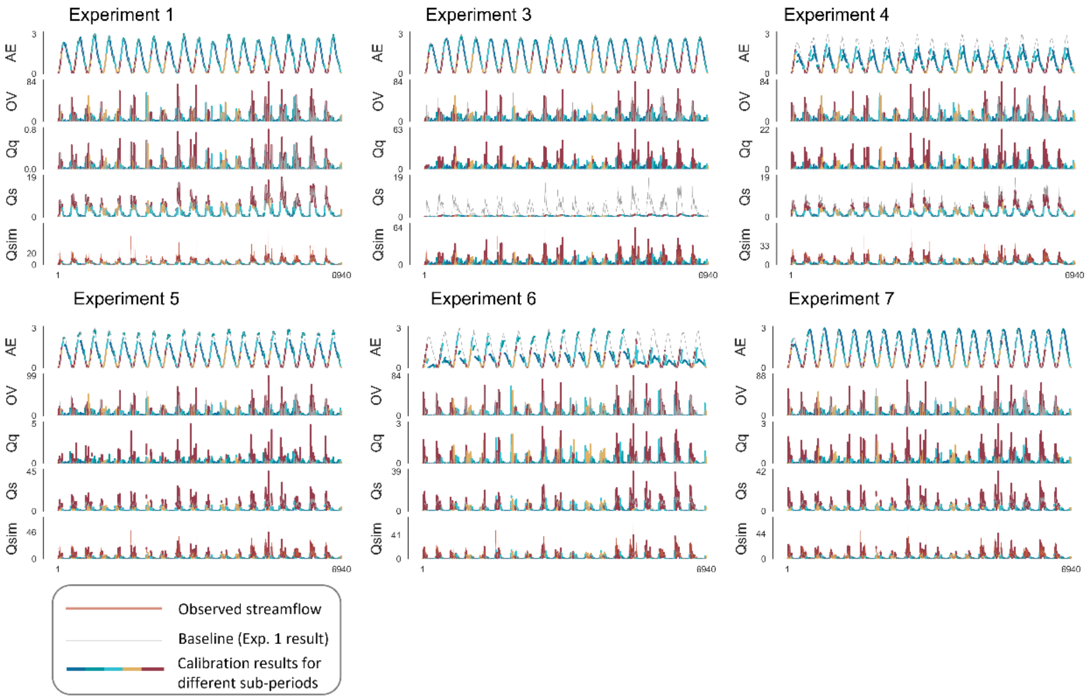

Figure S11. Flux simulation results of experiments over the entire study period for case A. The figure shows the flux simulation results from Experiments 1 to 7, with different colors representing different sub-periods. In Experiment 7, five separate calibrations are performed for five sub-periods, and the results are then aggregated to obtain the final simulation.

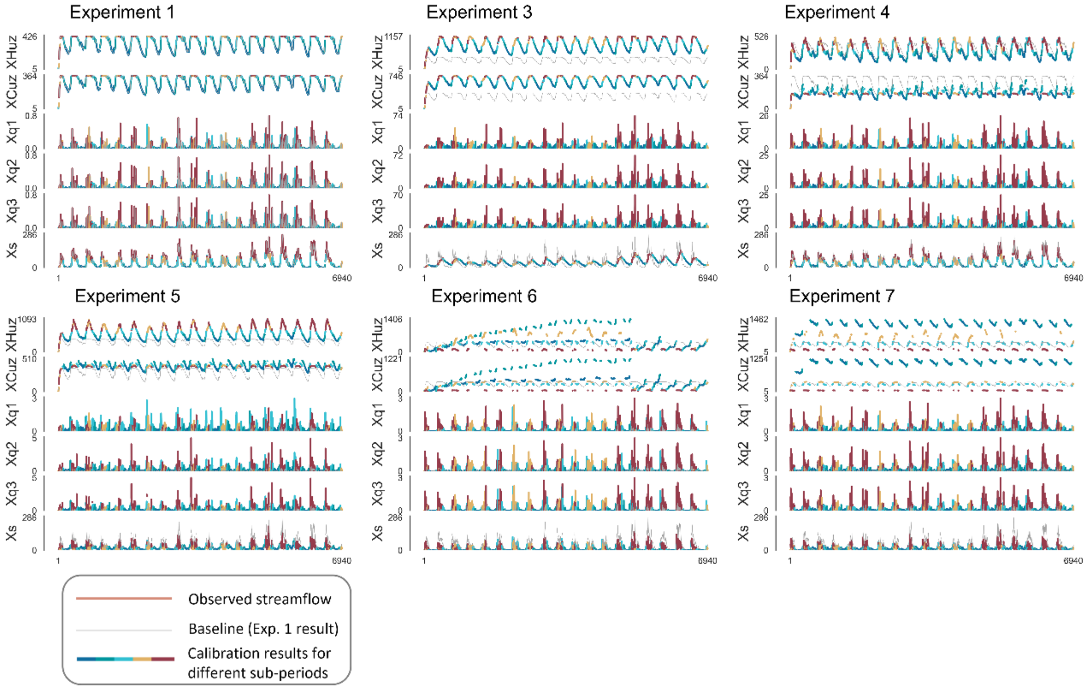

Figure S12. State variables simulation results of experiments over the entire study period for case A. The figure shows the state variable simulation results from Experiments 1 to 7, with different colors representing different sub-periods. In Experiment 7, five separate calibrations are performed for five sub-periods, and the results are then aggregated to obtain the final simulation.

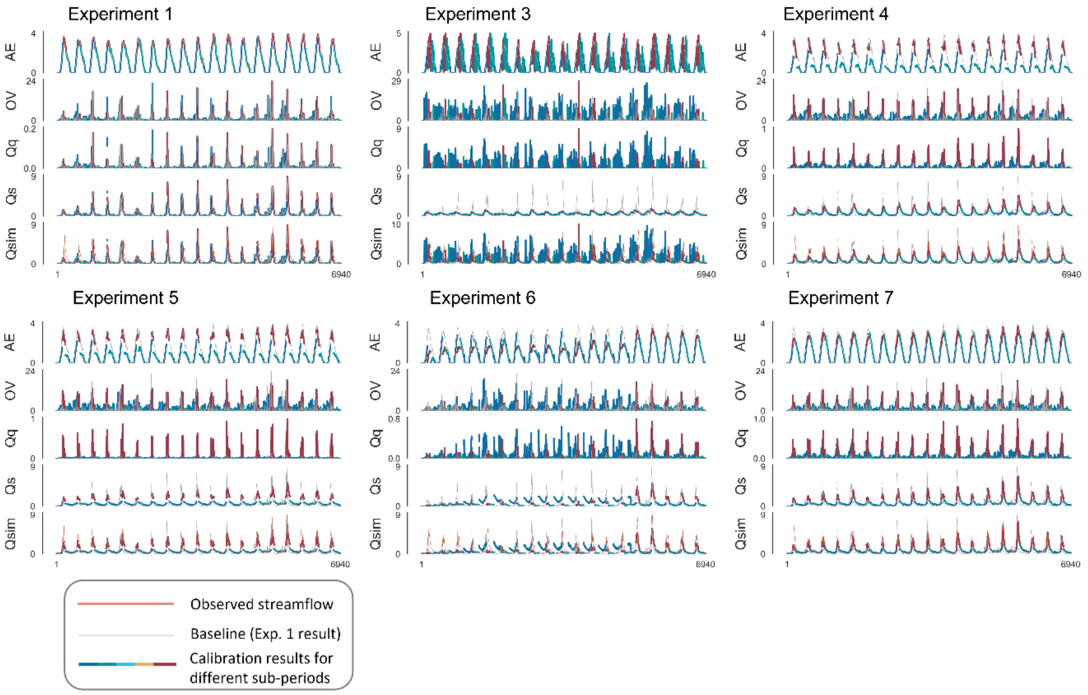

Figure S13. Fluxes simulation results of experiments over the entire study period for case B.

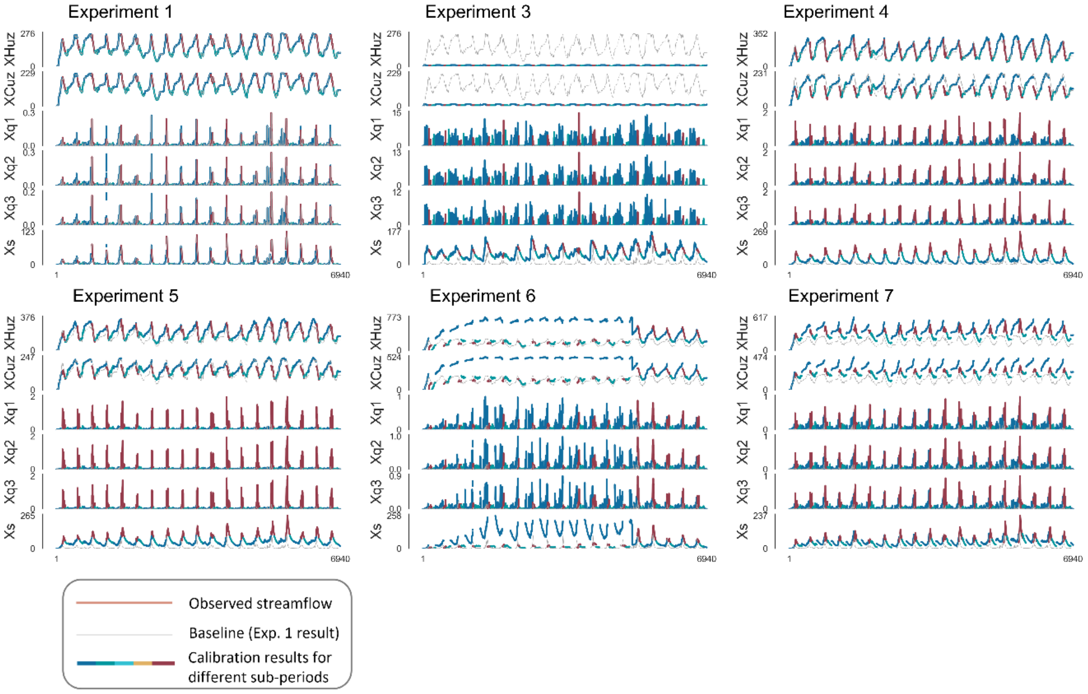

Figure S14. State variables simulation results of experiments over the entire study period for case B.

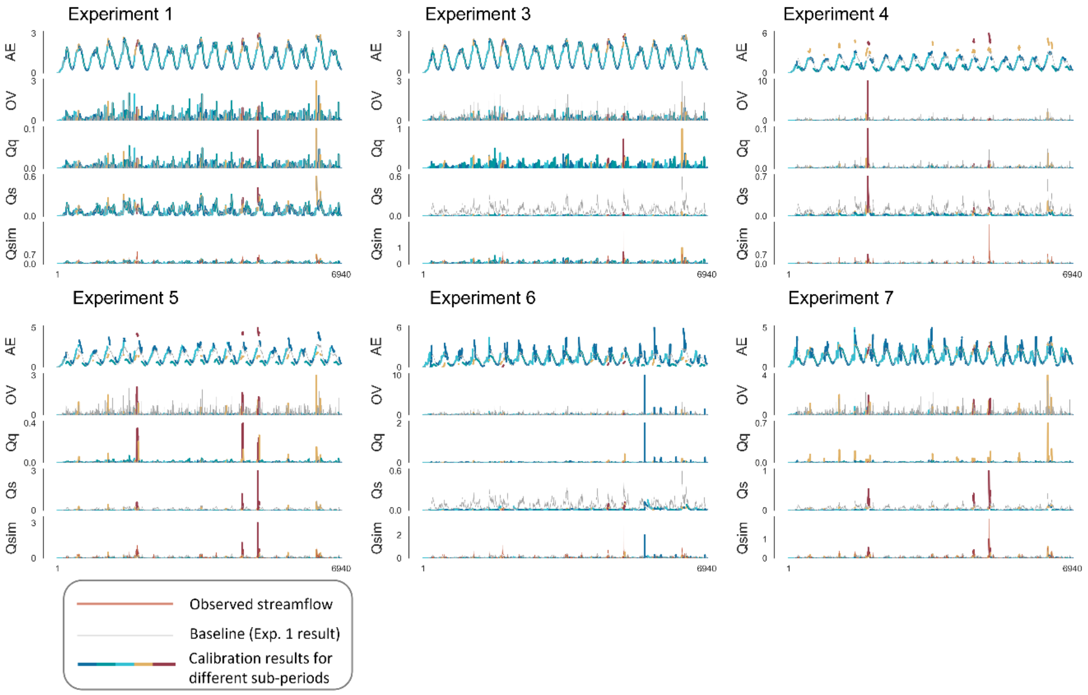

Figure S15. Fluxes simulation results of experiments over the entire study period for case C.

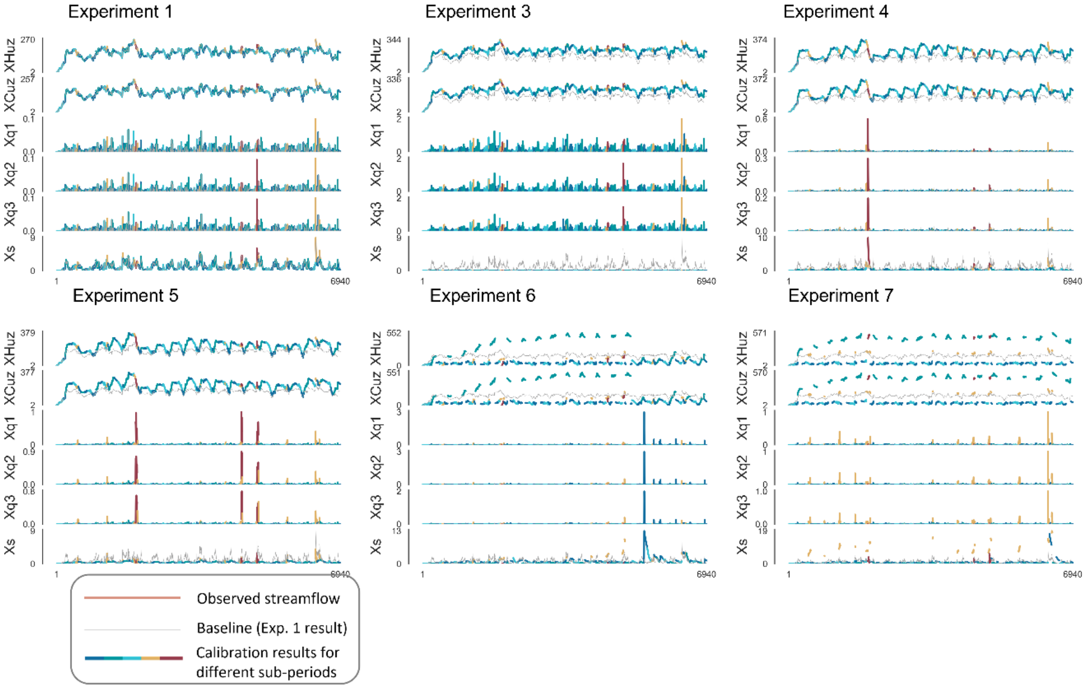

Figure S16. State variables simulation results of experiments over the entire study period for case C.

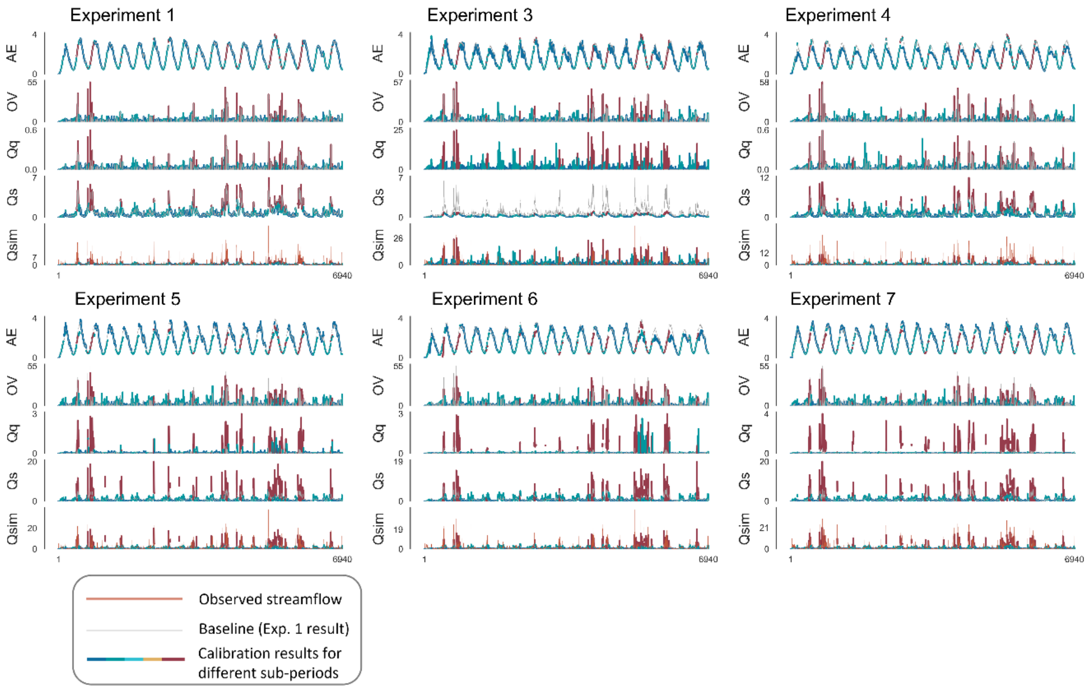

Figure S17. Fluxes simulation results of experiments over the entire study period for case D.

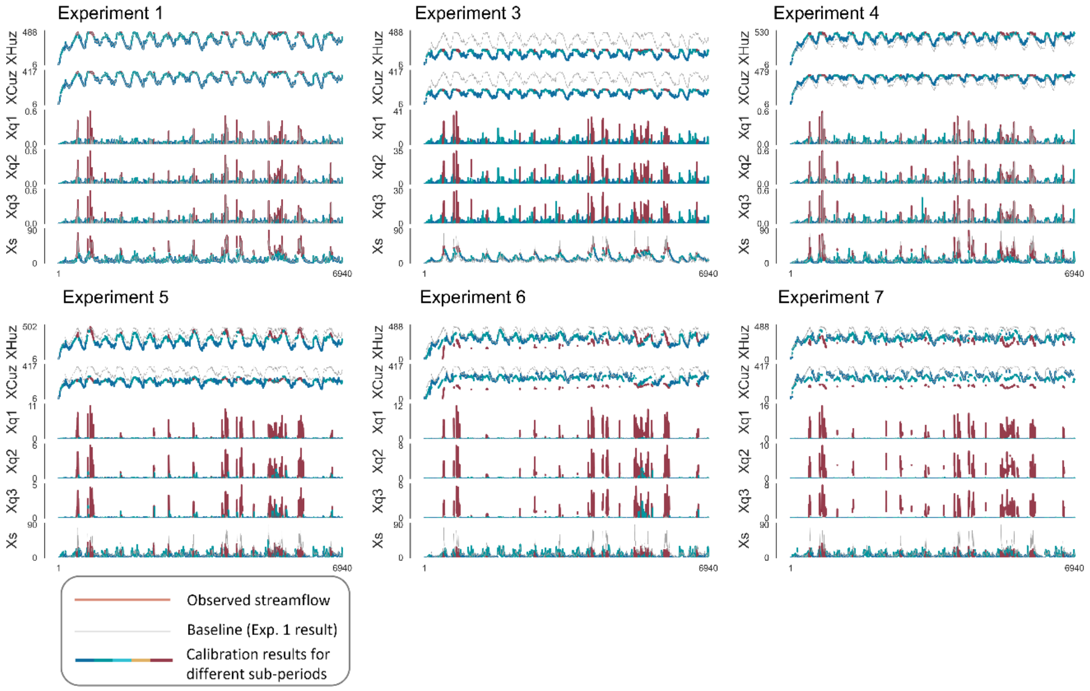

Figure S18. State variables simulation results of experiments over the entire study period for case D.

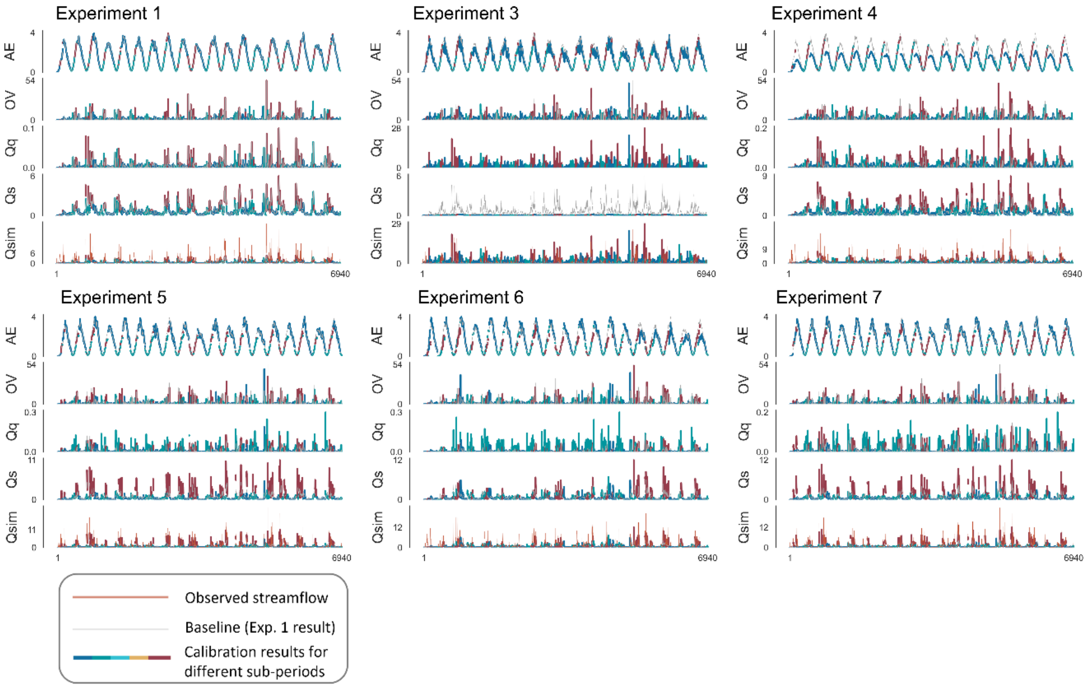

Figure S19. Fluxes simulation results of experiments over the entire study period for case E.

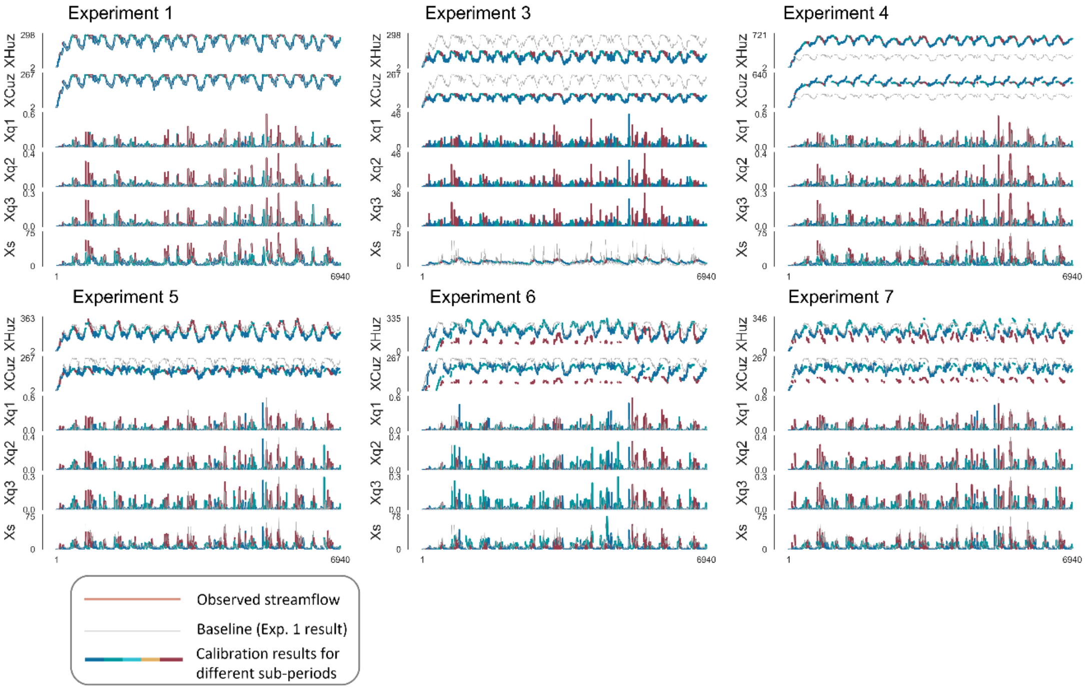

Figure S20. State variables simulation results of experiments over the entire study period for case E.

The comparative analysis of Experiments 1, 5, and 7 further illustrates the performance improvements introduced by Experiments 5 and 7. Fig. S21-S25 illustrates the flux mapping of various sub-periods in five study cases, comparing Experiments 1, 5, and 7. To facilitate comparison, the results of Experiment 1 are also presented by the same sub-periods as Experiments 5 and 7. In case A, compared to Experiment 1, Experiment 7 consistently identified more optimal parameter sets with smaller objective function values within the same period (Fig. S21). 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. In Experiment 7, clusters of scatter points of the same colour 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, Experiment 7 possesses a narrower range of optimal equifinality parameters during the parameter evolution process, reducing the model's internal fluxes equifinality and uncertainty. Further comparison with Experiment 5 (Fig. S21c) 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 (green) 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). 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. S21b) 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. S24). 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.

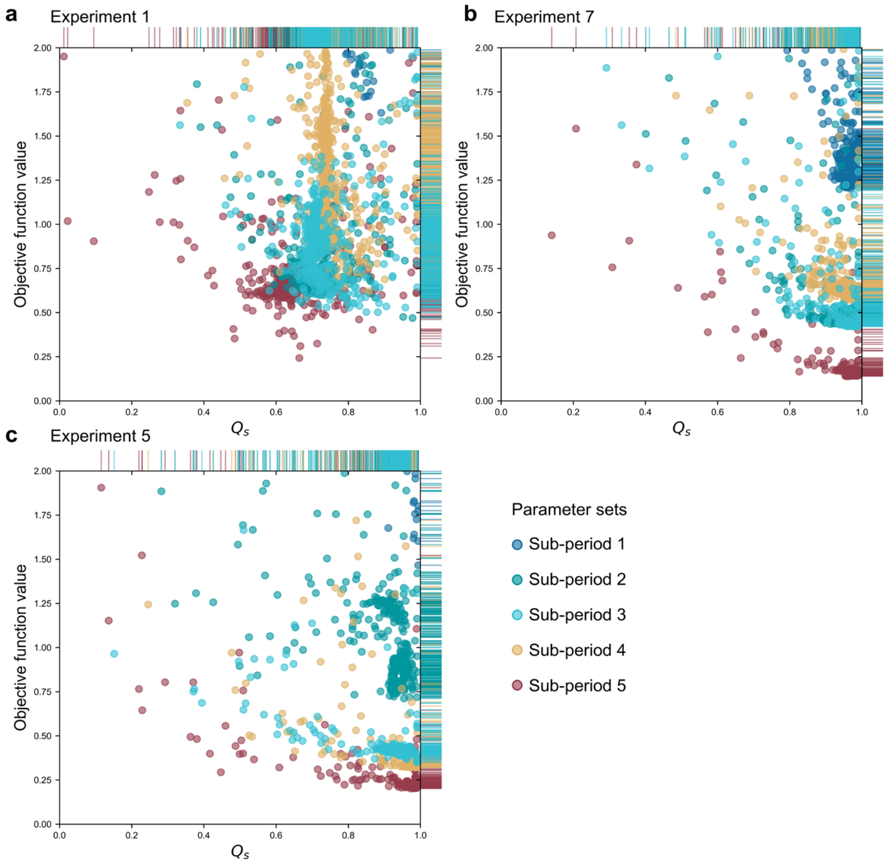

Figure S21. a, Flux mapping for case A in the conventional scheme, **b,** Experiment 7, and **c,** Experiment 5, where the horizontal axis represents the proportion of Qs in the runoff. Each scatter point in the figures represents a parameter set generated during the SCE-UA algorithm optimization process. The colour and relative position of each scatter point on the

395 axes illustrate the variation in runoff components for sub-periods under specific parameter sets, as well as the corresponding objective function value.

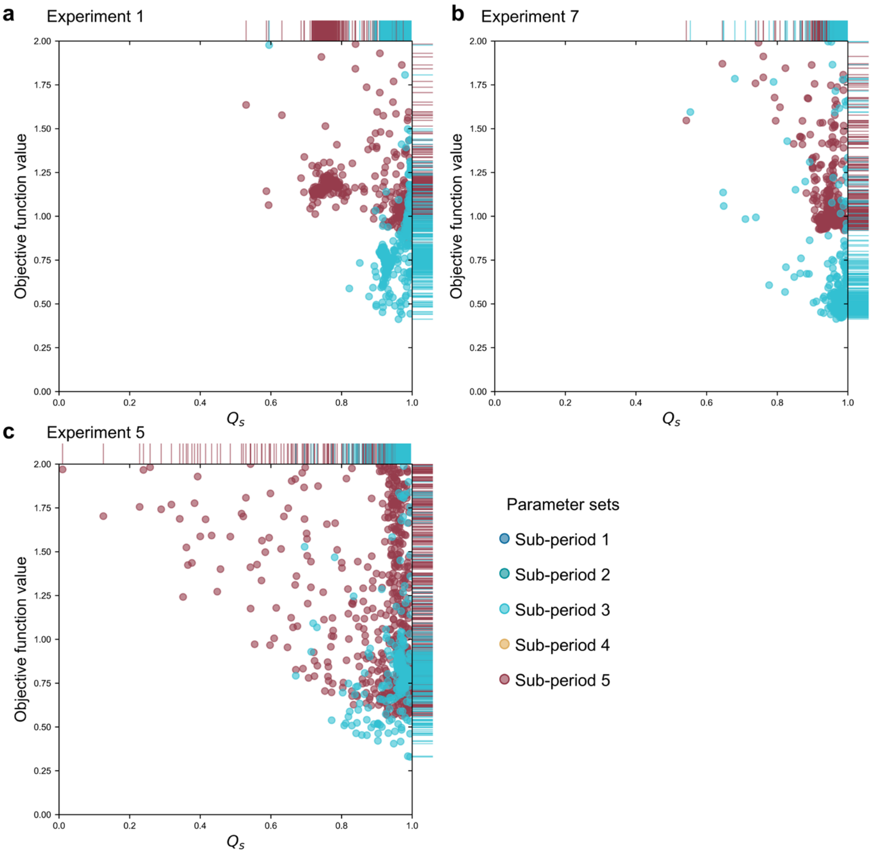

Figure S22. Flux mapping of the study case B. The left image is Experiment 1, the right image represents Experiment 7, and the lower left image is Experiment 5.

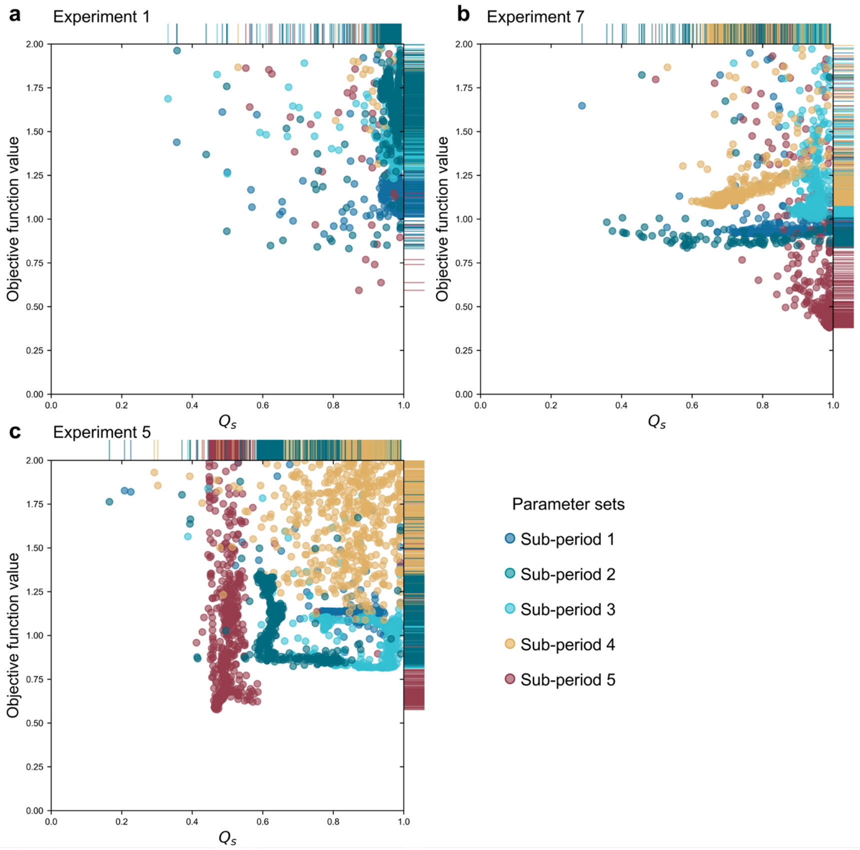

Figure S23. Flux mapping of the study case C. The left image is Experiment 1, the right image represents Experiment 7, and the lower left image is Experiment 5.

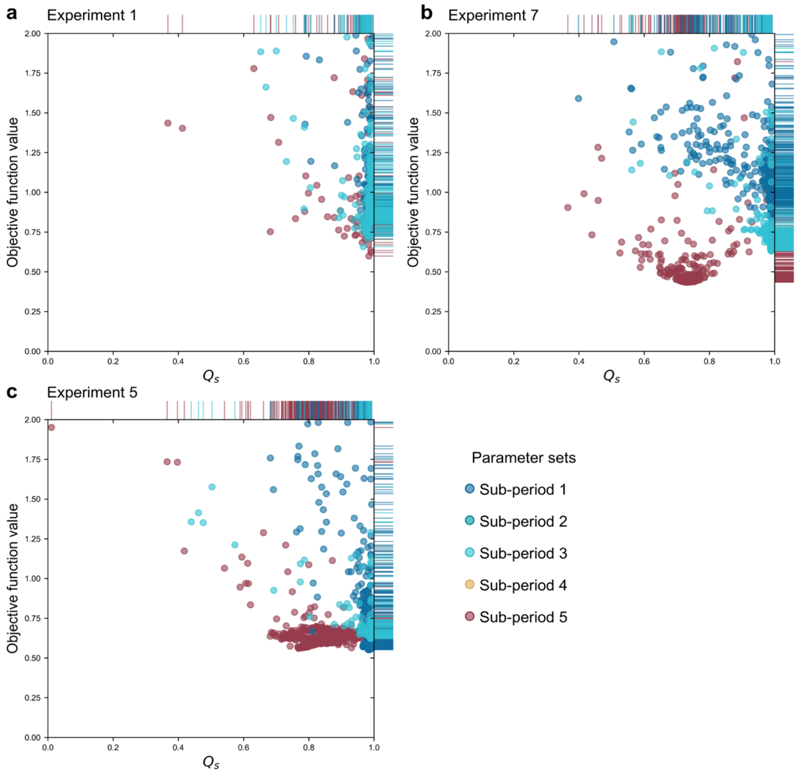

Figure S24. Flux mapping of study case D. The left image is Experiment 1, the right image represents Experiment 7, and the lower left image is Experiment 5.

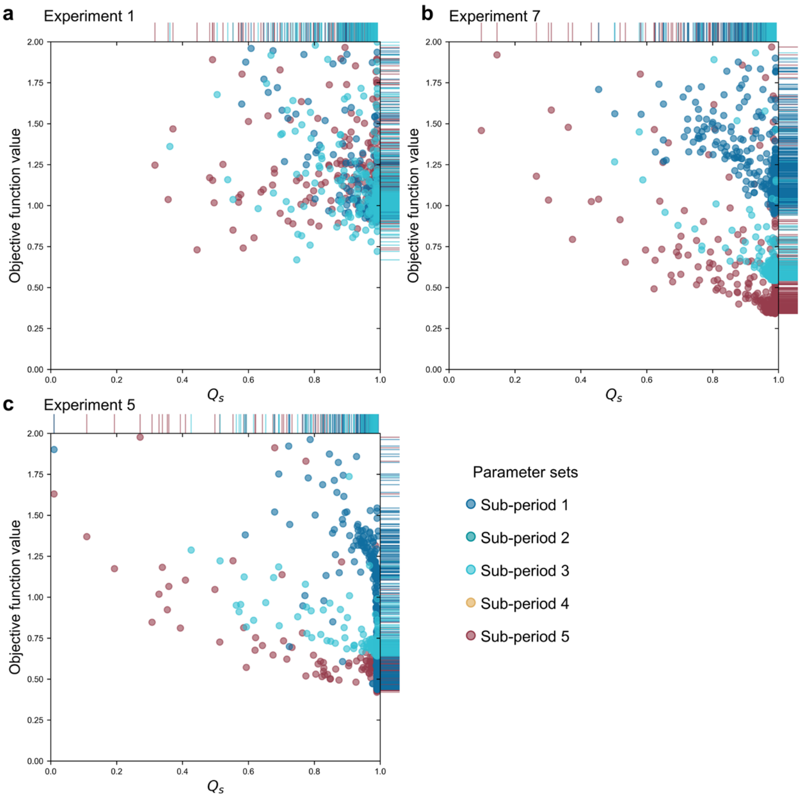

Figure S25. Flux mapping of the study case E. The left image is Experiment 1, the right image represents Experiment 7, and the lower left image is Experiment 5.

S7 Parameters

The dynamic parameter sets, optimized by various calibration experiments across five case studies, are depicted in Fig. S26. 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 five case studies. In Experiment 5, all parameters vary across sub-periods, as illustrated in Fig. S26. 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 6 and 7, certain parameters, such as K_s, exhibited minimal correlation with sub-period characteristics within catchments. As indicated in Fig. S26, 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-periods. However, this correlation is not significant. Furthermore, K_q 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. The phenomenon reflects the uncertainty inherent in model parameters and structure. In sum, compared to time-invariant schemes, Experiments 5 and 7 exhibit superior performance in identifying key parameters and their responses to catchment dynamics. The dynamic characteristic of parameters emphasizes the importance of calibration across sub-periods. Although the dynamic parameter set enhanced the model's response capability to catchment dynamics, the overall response of the entire dynamic parameter set to catchment dynamics remains relatively poor. The reasons for the improved simulation performance of the dynamic parameter set will be explored in the discussion section.

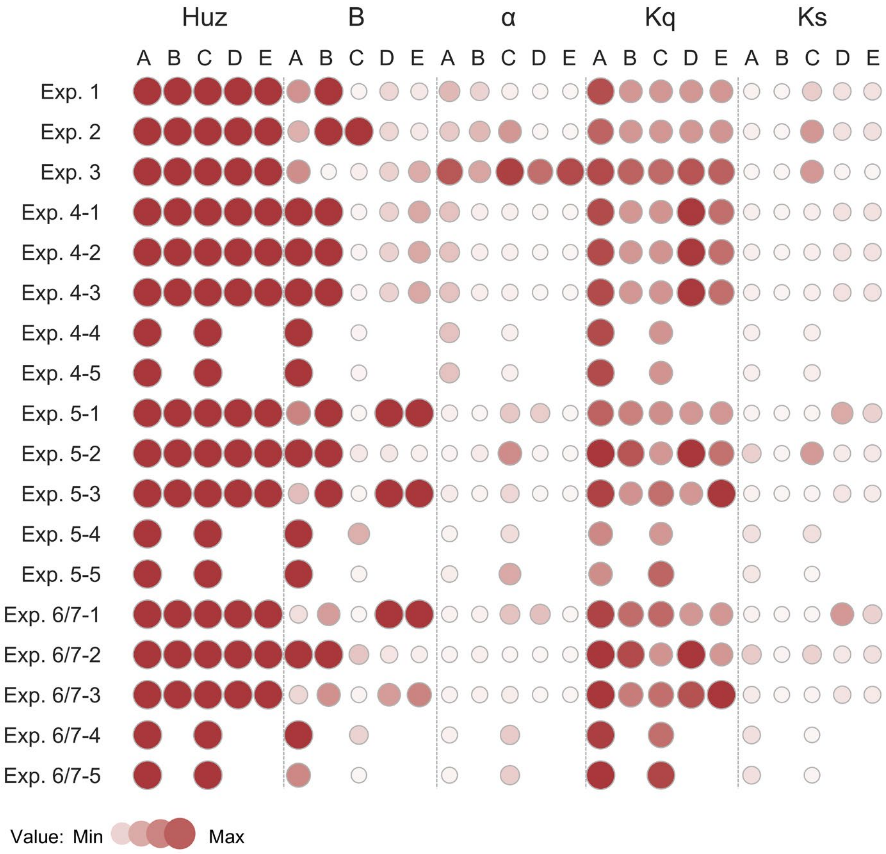

Figure S26. Assessment of dynamic parameter sets across various calibration experiments. The parameter boundaries shown in the figure are H_{uz} (0-1500), B (0-2), α (0-1), K_q (0.5-1), and K_s (0-0.5).

S8 Correlation between model parameters

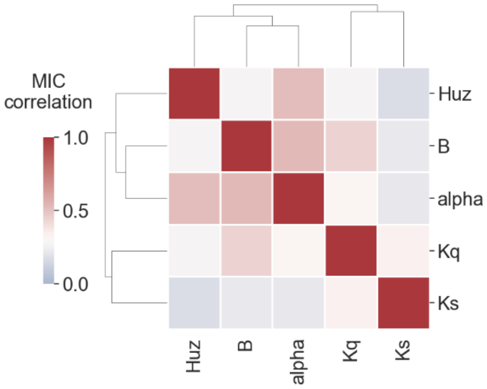

Figure S27. Correlation between model parameters in the study case B.

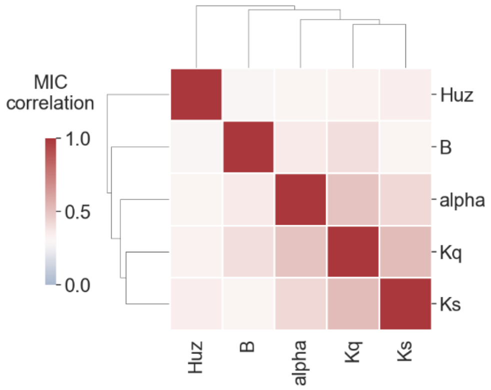

Figure S28. Correlation between model parameters in the study case C.

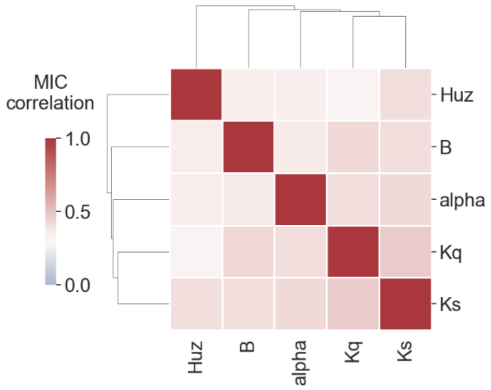

Figure S29. Correlation between model parameters in the study case D.

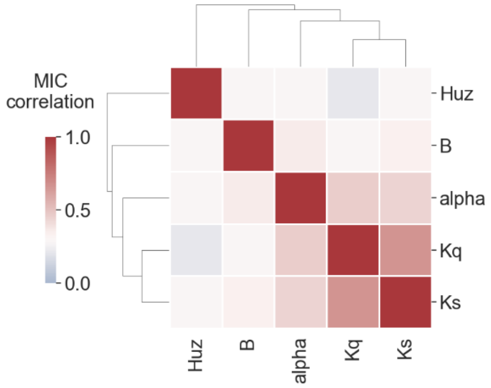

Figure S30. Correlation between model parameters in the study case E.

S9 Violin plots

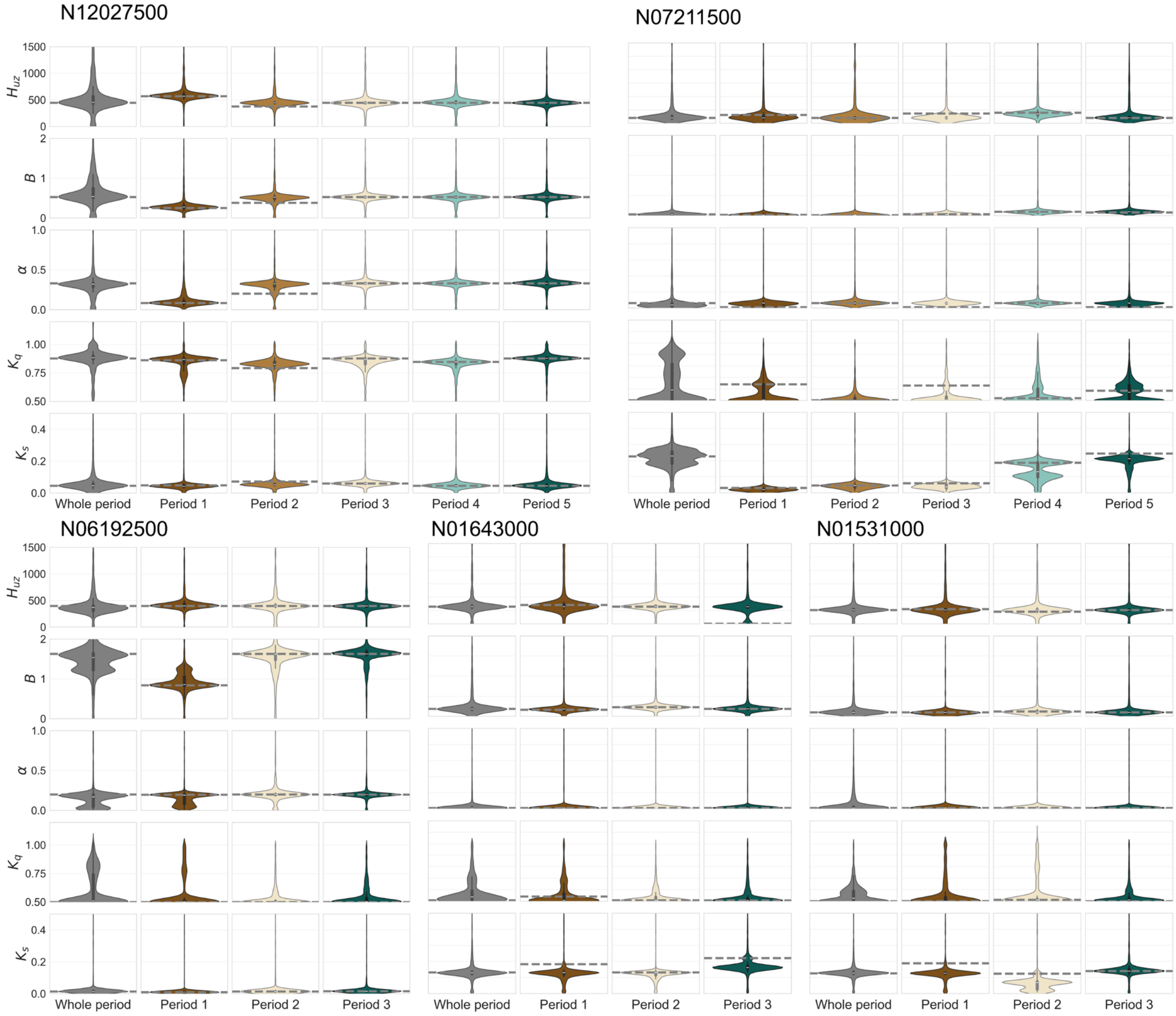

Figure S31. The distributions of the optimal parameter spaces (Experiment 7 with added parameter constraints) in the subperiods for different climatic and land-surface conditions in the five study cases. Each violin plot represents one parameter space, with parameter values plotted on the y-axis. The width of the violin plot indicates the distribution of the parameter. The parameter boundaries shown in the figure are H_{uz} (0-1500), B (0-2), α (0-1), K_q (0.5-1), and K_s (0-0.5).

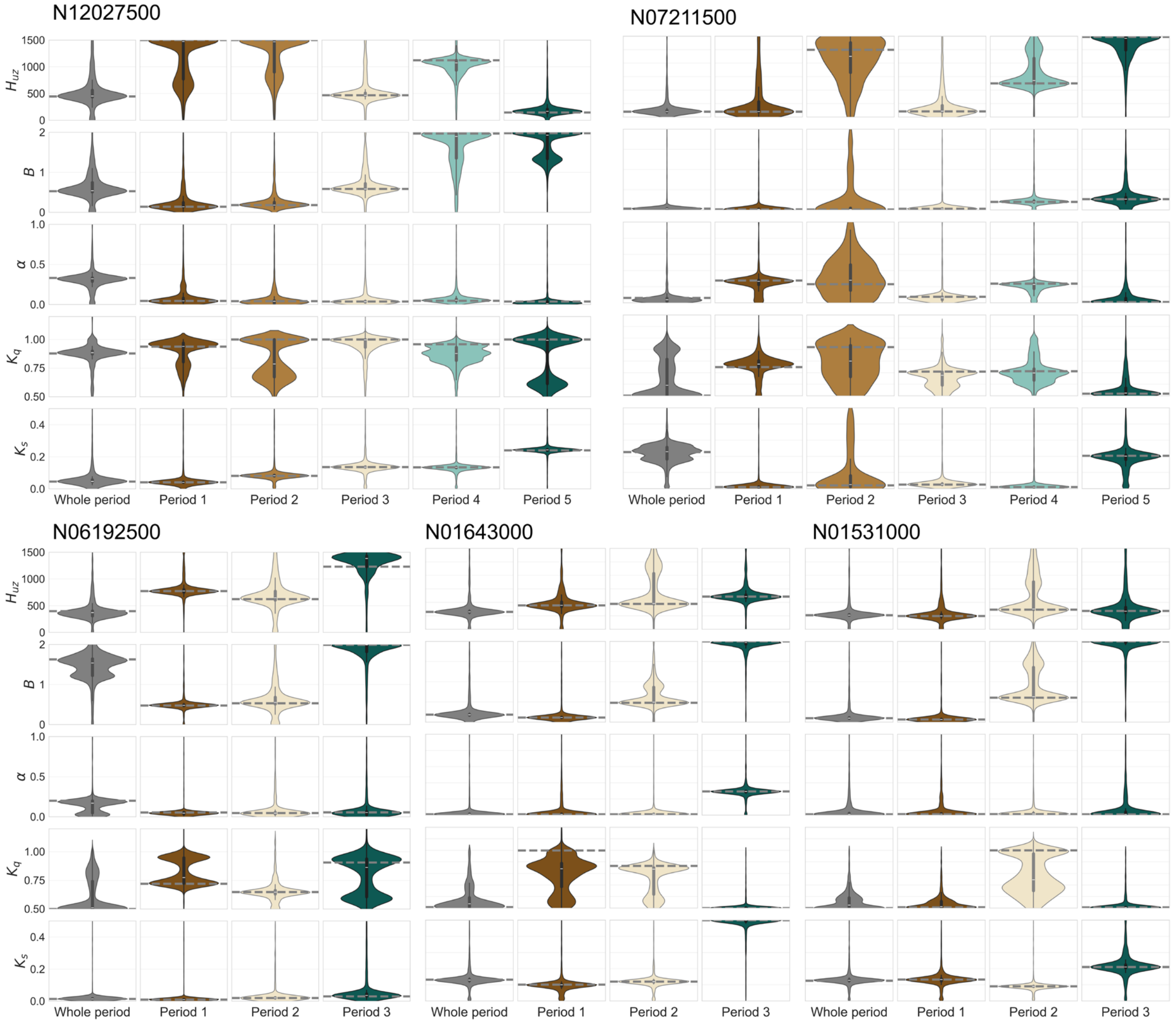

Figure S32. The distributions of the optimal parameter spaces (Experiment 7) in the subperiods for different climatic and land-surface conditions in the five study cases. Each violin plot represents one parameter space, with parameter values plotted on the y-axis. The width of the violin plot indicates the distribution of the parameter. The parameter boundaries shown in the figure are H_{uz} (0-1500), B (0-2), α (0-1), K_q (0.5-1), and K_s (0-0.5).

References

Arnold, J. G., Moriasi, D. N., Gassman, P. W., Abbaspour, K. C., White, M. J., Srinivasan, R., Santhi, C., Harmel, R. D., Van Griensven, A., and Van Liew, M. W.: SWAT: Model use, calibration, and validation, Transactions of the ASABE, 55, 1491-1508, https://doi.org/10.13031/2013.42256, 2012.

Beven, K.: TOPMODEL: A critique, Hydrological Processes, 11(9), 1069-1085, https://doi.org/10.1002/(SICI)1099-1085(199707)11:9<1069::AID-HYP545>3.0.CO;2-O, 1997.

Bezdek, J. C., Ehrlich, R., and Full, W.: FCM: The fuzzy c-means clustering algorithm, Computers & Geosciences, 10, 191-203, https://doi.org/10.1016/0098-3004(84)90020-7, 1984.

Duan, Q., Sorooshian, S., and Gupta, V.: Effective and efficient global optimization for conceptual rainfall-runoff models, Water Resources Research, 28, 1015-1031, https://doi.org/10.1029/91WR02985, 2010.

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.

Duan, Q. Y., Sorooshian, S., and Gupta, V. K.: Optimal use of the SCE-UA global optimization method for calibrating watershed models, Journal of Hydrology, 158, 265-284, https://doi.org/10.1016/0022-1694(94)90057-4, 1994.

Dunn, J. C.: A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters, Journal of Cybernetics 3(3), 32–57, https://doi.org/10.1080/01969727308546046, 1973.

Eckhardt, K. and Arnold, J. G.: Automatic calibration of a distributed catchment model, Journal of Hydrology, 251, 103-109, https://doi.org/10.1016/S0022-1694(01)00429-2, 2001.

Fan, Y. R., Huang, G. H., Li, Y. P., Wang, X. Q., Li, Z., and Jin, L.: Development of PCA-based cluster quantile regression (PCA-CQR) framework for streamflow prediction: Application to the Xiangxi river watershed, China, Applied Soft Computing, 51, 280-293, https://doi.org/10.1016/j.asoc.2016.11.039, 2017.

Frey, D. and Pimentel, R.: Principal component analysis and factor analysis, 1978.

Fuentes-Andino, D., K. Beven, S. Halldin, C. Y. Xu, J. E. Reynolds, and G. Di Baldassarre.: Reproducing an extreme flood with uncertain post-event information, Hydrology and Earth System Sciences, 21(7), 3597-3618, https://doi.org/10.5194/hess-21-3597-2017, 2017.

Ghumman, A. R., Jamaan, M., Ahmad, A., Shafiquzzaman, M., Haider, H., Al Salamah, I. S., and Ghazaw, Y. M.: Simulation of pan-evaporation using penman and hamon equations and artificial intelligence techniques, Water, 13, 793, https://doi.org/10.3390/w13060793, 2021.

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.

Hamon, W. R.: Estimating potential evapotranspiration, Journal of the Hydraulics Division, 87, 107-120, https://doi.org/10.1061/JYCEAJ.0000599, 1961.

Hathaway, R. J. and Bezdek, J. C.: Fuzzy c-means clustering of incomplete data, IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 31, 735-744, https://doi.org/10.1109/3477.956035, 2001.

Hock, R.: Temperature index melt modelling in mountain areas, Journal of Hydrology, 282, 104-115, https://doi.org/10.1016/S0022-1694(03)00257-9, 2003.

Khakbaz, F. and Kazeminezhad, M.: Work hardening and mechanical properties of severely deformed AA3003 by constrained groove pressing, Journal of Manufacturing Processes, 14, 20-25, https://doi.org/10.1016/j.jmapro.2011.07.001, 2012.

Kunnath-Poovakka, A., & Eldho, T. I.: A comparative study of conceptual rainfall-runoff models GR4J, AWBM and Sacramento at catchments in the upper Godavari river basin, India. Journal of Earth System Science, 128(2), 33. https://doi.org/10.1007/s12040-018-1055-8, 2019.

McCabe, G. J., Hay, L. E., Bock, A., Markstrom, S. L., and Atkinson, R. D.: Inter-annual and spatial variability of Hamon potential evapotranspiration model coefficients, Journal of Hydrology, 521, 389-394, https://doi.org/10.1016/j.jhydrol.2014.12.006, 2015.

Morton, F. I.: Catchment evaporation and potential evaporation further development of a climatologic relationship, Journal of Hydrology, 12, 81-99, https://doi.org/10.1016/0022-1694(71)90102-8, 1971.

Perrin, C., Michel, C., & Andréassian, V.: Improvement of a parsimonious model for streamflow simulation. Journal of Hydrology, 279(1–4), 275–289. https://doi.org/10.1016/S0022-1694(03)00225-7, 2003.

Reshef, D. N., Reshef, Y. A., Finucane, H. K., Grossman, S. R., McVean, G., Turnbaugh, P. J., Lander, E. S., Mitzenmacher, M., and Sabeti, P. C.: Detecting novel associations in large data sets, Science, 334, 1518-1524, https://doi.org/10.1126/science.1205438, 2011.

Sorooshian, S., Duan, Q., and Gupta, V. K.: Calibration of rainfall-runoff models: Application of global optimization to the Sacramento Soil Moisture Accounting Model, Water Resources Research, 29, 1185-1194, https://doi.org/10.1029/92WR02617, 1993.

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, Journal of Hydrology: Regional Studies, 44, https://doi.org/10.1016/j.ejrh.2022.101215, 2022.

Wold, S., Esbensen, K., and Geladi, P.: Principal component analysis, Chemometrics and Intelligent Laboratory Systems, 2, 37-52, https://doi.org/10.1016/0169-7439(87)80084-9, 1987.

Wright, J., Ganesh, A., Rao, S., Peng, Y., and Ma, Y.: Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization, Advances in Neural Information Processing Systems, 22, ISBN: 9781615679119, 2009.

Zhang, Y., Jia, S., Huang, H., Qiu, J., and Zhou, C.: A novel algorithm for the precise calculation of the maximal information coefficient, Scientific Reports, 4, 6662, https://doi.org/10.1038/srep06662, 2014.