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

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

[1] School of Geography and Planning, Sun Yat-sen University, Guangzhou, China

[2] Carbon-Water Research Station in Karst Regions of Northern Guangdong, Sun Yat-sen University, Guangzhou, China

[3] School of Atmospheric Science, Sun Yat-sen University, Zhuhai, Guangdong, China

[4] State Key Laboratory of Climate System Prediction and Risk Management, Nanjing University of Information Science and Technology, Nanjing, China

[5] Department of Land Surveying and Geo-informatics, Hong Kong Polytechnic University, Kowloon, Hong Kong Special Administrative Region, China

[6] Earth Sciences New Zealand, Hamilton 3216, New Zealand

[7] Department of Civil and Environmental Engineering, The University of Auckland, Auckland 1010, New Zealand

[8] Department of Civil and Environmental Engineering, University of Connecticut, Storrs, USA

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

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 default to 5 days. 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 ~~0.49, 0.50, and 0.54~~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~~exponentially~~ 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

($K_{sat} > 25$ mm/h and $f_{SWE} > 0.4$) and smaller area ($A < 100$ 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 (Stoelzle et al., 2020; Wang et al., 2022; Xie 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 (Tan et al., 2020; Yan et al., 2023). 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., 2024a). 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., 2024a). 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 (Hou et al., 2024; Thorslund and van Vliet, 2020). This limitation hinders accurate quantification of baseflow in most global catchments, despite their long-term streamflow data.

To optimize baseflow separation for gaged catchment lacking continuous environmental tracer data, a viable approach is to transfer optimized parameters from other catchments (Feigl et al., 2020; Klotz et al., 2017). 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 (Gnann et al., 2021; McMillan et al., 2022; Price, 2011; Zhang et al., 2020). While this parameter regionalization approach is widely used for transferring calibrated parameters to un-gaged catchments in hydrological modeling (Feigl et al., 2020; Klotz et al., 2017), 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 days 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~~remains a "black box" with no explicit 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 (Makke and Chawla, 2024; Rudin, 2019). 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).~~Although partial dependence plots may be adopted to explore the functional relationships between catchment attributes and N, they cannot mathematically formulate these relationship (Makke and Chawla, 2024; Rudin, 2019).~~ 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~~equation discovery~~ 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 ungagged catchments (Feigl et al., 2022; Li et al., 2024). 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 (Feigl et al., 2020; Klotz et al., 2017; 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 ~~into~~ 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. (2024b), which contains the segment length parameter of SMM optimized by ~~specific electrical conductance (~~SEC~~)~~ for 987 catchments across the Conterminous~~Continental~~ United States (CONUS). To ensure the reliability of baseflow separation using the SMM method~~To reduce uncertainty due to the limited suitability of SMM in certain catchments~~, catchments~~those~~ exhibiting suboptimal baseflow separation performance with ~~operationalized as~~ a Kling Gupta Efficiency between estimated and observed SEC below 0.5 ~~(Mei et al., 2024b)~~ were excluded ~~from the study~~(Mei et al., 2024b). ~~-~~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 **Figure 1**a-c, showing significant diversity. The optimal N parameters for these catchments are depicted in **Figure 1**d with most values smaller than 17 days. Larger N values tend to occur in

mountainous regions characterized by drier climates and greater snow persistence with implicit~~and no clear~~ 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 (**Figure 1**e, f).

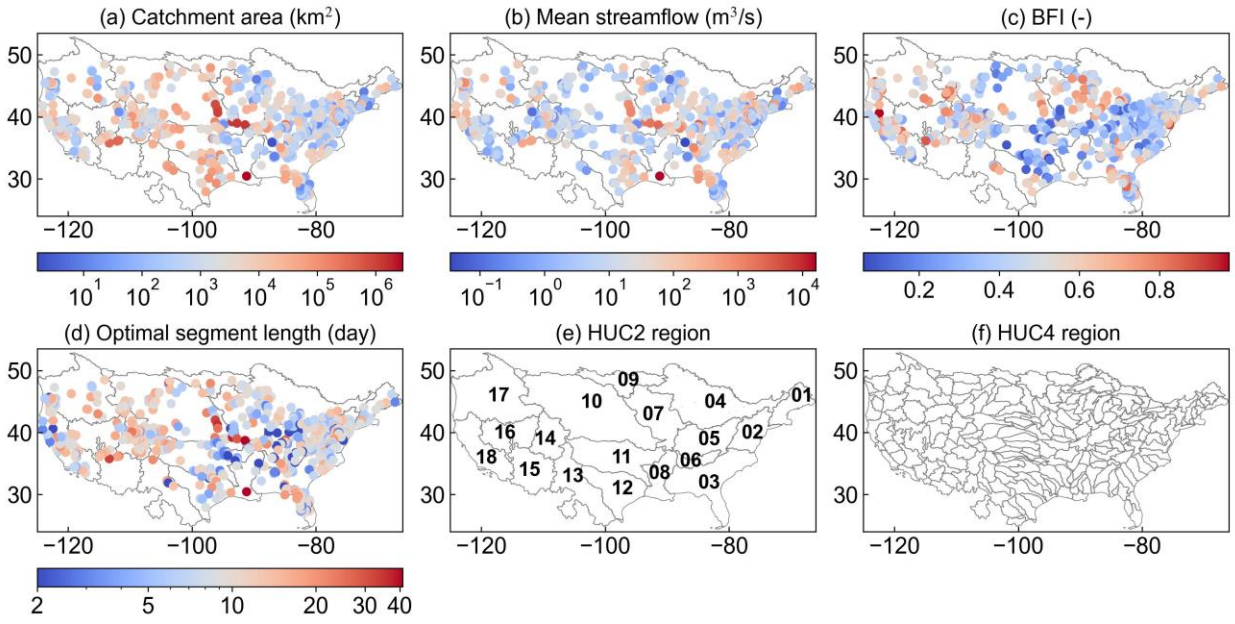


Figure 1. Spatial distribution of the 855 selected catchments with their drainage area (a), mean streamflow (b), BFI (c), and optimal segment length parameter *N* (d) superimposed on HUC2 regions. The HUC2 and HUC4 region maps are provided for referencing (e, f).

Lin et al. (2026) identified~~revealed that~~ 13 catchment characteristics representing the geomorphology, climate, soil hydraulic properties~~condition~~, and water usage, that significantly influence~~are influential to~~ the SMM parameter~~prediction of~~ *N* (**Table 1**, rows 1-13). In addition to these predictors~~13 characteristics~~, we added two more related to snow processes and vegetation dynamics ~~as inputs to SR~~ (**Table 1**, rows 14-15)~~.~~, as previous~~This is because some~~ studies have shown~~found~~ that these factors~~snow processes and vegetation dynamics~~ can substantially affect~~have nonnegligible impacts on~~ 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). ~~The Pearson correlation coefficients between these characteristics and the segment length parameter range from -0.45 to 0.53 (**Figure B1**).~~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 (**Figure S1**). 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.

Table 1. Catchment characteristics used as inputs to the SR method for predicting parameter *N*. Catchment characteristics highlighted in blue are selected for the development of SR based on mutual information approach.

	Name	Description (unit)
1	PE_x	Maximum daily potential evapotranspiration (mm/day)
2	TX_x	Maximum daily maximum temperature (°C)
3	P_{95}	The 95-th percentile of daily precipitation larger than 0.01 mm (mm/day)
4	P_{50}	Median of daily precipitation larger than 0.01 mm (mm/day)
5	P_5	The 5-th percentile of daily precipitation larger than 0.01 mm (mm/day)
6	f_P	Proportion of days with precipitation larger than 0.01 mm (-)
7	f_{cly}	Volumetric fraction of clay (-)
8	W_{sat}	Saturated water capacity (%)
9	k_{sat}	Saturated hydraulic conductivity (~~$\times 10^{-2}$~~ mm/hour)
10	A	Catchment area (km^2)
11	R_E	Elongation ratio (-)
12	$\bar{S}$	Mean daily storage of reservoirs ($\times 10^6$ m^3)
13	S_{SD}	Standard deviation of daily storage of reservoirs ($\times 10^6$ m^3)
14	f_{SWE}	Proportion of days with snow water equivalent larger than 0.05 mm (-)
15	$\overline{LAI}$	Mean 16-daily leaf area index (-)

3. Methods

3.1. ~~Recaps on~~ Smooth Minima Baseflow Separation ~~(SMM)~~

SMM is a widely used baseflow separation method (Aksoy et al., 2008; Piggott et al., 2005; Tan et al., 2020; Xie 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 result in a higher proportion of baseflow in streamflow, implying shorter surface flow duration. In the literature, N is often default to 5 days or 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 (Aksoy et al., 2008; Zhang et al., 2017). These two formulations of N are included in the ~~subsequent~~ 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 (Aksoy et al., 2008; Stoelzle et al., 2020).

3.2. **Symbolic regression modelling for the segment length parameter**

~~In this study, the~~ S~~s~~ymbolic 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~~For example,~~ **Figure 2**a ~~illustrates five SR trees representing different expressions within the dotted box~~. In this study, t~~T~~he 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 ~~fifteen~~ nine representative catchment attributes~~characteristics~~ in original units (**Table 1**) were~~are~~ used as predictors~~inputs~~ without normalization~~.~~ to maintain interpretability for the SR expressions. The function space is consisted of the catchment attributes,

free constants, and a set of mathematical operators: addition (+), subtraction (−), multiplication (×), division (÷), power-law (power), and logarithm (log).

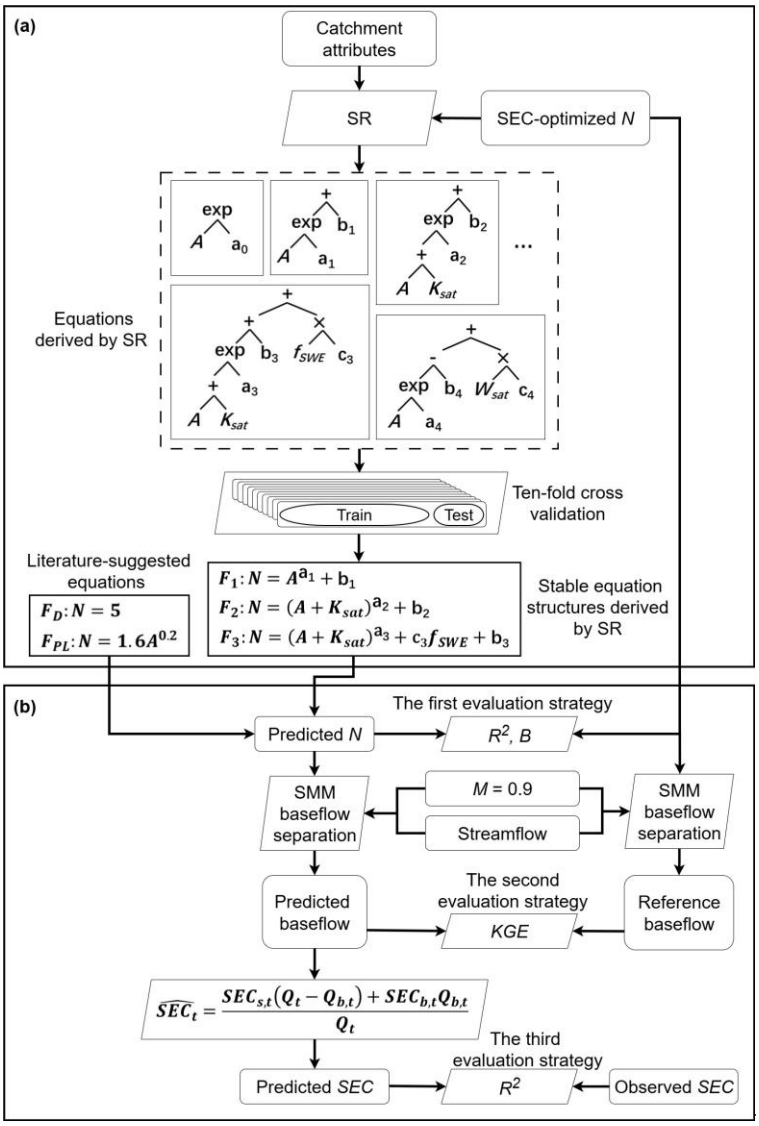
To control the search space and ensure physically interpretable expressions, several structural constraints were imposed in SR model training. Multiplication, division, power-law, and logarithmic operators were not allowed to be nested within operators of the same type. The internal complexity of expressions inside power-law and logarithmic operators was restricted to a maximum value of 3. The maximum allowable total complexity was set to 20. ~~The mathematical operators for training the SR models includes input variables, constants, addition, subtraction, multiplication, division, exponential, and logarithmic operators. The SR method evaluates both accuracy and complexity of each candidate expression to select the most accurate formulation under different levels of complexity allowance.~~ Expression complexity is defined as the sum of the complexity index assigned to each component in the equation. Take $N = 1.6 \times A^{0.2}$ as an example, if multiplication and power-law~~exponential~~ 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. ~~To train SR models, complexity index for each possible mathematical operator is set to 1 to avoid bias among the operators; the maximum allowable complexity for the resulting expressions is set to 20.~~

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 *N*s is evaluated using the mean squared error (MSE):

$$MSE = \frac{1}{C}\sum_{i=1}^{C}\left(N_i - \widehat{N}_i\right)^2 \quad , \qquad (1)$$

where C is the total number of catchments, and N_i and $\widehat{N}_i$ 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 (**Figure 2**a). 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 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.



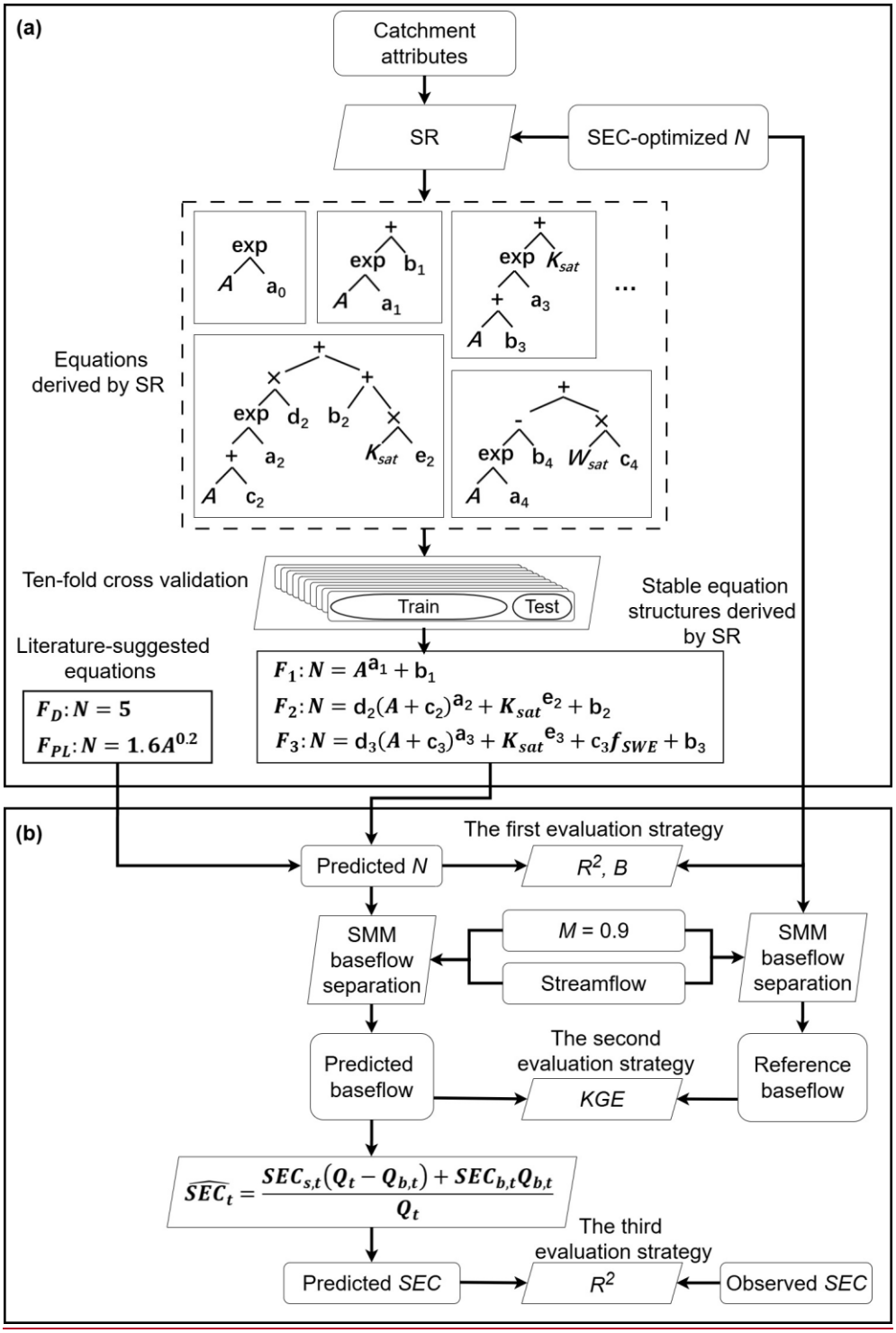


Figure 2. Flowchart of the SR-based prediction framework for the parameter *N* of SMM (a) and the three performance evaluation strategies (b).

3.3. **Evaluation strategies**

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

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

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

where $\bar{N}_r = \frac{1}{C}\sum_{i=1}^{C} N_i$ 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):

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

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:

$$\widehat{SEC_t} = \frac{SEC_{s,t}(Q_t - B_t) + SEC_{b,t}B_t}{Q_t} \quad , \qquad (5)$$

where Q_t and B_t are the observed streamflow and the predicted baseflow at time t, respectively; $SEC_{s,t}$ and $SEC_{b,t}$ are the surface flow and baseflow SEC concentrations, respectively. The values of $SEC_{s,t}$ and $SEC_{b,t}$ 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., 2024a). The derivation of Eq. 5 is documented in ~~Appendix A~~**Text S1** of the Supporting Information. 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 **Text S2**. 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 $F_2: N = d_2(A + c_2)^{a_2} + (K_{sat})^{e_2} + b_2$, and the most complex formulation $F_3: N = d_3(A + c_3)^{a_3} + (K_{sat})^{e_3} + f_3 \cdot f_{SWE} + b_3$. 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. ~~Three SR formulasexpressions with identical structures and increasing complexity of 5, 7, and 11 occurred consistently across the ten iterations of the cross-validation. They are $N = A^{a_1} + b_1$, $N = (A + K_{sat})^{a_2} + b_2$, and $N = (A + K_{sat})^{a_3} + c_3 \cdot f_{SWE} + b_3$, where a, b, c, d, e, and f are coefficientsconstants that exhibit minor variation among iterations (see **Table 2** for details). XXX. This consistency suggests that these three forms are robust for the predictions of N. Therefore, our analysis focuses on these formulas, which are hereafter referred to as F_1, F_2, and F_3, respectively.~~

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

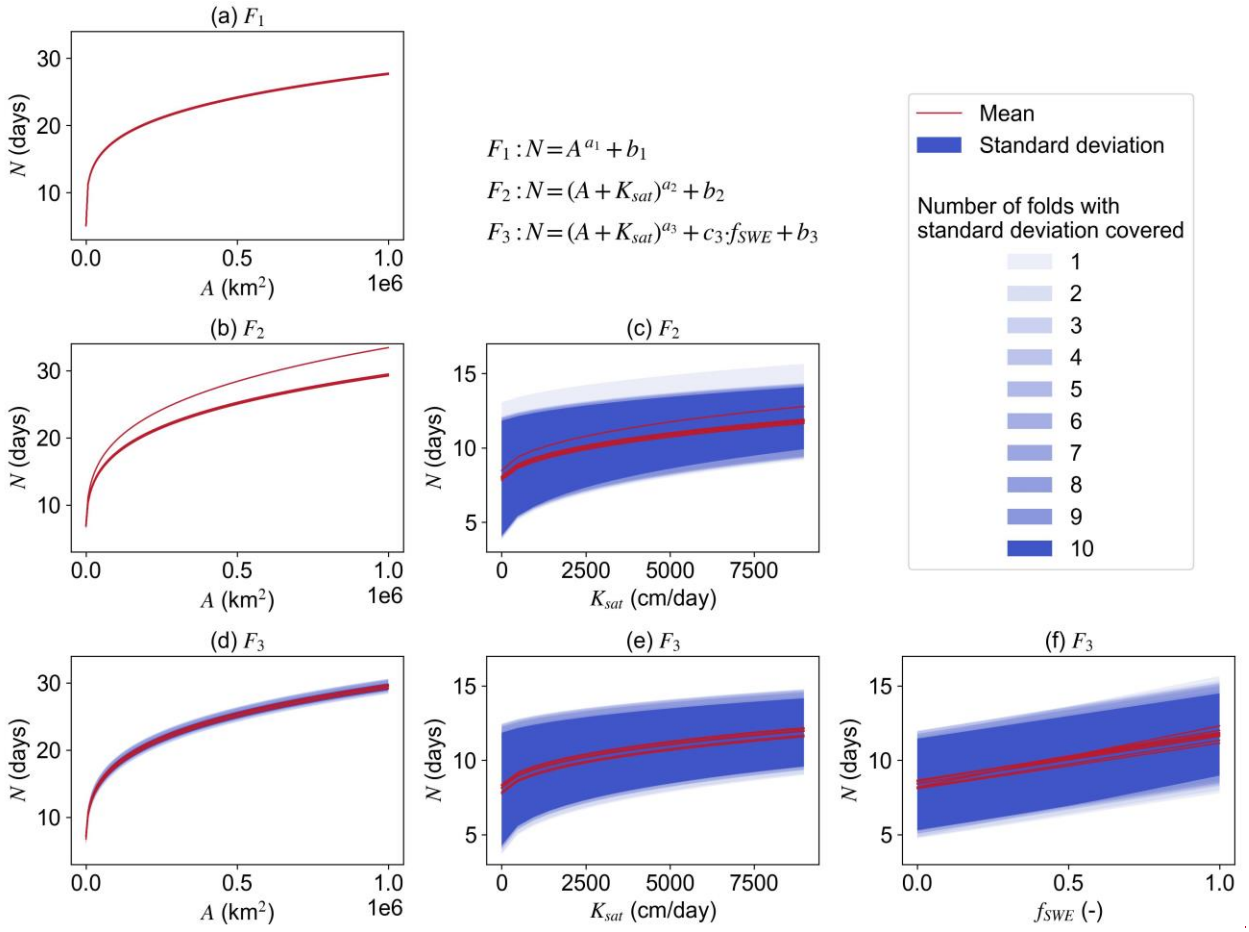
~~Iteration~~	~~F_1 (5)~~	~~F_2 (7)~~	~~F_3 (11)~~
~~1~~	~~$A^{0.23} + 3.77$~~	~~$(A + K_{sat})^{0.24} + 1.66$~~	~~$(A + K_{sat})^{0.24} + 3.05 f_{SWE} + 0.70$~~
~~2~~	~~$A^{0.23} + 3.70$~~	~~$(A + K_{sat})^{0.24} + 1.79$~~	~~$(A + K_{sat})^{0.24} + 3.12 f_{SWE} + 1.22$~~
~~3~~	~~$A^{0.23} + 3.80$~~	~~$(A + K_{sat})^{0.24} + 1.76$~~	~~$(A + K_{sat})^{0.24} + 3.18 f_{SWE} + 0.76$~~
~~4~~	~~$A^{0.23} + 3.67$~~	~~$(A + K_{sat})^{0.24} + 1.95$~~	~~$(A + K_{sat})^{0.24} + 3.18 f_{SWE} + 1.03$~~
~~5~~	~~$A^{0.23} + 3.70$~~	~~$(A + K_{sat})^{0.24} + 1.81$~~	~~$(A + K_{sat})^{0.24} + 3.39 f_{SWE} + 1.01$~~
~~6~~	~~$A^{0.23} + 3.74$~~	~~$(A + K_{sat})^{0.25} + 1.79$~~	~~$(A + K_{sat})^{0.24} + 4.12 f_{SWE} + 0.80$~~
~~7~~	~~$A^{0.23} + 3.65$~~	~~$(A + K_{sat})^{0.24} + 1.90$~~	~~$(A + K_{sat})^{0.24} + 3.20 f_{SWE} + 1.25$~~
~~8~~	~~$A^{0.23} + 3.61$~~	~~$(A + K_{sat})^{0.24} + 1.90$~~	~~$(A + K_{sat})^{0.24} + 3.44 f_{SWE} + 1.17$~~
~~9~~	~~$A^{0.23} + 3.73$~~	~~$(A + K_{sat})^{0.24} + 1.87$~~	~~$(A + K_{sat})^{0.24} + 3.26 f_{SWE} + 1.20$~~
~~10~~	~~$A^{0.23} + 3.75$~~	~~$(A + K_{sat})^{0.24} + 1.84$~~	~~$(A + K_{sat})^{0.24} + 3.23 f_{SWE} + 1.02$~~

285

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

9	$3.13 \times A^{0.15}$	$A^{0.23} + 3.81$	$0.46(A + 481)^{0.30} + {K_{sat}}^{0.28} + 2.42$	$0.30(A + 525)^{0.32} + {K_{sat}}^{0.33} + 3.38 f_{SWE} + 2.06$
10	$3.14 \times A^{0.15}$	$A^{0.23} + 3.79$	$0.43(A + 545)^{0.30} + {K_{sat}}^{0.31} + 2.29$	$0.29(A + 627)^{0.32} + {K_{sat}}^{0.35} + 3.57 f_{SWE} + 1.86$

Figure 3 further illustrates the behavior of these formulas. For F_1, the nearly identical exponents (~0.23) and intercepts (3.66–3.86 days) result in almost overlapping curves (**Figure 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 days) produce tightly clustered response curves (**Figure 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 days) and intercepts (1.21–2.32 days), which explains the clustering of curves in **Figure 3**d–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 days 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 **Figure 3**b and d also suggest that A exerts greater influence than K_{sat} and f_{SWE}. ~~Formula F_1 predicts N by A with a power-law function with identical exponents (0.23) and nearly identical intercepts (3.61–3.80 days, **Table 2**). These consistent coefficients result in nearly identical relationship across the ten F_1s that N increases exponentially with A at a diminishing rate (**Figure 3**a). Formula F_2 incorporates both A and K_{sat} in similar power-law functions with nearly identical exponents (0.24–0.25) and intercepts (1.66–1.95 days, **Table 2**). Increases in either variable contribute to higher values of N (**Figure 3**b and c). However, the relative importance of these predictors is modulated by their interaction: when A is large, the influence of K_{sat} becomes minimal; conversely, when A is small, the contribution of K_{sat} is more pronounced. This interaction also explains the observed patterns in **Figure 3**a and b, where F_1 and F_2 show clear dependence on catchment area when $A < 1000$ km^2. Formula F_3 extends F_2 by adding f_{SWE} as a linear term. The 10 replicated formulas still reveal nearly identical exponents (0.24), slopes (3.05–4.12 days, **Table 2**), and intercepts (0.70–1.25 days). The marginal relationships of N with A and K_{sat} mirror those of F_2 (**Figure 3**d and e). As for f_{SWE}, larger values lead to higher prediction with a constant rate of 0.3–0.4 days per 0.1 fraction increases as indicated by the linear relationship (**Figure 3**f). 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 **Figure 3**b and d also suggest that A exerts greater influence than K_{sat} and f_{SWE}.~~


(a) F_1
(b) F_2
(c) F_2
(d) F_3
(e) F_3
(f) F_3
$F_1 : N = A^{a_1} + b_1$
$F_2 : N = (A + K_{sat})^{a_2} + b_2$
$F_3 : N = (A + K_{sat})^{a_3} + c_3 : f_{SWE} + b_3$
Mean
Standard deviation
Number of folds with standard deviation covered
1
2
3
4
5
6
7
8
9
10
N (days)
A (km^2)
1e6
K_{sat} (cm/day)
f_{SWE} (-)

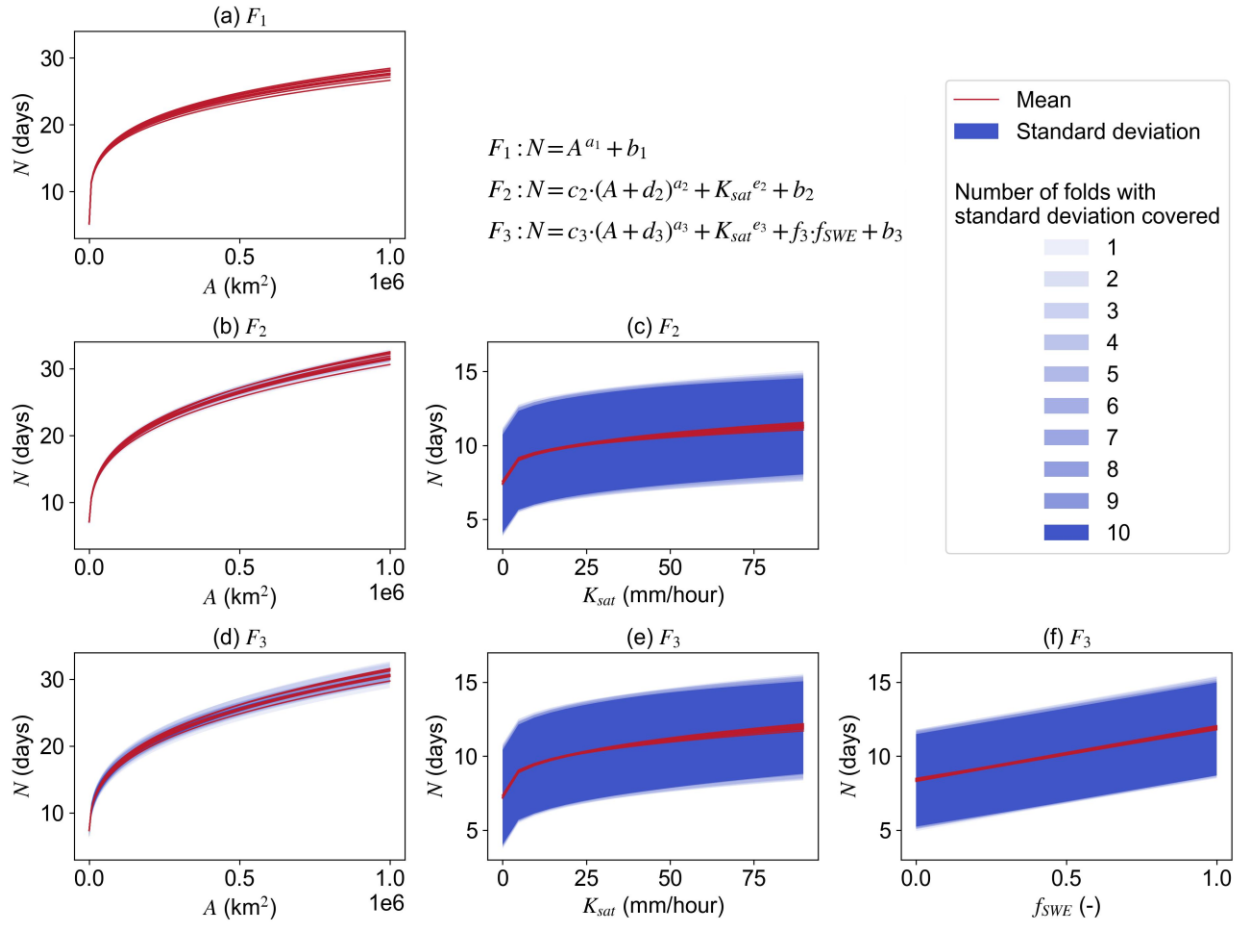
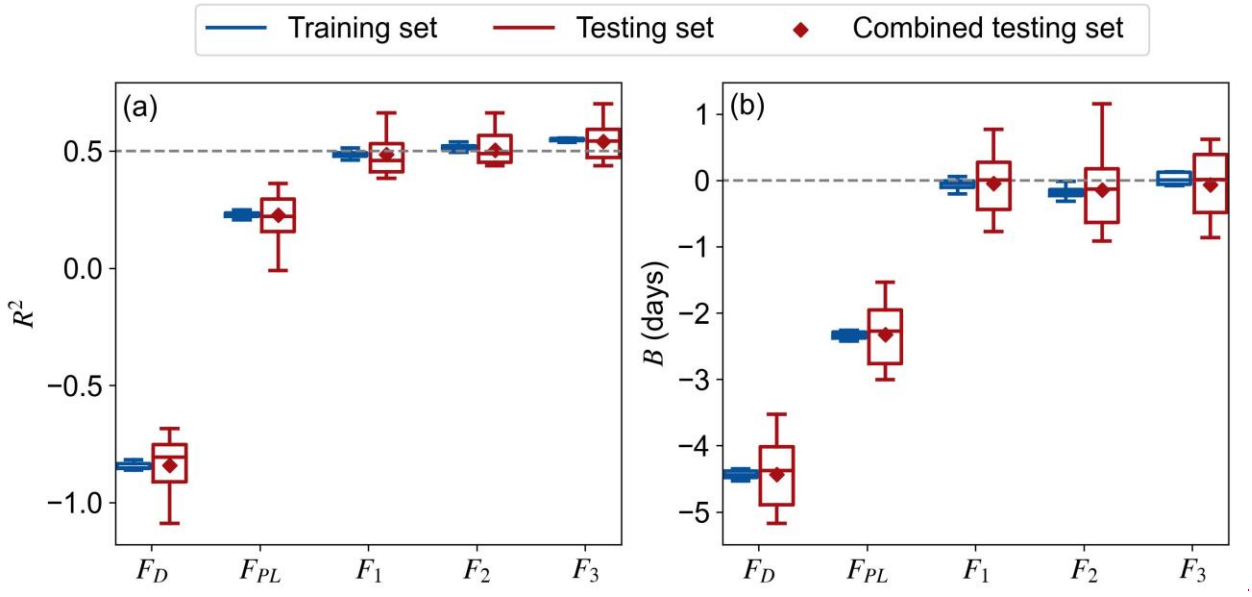


Figure 3. Marginal relationship of N on different predictors (A, K_{sat}, and f_{SWE}) that consist of the SR expressions (F_1, F_2, and F_3). Each line represents one of the ten instances of F_1, F_2, and F_3. Panels a, b, and d are for A; panels c and e are for K_{sat}; panel f is for f_{SWE}.

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 performance of formula F_D is poor, with negative R^2 values less than -0.6. F_{PL} shows higher~~improved~~ 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~~The~~ predictions from~~of N based on~~ the three SR-derived formulas, we observed higher

performance than~~significantly outperform those based on~~ the conventional~~two literature-suggested~~ formulas. The R^2 medians for F_1, F_2, and F_3 are 0.49, ~~0.52~~0.53, and ~~0.55~~0.56 for the training sets, and ~~0.46~~0.45, 0.49, and ~~0.54~~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 ~~0.49, 0.50, and 0.54~~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~~literature-suggested~~ formulas (**Figure 4**b). The medians B for F_1, F_2, and F_3 are ~~–0.05~~0.00/0.0~~61~~ days, ~~–0.18~~0.00/~~–0.13~~ 0.08 days, and 0.00/-0.0~~31~~ days for the training/testing sets, respectively, comparing to F_D and F_{PL} at -4.44/-4.38 days and -2.34/-2.28 days. These values for the combined testing set are -4.44, -2.33, ~~–0.05~~0.00, ~~–0.14~~0.00 and ~~–0.06~~0.00 days 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**).



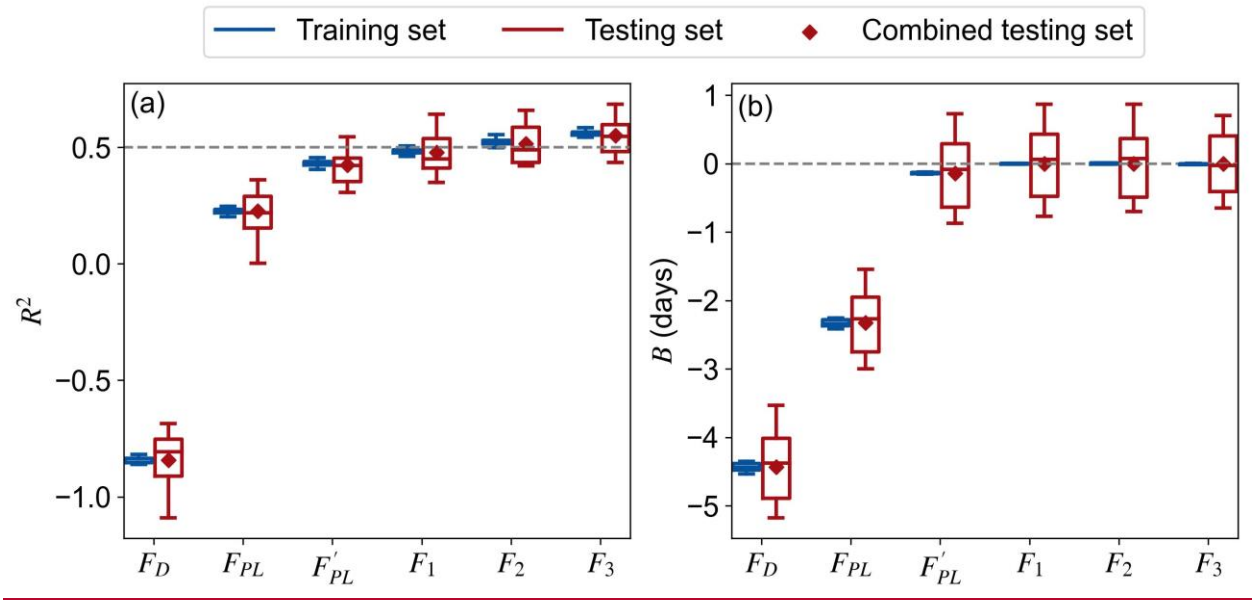


Figure 4. Performance of N predictions using the constant (F_D), power-law (F_{PL}), calibrated power-law (F'_{PL}), and the three SR formulas (F_1, F_2, and F_3) for the ten training and testing sets of the 10-fold cross-validation: coefficient of determination R^2 (a) and mean bias B (b).

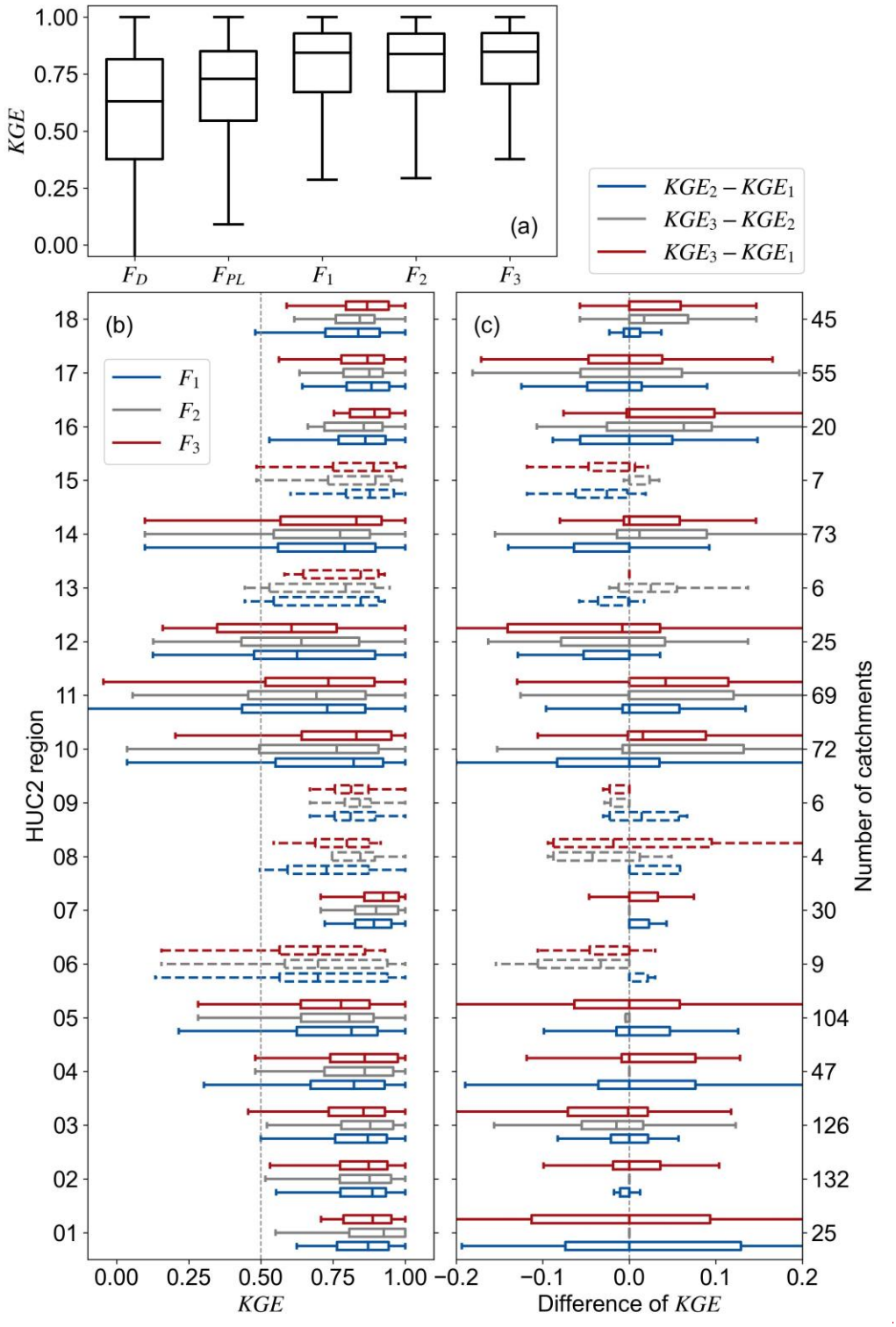
4.3. Application of N predictions in baseflow separation

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 (**Figure 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 $KGE < 0$~~KGE < 0~~, while 64% and 78% report $KGE > 0.5$~~KGE > 0.5~~, respectively. In contrast, the percentage of catchments with $KGE < 0$~~KGE < 0~~ drops to 1% for F'_{PL}, F_1, F_2, and F_3, and those with $KGE > 0.5$~~KGE > 0.5~~ rise to more than ~~87~~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 ~~0.61~~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 (Feng et al., 2020; Kratzert et al., 2019), 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 ~~06 and~~ 12. ~~This is because these SR formulas were fitted using data from all catchments, aiming to achieve optimal performance at a global scale. As a result, it may involve trade-offs that reduce its effectiveness in specific catchments.~~This may be because the SR formulas were calibrated using all catchments to optimize performance at the CONUS scale, 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.


1.00
0.75
0.50
0.25
0.00
KGE
(a)
F_D
F_{PL}
F_1
F_2
F_3
$KGE_2 - KGE_1$
$KGE_3 - KGE_2$
$KGE_3 - KGE_1$
(b)
(c)
F_1
F_2
F_3
HUC2 region
18
17
16
15
14
13
12
11
10
09
08
07
06
05
04
03
02
01
45
55
20
7
73
6
25
69
72
6
4
30
9
104
47
126
132
25
Number of catchments
0.00
0.25
0.50
0.75
1.00
KGE
−0.2
−0.1
0.0
0.1
0.2
Difference of KGE

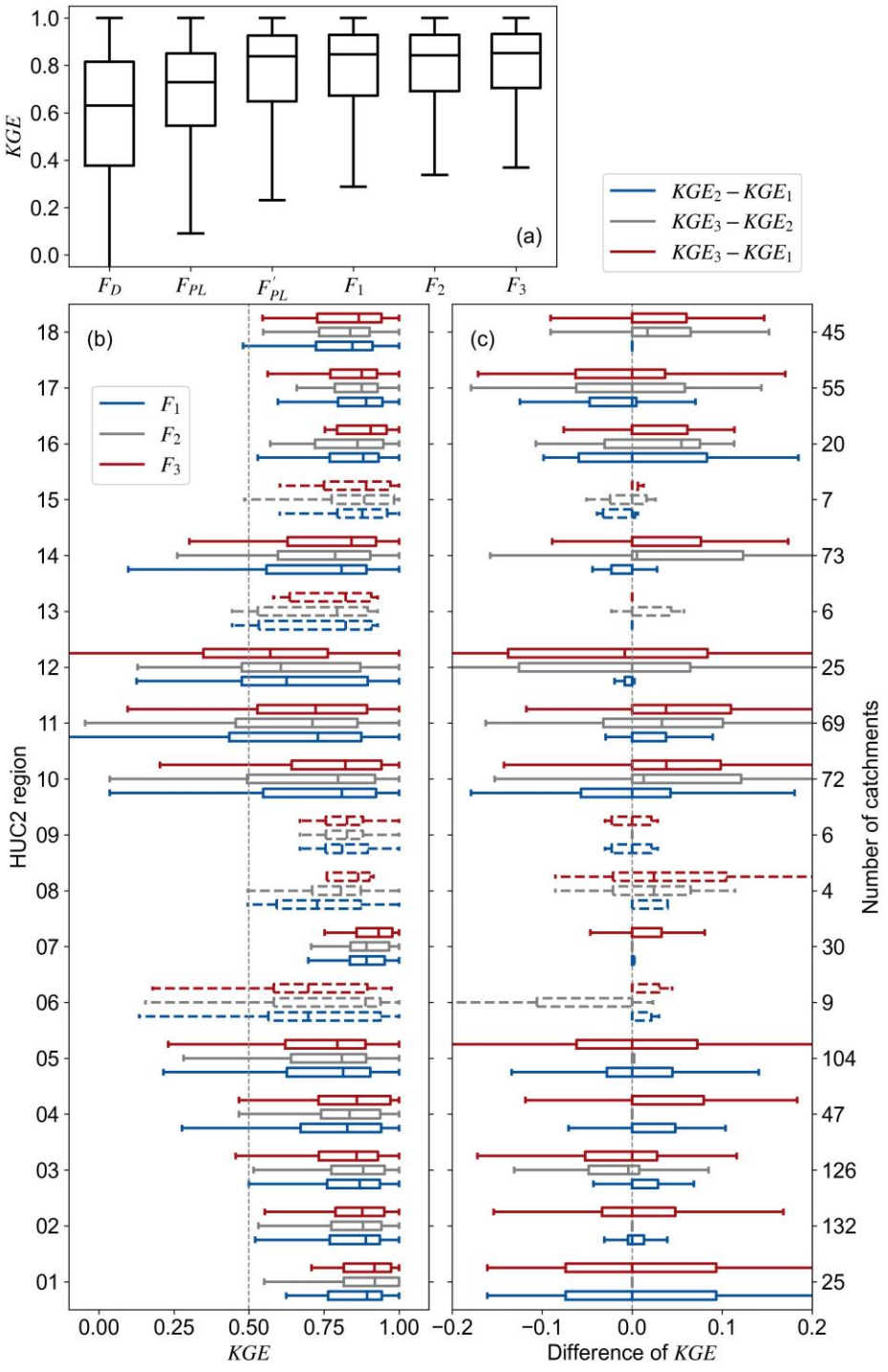
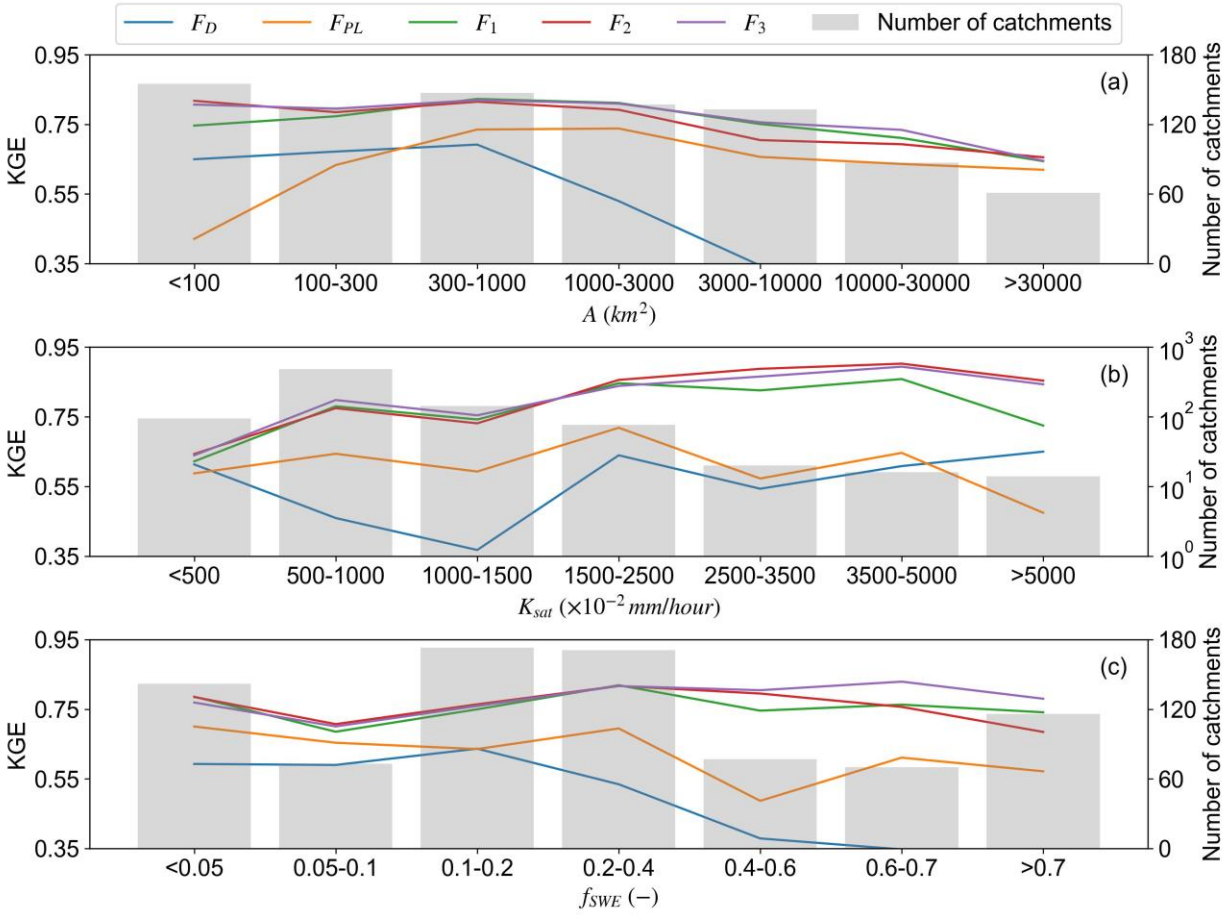


Figure 5. Baseflow separation performance based on different *NN* predictions for all catchments (a) and for catchments of the 18 HUC2 regions (b), and relative performance between each pair of the three SR cases for the HUC2 regions (c). Regions with fewer than 10 catchments are plotted~~indicated~~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

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 **Figure 6**a, when A exceeds 300 km^2, corresponding to $N = 5$ days by F_{PL}, the performance of F_D deteriorates markedly. This suggests that using $N = 5$ days 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/hour, 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 these regions~~. ~~However, when $K_{sat} > 25$ mm/hour, incorporating K_{sat} substantially improves the predictions, highlighting its influence for catchments with high averaged permeability.~~ In the intermediate ranges (5–25 mm/hour), the two regionally fitted power-law formulas (F_1 and F'_{PL}) outperform~~s~~ the conventional~~literature-suggested~~ power-law formula (F_{PL}). Although F_1 and F'_{PL} do~~es~~ not explicitly include K_{sat}, their~~its~~ 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. ~~However, w~~When $K_{sat} > 25$ mm/hour, incorporating K_{sat} substantially improves the predictions, highlighting the increasing importance of subsurface processes in highly permeable catchments~~highlighting its influence for catchments with high averaged permeability.~~ **Figure 6**c reveals the influence of f_{SWE}. F_3 outperforms all formulas~~F_1 and F_2~~ when $f_{SWE} > 0.4$, indicating the importance of f_{SWE} for~~that~~ catchments with strong~~erly influenced by~~ snow persistency~~exhibit lower performance if f_{SWE} is not accounted for, thereby underscoring its substantial effect on N~~. 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.



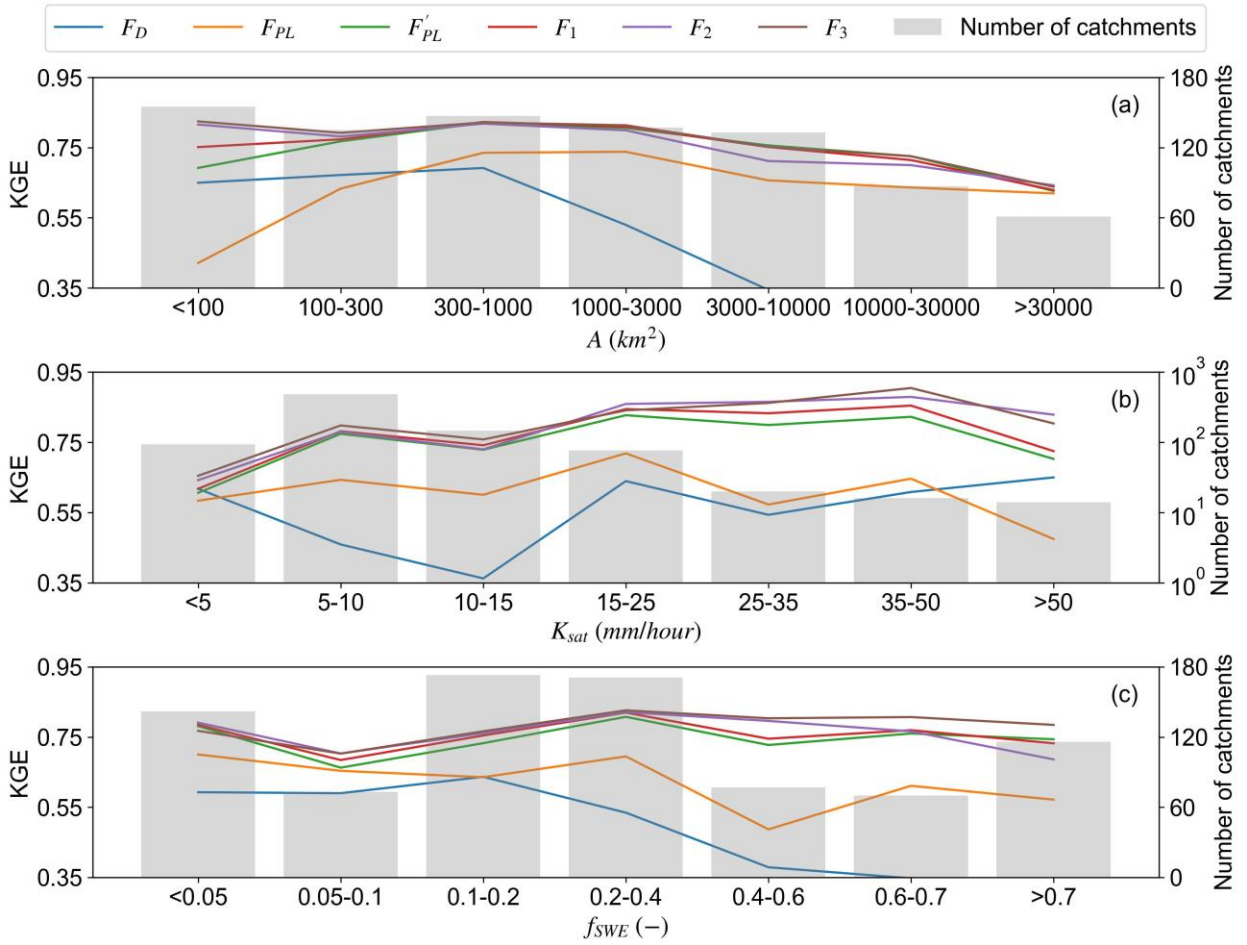


Figure 6. Performance of baseflow separation for different ranges of catchment area (a), catchment-averaged saturated hydraulic conductivity (b), and snow day fraction (c). The right y-axis for panel b is in logarithmic scale.

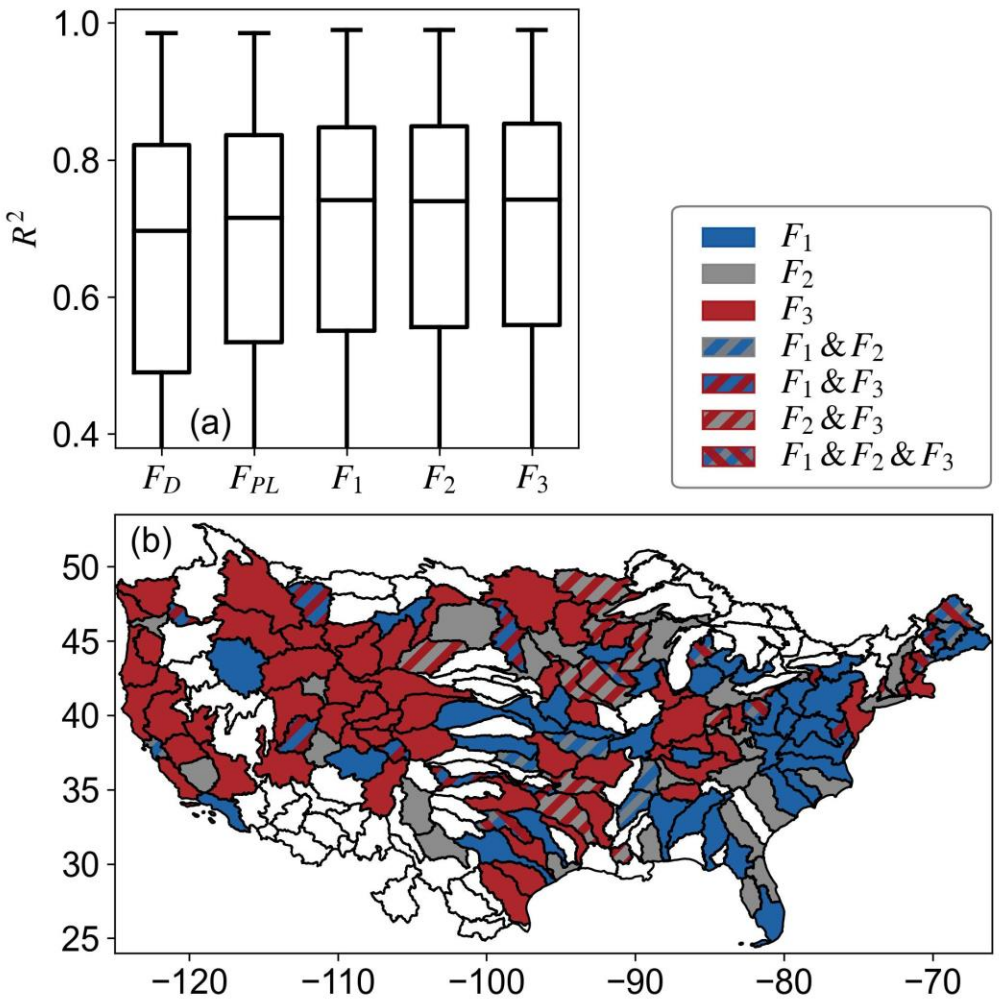
4.4. Estimation of SEC variation based on predicted *N*

~~**Figure 7** reveals the performance of SEC estimations by different N predictions for the testing sets. The overall distribution (**Figure 7**a) reveals that the ranking of performance follows the order $F_D < F_{PL} < F_1 < F_2 < F_3$, consistent with the NN predictions (**Figure 4**) and baseflow separation results (**Figure 5**). The medians of R^2 stay closed from 0.70 to 0.74, with similar interquartile ranges from 0.29 to 0.33.~~ The SEC estimation performance by different N predictions for the testing set is evaluated using R^2. The median R^2 values 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 (**Figure 4**) and the baseflow

separation results (**Figure 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~~b~~ 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 (**Figure 1**c), indicating surface runoff-dominated streamflow with limited relevance of delayed-flow processes (McMillan, 2020; Wu et al., 2021).

F_{1}~~2~~ 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, ~~most frequently F_2 and F_3,~~ suggesting a more complex interplay of hydrological drivers in these areas.


1.0
0.8
0.6
0.4
R^2
(a)
F_D
F_{PL}
F_1
F_2
F_3
F_1
F_2
F_3
F_1 & F_2
F_1 & F_3
F_2 & F_3
F_1 & F_2 & F_3
(b)
50
45
40
35
30
25
−120
−110
−100
−90
−80
−70

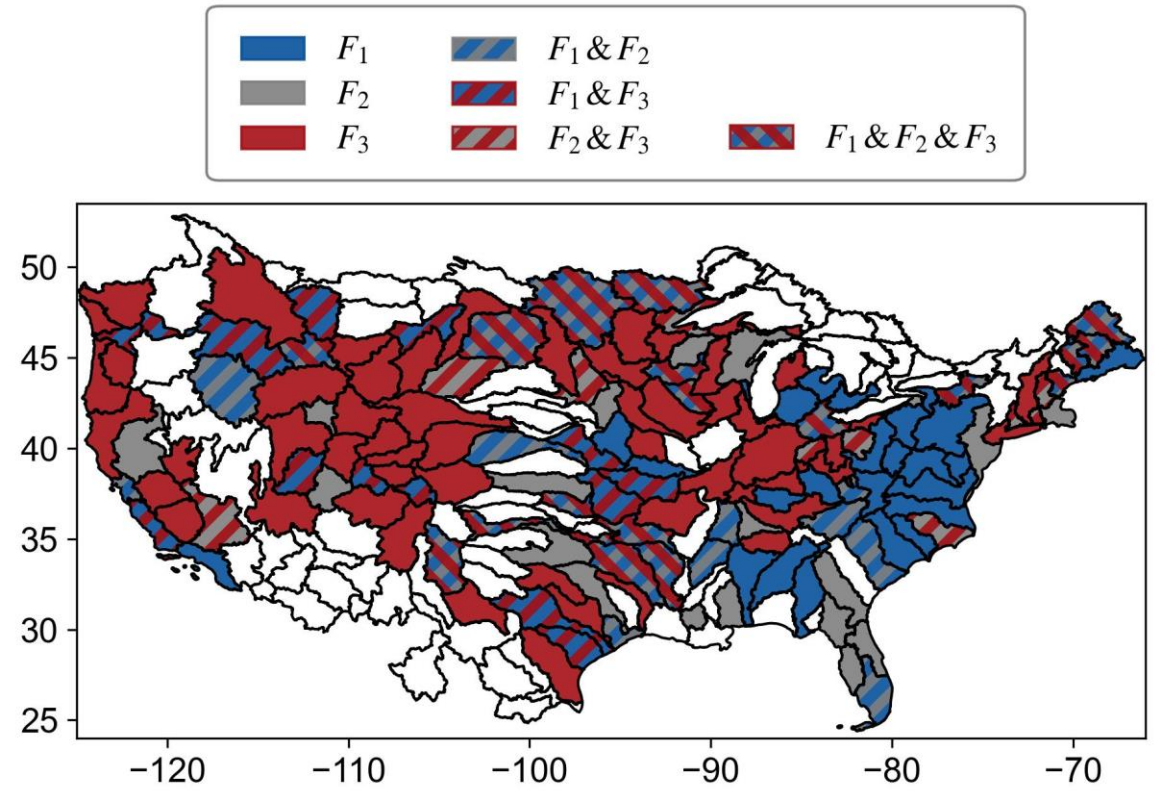


Figure 7. ~~**Performance of SEC estimations by different *NN* predictions for all catchments (a) and the S**~~**spatial distribution of the best-performing formulas for SEC estimations across different HUC4 regions** ~~**(b)**~~**. At the HUC4 scale, the best-performing formula is the one(s) appears most frequently as the best among all catchments within the region.** ~~**The best-performing formula for a HUC4 region is determined based on the average R^2 of all sites within the region.**~~

5. Discussions

5.1. Regionalization of baseflow separation parameter by SR

In this study, we used SR to derive~~identify three~~ mathematical expressions for the predictions of N using 9 catchment attributes. Across ten cross-validation iterations, the identified expressions exhibited consistent~~with identical~~ structures, ~~and~~ predictors, and nearly~~almost~~ identical regression coefficients, indicating that SR can yield stable functional relationships between catchment attributes and N~~using SR~~. Compared to the RF-based predictions reported by~~of N~~ Lin et al. (2026), the SR-based approach showed~~provides~~ 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~~but allows us to quantify how catchment attributes (i.e., catchment area, saturated hydraulic conductivity, and snow day fraction) influence baseflow separation~~ (Häfner et al., 2023; Karpatne et al., 2024). This explicit representation enables rigorous sensitivity analysis via differentiation; f~~F~~or instance, ~~by differentiating F_3, we obtain~~ the marginal effects derived from equation F_3 quantify~~of each attribute on N: $\frac{\partial N}{\partial A} = a_3 (A + K_{sat})^{a_3 - 1}$, $\frac{\partial N}{\partial K_{sat}} = a_3 (A + K_{sat})^{a_3 - 1}$, and $\frac{\partial N}{\partial f_{SWE}} = c_3$. These derivatives quantify the sensitivity of N to individual catchment attributes and facilitate a clearer interpretation of the underlying functional relationships.~~ 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.

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 (Elsken et al., 2019; Li et al., 2025). 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 ~~exponential~~ power-law operator in the SR operator set. The fact that all SR-derived formulas captured this power-law relationship indicates ~~that~~ the effectiveness~~use~~ of using this domain knowledge~~ was effective~~. Second, cross-validation was applied to assess model generalizability and reduce overfitting, ensuring representativeness of the SR expressions across the ~~diverse~~ 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 ~~0.23-0.25~~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 ~~0.24-0.25~~0.28-0.36 with respect to ~~to the sum of~~ K_{sat} ~~and A~~ (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 rout 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 (Barnhart et al., 2016; Godsey et al., 2014; Noor et al., 2023).

5.3. Trade-off between formula complexity and prediction accuracy

Our experiments reveal that formulas for baseflow separation with higher structural complexity generally~~tend to~~ 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.~~in some cases, simpler formulas yield comparable performance to more complex formulas.~~ For example, when K_{sat} and f_{SWE} take relatively small values ($K_{sat} < 25$ mm/hour and $f_{SWE} < 0.4$), the performance differences among F_1, F_2, and F_3 are minimal (**Figure 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}

($K_{sat} > 25$ mm/hour and $f_{SWE} > 0.4$), where their influence becomes more pronounced (Beven and Germann, 2013; Jenicek and Ledvinka, 2020). Moreover, the effects of K_{sat} and f_{SWE} are more evident when ~~A catchment area~~ is smaller than 100 ~~($A < 100$~~ km^2 (**Figure 6a**). As catchment area increases ($A > 100$ km^2), their influence is outweighed by A, highlighting the dominant role of drainage area in shaping the streamflow response time ~~(**Figure 6a**)~~(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: $N = A^{a_1} + b_1$, $N = d_2(A + c_2)^{a_2} + (K_{sat})^{e_2} + b_2$, and $N = d_3(A + c_3)^{a_3} + (K_{sat})^{e_3} + f_3 \cdot f_{SWE} + b_3$~~$N = A^{a_1} + b_1$, $N = (A + K_{sat})^{a_2} + b_2$, and $N = (A + K_{sat})^{a_3} + c_3 \cdot f_{SWE} + b_3$~~, 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 (**Figure 7**). 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.

Appendix A: Calculate SEC based on chemical and water balance relationship

For a two-component streamflow composition scenario, in which total streamflow is composed of a rapid surface/shallow subsurface flow component and a slower baseflow component, the mass balance relationships can be expressed as:

$$SEC_t Q_t = SEC_{s,t} S_t + SEC_{b,t} B_t \quad (A1)$$

$$Q_t = S_t + B_t \quad (A2)$$

where Q_t and SEC_t denote the total streamflow and the corresponding environmental tracer concentration at time step t, respectively. S_t and B_t are the surface flow and baseflow contributions at time step t. $SEC_{s,t}$ and $SEC_{b,t}$ are the tracer concentration of surface flow and baseflow.

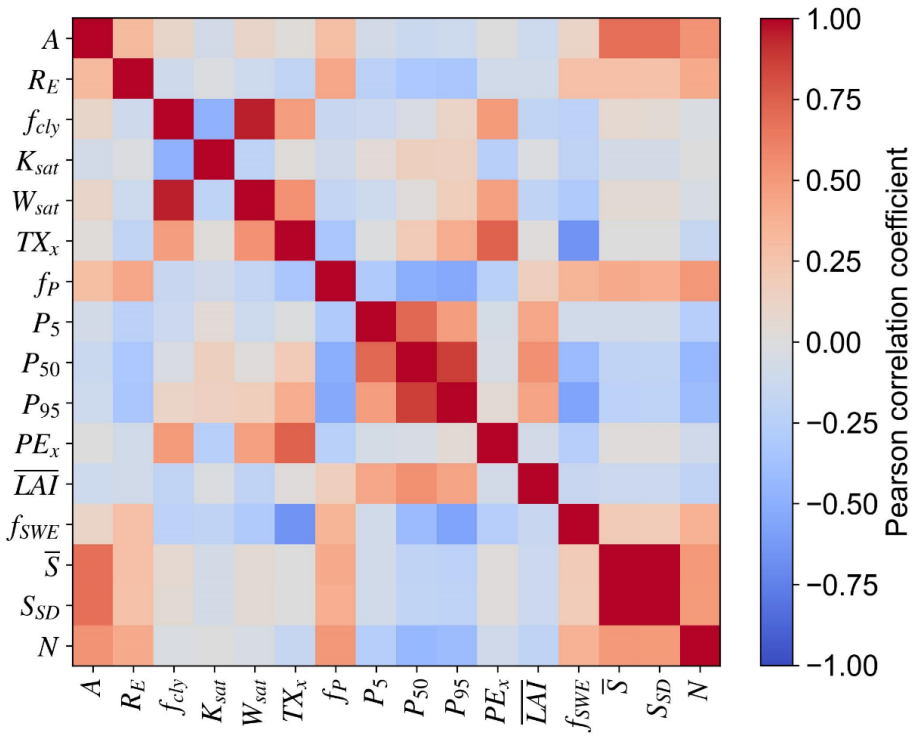
Combining equations (A1) and (A2), we obtain:

$$SEC_t Q_t = SEC_{s,t}(Q_t - B_t) + SEC_{b,t} B_t \quad (A3)$$

Dividing both sides by Q_t yields:

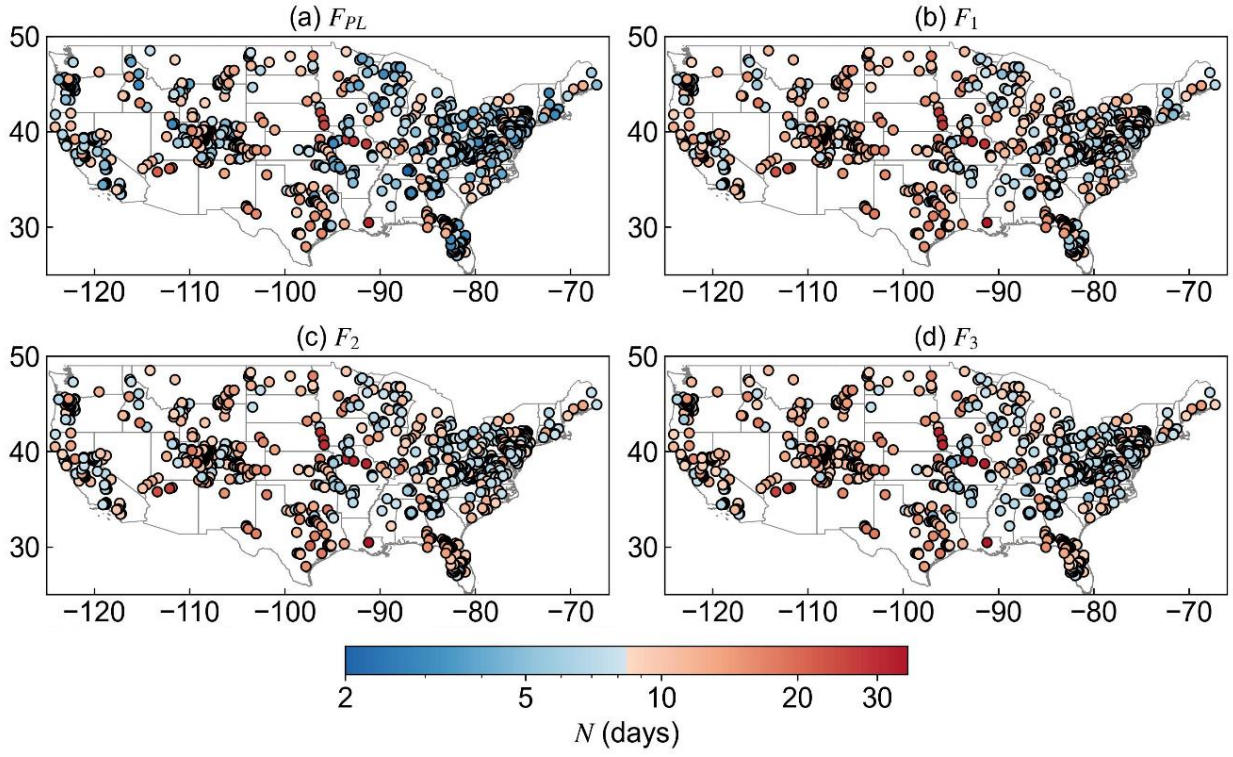
$$SEC_t = \frac{SEC_{s,t}(Q_t - B_t) + SEC_{b,t} B_t}{Q_t}, \quad (A4)$$

Appendix B:

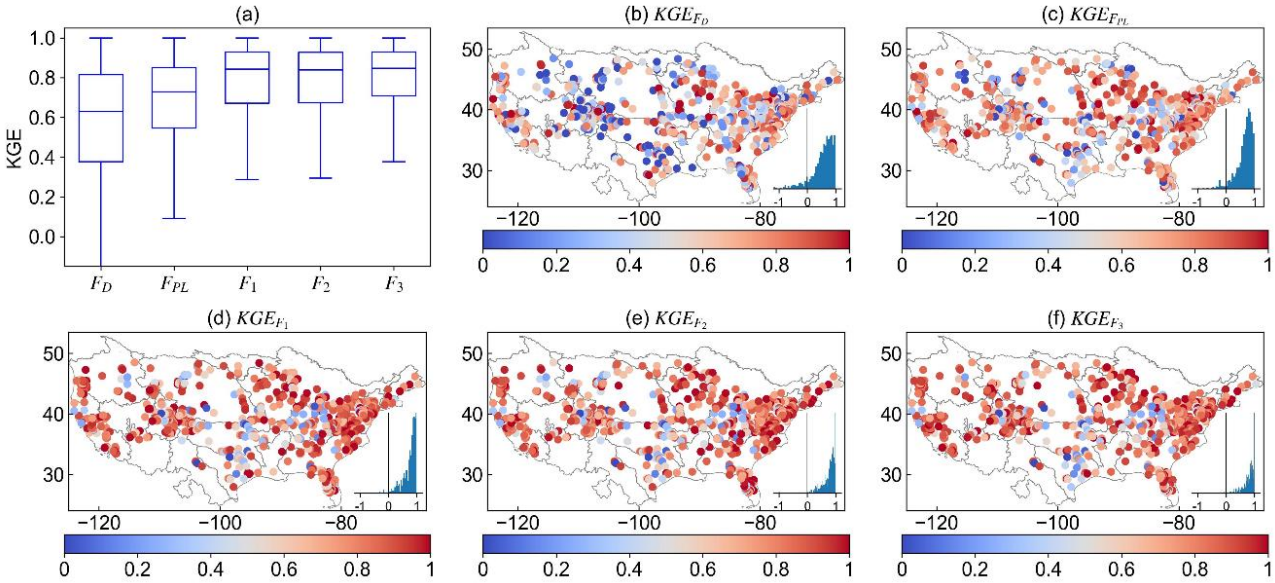


585

~~Figure B1. Pearson correlation coefficients between 15 catchment characteristics and the segment length parameter *N* of SMM.~~



~~**Figure B2. Spatial distributions of the *N* parameter predicted by the power-law function relationship with drainage area (a) and three SR-derived formulas (b-d).**~~



~~**Figure B3. Spatial distributions of baseflow separation performance measured by KGE for SMM with different parameterizations of *N*.**~~

Financial support. This work is supported by the National Natural Science Foundation of China (52579030) and the Guangdong Natural Science Foundation (2025A1515011666 and 2025A1515012264).

Competing interests. The authors declare that they have no conflict of interest.

Code and data availability. The streamflow and baseflow time series and optimal baseflow filters parameters are available on Mei et al. (2024b). 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).

Author contributions. Yongen Lin: Conceptuation, Methodology, Data curation, Formal analysis, Visualization, investigation, validation, writing–original draft; Dagang Wang: Conceptuation, Methodology, Writing–review & editing, supervision, Resources, funding acquisition, project administration; Yiwen Mei: Conceptuation, Methodology, Writing–review & editing, supervision, Resources, funding acquisition; Jinxin Zhu: Methodology, Writing–review & editing; Huan Wu: Funding acquisition, 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.

References

Aksoy, H., Unal, N.E., Pektas, A.O., 2008. Smoothed minima baseflow separation tool for perennial and intermittent streams. Hydrological Processes, 22(22): 4467-4476. DOI:https://doi.org/10.1002/hyp.7077

Apley, D.W., Zhu, J., 2020. Visualizing the Effects of Predictor Variables in Black Box Supervised Learning Models. Journal of the Royal Statistical Society Series B: Statistical Methodology, 82(4): 1059-1086. DOI:10.1111/rssb.12377

Barnhart, T.B. et al., 2016. Snowmelt rate dictates streamflow. Geophysical Research Letters, 43(15): 8006-8016. DOI:https://doi.org/10.1002/2016GL069690

Beven, K., Germann, P., 2013. Macropores and water flow in soils revisited. Water Resources Research, 49(6): 3071-3092. DOI:https://doi.org/10.1002/wrcr.20156

Cartwright, I., 2022. Implications of variations in stream specific conductivity for estimating baseflow using chemical mass balance and calibrated hydrograph techniques. Hydrology and Earth System Sciences, 26(1): 183-195. DOI:10.5194/hess-26-183-2022

Chadalawada, J., Herath, H.M.V.V., Babovic, V., 2020. Hydrologically Informed Machine Learning for Rainfall-Runoff Modeling: A Genetic Programming-Based Toolkit for Automatic Model Induction. Water Resources Research, 56(4): e2019WR026933. DOI:https://doi.org/10.1029/2019WR026933

Cranmer, M., 2023. Interpretable machine learning for science with PySR and SymbolicRegression. jl. arXiv preprint arXiv:2305.01582.

Diebold, F.X., Mariano, R.S., 1995. Comparing Predictive Accuracy. Journal of Business & Economic Statistics, 13(3): 253-263. DOI:10.1080/07350015.1995.10524599

Elsken, T., Metzen, J.H., Hutter, F., 2019. Neural architecture search: a survey. J. Mach. Learn. Res., 20(1): 1997–2017.

Feigl, M., Herrnegger, M., Klotz, D., Schulz, K., 2020. Function Space Optimization: A Symbolic Regression Method for Estimating Parameter Transfer Functions for Hydrological Models. Water Resour Res, 56(10): e2020WR027385. DOI:10.1029/2020WR027385

Feigl, M. et al., 2022. Automatic Regionalization of Model Parameters for Hydrological Models. Water Resources Research, 58(12): e2022WR031966. DOI:https://doi.org/10.1029/2022WR031966

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

Gnann, S.J., McMillan, H.K., Woods, R.A., Howden, N.J.K., 2021. Including Regional Knowledge Improves Baseflow Signature Predictions in Large Sample Hydrology. Water Resources Research, 57(2): e2020WR028354. DOI:https://doi.org/10.1029/2020WR028354

Godsey, S.E., Kirchner, J.W., Tague, C.L., 2014. Effects of changes in winter snowpacks on summer low flows: case studies in the Sierra Nevada, California, USA. Hydrological Processes, 28(19): 5048-5064. DOI:https://doi.org/10.1002/hyp.9943

Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. Journal of Hydrology, 377(1-2): 80-91. DOI:10.1016/j.jhydrol.2009.08.003

Gustard, A., Bullock, A.D., Dixon, J.M., 1992. Low Flow Estimation in the United Kingdom.

Häfner, D., Gemmrich, J., Jochum, M., 2023. Machine-guided discovery of a real-world rogue wave model. Proceedings of the National Academy of Sciences, 120(48): e2306275120. DOI:doi:10.1073/pnas.2306275120

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

Hou, X., Xie, D., Feng, L., Shen, F., Nienhuis, J.H., 2024. Sustained increase in suspended sediments near global river deltas over the past two decades. Nature Communications, 15(1): 3319. DOI:10.1038/s41467-024-47598-6

Humphrey, C.E. et al., 2022. Using Automated Seepage Meters to Quantify the Spatial Variability and Net Flux of Groundwater to a Stream. Water Resources Research, 58(6): e2021WR030711. DOI:https://doi.org/10.1029/2021WR030711

Jenicek, M., Ledvinka, O., 2020. Importance of snowmelt contribution to seasonal runoff and summer low flows in Czechia. Hydrol. Earth Syst. Sci., 24(7): 3475-3491. DOI:10.5194/hess-24-3475-2020

Karpatne, A., Jia, X., Kumar, V., 2024. Knowledge-guided machine learning: Current trends and future prospects. arXiv preprint arXiv:2403.15989.

Klotz, D., Herrnegger, M., Schulz, K., 2017. Symbolic Regression for the Estimation of Transfer Functions of Hydrological Models. Water Resources Research, 53(11): 9402-9423. DOI:10.1002/2017wr021253

Koza, J.R., 1994. Genetic programming as a means for programming computers by natural selection. Statistics and Computing, 4(2): 87-112. DOI:10.1007/BF00175355

Kronberger, G., de Franca, F.O., Burlacu, B., Haider, C., Kommenda, M., 2022. Shape-Constrained Symbolic Regression—Improving Extrapolation with Prior Knowledge. Evolutionary Computation, 30(1): 75-98. DOI:10.1162/evco_a_00294

Li, Q. et al., 2024. Advancing symbolic regression for earth science with a focus on evapotranspiration modeling. npj Climate and Atmospheric Science, 7(1). DOI:10.1038/s41612-024-00861-5

Li, X. et al., 2025. UniSymNet: A Unified Symbolic Network Guided by Transformer. arXiv preprint arXiv:2505.06091.

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

Lin, Y. et al., 2026. Regionalization of Optimal Baseflow Separation using Catchment-scale Characteristics. Water Resources Research, 62(3).

Makke, N., Chawla, S., 2024. Interpretable scientific discovery with symbolic regression: a review. Artificial Intelligence Review, 57(1). DOI:10.1007/s10462-023-10622-0

McGlynn, B.L., McDonnell, J.J., Seibert, J., Kendall, C., 2004. Scale effects on headwater catchment runoff timing, flow sources, and groundwater-streamflow relations. Water Resources Research, 40(7). DOI:https://doi.org/10.1029/2003WR002494

McMahon, T.A., Nathan, R.J., 2021. Baseflow and transmission loss: A review. WIREs Water, 8(4): e1527. DOI:https://doi.org/10.1002/wat2.1527

McMillan, H., 2020. Linking hydrologic signatures to hydrologic processes: A review. Hydrological Processes, 34(6): 1393-1409. DOI:https://doi.org/10.1002/hyp.13632

McMillan, H.K., Gnann, S.J., Araki, R., 2022. Large Scale Evaluation of Relationships Between Hydrologic Signatures and Processes. Water Resources Research, 58(6): e2021WR031751. DOI:https://doi.org/10.1029/2021WR031751

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

Mei, Y. et al., 2024a. Optimal Baseflow Separation Through Chemical Mass Balance: Comparing the Usages of Two Tracers, Two Concentration Estimation Methods, and Four Baseflow Filters. Water Resources Research, 60(7). DOI:10.1029/2023wr036386

Mei, Y., Zhu, J., Wang, D., 2024b. 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", [Dataset]. Zenodo. DOI:https://doi.org/10.5281/zenodo.8388365

Nagy, E.D., Szilagyi, J., Torma, P., 2024. Calibrating the lyne-hollick filter for baseflow separation based on catchment response time. Journal of Hydrology, 638. DOI:10.1016/j.jhydrol.2024.131483

Noor, K. et al., 2023. Snow sampling strategy can bias estimation of meltwater fractions in isotope hydrograph separation. Journal of Hydrology, 627: 130429. DOI:https://doi.org/10.1016/j.jhydrol.2023.130429

Pelletier, A., Andréassian, V., 2020. Hydrograph separation: an impartial parametrisation for an imperfect method. Hydrol. Earth Syst. Sci., 24(3): 1171-1187. DOI:10.5194/hess-24-1171-2020

Piggott, A.R., Moin, S., Southam, C., 2005. 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. Hydrological Sciences Journal, 50(5): null-920. DOI:10.1623/hysj.2005.50.5.911

Price, K., 2011. Effects of watershed topography, soils, land use, and climate on baseflow hydrology in humid regions: A review. Progress in Physical Geography: Earth and Environment, 35(4): 465-492. DOI:10.1177/0309133311402714

Rudin, C., 2019. Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. Nat Mach Intell, 1(5): 206-215. DOI:10.1038/s42256-019-0048-x

Samaniego, L., Kumar, R., Attinger, S., 2010. Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. Water Resources Research, 46(5). DOI:https://doi.org/10.1029/2008WR007327

Sheta, A., Abdel-raouf, A., Fraihat, K., Baareh, A.K., 2023. 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. DOI:10.14569/IJACSA.2023.0140489

Sólyom, P.B., Tucker, G.E., 2007. The importance of the catchment area–length relationship in governing non-steady state hydrology, optimal junction angles and drainage network pattern. Geomorphology, 88(1): 84-108. DOI:https://doi.org/10.1016/j.geomorph.2006.10.014

Song, W. et al., 2024. Towards data-driven discovery of governing equations in geosciences. Communications Earth & Environment, 5(1): 589. DOI:10.1038/s43247-024-01760-6

Stewart, M.K., 2015. Promising new baseflow separation and recession analysis methods applied to streamflow at Glendhu Catchment, New Zealand. Hydrol. Earth Syst. Sci., 19(6): 2587-2603. DOI:10.5194/hess-19-2587-2015

Stoelzle, M., Schuetz, T., Weiler, M., Stahl, K., Tallaksen, L.M., 2020. Beyond binary baseflow separation: a delayed-flow index for multiple streamflow contributions. Hydrology and Earth System Sciences, 24(2): 849-867. DOI:10.5194/hess-24-849-2020

Sundararajan, M., Najmi, A., 2020. The Many Shapley Values for Model Explanation. In: Hal, D., III, Aarti, S. (Eds.), Proceedings of the 37th International Conference on Machine Learning. PMLR, Proceedings of Machine Learning Research, pp. 9269--9278.

Tan, X., Liu, B., Tan, X., 2020. Global Changes in Baseflow Under the Impacts of Changing Climate and Vegetation. Water Resources Research, 56(9): e2020WR027349. DOI:https://doi.org/10.1029/2020WR027349

Tarasova, L., Gnann, S., Yang, S., Hartmann, A., Wagener, T., 2024. Catchment characterization: Current descriptors, knowledge gaps and future opportunities. Earth-Science Reviews, 252. DOI:10.1016/j.earscirev.2024.104739

Thorslund, J., van Vliet, M.T.H., 2020. A global dataset of surface water and groundwater salinity measurements from 1980–2019. Scientific Data, 7(1): 231. DOI:10.1038/s41597-020-0562-z

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

Wilstrup, C., Kasak, J., 2021. Symbolic regression outperforms other models for small data sets. arXiv preprint arXiv:2103.15147.

Wu, S., Zhao, J., Wang, H., Sivapalan, M., 2021. Regional Patterns and Physical Controls of Streamflow Generation Across the Conterminous United States. Water Resources Research, 57(6). DOI:10.1029/2020wr028086

Xie, J. et al., 2024. Majority of global river flow sustained by groundwater. Nature Geoscience. DOI:10.1038/s41561-024-01483-5

Xie, J. et al., 2022. Estimating Gridded Monthly Baseflow From 1981 to 2020 for the Contiguous US Using Long Short‐Term Memory (LSTM) Networks. Water Resources Research, 58(8). DOI:10.1029/2021wr031663

Xie, J. et al., 2020. Evaluation of typical methods for baseflow separation in the contiguous United States. Journal of Hydrology, 583. DOI:10.1016/j.jhydrol.2020.124628

Yan, X. et al., 2023. Detecting and attributing the changes in baseflow in China's Loess Plateau. Journal of Hydrology, 617: 128957. DOI:https://doi.org/10.1016/j.jhydrol.2022.128957

Zhang, J., Zhang, Y., Song, J., Cheng, L., 2017. Evaluating relative merits of four baseflow separation methods in Eastern Australia. Journal of Hydrology, 549: 252-263. DOI:10.1016/j.jhydrol.2017.04.004

Zhang, J. et al., 2020. Large-scale baseflow index prediction using hydrological modelling, linear and multilevel regression approaches. Journal of Hydrology, 585: 124780. DOI:https://doi.org/10.1016/j.jhydrol.2020.124780