Hydrol. Earth Syst. Sci., 30, 3425–3438, 2026
https://doi.org/10.5194/hess-30-3425-2026

Hydrology and Earth System Sciences
Open Access

Symbolic regression-based regionalization of baseflow separation parameter using catchment-scale characteristics

Yongen Lin[1,3], Dagang Wang[1,2,3], Yiwen Mei[3], Jinxin Zhu[1,3], Huan Wu[4,5], Shuo Wang[6], Zhonghou Xu[7], Asaad Y. Shamseldin[8,✝], and Emmanouil N. Anagnostou[9]

[1]School of Geography and Planning, Sun Yat-sen University, Guangzhou, China
[2]Guangdong Key Laboratory for Urbanization and Geo-simulation, Sun Yat-sen University, Guangzhou, China
[3]Carbon-Water Research Station in Karst Regions of Northern Guangdong, Sun Yat-sen University, Guangzhou, China
[4]School of Atmospheric Science, Sun Yat-sen University, Zhuhai, Guangdong, China
[5]State Key Laboratory of Climate System Prediction and Risk Management, Nanjing University of Information Science and Technology, Nanjing, China
[6]Department of Land Surveying and Geo-informatics, Hong Kong Polytechnic University, Kowloon, Hong Kong Special Administrative Region, China
[7]Earth Sciences New Zealand, Hamilton 3216, New Zealand
[8]Department of Civil and Environmental Engineering, The University of Auckland, Auckland 1010, New Zealand
[9]Department of Civil and Environmental Engineering, University of Connecticut, Storrs, USA
✝deceased

Correspondence: Dagang Wang (wangdag@mail.sysu.edu.cn)

Received: 3 December 2025 – Discussion started: 12 December 2025
Revised: 17 May 2026 – Accepted: 18 May 2026 – Published: 3 June 2026

Abstract. Accurate separation of baseflow from streamflow is of utmost importance for understanding catchment hydrological processes and supporting effective water resource management. The Smooth Minima Method is a common baseflow separation technique with a segment length parameter (N) representing the catchment average flow event duration. N is usually predicted by a power function with catchment area or defaults to 5 d. Yet these estimations are insufficient given the multivariate nature of N with other catchment attributes. In this study, we employ symbolic regression (SR) to search for possible formulation of N with a range of catchment attributes based on 855 catchments across the Contiguous United States. We ultimately identify three mathematical expressions of increasing complexity, achieving R^2 values of 0.48, 0.52, and 0.55, compared to 0.23 and −0.84 for the power function and constant values. The three expressions reveal that N increases following a power-law relationship with catchment area (A) and catchment-averaged soil saturated hydraulic conductivity (K_{sat}) with decreasing rates, while it increases linearly with snow day fraction (f_{SWE}). The effects of K_{sat} and f_{SWE} on N are particularly pronounced for larger values ($K_{\mathrm{sat}} > 25\,\mathrm{mm\,h^{-1}}$ and $f_{\mathrm{SWE}} > 0.4$) and smaller area ($A < 100\,\mathrm{km^2}$). The different calculations of N are also evaluated in baseflow separation, revealing higher medians of Kling-Gupta Efficiency of at least 0.84, outperforming the literature-suggested formulas for a maximum increment of 0.22. This study highlights the potential of SR for uncovering physically meaningful formulas in optimal baseflow separation.

1 Introduction

Baseflow is an essential component of streamflow, primarily originating from groundwater, deep interflow, snow melting, and other delayed sources (Xie et al., 2020; Wang et al., 2022; Stoelzle et al., 2020). The proportion of baseflow in streamflow reflects the complex interactions between surface water and groundwater systems (Pelletier and Andréassian, 2020; Xie et al., 2024), and understanding this proportion can aid in water resources management and riverine ecosystem conservation (Yan et al., 2023; Tan et al., 2020). Baseflow is difficult to measure directly and it is usually estimated using baseflow separation methods (Humphrey et al., 2022; Stewart, 2015), which take continuous streamflow data as the only inputs. The performance of baseflow separation is sensitive to parameters of the separation methods, which, if not optimized, may lead to unrealistic baseflow dynamics (Mei et al., 2024b). Incorporating environmental tracer data for parameter optimization is a common practice, as it ensures reliable baseflow separation by maintaining dual mass balance for both tracer concentration and streamflow volume (Cartwright, 2022; Hagedorn, 2020; McMahon and Nathan, 2021). Commonly used tracers include specific electrical conductivity (SEC), turbidity, and stable isotopes, among which SEC is the most widely applied due to its routine availability in many monitoring programs (Mei et al., 2024b). However, a critical challenge arises as this method is not applicable for catchments without continuous tracer data, which unfortunately constitutes the majority of gaged catchments worldwide (Thorslund and van Vliet, 2020; Hou et al., 2024). This limitation hinders accurate quantification of baseflow in most global catchments, despite their long-term streamflow data.

To optimize baseflow separation for gaged catchments lacking continuous environmental tracer data, a viable approach is to transfer optimized parameters from other catchments (Klotz et al., 2017; Feigl et al., 2020). Specifically, prediction models can be developed for these optimized parameters based on factors representing catchment physical conditions. This approach is fundamentally rooted in the hydrological similarity theory, which assumes that baseflow parameters reflect the catchment's hydrological signatures and relate to its physical characteristics (Zhang et al., 2020; Gnann et al., 2021; McMillan et al., 2022; Price, 2011). While this parameter regionalization approach is widely used for transferring calibrated parameters to un-gaged catchments in hydrological modeling (Klotz et al., 2017; Feigl et al., 2020), its application in the context of baseflow parameters is unexplored. The smooth minima method (SMM) is a widely used method for baseflow separation, with a segment length parameter (N) representing the average streamflow delay and catchment response (Stoelzle et al., 2020). This parameter is often defaulted to 5 d or estimated by an empirical power-law relationship with drainage area (A) as $N = 1.6 \times A^{0.2}$ (Aksoy et al., 2008). However, Lin et al. (2026) found that while A is the most influential predictor, incorporating additional factors representing geomorphology, climate, soil hydraulics properties, and human activities into the nonparametric random forest (RF) model yielded more accurate estimates than the power function. This highlights the complex interactions between streamflow delay and the diverse catchment characteristics (Price, 2011; Stoelzle et al., 2020).

Despite higher prediction accuracy, the RF-based regionalization model does not provide an explicit analytical expression linking catchment attributes to parameter N (Rudin, 2019). Although tree-based models can be interpreted using post-hoc tools such as SHAP values and partial dependence plots, they provide only approximate, model-dependent interpretations rather than explicit functional relationships (Rudin, 2019; Makke and Chawla, 2024). Moreover, they typically describe average effects while overlooking higher-order interactions and may be sensitive to feature correlation, potentially leading to unstable or misleading explanations (Apley and Zhu, 2020; Sundararajan and Najmi, 2020). To develop interpretable regionalization models with structural transparency for the optimal baseflow parameters N, this study employs an emerging machine learning technique called symbolic regression (SR), which focuses on data-driven identification of explicit mathematical expressions to describe relationships between predictors and target variable (Koza, 1994; Song et al., 2024). In recent years, SR has been increasingly applied in hydrology to uncover governing relationships in complex environmental systems, owing to its ability to balance predictive performance and interpretability (Chadalawada et al., 2020; Sheta et al., 2023). One example is the use of SR to extract explicit functional relationships between catchment attributes and hydrological model parameters for ungaged catchments (Li et al., 2024; Feigl et al., 2022). Unlike a "black-box" model such as RF, SR derives explicit and concise equations that identify underlying data patterns, while mitigating overfitting through complexity control (Kronberger et al., 2022; Wilstrup and Kasak, 2021). This structural transparency enables direct interpretation of how catchment attributes govern baseflow parameter values, which in terms influence the partitioning of streamflow (Klotz et al., 2017; Feigl et al., 2020; Sheta et al., 2023).

To evaluate the effectiveness for regionalization of baseflow parameters, this study applies SR to model the segment length parameter of SMM and addresses three objectives: (a) assess the complexity-performance trade-off of SR-derived formulas for N; (b) explore the functional relationships between N and catchment attributes in the SR formulas; and (c) evaluate the different Ns calculated by SR in baseflow separation. This study should not be viewed as an effort to assert a superior utility of SR over other machine learning models in the regionalization of baseflow parameters. Instead, the SR formulas serve as post-hoc interpretability tools to complement other black box models, enhancing the transparency of the underlying relationship between hydrological signatures and catchment attributes (Rudin, 2019).

2 Study catchments and data

This study used the baseflow dataset produced by Mei et al. (2024a), which contains the segment length parameter of SMM optimized by SEC for 987 catchments across the Conterminous United States (CONUS). To ensure the reliability of baseflow separation using the SMM method, catchments exhibiting suboptimal baseflow separation performance with a Kling Gupta Efficiency between estimated and observed SEC below 0.5 were excluded (Mei et al., 2024a). Additionally, catchments with incomplete attribute data (see the next paragraph for details) were also eliminated. After applying these criteria, a total of 855 catchments remained. The spatial distribution of these catchments, overlaid with drainage area, mean streamflow, and baseflow index (BFI), is depicted in Fig. 1a–c, showing significant diversity. The optimal N parameters for these catchments are depicted in Fig. 1d with most values smaller than 17 d. Larger N values tend to occur in mountainous regions characterized by drier climates and greater snow persistence with implicit spatial patterns, which may be attributed to the multifactor controls of N (Lin et al., 2026). The level 2 and level 4 Hydrological Unit Codes (HUC2 and HUC4), which represent large and sub-regional hydrologic basins in the United States, are provided for better referencing the spatial distributions of results in the analysis (Fig. 1e, f).

Lin et al. (2026) identified 13 catchment characteristics representing the geomorphology, climate, soil hydraulic properties, and water usage, that significantly influence the SMM parameter N (Table 1, rows 1–13). In addition to these predictors, we added two more related to snow processes and vegetation dynamics (Table 1, rows 14–15), as previous studies have shown that these factors can substantially affect baseflow generation (Price, 2011; Stoelzle et al., 2020; Xie et al., 2022). The data of these 15 variables are all obtained from Lin et al. (2025). To reduce the dimensionality of the predictor space, we examined the mutual information (MI) between candidate predictors. Variable pairs exhibiting MI values greater than 0.5 were considered to contain substantial shared information (Fig. S1 in the Supplement). In such cases, only one representative variable was retained to prevent redundant contributions in training the symbolic regression models. After this screening step, the number of predictors was reduced to nine (highlighted in blue in Table 1), which were then used as inputs for the SR analysis.

3 Methods

3.1 Smooth Minima Baseflow Separation

SMM is a widely used baseflow separation method (Aksoy et al., 2008; Piggott et al., 2005; Xie et al., 2020; Tan et al., 2020). It assumes that total streamflow is partitioned into baseflow and event-flow components and baseflow constitutes 100 % of streamflow during low-flow periods (Gustard et al., 1992). The SMM procedure involves partitioning daily streamflow into non-overlapping N-day intervals and identifying the minimum value within each segment. These minimum points (Q_1, Q_2, Q_i, ...) are then screened using a filtering coefficient (M): a point is discarded if $M \cdot Q_i$ exceeds the value of either adjacent minimum. Finally, the baseflow series is constructed by linearly interpolating the remaining minima. The method involves two key parameters: the segment length parameter (N) and the filtering coefficient parameter (M). The segment length parameter N is a proxy of the flow event duration (Stoelzle et al., 2020). Generally, a smaller N results in a higher proportion of baseflow in streamflow, implying shorter surface flow duration. In the literature, N often defaults to 5 d or is predicted using a power-law relationship with catchment area, namely $N = 1.6 \times A^{0.2}$, where A is the catchment area in km^2 and N is expressed in days (Zhang et al., 2017; Aksoy et al., 2008). These two formulations of N are included in the comparison and are denoted as F_{D} and F_{PL}, respectively. To further examine the explanatory capacity of the areal-based power-law relationship, we considered a calibrated form of $N = a \cdot A^b$ (denoted as F'_{PL}), where the coefficients a and b are estimated from the data by minimizing the squared error between the reference and predicted Ns.

The filtering coefficient parameter M is used to determine if a streamflow minimum qualifies as a strict baseflow point. Higher values of M (typically not exceeding 1) correspond to more stringent criteria for identifying pure baseflow conditions. Unlike N, the parameter M is less sensitive to the baseflow separation results and is commonly assigned a constant value of 0.9 (Stoelzle et al., 2020; Aksoy et al., 2008).

3.2 Symbolic regression modelling for the segment length parameter

Symbolic regression (SR) is employed to derive expressions for the parameter N. SR represents an expression using a tree structure, where each node corresponds to a mathematical operator, and each leaf represents an input variable or constant. Structure of the tree evolves to identify expressions that best fit the inputted data through genetic programming (Koza, 1994). Five sample SR trees were demonstrated within the dotted box in Fig. 2a. In this study, the SR method is implemented using the PySR library in Python (Cranmer, 2023), which enforces syntactically valid mathematical structures through predefined operator sets and expression tree representations. The nine representative catchment attributes in original units (Table 1) were used as predictors without normalization to maintain interpretability for the SR expressions. The function space consists of the catchment attributes, free constants, and a set of mathematical operators: addition (+), subtraction (−), multiplication (×), division (/), power-law (power), and logarithm (log).

To control the search space and ensure physically interpretable expressions, several structural constraints were imposed in SR model training. Multiplication, division, power-law, and logarithmic operators were not allowed to be nested within operators of the same type. The internal complexity of expressions inside power-law and logarithmic operators was restricted to a maximum value of 3. The maximum allowable total complexity was set to 20. Expression complexity is defined as the sum of the complexity index assigned to

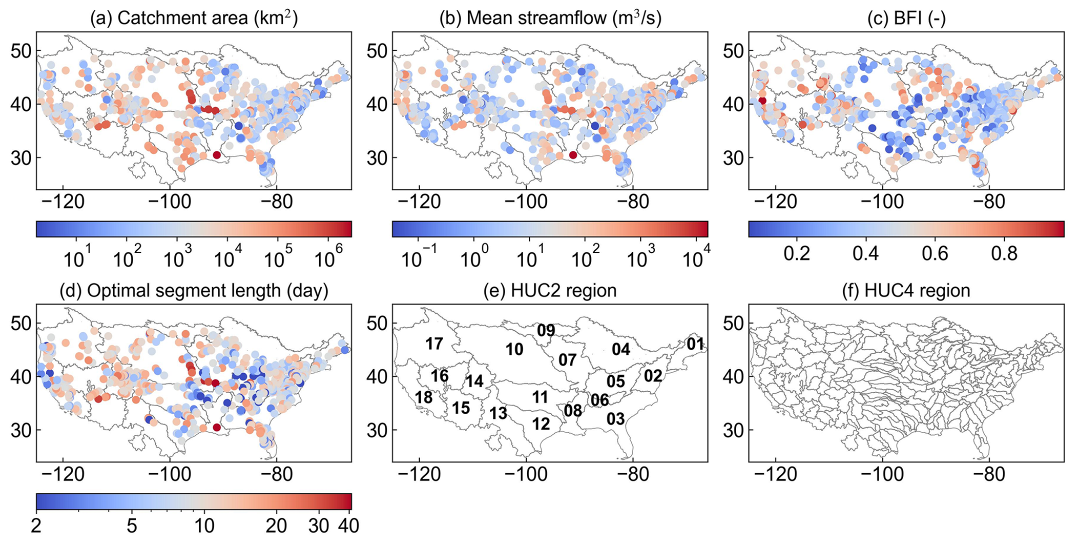

Figure 1. Spatial distribution of the 855 selected catchments with their drainage area **(a)**, mean streamflow **(b)**, BFI **(c)**, and optimal segment length parameter N **(d)** superimposed on HUC2 regions. The HUC2 and HUC4 region maps are provided for referencing **(e, f)**.

Table 1. Catchment characteristics used as inputs to the SR method for predicting parameter N. Catchment characteristics marked with an asterisk are selected for the development of SR based on the mutual information approach.

	Name	Description (unit)
1	$^*\mathrm{PE}_x$	Maximum daily potential evapotranspiration ($\mathrm{mm\,d^{-1}}$)
2	TX_x	Maximum daily maximum temperature (°)
3	P_{95}	The 95th percentile of daily precipitation larger than 0.01 mm ($\mathrm{mm\,d^{-1}}$)
4	$^*P_{50}$	Median of daily precipitation larger than 0.01 mm ($\mathrm{mm\,d^{-1}}$)
5	P_5	The 5th percentile of daily precipitation larger than 0.01 mm ($\mathrm{mm\,d^{-1}}$)
6	$^*f_\mathrm{P}$	Proportion of days with precipitation larger than 0.01 mm (–)
7	f_cly	Volumetric fraction of clay (–)
8	$^*W_\mathrm{sat}$	Saturated water capacity (%)
9	$^*K_\mathrm{sat}$	Saturated hydraulic conductivity ($\mathrm{mm\,h^{-1}}$)
10	$^*A$	Catchment area ($\mathrm{km^2}$)
11	$^*R_\mathrm{E}$	Elongation ratio (-)
12	$^*\overline{S}$	Mean daily storage of reservoirs ($\times 10^6\ \mathrm{m^3}$)
13	S_SD	Standard deviation of daily storage of reservoirs ($\times 10^6\ \mathrm{m^3}$)
14	$^*f_\mathrm{SWE}$	Proportion of days with snow water equivalent larger than 0.05 mm (–)
15	$\overline{\mathrm{LAI}}$	Mean 16-daily leaf area index (–)

each component in the equation. Take $N = 1.6 \times A^{0.2}$ as an example, if multiplication and power-law operators are each assigned a complexity of 2 and constants and input variables are assigned a complexity of 1, the total complexity of the expression is calculated as $2+2+1+1+1=7$. In this study, all operators were assigned a uniform complexity index of 1 to avoid bias toward specific functional forms. Recursive formulations (i.e., expressions where the output variable appears as an input to itself) were not permitted to ensure model interpretability and avoid trivial or ill-posed solutions.

The SR search process was configured with the following hyperparameters: a population size of 33, populations of 15, the crossover rate of 0.066, and evolved over 40 generations. The goodness of fit between the reference and the predicted Ns is evaluated using the mean squared error (MSE):

$$\mathrm{MSE} = \frac{1}{C}\sum_{i=1}^{C}\left(N_i - \hat{N}_i\right)^2, \tag{1}$$

where C is the total number of catchments, and N_i and $\hat{N}_i$ represent the reference and predicted values of N for catchment i, respectively. To evaluate the robustness of the SR models, a ten-fold cross-validation strategy is employed (Fig. 2a). The 855 catchments are randomly partitioned into ten subsets of approximately equal size. In each iteration, the model is trained on nine subsets and tested on the remaining one to estimate the generalization error. This process is

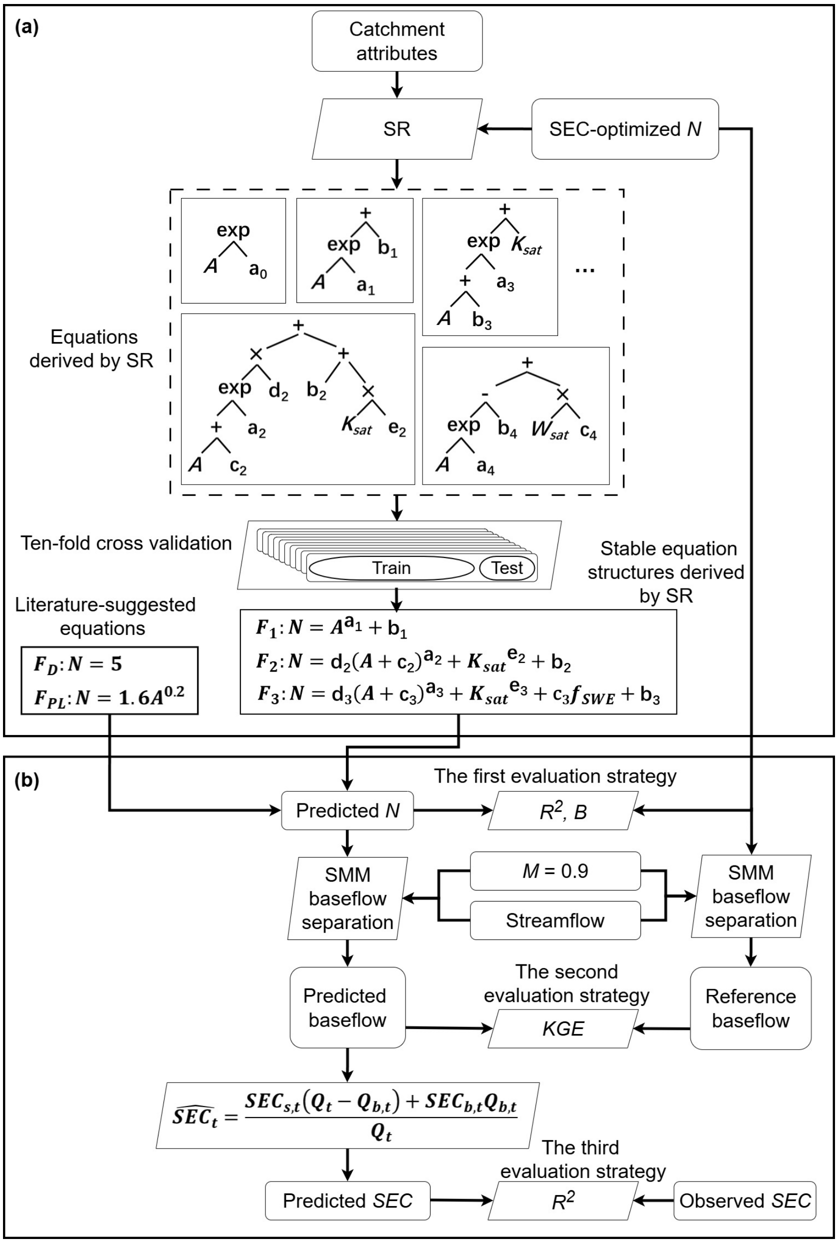

Figure 2. Flowchart of the SR-based prediction framework for the parameter N of SMM **(a)** and the three performance evaluation strategies **(b)**.

repeated ten times so that each subset serves once as the testing set. In each iteration, 7 to 10 expressions with varying levels of complexity are generated, resulting in a total of 91 expressions. Among these expressions, we identified recurrent equation forms across all ten iterations.

3.3 Evaluation strategies

Three strategies are employed to evaluate the performance of different Ns calculated by the SR formulas (Fig. 2b). The first strategy compares the predicted and reference Ns using the mean bias (B) and coefficient of determination (R^2):

$$B = \frac{1}{C}\sum_{i=1}^{C}\left(N_i - \hat{N}_i\right), \tag{2}$$

$$R^2 = \frac{\sum_{i=1}^{C}\left(N_i - \overline{N}_r\right)^2 - \sum_{i=1}^{C}\left(N_i - \hat{N}_i\right)^2}{\sum_{i=1}^{C}\left(N_i - \overline{N}_r\right)^2}, \tag{3}$$

where $\overline{N}_r = \frac{1}{C}\sum\limits_{i=1}^{C} N_i$ is the mean of the reference N over all catchments. Positive (negative) Bs indicate overestimation (underestimation), with lower magnitudes suggesting better performance. R^2 measures the proportion of variance explained by the prediction; it ranges from 0 to 1, with higher values indicating better agreement.

The second strategy evaluates the SR-derived N values in the context of baseflow separation using the SMM method. Specifically, the SMM baseflow time series are calculated using different N values with the other parameter M fixed at 0.9 to eliminate its influence. Baseflow time series calculated with the reference N for each catchment (N_i) is used as the reference. The similarity between the calculated and reference baseflow is assessed using the Kling-Gupta Efficiency (KGE) metric (Gupta et al., 2009):

$$\mathrm{KGE} = 1 - \sqrt{(\rho - 1)^2 + \left(\frac{\sigma_p}{\sigma_r} - 1\right)^2 + \left(\frac{\mu_p}{\mu_r} - 1\right)^2}, \tag{4}$$

where ρ is the correlation coefficient between the predicted and reference baseflow, and μ and σ denote the mean and standard deviation of baseflow. KGE ranges from negative infinity to 1, with higher values indicating better agreement.

The third strategy compares SEC calculated using the SMM baseflow with the observed SEC. Specifically, baseflow time series generated by SMM using Ns derived by the constant, the power-law, and the three SR formulas are used to estimate SEC based on the chemical and water balance relationship:

$$\hat{\mathrm{SEC}}_t = \frac{\mathrm{SEC}_{\mathrm{s},t}\,(Q_t - B_t) + \mathrm{SEC}_{\mathrm{b},t} B_t}{Q_t}, \tag{5}$$

where Q_t and B_t are the observed streamflow and the predicted baseflow at time t, respectively; $\mathrm{SEC}_{\mathrm{s},t}$ and $\mathrm{SEC}_{\mathrm{b},t}$ are the surface flow and baseflow SEC concentrations, respectively. The values of $\mathrm{SEC}_{\mathrm{s},t}$ and $\mathrm{SEC}_{\mathrm{b},t}$ are derived using the extreme value interpolation method, which connects the monthly maxima and minima of the observed SEC with spline interpolation to represent the stable variation of SEC in each of the flow components (Mei et al., 2024b). The derivation of Eq. (5) is documented in Supplement Sect. S1 in the Supplement. The agreement between observed and predicted streamflow SEC is assessed using the R^2 metric (Eq. 1).

To assess whether the differences in SR-based SEC prediction performance are statistical significant, we apply the Diebold–Mariano (DM) test (Diebold and Mariano, 1995). Details on procedure and test statistics are provided in Section S2 in the Supplement. For each catchment, pairwise DM tests are performed among the three SR formulas to determine whether their SEC R^2 differences are statistical significance at the 0.01 significance level. If the largest R^2 is significantly different from the second-largest R^2, the SR formula associated with the largest R^2 is the single best formula of the catchment. If the difference between the largest and the second-largest R^2s is insignificant but that between the second and the third ones is significant, the first two SR formulas are tied. Otherwise, the three SR are tied. Based on these catchment-scale rankings, the best-performing formula for the HUC2 and HUC4 regions is determined as the one most frequently ranked as the best among all catchments within the regions.

4 Results

4.1 SR expressions for N predictions

Three SR formulas with identical functional structures and increasing complexity levels of 5, 13, and 17 were consistently identified across the ten cross-validation iterations. As summarized in Table 2, the simplest formulation is $F_1 : N = A^{a_1} + b_1$, followed by the intermediate formulation $F_2 : N = d_2(A + c_2)^{a_2} + (K_{\mathrm{sat}})^{e_2} + b_2$, and the most complex formulation $F_3 : N = d_3(A + c_3)^{a_3} + (K_{\mathrm{sat}})^{e_3} + f_3 \cdot f_{\mathrm{SWE}} + b_3$. The optimized coefficients of these replicated formulas vary only within narrow ranges (Table 2), indicating a high degree of consistency in parameterization across folds. From F_1 to F_3, the SR results progressively incorporate A, K_{sat}, and f_{SWE}, suggesting that catchment size, subsurface permeability, and snow processes jointly govern the variability of N. The repeated identification of these three formulations across all iterations further demonstrates their structural robustness for predicting N. Accordingly, subsequent analyses focus on these three formulas.

Figure 3 further illustrates the behavior of these formulas. For F_1, the nearly identical exponents (~ 0.23) and intercepts (3.66–3.86 d) result in almost overlapping curves (Fig. 3a), indicating a stable power-law relationship between N and A, with a diminishing rate of increase. For F_2, the similarly constrained exponents (0.28–0.32) and intercepts (1.37–2.78 d) produce tightly clustered response curves (Fig. 3b–c), showing that both A and K_{sat} contribute positively to N. F_3 extends F_2 by introducing f_{SWE} as a linear term. The replicated formulas still exhibit closely grouped slopes (3.28–3.71 d) and intercepts (1.21–2.32 d), which explains the clustering of curves in Fig. 3d–f. The marginal relationships of N with A and K_{sat} in F_3 remain consistent with those in F_2, whereas increasing f_{SWE} leads to an approximately linear increase in N, at a rate of about 0.3–0.4 d per 0.1 increment in snow fraction. Overall, SR identifies A as the most influential factor in predicting N, as evidenced by its presence in all SR-derived formulas. The narrower ranges of predicted Ns in Fig. 3b and d also suggest that A exerts greater influence than K_{sat} and f_{SWE}.

4.2 Evaluation of N predictions

Figure 4 presents the performance of N predicted using the two literature-suggested formulas (F_{D} and F_{PL}) and the three SR formulas (F_1, F_2, and F_3) for the ten training and testing sets. It should be noted that the literature-suggested formulas do not require training; instead, they are directly applied to the training and testing sets separately. Overall, the

Table 2. The calibrated F'_{PL} and the three stable SR expressions across all ten iterations of the cross-validation. The numbers within the brackets of the first row are the complexity of the expressions.

Iteration	F'_{LR} (5)	F_1 (5)	F_2 (13)	F_3 (17)
1	$3.00 \times A^{0.15}$	$A^{0.23} + 3.66$	$0.40(A+529)^{0.31} + K_{\mathrm{sat}}^{0.30} + 2.39$	$0.27(A+640)^{0.33} + K_{\mathrm{sat}}^{0.35} + 3.71 f_{\mathrm{SWE}} + 1.79$
2	$3.10 \times A^{0.15}$	$A^{0.23} + 3.76$	$0.62(A+624)^{0.28} + K_{\mathrm{sat}}^{0.32} + 1.37$	$0.40(A+652)^{0.30} + K_{\mathrm{sat}}^{0.36} + 3.38 f_{\mathrm{SWE}} + 1.26$
3	$3.10 \times A^{0.15}$	$A^{0.23} + 3.74$	$0.47(A+526)^{0.30} + K_{\mathrm{sat}}^{0.28} + 2.19$	$0.32(A+575)^{0.32} + K_{\mathrm{sat}}^{0.33} + 3.55 f_{\mathrm{SWE}} + 1.80$
4	$3.25 \times A^{0.14}$	$A^{0.23} + 3.86$	$0.39(A+334)^{0.30} + K_{\mathrm{sat}}^{0.31} + 2.78$	$0.26(A+361)^{0.33} + K_{\mathrm{sat}}^{0.35} + 3.50 f_{\mathrm{SWE}} + 2.32$
5	$3.17 \times A^{0.15}$	$A^{0.23} + 3.79$	$0.39(A+371)^{0.31} + K_{\mathrm{sat}}^{0.31} + 2.64$	$0.35(A+547)^{0.31} + K_{\mathrm{sat}}^{0.35} + 3.60 f_{\mathrm{SWE}} + 1.61$
6	$3.03 \times A^{0.15}$	$A^{0.23} + 3.76$	$0.56(A+602)^{0.28} + K_{\mathrm{sat}}^{0.29} + 1.77$	$0.41(A+754)^{0.30} + K_{\mathrm{sat}}^{0.34} + 3.63 f_{\mathrm{SWE}} + 1.21$
7	$3.14 \times A^{0.15}$	$A^{0.23} + 3.83$	$0.44(A+401)^{0.30} + K_{\mathrm{sat}}^{0.29} + 2.58$	$0.30(A+443)^{0.32} + K_{\mathrm{sat}}^{0.33} + 3.28 f_{\mathrm{SWE}} + 2.18$
8	$3.16 \times A^{0.15}$	$A^{0.23} + 3.85$	$0.44(A+421)^{0.30} + K_{\mathrm{sat}}^{0.31} + 2.49$	$0.31(A+491)^{0.32} + K_{\mathrm{sat}}^{0.35} + 3.54 f_{\mathrm{SWE}} + 1.95$
9	$3.13 \times A^{0.15}$	$A^{0.23} + 3.81$	$0.46(A+481)^{0.30} + K_{\mathrm{sat}}^{0.28} + 2.42$	$0.30(A+525)^{0.32} + K_{\mathrm{sat}}^{0.33} + 3.38 f_{\mathrm{SWE}} + 2.06$
10	$3.14 \times A^{0.15}$	$A^{0.23} + 3.79$	$0.43(A+545)^{0.30} + K_{\mathrm{sat}}^{0.31} + 2.29$	$0.29(A+627)^{0.32} + K_{\mathrm{sat}}^{0.35} + 3.57 f_{\mathrm{SWE}} + 1.86$

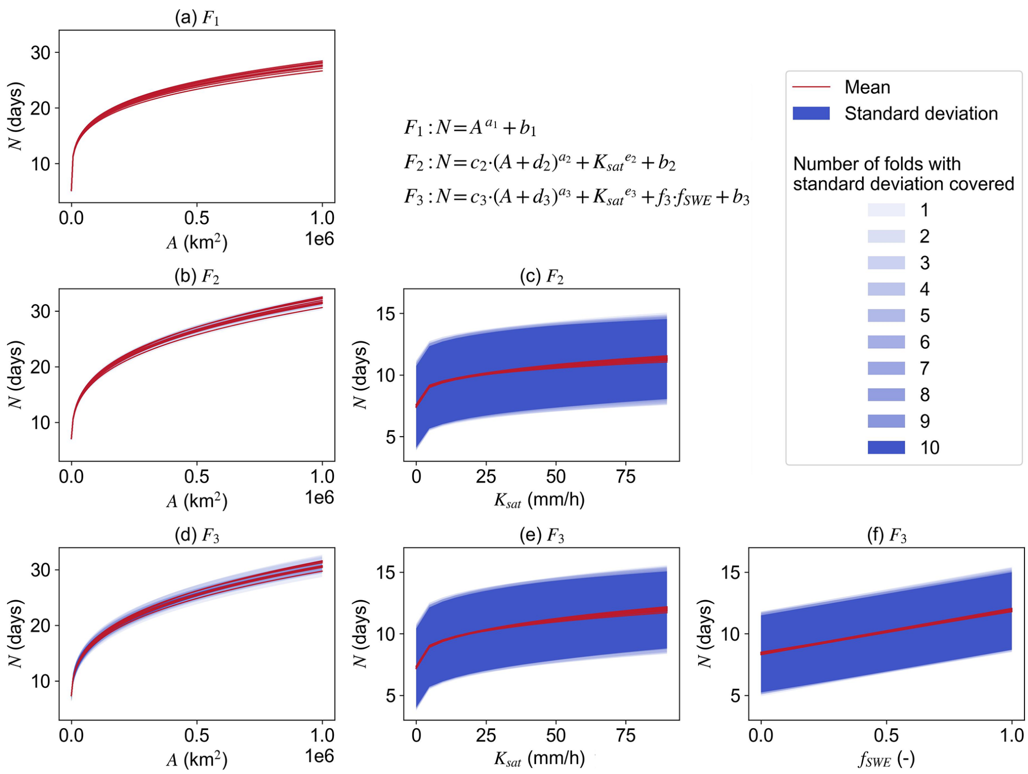

Figure 3. Marginal relationship of N on different predictors (A, K_{sat}, and f_{SWE}) that consist of the SR expressions (F_1, F_2, and F_3). Each line represents one of the ten instances of F_1, F_2, and F_3. Panels a, b, and d are for A; panels **(c)** and **(e)** are for K_{sat}; panel f is for f_{SWE}.

performance of formula F_{D} is poor, with negative R^2 values less than −0.6. F_{PL} shows higher performance but still yields modest R^2 medians of 0.23 and 0.22 for the training and testing sets, respectively. The calibration with respect to SEC data improves the performance for F'_{PL} to median R^2 values of 0.43 and 0.42 for the training and testing sets, respectively, and R^2 of 0.42 for the combined testing set. For predictions from the three SR-derived formulas, we observed higher performance than the conventional formulas. The R^2 medians for F_1, F_2, and F_3 are 0.49, 0.53, and 0.56 for the training sets, and 0.45, 0.49, and 0.55 for the testing sets, respectively. The overall R^2 values for the ten testing sets together for F_1, F_2, and F_3 are 0.48, 0.52, and 0.55, respectively. In terms of the bias, the SR-derived formulas significantly reduce the underestimation of N by the conventional formulas (Fig. 4b). The medians B for F_1, F_2, and F_3 are 0.00/0.06, 0.00/0.08, and 0.00/ − 0.03 d for the training/testing sets, respectively, comparing to F_D and F_{PL} at −4.44/ − 4.38 and −2.34/ − 2.28 d. These values for the combined testing set are −4.44, −2.33, 0.00, 0.00 and 0.00 d for F_D, F_{PL}, F_1, F_2, and F_3, respectively. It is noteworthy that the testing sets exhibit wider ranges of R^2 and B values compared to the training sets. This variability is primarily attributed to the differences in catchment attributes of the testing samples rather than the discrepancies among the 10 replicated formulas. This is supported by the nearly identical coefficients (i.e., a, b, and c) in the SR formulas (Table 2).

4.3 Application of N predictions in baseflow separation

Figure 5a reveals the baseflow separation performance of different N predictions measured by KGE for the testing sets. The baseflow separation performance roughly follows the ranking of $F_D < F_{PL} < F'_{PL} < F_1 < F_2 < F_3$, in line with the N predictions (Fig. 4). The median KGEs for F_D, F_{PL}, F'_{PL}, F_1, F_2, and F_3 are 0.63, 0.73, 0.83, 0.84, 0.84, and 0.85, respectively. For F_D and F_{PL}, 12 % and 5 % of catchments exhibit KGE < 0, while 64 % and 78 % report KGE > 0.5, respectively. In contrast, the percentage of catchments with KGE < 0 drops to 1 % for F'_{PL}, F_1, F_2, and F_3, and those with KGE > 0.5 rise to more than 85 %. These observations indicate the benefits of performing regional calibration for the N prediction.

Figure 5b shows the performance of the three SR formulas across the 18 HUC2 regions. Overall, all formulas perform well for the HUC2 regions, with median KGEs ranging from 0.57 to 0.93. The best performing regions for F_1, F_2, and F_3 are HUC 02, 01, and 07, respectively, exhibiting median KGE values above 0.89 and over 95 % of catchments achieving KGE values greater than 0.5. In contrast, the lowest performance for all three formulas occurs in HUC 12, where median KGE values fall below 0.65 and more than 25 % of catchments show KGE values below 0.5. This may be related to the relative arid climate and flashy hydrological response of HUC 12 (Kratzert et al., 2019; Feng et al., 2020), which is difficult for SMM to capture. Note that SMM is more skillful for smooth baseflow dynamics (Stewart, 2015). All three formulas exhibit larger performance variability in HUC 10–12 and 14 compared to other regions. This can be attributed to the fact that these regions are mountainous and plain areas, characterized by relatively high heterogeneity in catchment characteristics.

Figure 5c compares the relative performance of F_1, F_2, and F_3 across regions. F_2 outperforms F_1 for HUC 04-09 located in the northern and northeastern CONUS, while it performs worse than F_1 in HUC 12–15 on the south. F_3 generally outperforms F_2 in the mountainous region spanning HUC 10–11 and 13–18, and shows comparable performance to F_2 in HUC 01, 02, 04, 05, and 07 with mild topography. In HUC 02, 04, 07, 08, 10, 11, 13, 14, 16, and 18, the most complex F_3 outperforms the other SR formulas. However, F_3 performs worse than F_1 and F_2 in HUC 12. This may be because the SR formulas were calibrated using all catchments to optimize performance at the CONUS scale; thus, their region-specific performance inevitably reflects trade-offs. In relatively warm regions such as HUC 12, where snow influence is weak, the SWE term included in F_3 may add unnecessary complexity rather than useful information, resulting in slightly poorer performance than F_1 and F_2.

Figure 6 presents the average KGE of baseflow separation as functions of the three influential catchment attributes. Overall, the SR formulas consistently outperform the literature-suggested formulas for most of the predictor value ranges, with relatively minor performance differences among F_1, F_2, and F_3, highlighting their robustness across diverse catchment conditions. According to Fig. 6a, when A exceeds 300 km^2, corresponding to $N = 5$ d by F_{PL}, the performance of F_D deteriorates markedly. This suggests that using $N = 5$ d cannot adequately represent its variation for larger As. Conversely, for $A < 300$ km^2, the performance of F_{PL} declines, indicating that the power function coefficients are not suitable for smaller catchments over CONUS. The calibrated power-function (F'_{PL}) reveals improved performance to a similar level with the SR formulas except for the $A < 100$ km^2 bin, indicating the necessity regional calibration. Across the full range of A, the SR formulas consistently achieve higher KGE values, emphasizing the benefits of regional calibration. For small catchment ($A < 100$ km^2), accounting for f_{SWE} and K_{sat} leads to performance gains. In contrast, for larger basins ($A > 100$ km^2), the additional contribution of these factors becomes negligible.

Figure 6b examines the influence of K_{sat}. When $K_{sat} <$ 5 mm h^{-1}, all predictions show similar level of baseflow prediction performance, suggesting that K_{sat} is not a major factor for catchments with low soil hydraulic conductivity. In the intermediate ranges (5–25 mm h^{-1}), the two regionally fitted power-law formulas (F_1 and F'_{PL}) outperform the conventional power-law formula (F_{PL}). Although F_1 and F'_{PL} do not explicitly include K_{sat}, their performance remains comparable to F_2 and F_3, which incorporate K_{sat}. This indicates that regional fitting may compensate the effects of K_{sat} for small and medium values. When $K_{sat} > 25$ mm h^{-1}, incorporating K_{sat} substantially improves the predictions, highlighting the increasing importance of subsurface processes in highly permeable catchments. Figure 6c reveals the influence of f_{SWE}. F_3 outperforms all formulas when $f_{SWE} > 0.4$, indicating the importance of f_{SWE} for catchments with stronger snow persistency. In contrast, for $f_{SWE} < 0.4$, F'_{PL}, F_1, and F_2 perform similarly to F_3, demonstrating that regionally calibrated coefficients can partially offset the lack of explicit consideration of snow-related variables.

4.4 Estimation of SEC variation based on predicted N

The SEC estimation performance by different N predictions for the testing set is evaluated using R^2. The median R^2 val-

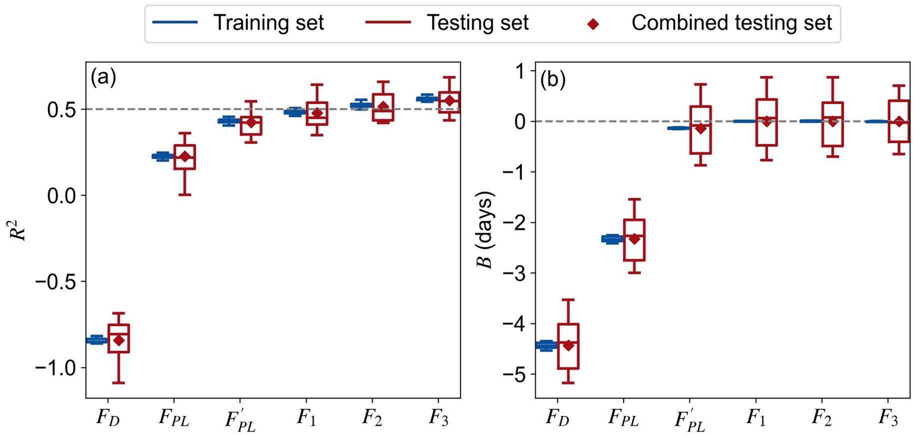

Figure 4. Performance of N predictions using the constant (F_D), power-law (F_{PL}), calibrated power-law (F'_{PL}), and the three SR formulas (F_1, F_2, and F_3) for the ten training and testing sets of the 10-fold cross-validation: coefficient of determination R^2 **(a)** and mean bias B **(b)**.

ues range from 0.70 to 0.74 for the F_D, F_{PL}, F'_{PL}, F_1, F_2, and F_3 formulas, with similar interquartile ranges between 0.29 and 0.33, indicating relatively small overall differences among formulas. Despite these modest differences, both the medians and interquartile ranges of R^2 follow a clear order of $F_D < F_{PL} < F'_{PL} < F_1 < F_2 < F_3$. This ranking is consistent with both the N predictions (Fig. 4) and the baseflow separation results (Fig. 5). The mean R^2 differences between most formulas are statistically significant, except for the comparison between F_1 and F_2 ($p \geq 0.01$).

Figure 7 illustrates the spatial distribution of the best-performing formulas across the HUC4 regions. F_3 demonstrates superior performance across most of the mountainous regions, including the Rocky Mountain foothills, Cascade Range, Sierra Nevada, and some regions within the Great Plains. The improvement for the mountainous regions is likely due to the explicit incorporation of f_{SWE} in the SR formula, capturing the influence of snow processes; for the relatively flat Great Plains regions, the improvement is primarily driven by the consideration of K_{sat}, which better represents subsurface processes such as infiltration capacity, percolation, and lateral subsurface flow that regulate groundwater recharge and baseflow contributions. In contrast, F_1 performs best in the eastern and southeastern CONUS, including the Appalachian Mountains, coastal plain, and Florida peninsular, where the effects of K_{sat} and f_{SWE} are minimal. These regions are generally characterized by low BFI values (Fig. 1c), indicating surface runoff-dominated streamflow with limited relevance of delayed-flow processes (Wu et al., 2021; Mcmillan, 2020). F_1 achieves the highest performance in fewer HUC4 regions, indicating that the effects of K_{sat} may be compensated by the regional calibration of F_1. Interestingly, several regions located in the Great Plains exhibit mixed optimal performance for multiple formulas, suggesting a more complex interplay of hydrological drivers in these areas.

5 Discussions

5.1 Regionalization of baseflow separation parameter by SR

In this study, we used SR to derive mathematical expressions for the predictions of N using 9 catchment attributes. Across ten cross-validation iterations, the identified expressions exhibited consistent structures, predictors, and nearly identical regression coefficients, indicating that SR can yield stable functional relationships between catchment attributes and N. Compared to the RF-based predictions reported by Lin et al. (2026), the SR-based approach showed lower predictive skill ($R^2 = 0.54$ vs. 0.80), reflecting the trade-off between predictive accuracy and interpretability. While RF achieves superior predictive performance, it functions as a "black-box" ensemble, offering no explicit functional form to clarify whether environmental controls operate additively, multiplicatively, or through nonlinear transformations. In contrast, SR provides structural transparency by yielding a closed-form equation, facilitating direct analytical insights (Karpatne et al., 2024; Häfner et al., 2023). This explicit representation enables rigorous sensitivity analysis via differentiation; for instance, the marginal effects derived from equation F_3 quantify how geomorphic and climatic factors jointly govern N. By trading a degree of predictive skill for parsimony, SR transforms the problem from simple estimation into a hypothesis-generating exercise, providing compact transfer functions that are easily integrated into regionalization frameworks (Feigl et al., 2022; Samaniego et al., 2010). Therefore, RF and SR should be viewed as complementary rather than competing approaches: RF provides a benchmark for predictive performance, while SR offers structural transparency that facilitates theoretical interpretation and model integration.

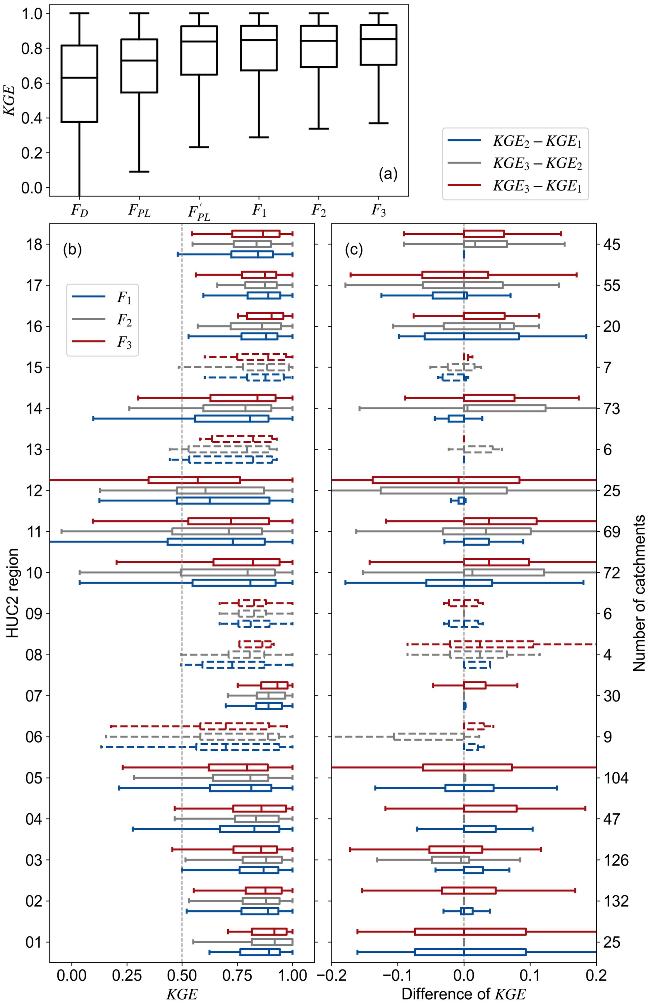

Figure 5. Baseflow separation performance based on different N predictions for all catchments **(a)** and for catchments of the 18 HUC2 regions **(b)**, and relative performance between each pair of the three SR cases for the HUC2 regions **(c)**. Regions with fewer than 10 catchments are plotted by dashed lines in panels **(b)** and **(c)**.

One important aspect to consider when training SR models is the diversity of mathematical operators. A wide variety of operators for model training enables the discovery of complex relationships among variables, but it also enlarges the search space, increasing training time and risk of overfitting (Li et al., 2025; Elsken et al., 2019). To address these challenges, we adopted two strategies. First, we incorporated domain knowledge to guide the choice of operators. Specifically, the well-established power-law relationship between streamflow response time and catchment area motivated the inclusion of the power-law operator in the SR operator set. The fact that all SR-derived formulas captured this power-law relationship indicates the effectiveness of using this domain knowledge. Second, cross-validation was applied to assess model generalizability and reduce overfitting, ensuring representativeness of the SR expressions across the catchments. Although SR can generate more complex formulas than F_3 with better training fit, these structures are inconsistent across folds. This suggests that they may overfit specific subsets of the data and lack generalizability, and are therefore excluded from our analysis.

5.2 Influential catchment attributes for N predictions

The segment length parameter N is a proxy of the average duration of surface flow, and larger values indicate surface flow of event sustain for longer time on average (Stoelzle et al., 2020). Our findings indicate that N is a power function of catchment area with exponent of 0.22–0.23 (F_1, Table 2), indicating that larger catchments are characterized by longer average surface flow duration (Garzon et al., 2023; Mei and Anagnostou, 2015), and the increasing rate in duration slows down for larger As. This could be attributed to the longer averaged water paths for larger catchments (i.e., the Hack's Law), increasing the average traveling time for surface flow of event (Tarasova et al., 2024). Saturated hydraulic conductivity is another key factor in the predictions of N: N increases according to the power-law functions with exponents of 0.28–0.36 with respect to K_{sat} (F_2 and F_3 in Table 2). This indicates that higher K_{sat}values tend to increase the average duration of surface flow with decreasing rate. A possible explanation is that catchments with higher K_{sat} tend to promote infiltration, making more rainfall excess to route through the slow subsurface flow paths than the rapid overland ones (Nagy et al., 2024). F_3 reveals that prolonged snow cover (higher fraction of days covered by snow) is linearly associated with longer surface flow duration. This is due to the snowpack acting as a seasonal storage that modulates streamflow timing (Stoelzle et al., 2020). The gradual melting of snowpack slowly released meltwater to the stream networks, increasing the time needed to leave the catchment (Noor et al., 2023; Barnhart et al., 2016; Godsey et al., 2014). Overall, the additive structure among A, K_{sat}, and f_{SWE} reflect the separate contributions of different catchment-scale delay processes to the event flow through different pathways and at different time scales. Their additive combination therefore provides a parsimonious empirical approximation of the integrated flow-duration response.

5.3 Trade-off between formula complexity and prediction accuracy

Our experiments reveal that formulas for baseflow separation with higher structural complexity generally achieve better predictive performance. This is because more complex formulas can incorporate additional predictors for more detailed descriptions of the underlying hydrological processes. However, the predictive gain from increased complexity is not uniform across all hydro-climatic conditions. For example, when K_{sat} and f_{SWE} take relatively small values

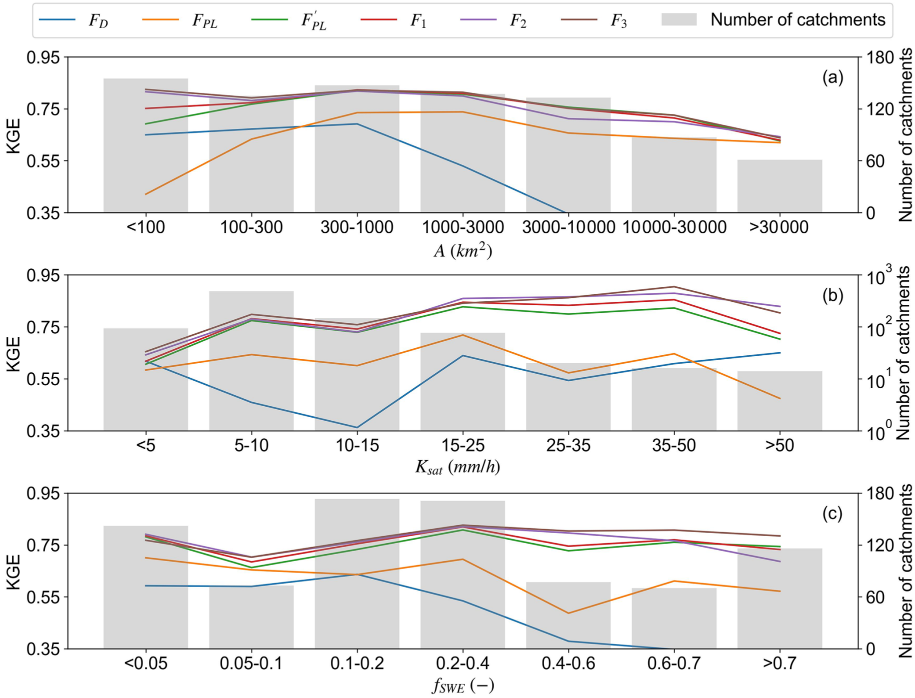

Figure 6. Performance of baseflow separation for different ranges of catchment area **(a)**, catchment-averaged saturated hydraulic conductivity **(b)**, and snow day fraction **(c)**. The right y-axis for panel **(b)** is in logarithmic scale.

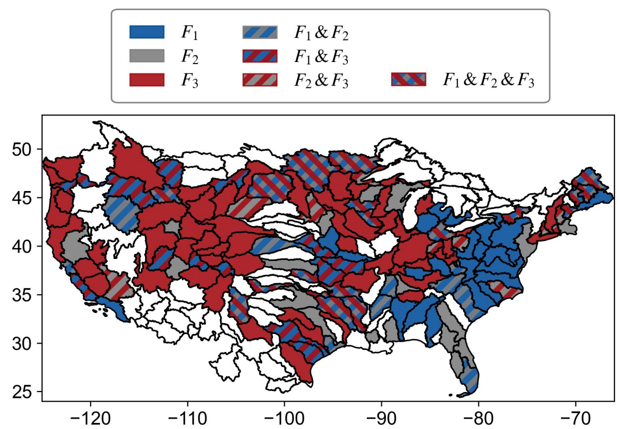

Figure 7. Spatial distribution of the best-performing formulas for SEC estimations across different HUC4 regions. At the HUC4 scale, the best-performing formula is the one(s) appears most frequently as the best among all catchments within the region.

($K_{\text{sat}} < 25\,\text{mm}\,\text{h}^{-1}$ and $f_{\text{SWE}} < 0.4$), the performance differences among F_1, F_2, and F_3 are minimal (Fig. 6). This is likely because small variations in K_{sat} and f_{SWE} have limited impact to the prediction, allowing the regional calibration to partially compensate for the absence of these variables in the formulas. However, such compensation is limited for large K_{sat} and f_{SWE} ($K_{\text{sat}} > 25\,\text{mm}\,\text{h}^{-1}$ and $f_{\text{SWE}} > 0.4$), where their influence becomes more pronounced (Jenicek and Ledvinka, 2020; Beven and Germann, 2013). Moreover, the effects of K_{sat} and f_{SWE} are more evident when A is smaller than $100\,\text{km}^2$ (Fig. 6a). As catchment area increases ($A > 100\,\text{km}^2$), their influence is outweighed by A, highlighting the dominant role of drainage area in shaping the streamflow response time (McGlynn et al., 2004; Sólyom and Tucker, 2007).

6 Conclusions and future work

In this study, we applied symbolic regression to derive formulas for the prediction of segment length parameter N (a proxy of the average duration of surface flow) of the smooth minima baseflow separation method across 855 CONUS catchments. Three stable formulas with increasing complexity were identified: $N = A^{a_1} + b_1$, $N = d_2(A + c_2)^{a_2} + (K_{\text{sat}})^{e_2} + b_2$, and $N = d_3(A+c_3)^{a_3} + (K_{\text{sat}})^{e_3} + f_3 \cdot f_{\text{SWE}} + b_3$, where catchment area ($A$), saturated hydraulic conductivity (K_{sat}), and snow day fraction (f_{SWE}) are identified as important predictors. These SR formulas showed substantial improvements in predictive accuracy of N comparing to the constant ($N = 5$) and power-law formula ($N = 1.6A^{0.2}$). The SR-derived Ns also reveal better performance in baseflow separation and in estimation of electrical conductance dynamics. Among the three SR formulas, F_3 performs better in regions influenced by snow, such as the mountainous mid-west and the northern CONUS; F_1 shows better performance in the eastern and southeastern CONUS, where both the climate and terrain are mild and infiltration rate is low. Overall, formulas that consider K_{sat} and f_{SWE} tend to yield higher predictive accuracy, but simpler formulas without K_{sat} or f_{SWE} can still achieve comparable performance when the values of these key predictors are relatively small. This study presents a new paradigm for regionalization of optimal baseflow parameters using symbolic regression and demonstrates its potential to improve model interpretability and transferability across diverse catchments.

This study used all gages across CONUS to train SR. However, the influence of catchment attributes on N varies across regions. Future work could explore the benefits of developing region-specific SR formulas for different catchment clusters to improve the prediction performance. Furthermore, this study investigated SR-based modeling of only one parameter of a baseflow separation method. Future research could explore applying SR to other baseflow separation methods to identify the governing equations relating catchment attributes to the parameters of these methods. This may help to understand how catchment attributes influence the partitioning of streamflow.

Code and data availability. The streamflow and baseflow time series and optimal baseflow filters parameters are available on Mei et al. (2024a, https://doi.org/10.5281/zenodo.8388365). The HUC region maps are downloaded from Climate Mapping for Resilience & Adaptation (CMRA). The catchment attributes of the 855 catchments are available on Lin et al. (2025, https://doi.org/10.5281/zenodo.16924118).

Supplement. The supplement related to this article is available online at https://doi.org/10.5194/hess-30-3425-2026-supplement.

Author contributions. Yongen Lin: Conceptualization, Methodology, Data curation, Formal analysis, Visualization, investigation, validation, writing – original draft; Dagang Wang: Conceptualization, Methodology, Writing – review & editing, supervision, Resources, funding acquisition, project administration; Yiwen Mei: Conceptualization, Methodology, Writing – review & editing, supervision, Resources, funding acquisition; Jinxin Zhu: Methodology, Writing – review & editing; Huan Wu: Writing – review & editing; Shuo Wang: Writing – review & editing; Zhonghou Xu: Resources, Writing – review & editing; Asaad Y. Shamseldin: Resources, Writing – review & editing; Emmanouil N. Anagnostou: Writing – review & editing.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. The authors bear the ultimate responsibility for providing appropriate place names. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.

Acknowledgements. This paper is dedicated to the memory of Professor Asaad Y. Shamseldin, whose guidance were invaluable to this work. He passed away in February 2026 and is deeply missed.

Financial support. This work is supported by the National Natural Science Foundation of China (grant no. 52579030) and the Guangdong Natural Science Foundation (grant nos. 2025A1515011666 and 2025A1515012264).

Review statement. This paper was edited by Rohini Kumar and reviewed by four anonymous referees.

References

Aksoy, H., Unal, N. E., and Pektas, A. O.: Smoothed minima baseflow separation tool for perennial and intermittent streams, Hydrol. Process., 22, 4467–4476, https://doi.org/10.1002/hyp.7077, 2008.

Apley, D. W. and Zhu, J.: Visualizing the Effects of Predictor Variables in Black Box Supervised Learning Models, J. Roy. Stat. Soc. B, 82, 1059–1086, 10.1111/rssb.12377, 2020.

Barnhart, T. B., Molotch, N. P., Livneh, B., Harpold, A. A., Knowles, J. F., and Schneider, D.: Snowmelt rate dictates streamflow, Geophys. Res. Lett., 43, 8006–8016, https://doi.org/10.1002/2016GL069690, 2016.

Beven, K. and Germann, P.: Macropores and water flow in soils revisited, Water Resour. Res., 49, 3071–3092, https://doi.org/10.1002/wrcr.20156, 2013.

Cartwright, I.: Implications of variations in stream specific conductivity for estimating baseflow using chemical mass balance and calibrated hydrograph techniques, Hydrol. Earth Syst. Sci., 26, 183–195, https://doi.org/10.5194/hess-26-183-2022, 2022.

Chadalawada, J., Herath, H. M. V. V., and Babovic, V.: Hydrologically Informed Machine Learning for Rainfall-Runoff Modeling: A Genetic Programming-Based Toolkit for Automatic Model Induction, Water Resour. Res., 56, e2019WR026933, https://doi.org/10.1029/2019WR026933, 2020.

Cranmer, M.: Interpretable machine learning for science with PySR and SymbolicRegression.jl, arXiv [preprint], https://doi.org/10.48550/arXiv.2305.01582, 2023.

Diebold, F. X. and Mariano, R. S.: Comparing Predictive Accuracy, J. Bus. Econ. Stat., 13, 253–263, 10.1080/07350015.1995.10524599, 1995.

Elsken, T., Metzen, J. H., and Hutter, F.: Neural architecture search: a survey, J. Mach. Learn. Res., 20, 1997–2017, 2019.

Feigl, M., Herrnegger, M., Klotz, D., and Schulz, K.: Function Space Optimization: A Symbolic Regression Method for Estimating Parameter Transfer Functions for Hydrological Models, Water Resour. Res., 56, e2020WR027385, 10.1029/2020WR027385, 2020.

Feigl, M., Thober, S., Schweppe, R., Herrnegger, M., Samaniego, L., and Schulz, K.: Automatic Regionalization of Model Parameters for Hydrological Models, Water Resour. Res., 58, e2022WR031966, https://doi.org/10.1029/2022WR031966, 2022.

Feng, D., Fang, K., and Shen, C.: Enhancing Streamflow Forecast and Extracting Insights Using Long-Short Term Memory Networks With Data Integration at Continental Scales, Water Resour. Res., 56, e2019WR026793, https://doi.org/10.1029/2019wr026793, 2020.

Garzon, L. F. L., Johnson, M. F., Mount, N., and Gomez, H.: Exploring the effects of catchment morphometry on overland flow response to extreme rainfall using a 2D hydraulic-hydrological model (IBER), J. Hydrol., 627, 130405, https://doi.org/10.1016/j.jhydrol.2023.130405, 2023.

Gnann, S. J., McMillan, H. K., Woods, R. A., and Howden, N. J. K.: Including Regional Knowledge Improves Baseflow Signature Predictions in Large Sample Hydrology, Water Resour. Res., 57, e2020WR028354, https://doi.org/10.1029/2020WR028354, 2021.

Godsey, S. E., Kirchner, J. W., and Tague, C. L.: Effects of changes in winter snowpacks on summer low flows: case studies in the Sierra Nevada, California, USA, Hydrol. Process., 28, 5048–5064, https://doi.org/10.1002/hyp.9943, 2014.

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

Gustard, A., Bullock, A. D., and Dixon, J. M.: Low Flow Estimation in the United Kingdom, Institute of Hydrology, Wallingford, UK, ISBN 0948540451, 1992.

Häfner, D., Gemmrich, J., and Jochum, M.: Machine-guided discovery of a real-world rogue wave model, P. Natl. Acad. Sci. USA, 120, e2306275120, https://doi.org/10.1073/pnas.2306275120, 2023.

Hagedorn, B.: Hydrograph separation through multi objective optimization: Revealing the importance of a temporally and spatially constrained baseflow solute source, J. Hydrol., 590, 125349, https://doi.org/10.1016/j.jhydrol.2020.125349, 2020.

Hou, X., Xie, D., Feng, L., Shen, F., and Nienhuis, J. H.: Sustained increase in suspended sediments near global river deltas over the past two decades, Nat. Commun., 15, 3319, https://doi.org/10.1038/s41467-024-47598-6, 2024.

Humphrey, C. E., Solomon, D. K., Genereux, D. P., Gilmore, T. E., Mittelstet, A. R., Zlotnik, V. A., Zeyrek, C., Jensen, C. R., and MacNamara, M. R.: Using Automated Seepage Meters to Quantify the Spatial Variability and Net Flux of Groundwater to a Stream, Water Resour. Res., 58, e2021WR030711, https://doi.org/10.1029/2021WR030711, 2022.

Jenicek, M. and Ledvinka, O.: Importance of snowmelt contribution to seasonal runoff and summer low flows in Czechia, Hydrol. Earth Syst. Sci., 24, 3475–3491, https://doi.org/10.5194/hess-24-3475-2020, 2020.

Karpatne, A., Jia, X., and Kumar, V.: Knowledge-guided machine learning: Current trends and future prospects, arXiv [preprint], https://doi.org/10.48550/arXiv.2403.15989, 2024.

Klotz, D., Herrnegger, M., and Schulz, K.: Symbolic Regression for the Estimation of Transfer Functions of Hydrological Models, Water Resour. Res., 53, 9402–9423, https://doi.org/10.1002/2017wr021253, 2017.

Koza, J. R.: Genetic programming as a means for programming computers by natural selection, Stat. Comput., 4, 87–112, https://doi.org/10.1007/BF00175355, 1994.

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, 2019.

Kronberger, G., de Franca, F. O., Burlacu, B., Haider, C., and Kommenda, M.: Shape-Constrained Symbolic Regression – Improving Extrapolation with Prior Knowledge, Evol. Comput., 30, 75–98, https://doi.org/10.1162/evco_a_00294, 2022.

Li, Q., Zhang, C., Wei, Z., Jin, X., Shangguan, W., Yuan, H., Zhu, J., Li, L., Liu, P., Chen, X., Yan, Y., and Dai, Y.: Advancing symbolic regression for earth science with a focus on evapotranspiration modeling, npj Climate and Atmospheric Science, 7, https://doi.org/10.1038/s41612-024-00861-5, 2024.

Li, X., Zhang, J., Li, D., Liu, X., Xu, J., and Yin, J.: UniSymNet: A Unified Symbolic Network Guided by Transformer, arXiv [preprint], https://doi.org/10.48550/arXiv.2505.06091, 2025.

Lin, Y., Mei, Y., and Wang, D.: Data and code archive for "Regionalization of Optimal Baseflow Separation using Catchment-scale Characteristics", Zenodo [data set/code], https://doi.org/10.5281/zenodo.16924118, 2025.

Lin, Y., Mei, Y., Wang, D., Zhu, J., Wu, H., Wang, S., Yin, G., Gao, L., and N. Anagnostou, E.: Regionalization of Optimal Baseflow Separation using Catchment-scale Characteristics, Water Resour. Res., 62, https://doi.org/10.1029/2025WR040578, 2026.

Makke, N. and Chawla, S.: Interpretable scientific discovery with symbolic regression: a review, Artif. Intell. Rev., 57, https://doi.org/10.1007/s10462-023-10622-0, 2024.

McGlynn, B. L., McDonnell, J. J., Seibert, J., and Kendall, C.: Scale effects on headwater catchment runoff timing, flow sources, and groundwater-streamflow relations, Water Resour. Res., 40, https://doi.org/10.1029/2003WR002494, 2004.

McMahon, T. A. and Nathan, R. J.: Baseflow and transmission loss: A review, WIREs Water, 8, e1527, https://doi.org/10.1002/wat2.1527, 2021.

McMillan, H.: Linking hydrologic signatures to hydrologic processes: A review, Hydrol. Process., 34, 1393–1409, https://doi.org/10.1002/hyp.13632, 2020.

McMillan, H. K., Gnann, S. J., and Araki, R.: Large Scale Evaluation of Relationships Between Hydrologic Signatures and Processes, Water Resour. Res., 58, e2021WR031751, https://doi.org/10.1029/2021WR031751, 2022.

Mei, Y. and Anagnostou, E. N.: A hydrograph separation method based on information from rainfall and runoff records, J. Hydrol., 523, 636–649, https://doi.org/10.1016/j.jhydrol.2015.01.083, 2015.

Mei, Y., Zhu, J., and Wang, D.: Data and Code Archive for "Optimal Baseflow Separation through Chemical Mass Balance: Comparing the Usages of Two Tracers, Two Concentration Estimation Methods, and Four Baseflow Filters", Zenodo [data set/code], https://doi.org/10.5281/zenodo.8388365, 2024a.

Mei, Y., Wang, D., Zhu, J., Tang, G., Cai, C., Shen, X., Hong, Y., and Zhang, X.: Optimal Baseflow Separation Through Chemical Mass Balance: Comparing the Usages of Two Tracers, Two Concentration Estimation Methods, and Four Baseflow Filters, Water Resour. Res., 60, https://doi.org/10.1029/2023wr036386, 2024b.

Nagy, E. D., Szilagyi, J., and Torma, P.: Calibrating the lyne-hollick filter for baseflow separation based on catchment response time, J. Hydrol., 638, https://doi.org/10.1016/j.jhydrol.2024.131483, 2024.

Noor, K., Marttila, H., Welker, J. M., Mustonen, K.-R., Kløve, B., and Ala-aho, P.: Snow sampling strategy can bias estimation of meltwater fractions in isotope hydrograph separation, J. Hydrol., 627, 130429, https://doi.org/10.1016/j.jhydrol.2023.130429, 2023.

Pelletier, A. and Andréassian, V.: Hydrograph separation: an impartial parametrisation for an imperfect method, Hydrol. Earth Syst. Sci., 24, 1171–1187, https://doi.org/10.5194/hess-24-1171-2020, 2020.

Piggott, A. R., Moin, S., and Southam, C.: A revised approach to the UKIH method for the calculation of baseflow [Une approche améliorée de la méthode de l'UKIH pour le calcul de l'écoulement de base], Hydrolog. Sci. J., 50, 920, https://doi.org/10.1623/hysj.2005.50.5.911, 2005.

Price, K.: Effects of watershed topography, soils, land use, and climate on baseflow hydrology in humid regions: A review, Prog. Phys. Geog., 35, 465–492, https://doi.org/10.1177/0309133311402714, 2011.

Rudin, C.: Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead, Nat. Mach. Intell., 1, 206–215, https://doi.org/10.1038/s42256-019-0048-x, 2019.

Samaniego, L., Kumar, R., and Attinger, S.: Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale, Water Resour. Res., 46, https://doi.org/10.1029/2008WR007327, 2010.

Sheta, A., Abdel-raouf, A., Fraihat, K., and Baareh, A. K.: Evolutionary Design of a PSO-Tuned Multigene Symbolic Regression Genetic Programming Model for River Flow Forecasting, International Journal of Advanced Computer Science and Applications, 14, 2023, 10.14569/IJACSA.2023.0140489, 2023.

Sólyom, P. B. and Tucker, G. E.: The importance of the catchment area–length relationship in governing non-steady state hydrology, optimal junction angles and drainage network pattern, Geomorphology, 88, 84–108, https://doi.org/10.1016/j.geomorph.2006.10.014, 2007.

Song, W., Jiang, S., Camps-Valls, G., Williams, M., Zhang, L., Reichstein, M., Vereecken, H., He, L., Hu, X., and Shi, L.: Towards data-driven discovery of governing equations in geosciences, Commun. Earth Environ., 5, 589, https://doi.org/10.1038/s43247-024-01760-6, 2024.

Stewart, M. K.: Promising new baseflow separation and recession analysis methods applied to streamflow at Glendhu Catchment, New Zealand, Hydrol. Earth Syst. Sci., 19, 2587–2603, https://doi.org/10.5194/hess-19-2587-2015, 2015.

Stoelzle, M., Schuetz, T., Weiler, M., Stahl, K., and Tallaksen, L. M.: Beyond binary baseflow separation: a delayed-flow index for multiple streamflow contributions, Hydrol. Earth Syst. Sci., 24, 849–867, https://doi.org/10.5194/hess-24-849-2020, 2020.

Sundararajan, M. and Najmi, A.: The Many Shapley Values for Model Explanation, Proceedings of the 37th International Conference on Machine Learning, Proceedings of Machine Learning Research (PMLR), 119, https://proceedings.mlr.press/v119/sundararajan20b.html (last access: 29 May 2026), 2020.

Tan, X., Liu, B., and Tan, X.: Global Changes in Baseflow Under the Impacts of Changing Climate and Vegetation, Water Resour. Res., 56, e2020WR027349, https://doi.org/10.1029/2020WR027349, 2020.

Tarasova, L., Gnann, S., Yang, S., Hartmann, A., and Wagener, T.: Catchment characterization: Current descriptors, knowledge gaps and future opportunities, Earth-Sci. Rev., 252, https://doi.org/10.1016/j.earscirev.2024.104739, 2024.

Thorslund, J. and van Vliet, M. T. H.: A global dataset of surface water and groundwater salinity measurements from 1980–2019, Scientific Data, 7, 231, https://doi.org/10.1038/s41597-020-0562-z, 2020.

Wang, C., Gomez-Velez, J. D., and Wilson, J. L.: Dynamic coevolution of baseflow and multiscale groundwater flow system during prolonged droughts, J. Hydrol., 609, https://doi.org/10.1016/j.jhydrol.2022.127657, 2022.

Wilstrup, C. and Kasak, J.: Symbolic regression outperforms other models for small data sets, arXiv [preprint], https://doi.org/10.48550/arXiv.2103.15147, 2021.

Wu, S., Zhao, J., Wang, H., and Sivapalan, M.: Regional Patterns and Physical Controls of Streamflow Generation Across the Conterminous United States, Water Resour. Res., 57, https://doi.org/10.1029/2020wr028086, 2021.

Xie, J., Liu, X., Tian, W., Wang, K., Bai, P., and Liu, C.: Estimating Gridded Monthly Baseflow From 1981 to 2020 for the Contiguous US Using Long Short-Term Memory (LSTM) Networks, Water Resour. Res., 58, https://doi.org/10.1029/2021wr031663, 2022.

Xie, J., Liu, X., Wang, K., Yang, T., Liang, K., and Liu, C.: Evaluation of typical methods for baseflow separation in the contiguous United States, J. Hydrol., 583, https://doi.org/10.1016/j.jhydrol.2020.124628, 2020.

Xie, J., Liu, X., Jasechko, S., Berghuijs, W. R., Wang, K., Liu, C., Reichstein, M., Jung, M., and Koirala, S.: Majority of global river flow sustained by groundwater, Nat. Geosci., https://doi.org/10.1038/s41561-024-01483-5, 2024.

Yan, X., Sun, J., Huang, Y., Xia, Y., Wang, Z., and Li, Z.: Detecting and attributing the changes in baseflow in China's Loess Plateau, J. Hydrol., 617, 128957, https://doi.org/10.1016/j.jhydrol.2022.128957, 2023.

Zhang, J., Zhang, Y., Song, J., and Cheng, L.: Evaluating relative merits of four baseflow separation methods in Eastern Australia, J. Hydrol., 549, 252–263, https://doi.org/10.1016/j.jhydrol.2017.04.004, 2017.

Zhang, J., Zhang, Y., Song, J., Cheng, L., Kumar Paul, P., Gan, R., Shi, X., Luo, Z., and Zhao, P.: Large-scale baseflow index prediction using hydrological modelling, linear and multilevel regression approaches, J. Hydrol., 585, 124780, https://doi.org/10.1016/j.jhydrol.2020.124780, 2020.