ENVIRONMENTAL RESEARCH
COMMUNICATIONS

INCLUSIVE PUBLISHING
TRUSTED SCIENCE

Part of the
ENVIRONMENTAL RESEARCH
SERIES

PURPOSE-LED
PUBLISHING™

PAPER • OPEN ACCESS

Advanced prediction of river discharge: a systematic comparison of seven models and eight machine learning methods

To cite this article: Betul Mete *et al* 2026 *Environ. Res. Commun.* **8** 105013

View the article online for updates and enhancements.

You may also like

- Application of multilayer perceptron (MLP) method for streamflow forecasting (case study: Upper Citarum River, Indonesia)
Enung, Heri Kasyanto, Risna Rismiana Sari et al.
- Flood forecasting with Machine Learning in a scarce data layout
Théo Defontaine, Sophie Ricci, Corentin Lapeyre et al.
- Comparison of Artificial Neural Network and Support Vector Machine for Long-Term Runoff Simulation
Z Hassan, S Z Rosdi, A N Kamarudzaman et al.

This content was downloaded from IP address 194.27.118.1 on 24/10/2026 at 00:00

Environmental Research Communications

Check for updates

OPEN ACCESS

RECEIVED
24 June 2026

REVISED
23 September 2026

ACCEPTED FOR PUBLICATION
7 October 2026

PUBLISHED
15 October 2026

Original content from this work may be used under the terms of the Creative Commons Attribution 4.0 licence.

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



PAPER

Advanced prediction of river discharge: a systematic comparison of seven models and eight machine learning methods

Betul Mete[1,*], **Emirhan Mustafa Anık**[2], **Sinan Nacar**[3] and **Adem Bayram**[1]

[1] Department of Civil Engineering, Faculty of Engineering, Karadeniz Technical University, 61080 Trabzon, Türkiye
[2] Department of Civil Engineering, Faculty of Engineering, Bursa Uludağ University, 16059 Bursa, Türkiye
[3] Department of Civil Engineering, Faculty of Engineering and Architecture, Tokat Gaziosmanpaşa University, 60150 Tokat, Türkiye
[*] Author to whom any correspondence should be addressed.

E-mail: betulmete@ktu.edu.tr

Keywords: air temperature, empirical cumulative distribution function, support vector regression, precipitation, SHAP

Supplementary material for this article is available online

Abstract

This study investigates the predictability of daily river discharge (Q, using eight machine learning (ML) methods: multilayer perceptron (MLP), Kolmogorov–Arnold network, convolutional neural network, extreme gradient boosting, random forest, k-nearest neighbors (KNN), support vector regression (SVR), and linear regression (LR). Precipitation (P), air temperature (T), cumulative precipitation (P_C), and chronological order of months (α) were used as input parameters. The prediction models with different input combinations were developed and evaluated for three gaging stations in the Broad River Basin, South Carolina, USA. Model performance was assessed using various performance metrics and the empirical cumulative distribution function (ECDF). Furthermore, the SHAP technique was applied to examine the relationship between the input parameters and Q. The model, which incorporates multiple lagged P inputs along with T, P_C, and α, consistently achieved the lowest errors and the highest Nash–Sutcliffe efficiency coefficient values (0.546–0.785) across all stations in all datasets. Whilst SVR is the most successful method, the other methods (except for LR) have achieved results very close to SVR's. In the ECDF analysis, for all datasets and gaging stations, the SVR and MLP methods demonstrated high consistency between observed and predicted Q distributions at high values, whilst the MLP and KNN methods showed high consistency at low values. Nevertheless, when different flow regimes and bootstrap-based confidence intervals are considered, the SVR model has demonstrated more robust performance compared to other models. Based on SHAP values, P_C and α are the most influential features. This study demonstrates that ML methods can be effectively used for daily Q prediction and provide a validated basis for use with future meteorological inputs under climate change scenarios.

1. Introduction

Discharge (Q) is a fundamental hydrological variable that affects many processes directly, including the biological diversity of aquatic ecosystems, agricultural productivity, hydroelectric power generation, and water supply to urban areas (Venkataramana *et al* 2021, Bhattarai and Shrestha 2023, Jimenez *et al* 2025, Speir *et al* 2025). It may not be possible to measure Q directly across all desired areas due to the limited availability of gauging station and the physical and economic challenges of conducting observations (Muhebwa *et al* 2024). For this reason, it is important to make accurate and reliable Q predictions, both to address data gaps and to support decision-making for the sustainable management of limited water resources. Although empirical equations (Choo *et al* 2014, Shibata and Ito 2014), hydrographs (Wang *et al* 2009, Arico *et al* 2010), and rating curves (Di Baldassarre and Montanari 2009, Jalbert *et al* 2011) have been widely used in literature, machine learning (ML) based methods have recently gained increasing attention due to their ability to model the complex and nonlinear relationships in river systems

© 2026 The Author(s). Published by IOP Publishing Ltd

effectively. A general overview of the studies using ML reveals that the most commonly employed methods are the multilayer perceptron (MLP) (Enung *et al* 2023, Kisi *et al* 2024), extreme gradient boosting (XGB) (Szczepanek 2022, Alipour 2025), random forest (RF) (Talukdar *et al* 2023; Saleh *et al* 2024), long short-term memory (LSTM) (Ghimire *et al* 2021), gated recurrent units (GRU) (Workneh *et al*, 2025), and support vector regression (SVR) (Asadollahi *et al* 2024) for Q prediction. However, studies that combine methods from different classes (and thus different learning architectures) and compare them comprehensively remain quite limited (Kisi *et al* 2024, Rather *et al* 2025). Kaur and Chavan (2025) evaluated 11 deep-learning and ML models for one-day-ahead streamflow forecasting in four sub-basins of the Krishna River Basin, India. It was found that deep learning models generally outperformed classical ML approaches, but the best methods varied across sub-basins. However, it was also noted that ML methods such as RF and XGB exhibit performance comparable to deep learning architectures. Method performance can vary considerably across basins depending on local hydro-climatic characteristics.

The Broad River Basin plays a pivotal role in regional energy security, hosting several hydropower dams, hydroelectric power plants (HPPs), and a strategically important nuclear power station. Downstream operations are centered around the Parr Shoals Dam, which supports the 15 MW Parr HPP and a 576 MW pumped-storage facility. Furthermore, the reservoir serves as a vital source of cooling-water for the 973 MW Virgil C. Summer nuclear generating station (San *et al* 2026). Thus, potential shifts in river Q pose a direct threat to the reliability of energy supply. Consequently, reliable Q prediction is a fundamental requirement for optimizing operational processes and planning energy production in the basin. Despite the availability of long-term continuous hydro-meteorological datasets, as far as is known, no prior study that has comprehensively assessed a wide range of ML methods for Q-prediction across the Broad River Basin and its tributaries. This study aims to address this gap by providing a station-based modeling approach for the Broad River Basin to support possible future Q projections under changing climate conditions.

To comprehensively investigate the predictability of daily Q values in the Broad River Basin, this study uses eight ML-based methods, including MLP, Kolmogorov–Arnold network (KAN), convolutional neural networks (CNNs), XGB, RF, SVR, k-nearest neighbors (KNN), and linear regression (LR). The analyses were conducted using data from three United States Geological Survey (USGS) gaging stations: one on the main branch (02156500, the Broad River) and two on tributaries (02160105, the Tyger River, and 02160700, the Enoree River). The average precipitation (P) and air temperature (T) series were calculated for each station using the Thiessen polygon method over its drainage area. In addition to these primary meteorological variables, cumulative precipitation (P_C) and the chronological order of the months (1–12) were incorporated as inputs to the predictive models. The accuracy of the models and methods was assessed using various performance statistics. This study demonstrates the effectiveness of ML-based methods at the station level across the basin. It provides scientific insights to support the operational management of hydropower dams and energy generation systems.

2. Materials and methods

The overall workflow of the proposed methodology is illustrated in figure 1. First, the input variables were determined through the correlation analysis and cumulative precipitation (P_C) analysis. Based on the selected variables, seven different input combinations (QP1–QP7) were constructed and evaluated using eight ML methods. Model performance was subsequently assessed using statistical metrics, scatter plots, empirical cumulative distribution functions (ECDFs), flow-regime-based analyses, and bootstrap confidence intervals. Finally, SHAP-based analyses were performed to interpret the contribution of each input variable to the model predictions. The details of each stage are presented in the following sections.

2.1. Study area

The Broad River is located within the South Atlantic–Gulf hydrological region of the United States of America (USA). The river originates on the eastern slopes of the Blue Ridge Mountains and flows southeastwards into South Carolina. The main tributaries of the river are the First Broad, Second Broad, Pacolet, Tyger, and Enoree (Nacar *et al* 2019). The river has a drainage area of around 13 700 km^2. Most of the river basin is forested (60.7%), and built-up and dense short vegetation areas cover approximately 23.0 and 10.4% of the total area, respectively (Heberger 2022a).

The basin has gaging stations operated by the USGS and meteorological observation stations operated by the National Weather Service. To ensure the consistency and reliability of the analysis, this study selected stations with long-term (1986–2022) daily Q and meteorological measurements. The locations of the stations are shown in figure 2.

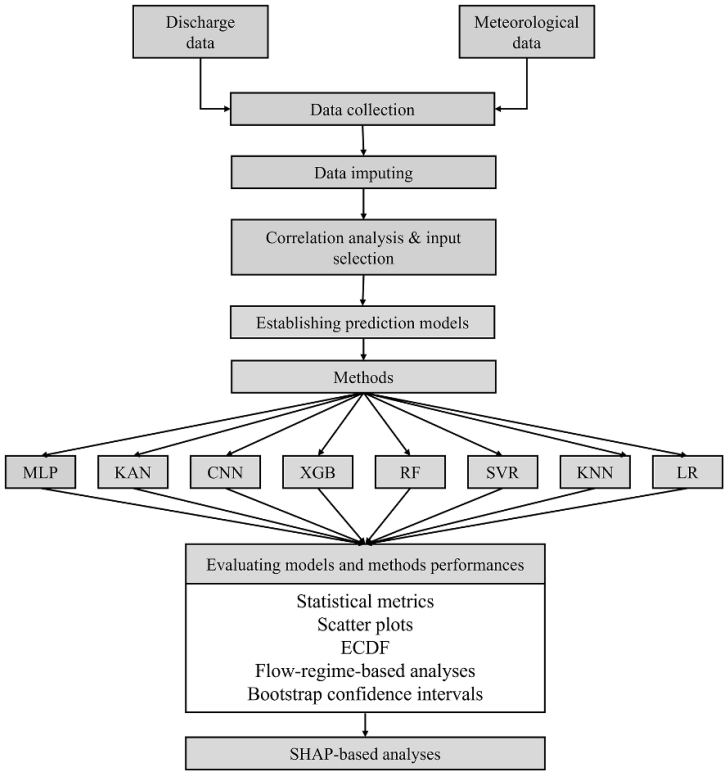


Figure 1. The flowchart of the study.

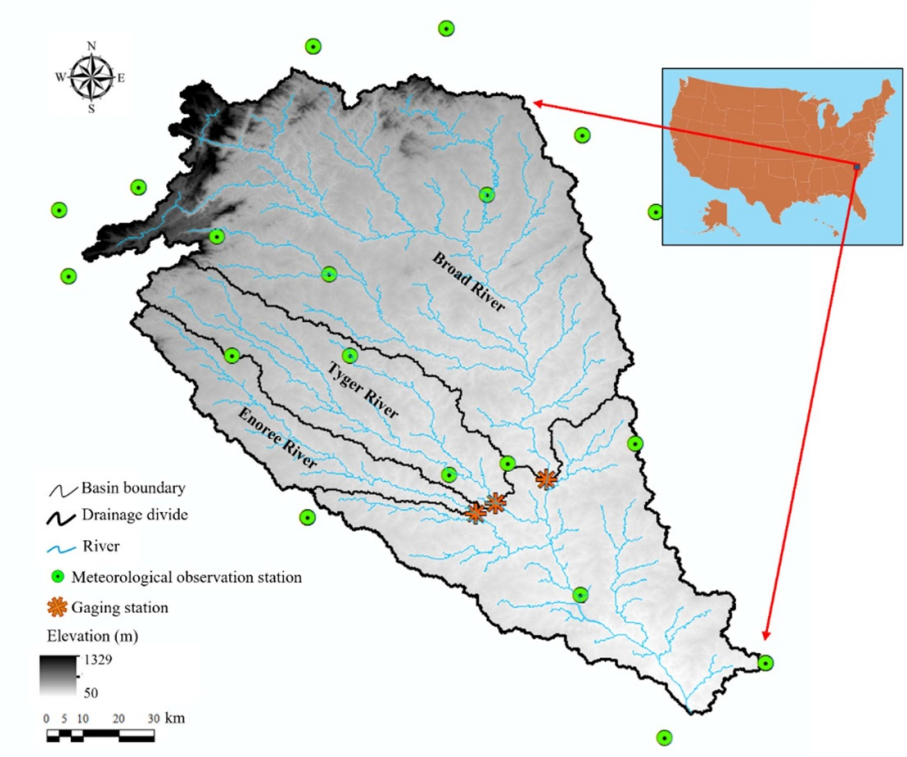


Figure 2. Meteorological observation and river gaging stations, the Broad River Basin, South Carolina, USA (adapted from Heberger 2022b).

2.2. ML methods

In this study, MLP, KAN, CNN, XGB, RF, SVR, KNN, and LR models were used to compare the predictive performance of different ML methods. Accordingly, ML methods that are more compatible with the data structure and can directly utilize the existing input structure were preferred. Among these methods, the MLP, comprising input, hidden, and output layers, is a type of artificial neural networks widely used in hydrological forecasting (Nacar *et al* 2018). Each layer consists of interconnected nodes that process information and can learn patterns from the training data, and these nodes are initially connected with random weight values (Moridnejad *et al* 2015). To minimize the difference between the actual and predicted outputs, the weight values are updated based on the errors (Gardner and Dorling 1998, Mete *et al* 2026). The KAN, developed based on the Kolmogorov–Arnold theorem, is a method proposed as an alternative to the MLP (Z. Liu *et al* 2025). The theorem assumes that any continuous multivariate function on a finite domain can be expressed as the sum of a finitely many univariate functions (Xu *et al* 2024). Building on this theorem, the KAN can transform multiple inputs into a single output using a series of univariate functions. Unlike the MLP, which uses fixed activation functions, the KAN employs learnable univariate functions parameterized by B-splines (C. Liu *et al* 2025, Anik *et al* 2026). The CNN is a deep learning architecture that transforms the input through a sequence of computational stages before producing the target output (Dehghani *et al* 2023). Convolution operations extract informative patterns from the input by applying learnable kernels, allowing the network to identify relevant local structures within the data. These extracted patterns can then be compressed through pooling operations, which retain the most informative characteristics while reducing the amount of information passed to subsequent stages and helping to control model complexity (Borovykh *et al* 2017, Zuo *et al* 2019). The resulting representations are finally processed by fully connected units to establish the relationship between the extracted features and the predicted variable (Shu *et al* 2021). The XGB (Chen and Guestrin 2016) employs a gradient-boosting framework that combines multiple base learners, uses an additive approach, and can function as a parallel tree booster (Naghibi *et al* 2020, Shen *et al* 2022). Using a more standard algorithm than gradient boosting reduces overfitting and improves performance (Bandara *et al* 2020). The RF is a prediction method that constructs an ensemble of randomized decision trees (Khan *et al* 2019, Rehamnia *et al* 2026). The RF employs a decision-tree construction methodology that utilizes randomized decision-making to reduce errors and improve accuracy (Wang *et al* 2022). The SVR (Cortes and Vapnik 1995) is an ML approach that aims to learn a regression function that adequately captures the relationship between the input dataset and the target values. To this end, the SVR maps the inputs into a high-dimensional space via a linear function and ensures separation of the data sets using an ideal hyperplane (Bargam *et al* 2024). The KNN is an ML method that uses a relevant variable with similar historical patterns to represent the current situation. The KNN method is generally non-parametric and makes no assumptions about the dataset distribution's (Sun and Trevor 2017). The LR is one of the simplest and most fundamental ML methods for determining the statistical relationship between a single dependent variable and one or more independent variables (Marill 2004, Forkuor *et al* 2017). The hyperparameters used in the ML methods are presented in supplementary table S1. For each ML method, a predefined and finite set of candidate values was specified for each hyperparameter, and all possible combinations of these candidate values were evaluated independently. Each candidate configuration was fitted separately using the training dataset, without using the test observations during model fitting. After fitting, the predictive performance of each configuration was calculated separately for the training, validation, and test datasets, where applicable, using the predefined evaluation metrics. The candidate configurations were then compared by jointly considering their predictive performance across the available datasets, with the Nash–Sutcliffe efficiency coefficient (NSEC) used as the primary criterion and the root mean square error (RMSE) and mean absolute error (MAE) additionally considered to assess prediction errors and ensure consistent predictive performance. The configuration selected through this systematic comparison was subsequently used as the final model configuration for the corresponding algorithm.

2.3. Data processing and model development

Before the development of Q prediction models, the missing values in meteorological datasets were imputed using regression analysis based on concurrent observations from neighboring stations. Missing data were randomly distributed across the 1986–2022 period and affected less than 10% of records across all stations. The Thiessen polygon method was used to determine the meteorological stations for the drainage areas of the gaging stations. The mean P and T series for each drainage area were calculated using the area-weighted ratios of the polygons. Spearman correlation analyses were performed between Q and meteorological data. The analysis considered the values of P and T over a time lag of 0–5 d (figure 3). An analysis of the results reveals that across all rivers, the input that correlates most strongly

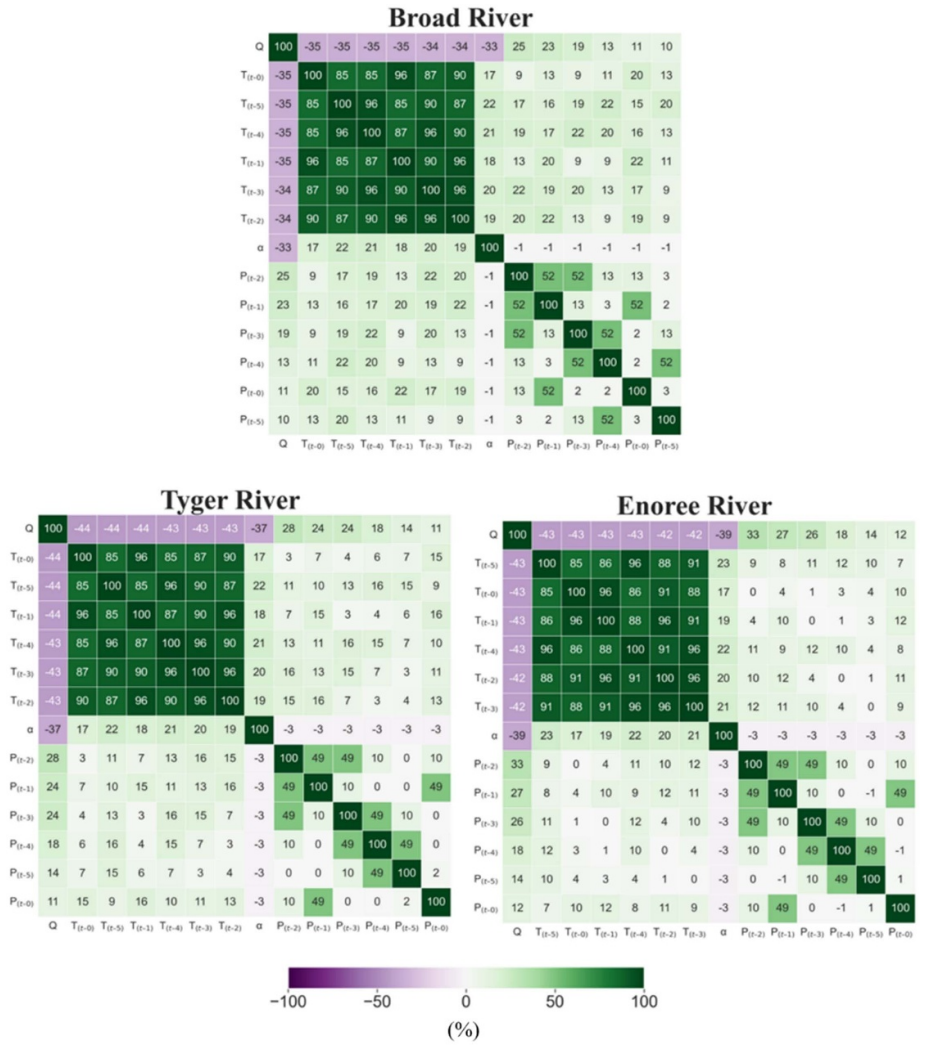


Figure 3. Spearman correlation coefficients between river discharge and meteorological variables, Broad River Basin, South Carolina.

with Q is T, with values ranging from approximately −35 to −45%. Following T, the inputs showing the highest correlation with Q are α, $P_{(t-2)}$, $P_{(t-1)}$, and $P_{(t-3)}$, respectively. Conversely, the variables with the lowest correlation with Q values are $P_{(t-4)}$, $P_{(t-5)}$, and $P_{(t-0)}$, respectively.

The direct use of all these variables in the ML models increases computational costs and reduces interpretability. Therefore, using as few input variables as possible is more efficient, as it reduces complexity, in line with the parsimonious modeling principle (Burnham and Anderson 2002). Consequently, T values generally exhibit high correlations with one another (above 85%). Li *et al* (2025) argued that a correlation coefficient exceeding 0.80 indicates a very strong relationship between two variables. In such cases, it is more appropriate to include only one variable in the model to reduce multicollinearity and prevent data redundancy. Consequently, given that the correlation coefficients among the different time-lagged T parameters exceeded 0.80, amongst themselves, it was decided to use only the $T_{(t-0)}$ value in the modeling. It is observed that the other variables do not exhibit strong correlations with one another, and it has been concluded that all variables, except $P_{(t-0)}$, which has the lowest correlation, should be included in the modeling. Additionally, P_C data was derived over periods of 2–30 d to examine the influence of accumulated precipitation over preceding days. Using the most appropriate P_C period rather than all P_C periods is more suitable for the modeling objective. Furthermore, extending the P_C period too far back in time requires a much larger number of P values for modeling. Consequently, it is necessary to select an appropriate P_C period that will provide a high contribution whilst going back as little as possible into the past. To this end, figure 4, which shows the variation in the correlation values between

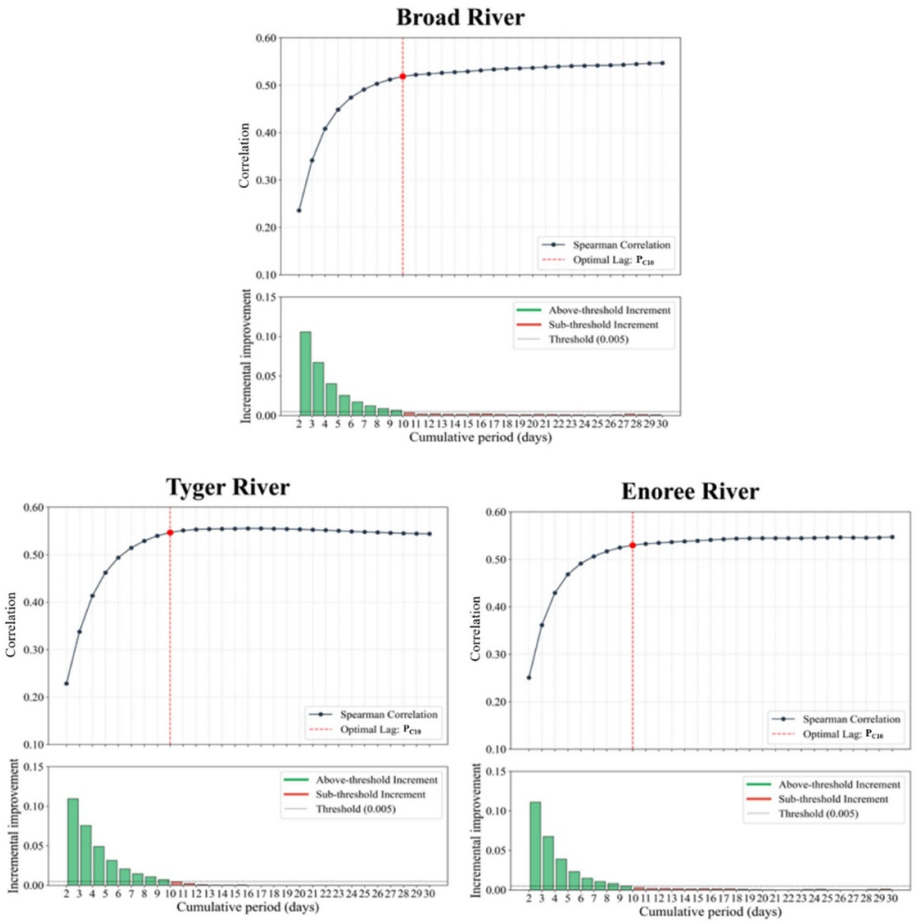


Figure 4. Spearman correlation coefficients between discharge and cumulative precipitation periods, Broad River Basin.

P_C and Q for different periods as a line graph, illustrates the incremental improvement by calculating the difference between the correlation values of each P_C period with Q and the next P_C period with Q.

The incremental improvement value decreases as the P_C period increases. Within the scope of this study, a threshold of 0.005 was set for incremental improvement, and using P_C values below this threshold was deemed unnecessary. When this approach is applied, the 10 d cumulative precipitation (P_{C10}) is determined for the river basin as a suitable value, and the fact that all rivers yield the same period demonstrates that the selected threshold value is consistent and balanced.

Finally, α, representing the seasonal cycle, has also been included. In addition, cyclic month encoding using sine and cosine transformations was also evaluated with the SVR method, and the corresponding performance results are presented in supplementary table S2. Although the cyclic representation yielded slight performance improvements for several model configurations, particularly for QP6 and QP7, these gains were not consistent across all rivers and input combinations. Overall, the predictive performance remained comparable to that obtained using the direct chronological month encoding. Furthermore, the sine/cosine representation complicated the interpretation of the seasonal variable in the SHAP analysis by distributing its contribution across two separate features. Therefore, the chronological month index (1–12) was retained throughout this study. Consequently, for each river basin, seven Q prediction models were established using different combinations of input parameters, based on the correlation analysis results. The same set of input combinations was applied across all basins, but the models were trained separately using station-specific data for each basin. The model contents are given in table 1.

The data were divided into training, validation, and testing datasets while preserving the chronological sequence of the time series. Specifically, the first 25 years of data were used for training (67.6%), the subsequent 6 years for validation (16.2%), and the final 6 years for testing (16.2%). This partitioning prevents data leakage from future observations and enables a realistic evaluation of the model's predictive performance. For each method, all predefined hyperparameter configurations were evaluated, and the configuration providing the best predictive performance was selected. For the neural network methods (MLP, CNN, and KAN), the validation set was used during training to tune hyperparameters and

Table 1. Discharge prediction models.

Model	Input parameter							
QP1	$P_{(t-2)}$	α						
QP2	$P_{(t-2)}$	$T_{(t-0)}$	α					
QP3	$P_{(t-2)}$	P_{C10}	α					
QP4	$T_{(t-0)}$	P_{C10}	α					
QP5	$P_{(t-2)}$	$T_{(t-0)}$	P_{C10}	α				
QP6	$P_{(t-1)}$	$P_{(t-2)}$	$P_{(t-3)}$	$T_{(t-0)}$	P_{C10}	α		
QP7	$P_{(t-1)}$	$P_{(t-2)}$	$P_{(t-3)}$	$P_{(t-4)}$	$P_{(t-5)}$	$T_{(t-0)}$	P_{C10}	α

implement early stopping to mitigate overfitting. In contrast, the other methods (XGB, RF, SVR, KNN, and LR) do not require a dedicated validation dataset during model development. Therefore, these models were developed using combined training and validation datasets, corresponding to 83.8% of the total data. In comparison, the same independent testing dataset (16.2%) was retained for performance evaluation across all methods. The descriptive statistics of the datasets is summarized in table 2. Since the lagged precipitation variables exhibited nearly identical descriptive statistics, only the statistics of $P_{(t-2)}$, together with $T_{(t-0)}$, P_{C10}, and Q, are reported.

According to table 2, the P with time lag and Q series show strong positive skewness (Cs) and high kurtosis (Ck). This indicates non-normal and heavy-tailed distributions. The presence of non-normal distributions and sensitivity to extreme values required the application of a scaling procedure before conducting the analysis. To minimize the effect of extreme values and enhance the model's robustness, all data were normalized using equation (1), where F_n, F_i, F_{min}, and F_{max} represent the normalized, raw, minimum, and maximum figures, respectively, while a and b control the target range. The F_{min} and F_{max} values were determined exclusively from the training dataset, and the same values were subsequently used to normalize the validation and test datasets. Particularly for neural network-based methods, the most used transfer function has an output range bounded between 0 and 1. Normalized values close to the boundary limits cause network training to slow down and become ineffective. For this reason, a and b were set to 0.8 and 0.1, respectively, yielding a normalized range of 0.1–0.9. This range was applied uniformly across all methods for consistency,

$$F_n = a \times \frac{F_i - F_{min}}{F_{max} - F_{min}} + b. \tag{1}$$

The predictive performance of the models and methods was quantified using the RMSE, MAE, and NSEC, given in equations (2)–(4). These metrics were used exclusively for model evaluation and hyperparameter selection. The best-performing model was determined based on achieving higher NSEC values together with lower RMSE and MAE values. The RMSE assigns greater weight to large deviations because of its squared-error structure; the MAE provides a more stable, less outlier-sensitive measure of average error. Therefore, both statistics were used to quantify the magnitude of residual errors. Additionally, the NSEC was used to evaluate the overall predictive efficiency and the models' explanatory capability in reproducing the observed Q values,

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(t_i - td_i)^2} \tag{2}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|(t_i - td_i)| \tag{3}$$

$$\mathrm{NSEC} = 1 - \frac{\sum_{i=1}^{N}(t_i - td_i)^2}{\sum_{i=1}^{N}(t_i - \bar{t})}. \tag{4}$$

In addition to the performance metrics, the ECDF was employed to assess the ability of the developed models to reproduce the statistical distribution of observed Q values. The ECDF is a nonparametric estimator constructed directly from the observed data, without distributional assumptions (Li *et al* 2023). For a dataset comprising n observations sorted in ascending order, it is given by equation (5),

$$F(x) = \frac{Number\ of\ observation\ \leqslant\ x}{n} \tag{5}$$

Table 2. Descriptive statistics of the datasets.

		Broad River				Tyger River				Enoree River			
		$P_{(t-2)}$	$T_{(t-0)}$	P_{C10}	Q	$P_{(t-2)}$	$T_{(t-0)}$	P_{C10}	Q	$P_{(t-2)}$	$T_{(t-0)}$	P_{C10}	Q
Training	D_{min}	0.00	−9.87	0.00	3.00	0.00	−8.87	0.00	0.52	0.00	−9.18	0.00	0.80
	D_{mean}	3.60	15.30	35.98	92.84	3.49	15.91	34.88	23.28	3.34	15.83	33.43	13.87
	D_{max}	108.72	31.12	225.88	2002.30	158.87	31.78	288.95	702.35	159.48	31.98	319.79	642.90
	D_{Sd}	7.53	8.23	29.54	116.58	7.95	8.20	30.42	32.27	7.49	8.34	29.70	19.38
	D_{Cs}	4.05	−0.26	1.40	6.86	4.75	−0.25	1.69	6.57	4.84	−0.23	1.90	8.89
	D_{Ck}	25.53	−1.00	2.90	71.72	41.61	−1.02	5.42	70.24	46.05	−1.04	8.08	172.43
Validation	D_{min}	0.00	−8.82	0.00	6.30	0.00	−7.14	0.00	0.59	0.00	−6.93	0.00	1.00
	D_{mean}	3.65	15.64	36.51	85.43	3.51	16.45	35.03	19.19	3.43	16.52	34.27	13.17
	D_{max}	71.33	30.31	218.76	1668.10	76.92	31.54	235.19	337.02	79.24	31.41	235.84	283.20
	D_{Sd}	7.66	8.33	30.50	102.15	7.75	8.22	31.59	27.49	7.50	8.27	32.18	19.07
	D_{Cs}	3.67	−0.31	1.55	5.87	3.93	−0.32	1.89	5.31	3.98	−0.29	2.24	6.18
	D_{Ck}	17.44	−0.98	3.34	54.44	20.74	−0.97	5.31	37.12	22.63	−1.01	7.68	52.99
Testing	D_{min}	0.00	−7.60	0.00	17.20	0.00	−7.08	0.00	2.79	0.00	−7.00	0.00	2.30
	D_{mean}	4.13	15.84	41.36	114.12	4.01	16.66	40.16	24.01	3.97	16.66	39.79	14.93
	D_{max}	77.23	29.08	173.34	2228.80	99.20	29.12	176.69	504.11	92.88	29.43	186.74	373.80
	D_{Sd}	8.74	8.23	31.39	125.95	8.33	8.04	30.41	29.94	8.18	8.11	30.24	19.09
	D_{Cs}	3.65	−0.30	1.03	5.95	3.83	−0.32	1.06	5.41	3.59	−0.30	1.08	6.57
	D_{Ck}	16.93	−1.02	0.85	62.00	21.98	−1.02	1.23	50.96	18.71	−1.05	1.48	77.07
All	D_{min}	0.00	−9.87	0.00	3.00	0.00	−8.87	0.00	0.52	0.00	−9.18	0.00	0.80
	D_{mean}	3.69	15.44	36.94	95.09	3.57	16.12	35.76	22.73	3.46	16.08	34.60	13.93
	D_{max}	108.72	31.12	225.88	2228.80	158.87	31.78	288.95	702.35	159.48	31.98	319.79	642.90
	D_{Sd}	7.76	8.25	30.07	116.29	7.98	8.19	30.67	31.21	7.61	8.30	30.29	19.29
	D_{Cs}	3.92	−0.28	1.36	6.57	4.46	−0.27	1.61	6.28	4.46	−0.25	1.82	8.09
	D_{Ck}	22.45	−1.00	2.55	68.25	34.84	−1.01	4.64	65.08	36.69	−1.04	6.79	138.82

Min: minimum, Max: maximum, Sd: standard deviation, Cs: skewness coefficient, Ck: kurtosis coefficient.

where x is any Q value in the dataset, and $F(x)$ represents the cumulative probability that an observed Q value does not exceed x. The ECDF was estimated using the Weibull plotting position formula (equation (6). To facilitate a scale-independent evaluation of model errors across the full Q distribution, normalized bias (NB) was computed using equation (7) at each 20th percentile interval of the ECDF. Normalizing the bias by the observed value at each percentile renders the error metric dimensionless and proportional, thereby eliminating the confounding influence of flow magnitude on error assessment,

$$F(x_i) = \frac{i}{n+1} \tag{6}$$

$$\mathrm{NB} = \frac{\sum_{i=1}^{N} (td_i - t_i)}{\sum_{i=1}^{N} (t_i)}. \tag{7}$$

To examine in greater detail how discharge prediction performance varies across discharge regimes, in addition to the ECDF analysis presented in the previous section, the observed discharge time series has been divided into five regimes based on the flow duration curve (FDC) segmentation proposed by Pfannerstill *et al* (2014). This approach enables evaluation not only of the models' overall performance but also of their performance under different flow conditions. In the FDC segmentation, the probability of exceedingly very high (VH) flows is 0–10%, high (H) flows correspond to a probability of 10–30%, medium (M) flows to 30–70%, low (L) flows to 70–90%, and very low (VL) flows to 90–100%. Model performance was assessed separately for each Q regime. In the assessment, the normalized RMSE (NRMSE), which penalizes large errors more heavily due to its quadratic nature and expresses the error relative to the mean observed flow within each regime, was used, whilst the normalized MAE (NMAE), which provides a more balanced representation of the typical prediction error relative to the mean observed flow, was also employed

The Shapley additive explanations (SHAP) technique was employed to interpret the relationships between the input parameters and the output parameter (Q) used. It is an ML interpretation technique that assumes independence among inputs (Lundberg and Lee 2017). It compares the effect of a variable's presence or absence on changes in the ML method's output by creating different combinations of input parameters. Although changes in SHAP values and inputs can be plotted for each data point to visualize the relationship between Q and the inputs, this remains limited in its ability to clarify the changes. To interpret this variation more precisely, the locally-weighted-regression scatter-plot smoothing (LOWESS) technique, which involves generating curves, was employed (Cleveland 1979, Moran 1984). In this case, by illustrating the differences within the range with curves makes the interrelationships between features clearer.

3. Results and discussion

For each river, seven distinct models were independently established, trained, validated, and tested. Although the models share the same identifier, the input datasets differ. The NSEC values of the ML methods and models on the test dataset for all rivers are presented as a success point plot in figure 5. The QP1 and QP2 models, which yield NSEC values of approximately 0.1–0.4, have the lowest performance. The fact that the QP2 model performs slightly better than the QP1 model indicates that $T_{(t-0)}$, although less significant, enhances the modeling performance. The QP4 model is the third-worst model following these. When examining the NSEC difference between the QP2 and QP4 models, the P_{C10} contributes significantly more than $P_{(t-2)}$. The QP3 and QP5 models, on the other hand, generally fall within the 0.40–0.65 NSEC range. However, the QP5 model generally achieves NSEC values that are slightly higher than those of the QP3 model. According to the NSEC metric, the best models are QP6 and QP7. Since the QP6 model has fewer inputs than the QP7 model and performs better in some methods, it has been determined to be the best model. Considering all these inputs, P_{C10}, $P_{(t-1)}$, $P_{(t-2)}$, and $P_{(t-3)}$ clearly contribute to modeling performance. $T_{(t-0)}$, while not as significant as the others, can still contribute to improved performance. However, $P_{(t-4)}$ and $P_{(t-5)}$ do not directly contribute to the model. From a methodological perspective, it is evident that the ML methods, except LR, yield NSEC values that are generally similar.

The prediction performances of the methods in QP6, with the best model selected, is given in table 3 for the training, validation, and test datasets. On the Broad River, the RF method achieved

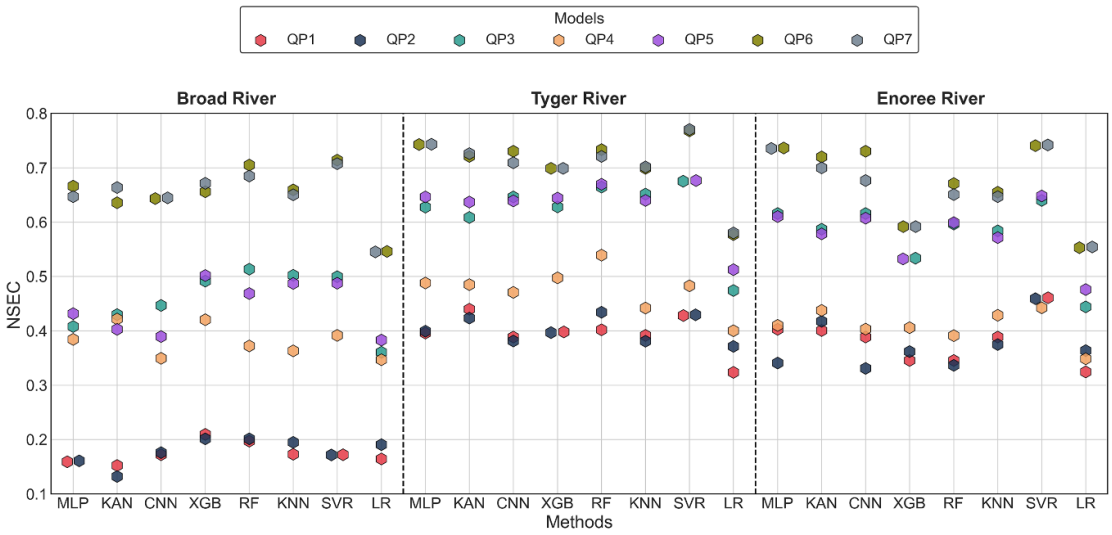


Figure 5. The success point plot of the models and methods, Broad River Basin.

the best performance on the training dataset (RMSE = 47.467 m^3 s^{-1}, MAE = 28.239 m^3 s^{-1}, and NSEC = 0.827), whilst the SVR method performed best on the test dataset (RMSE = 67.353 m^3 s^{-1}, MAE = 39.679 m^3 s^{-1}, and NSEC = 0.714). On the Tyger River, the RF method demonstrated the highest performance on the training dataset (RMSE = 14.582 m^3 s^{-1}, MAE = 8.146 m^3 s^{-1}, and NSEC = 0.785). In comparison, the SVR method demonstrated the highest performance on the testing (RMSE = 14.428 m^3 s^{-1}, MAE = 8.239 m^3 s^{-1}, and NSEC = 0.768). On the Enoree River, the RF method demonstrated the highest performance on the training dataset (RMSE = 8.985 m^3 s^{-1}, MAE = 4.773 m^3 s^{-1}, and NSEC = 0.784), whilst the SVR method performed best on the test dataset (RMSE = 9.716 m^3 s^{-1}, MAE = 5.581 m^3 s^{-1}, and NSEC = 0.741). However, on the test dataset, the MLP method, with an RMSE of 9.795 m^3 s^{-1}, MAE of 4.908 m^3 s^{-1}, and NSEC of 0.737, performed very closely to the SVR method. Upon examining table 3, it is observed that, except for LR, all methods generally achieve NSEC values above or close to 0.70, indicating that these methods achieve good-class performance, as reported by Moriasi *et al* (2007, Moriasi *et al* 2015). The NSEC values below 0.60 obtained from the LR method support the conclusion that there is no linear relationship between the input parameters and the discharge. To reduce the study's complexity, the LR method was not included in subsequent analyses.

The scatter plots generated for the QP6 model, based on observed vs predicted values in the test dataset, from the ML methods for the Broad River Basin, are presented in figure 6. All methods generally exhibit a similar scatter pattern. Whilst these methods produce suitable predictions for data within the ranges of 0–350 m^3 s^{-1} for the Broad River, 0–100 m^3 s^{-1} for the Tyger River, and 0–50 m^3 s^{-1} for the Enoree River, the scatter increases for values above these ranges. In general, they tend to underestimate, particularly when predicting high values. Consequently, the ML methods struggle to perform well at high values, which occur in only a small proportion of the data, whilst providing more accurate predictions within the range where most of the data is concentrated.

Whilst the wide dispersion of high values in a scatter plot allows us to see their distribution clearly, it is quite difficult to interpret the variation at low values, which are clustered within a very narrow range, solely from the scatter plot. The ECDF analysis was performed to evaluate the ability of the QP6 model, especially at low Q range values, to reproduce the statistical distribution of observed Q across all three gaging stations and ML methods (figure 7). To gain a clearer picture of the variation at low values, using the logarithm of the x-axis in the ECDF analyses makes the distribution of these low values more clearly observed. The ECDF curves exhibit strong agreement throughout the mid-to-high Q range, confirming that QP6 model accurately captures the dominant flow dynamics across the Broad River Basin. The NB values of Q in this range vary by approximately ±25%, that supports this finding. However, a systematic divergence between the two curves is evident in the low-Q region, where the predicted ECDF consistently shifts to the right of the observed curve, and the NB values range from −16.4–116.6%. This indicates that the model tends to overestimate Q for lower values. Despite this limitation in the lower tail of the distribution, the close correspondence between observed and predicted ECDFs across most of the Q range further supports the robustness of the QP6 model for daily Q prediction in

Table 3. The prediction performance metrics for the QP6, Broad River Basin.

River	Metric	Dataset	MLP	KAN	CNN	XGB	RF	KNN	SVR	LR
Broad	RMSE ($m^3\ s^{-1}$)	Training	64.009	68.425	65.265	61.359	**47.467**	58.082	58.286	77.641
		Validation	57.107	60.710	59.268					
		Testing	72.764	75.985	75.195	73.863	68.376	73.468	**67.353**	84.843
	MAE ($m^3\ s^{-1}$)	Training	32.312	37.976	32.698	34.536	**28.239**	29.410	31.279	44.138
		Validation	31.661	35.555	30.379					
		Testing	42.092	45.093	43.442	42.205	40.536	41.737	**39.679**	50.356
	NSEC	Training	0.699	0.656	0.687	0.710	**0.827**	0.740	0.739	0.536
		Validation	0.687	0.647	0.663					
		Testing	0.666	0.636	0.643	0.656	0.705	0.660	**0.714**	0.546
Tyger	RMSE ($m^3\ s^{-1}$)	Training	16.092	17.417	16.317	17.127	**14.582**	16.180	15.065	20.458
		Validation	13.591	14.582	13.534					
		Testing	15.182	15.817	15.538	16.424	15.461	16.427	**14.428**	19.461
	MAE ($m^3\ s^{-1}$)	Training	8.011	8.833	8.644	9.038	8.146	**7.443**	8.098	11.773
		Validation	7.993	8.736	7.330					
		Testing	8.130	9.305	8.769	9.435	8.914	8.385	**8.239**	12.267
	NSEC	Training	0.752	0.709	0.745	0.704	**0.785**	0.735	0.771	0.577
		Validation	0.755	0.719	0.758					
		Testing	0.743	0.721	0.731	0.699	0.733	0.699	**0.768**	0.577
Enoree	RMSE ($m^3\ s^{-1}$)	Training	9.907	10.199	9.949	11.769	**8.985**	10.296	9.914	12.625
		Validation	9.556	9.250	9.718					
		Testing	9.795	10.096	9.910	12.196	10.949	11.210	**9.716**	12.762
	MAE ($m^3\ s^{-1}$)	Training	4.692	5.136	4.635	5.854	4.773	**4.348**	5.380	7.040
		Validation	4.827	5.150	4.877					
		Testing	4.908	5.526	4.920	6.453	5.566	**5.063**	5.581	7.519
	NSEC	Training	0.739	0.723	0.737	0.629	**0.784**	0.716	0.737	0.573
		Validation	0.749	0.765	0.740					
		Testing	0.737	0.720	0.730	0.592	0.671	0.655	**0.741**	0.553

Bold: the best values.

the basin. Furthermore, whilst the MLP and KNN methods exhibit a more suitable distribution at lower values, the MLP and SVR methods demonstrate a more suitable distribution at higher values.

A comparison of the models' prediction performance across different *Q* regimes, in terms of NRMSE and NMAE, is presented in figure 8. An examination of figure 8 reveals that, across all models, both NRMSE and NMAE vary significantly with the flow regime. Although the highest error values were observed at the VL and L flow rates in all basins, these values remained nearly constant at other flow rates. This indicates that low flows are more difficult to predict for these rivers. Overall, while all methods exhibit similar error values in the VH, *H*, and *M* regimes, there are significant differences between the methods in the VL and L regimes. Although the MLP method has lower errors than all other methods on the Enoree River, the CNN method has the lowest error on nearly all rivers especially lower values. These findings indicate that the proposed models can produce more stable predictions under high- and medium-flow conditions; however, low-flow events continue to pose a fundamental challenge for all models except CNN.

Finally, to quantitatively assess uncertainty in the performance of the prediction models and the sensitivity of the metrics to sampling variability, the bootstrap resampling method was applied exclusively to the test data set (Efron 1979, Efron and Tibshirani 1993, Hastie *et al* 2009). As the test data in this study covered only six years, the monthly block bootstrap approach was preferred over the yearly block bootstrap to preserve the temporal dependence in the hydrometeorological time series. In this study, a yearly block bootstrap would provide only six independent blocks, which may be insufficient for reliably quantifying sampling uncertainty. In the monthly block bootstrap, the observations for each month were treated as a block, resulting in approximately 72 blocks over the six-year test period, and the monthly blocks were resampled with replacement. Whilst a single month could be selected more than once in the same bootstrap sample, some months might not be selected at all. The daily observations for the selected

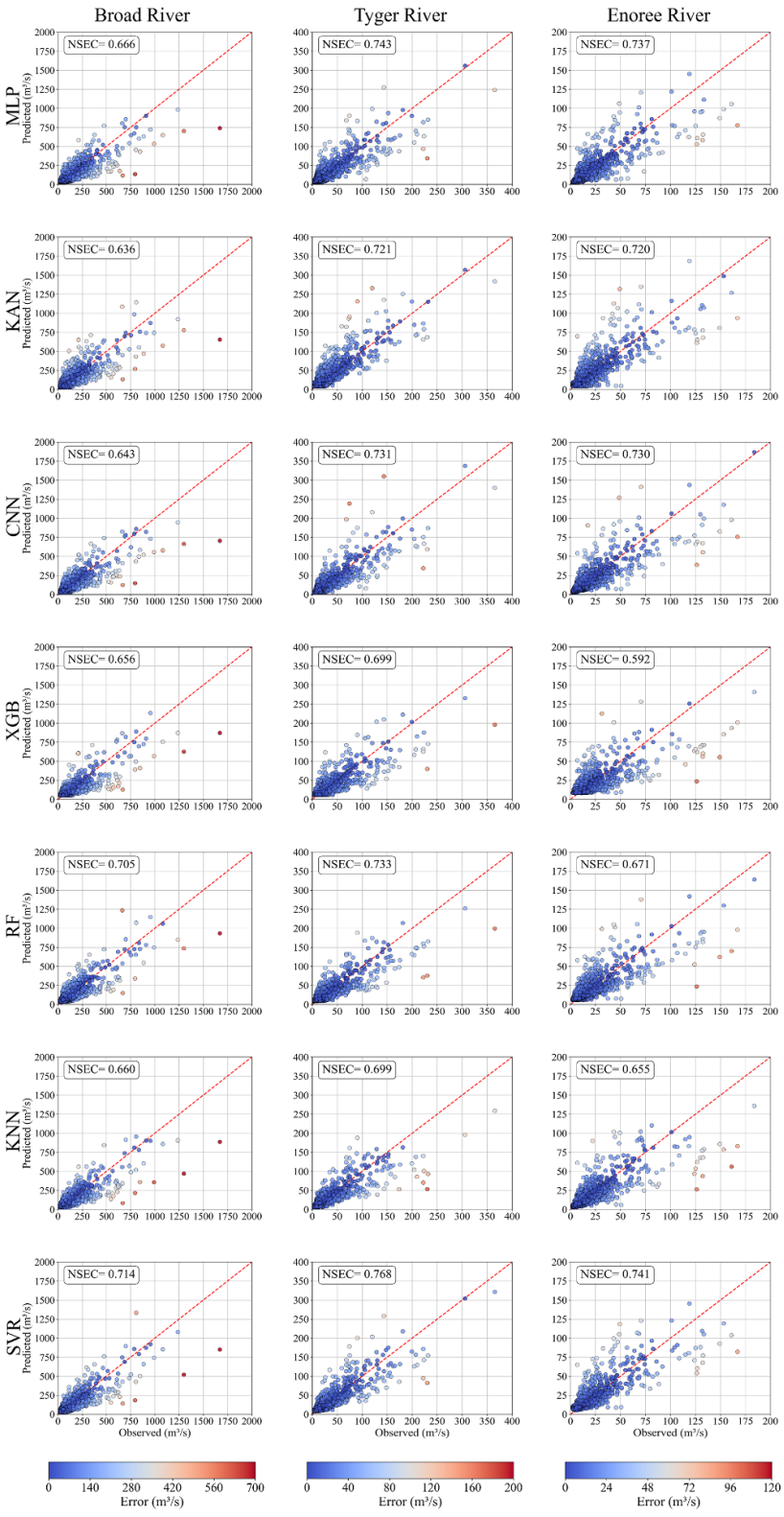


Figure 6. Scatter plots of observed vs predicted discharge values for the QP6 model.

months were combined whilst retaining their original chronological order, thereby creating a new bootstrap data set of the same length as the original data set in each iteration. This process was repeated 1000 times; for each bootstrap dataset, the RMSE, MAE, and NSEC values were calculated, yielding the mean values of the metrics along with their 95% confidence intervals (2.5th–97.5th percentiles).

The bootstrap results obtained are presented in figure 9. In the figure, the black dots represent the average performance derived from the bootstrap samples for each model, whilst the horizontal lines indicate the 95% confidence interval. This allows for a comparison not only of average performance but also

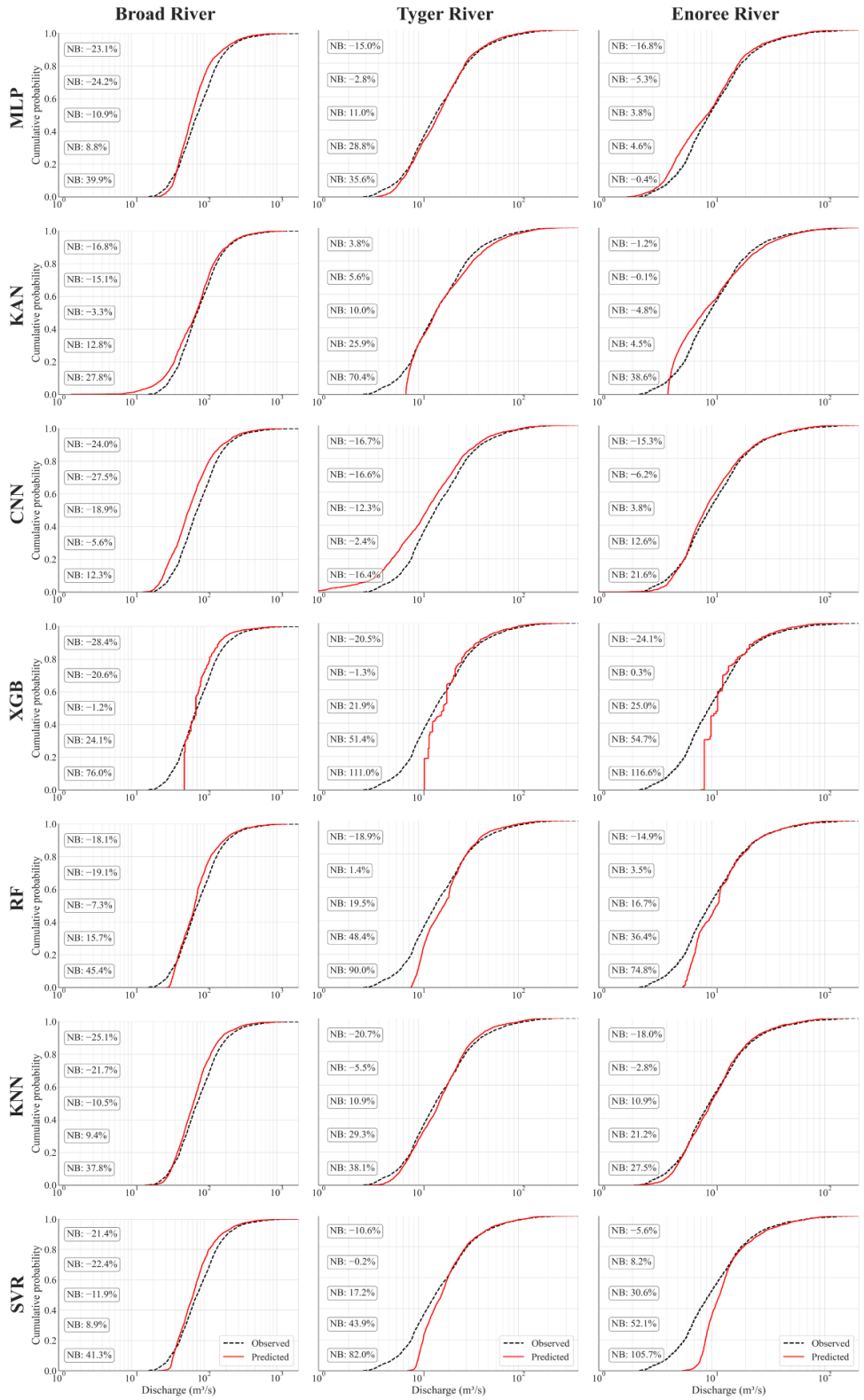


Figure 7. Empirical cumulative distribution functions of observed vs predicted discharge values for the QP6 model under different normalized bias conditions, the Broad River Basin.

of the models' performance stability and level of uncertainty. An examination of figure 9 reveals that across all three catchments, the SVR model produced the lowest mean error values (RMSE and MAE) whilst achieving the highest mean performance (NSEC). However, it is noteworthy that the confidence

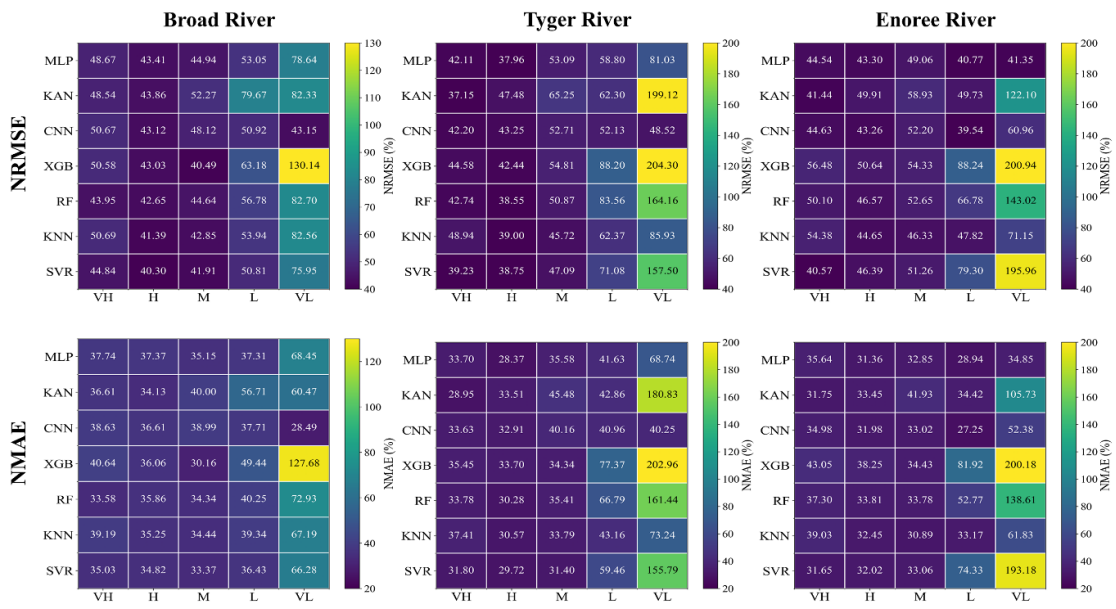


Figure 8. The heatmap of NRMSE and NMAE values for the prediction models across different flow regimes based on FDC segmentation for the QP6 model, Broad River Basin.

intervals of the MLP, KAN, CNN, and RF models largely overlap, and their performance is statistically very similar. In contrast, the XGB model exhibited higher error values and relatively wider confidence intervals, particularly in the Enoree catchments, demonstrating lower and more variable performance compared to the other methods. In general, the narrow confidence intervals indicate that the performance metrics did not vary significantly across bootstrap resamples and that the models produced robust results across different sample selections.

When table 3 and figures 6–9 are evaluated collectively, SVR consistently provides the most accurate and robust discharge predictions for all three rivers. According to table 3, SVR stands out across all evaluated performance metrics and rivers, generally achieving the highest NSEC values together with the lowest RMSE and MAE. However, figures 7 and 8 indicate that MLP and CNN also demonstrate highly competitive performance across different flow regimes, with their predictions remaining close to those of SVR. In particular, MLP and CNN emerge as the most competitive alternatives to SVR in the regime-based evaluation. Furthermore, figure 9, which presents the uncertainty analysis, shows that SVR generally exhibits a narrower range of error variation compared with the other methods, further supporting its overall robustness and consistency. Although some variations are observed among RF, KNN, and KAN, their overall predictive performances remain reasonably competitive, while MLP and CNN can be considered particularly comparable to SVR in terms of overall predictive capability. These results indicate that SVR provides the most consistent overall performance, while MLP and CNN also offer strong predictive capabilities across a wide range of hydrological conditions. In contrast, XGB consistently ranked below this group, whereas LR produced the weakest results across all rivers. The poor performance of LR is not unexpected considering the statistical characteristics of the datasets presented in table 2. The Q series exhibits strong positive Cs (6.28–8.09) and extremely high Ck (65.08–138.82), indicating that the discharge data are highly non-normal and characterized by substantial asymmetry and heavy tails. Such characteristics imply that the relationship between the predictor variables (P, T, and P_{C10}) and Q is highly nonlinear. Consequently, LR assumptions are violated, leading to systematic underestimation or overestimation of discharge, particularly during extreme discharge events.

Although XGB improves upon LR, its performance remains consistently below that of the other nonlinear ML methods. A possible explanation lies in the sequential boosting mechanism of XGB, which iteratively minimizes residual errors (Chen and Guestrin 2016). For highly skewed discharge datasets, this learning strategy may not represent the entire discharge distribution equally well, leading to uneven predictive performance across Q ranges. This interpretation is supported by the larger discrepancies between the observed and predicted cumulative distributions (figure 7) and the comparatively higher errors for the VH discharge class (figure 8), ultimately leading to lower overall predictive accuracy. In contrast, the superior performance of SVR can be attributed to its ability to model complex nonlinear relationships

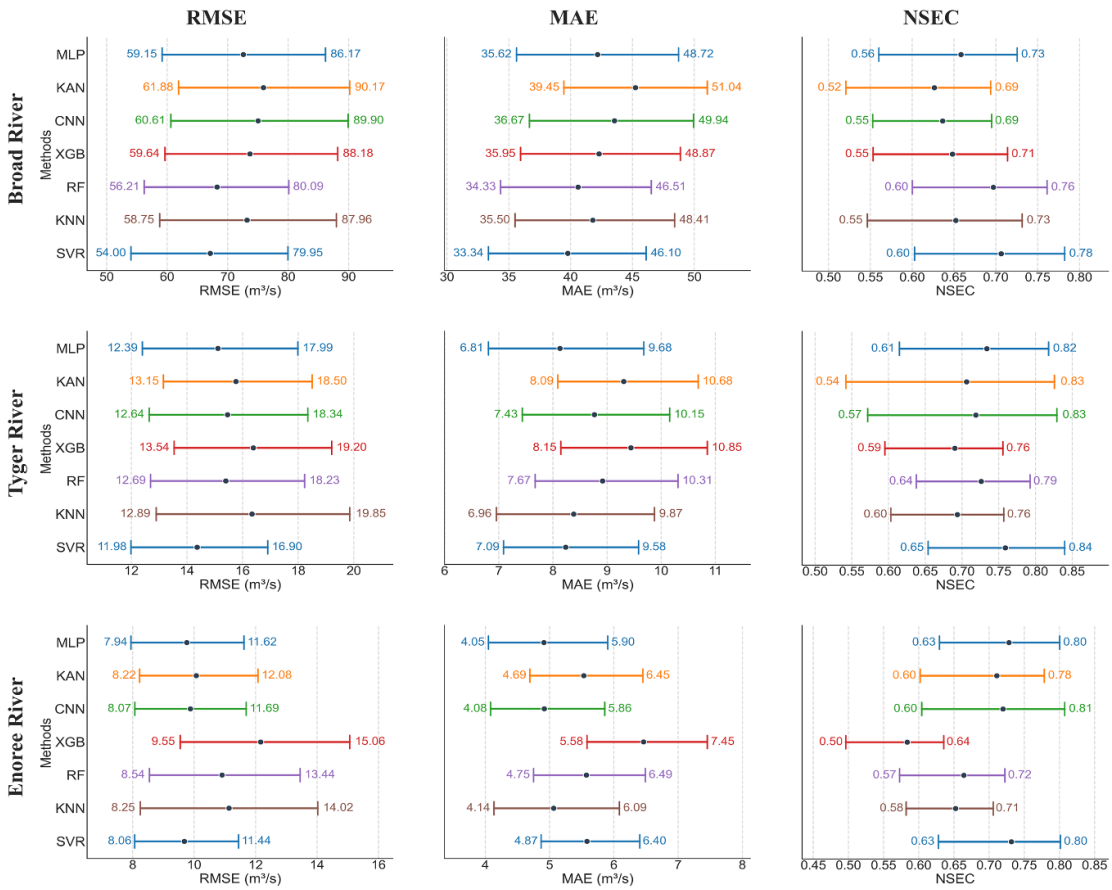


Figure 9. Forest plots of bootstrap-based 95% confidence intervals for RMSE, MAE, and NSEC of the evaluated prediction models for the QP6 model, the Broad River Basin.

via kernel functions while controlling model complexity through the structural risk minimization principle (Cortes and Vapnik 1995). Unlike boosting-based methods, SVR constructs a single optimal regression function by solving a global optimization problem using the structural risk minimization principle, rather than sequentially fitting residuals (Hu and Zheng 2015). This learning strategy generally provides strong generalization capability for complex nonlinear relationships. The scatter plots in figure 6 show that SVR predictions are more tightly clustered around the 1:1 line. In contrast, figure 7 demonstrates that the predicted cumulative discharge distributions closely reproduce the observed distributions across nearly all flow classes. These characteristics explain why SVR consistently achieves the highest NSEC values and the lowest RMSE and MAE values for all rivers.

The SHAP values were calculated for each river based on the trained ML methods. Taking the obtained SHAP values into account, the relative importance (RI) values representing the contributions of the input parameters as percentages are presented in figure 10 as a stacked bar chart. For the Broad River, the RI values of the input parameters are ranked from highest to lowest as P_{C10}, α, $P_{(t-1)}$, $P_{(t-2)}$, $T_{(t-0)}$, and $P_{(t-3)}$. Similarly, for the Tyger River, P_{C10} and α show the highest contributions, whilst the RI values of $P_{(t-1)}$ and $P_{(t-2)}$ are quite close to one another, followed by $T_{(t-0)}$ and $P_{(t-3)}$. Generally, the highest RI was shown by α and P_{C10}, whilst the lowest RI was shown by $P_{(t-3)}$. On the other hand, the fact that all methods exhibit similar RI values for the same inputs indicates that the ML methods have been appropriately trained. Rahimzad *et al* (2021) found that same-day P had no meaningful, and even slightly negative, effect on same-day Q prediction in a study conducted in the Kentucky River Basin, Kentucky, USA. The best-performing models in that basin excluded precipitation entirely and relied solely on Q at lags of one and two days. Khosravi *et al* (2022) reported that P was the most influential input for daily Q prediction in the Korkorsar Basin, Iran, because it reflected the basin's longer concentration time, including lagged P, substantially improving model performance. Kaur and Chavan (2025) identified lagged P and Q as the dominant predictors, while T was excluded as it did not improve predictive performance in the Krishna River Basin, India. Sanikhani *et al* (2025) found via SHAP analysis that current monthly P and one-month-lagged Q, which jointly accounted for 66%–83% of the total

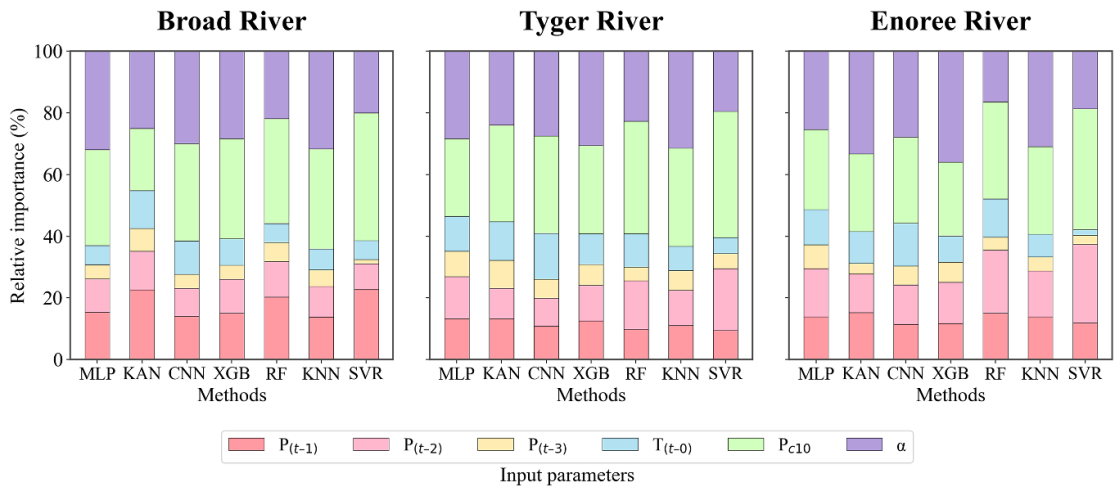


Figure 10. Stacked bar chart of relative importance based on the SHAP values for machine learning methods in the QP6 model, Broad River Basin.

predictive power, were the dominant predictors in the Boukan Dam Basin, Iran. These studies indicate that the RI of input variables in Q prediction is closely tied to basin-specific characteristics. This is consistent with the review by Mohammadi (2021), which compiled a wide range of comparative ML applications across diverse basins and reported that no single model or input configuration consistently outperforms others across varying hydrological conditions and basin characteristics. Notably, all these studies incorporated antecedent Q as an input, which typically enhances predictive performance owing to the strong autocorrelation inherent in Q data (Yaseen *et al* 2026). In contrast, the present study deliberately excludes antecedent Q values from all input combinations, relying exclusively on meteorological and temporal variables (P, T, P_C, and α). This choice reflects the study's objective of establishing a framework applicable to future discharge projection, for which observed future Q is inherently unavailable. The consistently strong performance of the QP6 across stations indicates that the meteorological input set alone provides substantial predictive skill for daily Q prediction.

The distribution of SHAP values for the SVR method in the QP6 model, as a function of the input parameters, along with the LOWESS curves, is presented in figure 11. In the Broad River, the P_{C10} value contributed the most at 32.4%, whilst the α contributed the second highest at 31.7%. In the Tyger River, the features that contributed most were P_{C10} and α, with RI values of 31.9 and 31.4%, respectively. In the Enoree River, however, unlike the other rivers, the features with the highest contributions were α and P_{C10}, with RI values of 28.4 and 30.9%, respectively. In all rivers, an increase in the Q was observed for the P_{C10} values greater than approximately 50 mm, and this increase was gradual. For the values below 50 mm, no significant change in the Q value was observed. When examining the α, an increase in the Q value was generally observed across all rivers during January–March, whilst a decrease was observed during June–November. Also, $P_{(t-1)}$ and $P_{(t-2)}$ were among the most significant input parameters. In the Broad River, $P_{(t-1)}$ shown a higher contribution compared to $P_{(t-2)}$, whereas the opposite is observed in the Tyger and Enoree rivers. $P_{(t-3)}$ is the feature showing the lowest contribution in all rivers. When considering these, all generally increases in Q after 5–10 mm, with no significant effect before that. $T_{(t-0)}$ is the second lowest with a higher RI than $P_{(t-3)}$. Generally, it increases the Q value below 15 °C–20 °C, whilst decreasing it at higher values. In addition, the SHAP contributions of precipitation variables (P_{C10}, $P_{(t-1)}$, $P_{(t-2)}$, and $P_{(t-3)}$) were evaluated collectively to determine their overall contribution to the model. Accordingly, the total contribution of precipitation variables was calculated as 61.5%, 60.9%, and 61.8% for the Broad, Tyger, and Enoree river basins, respectively. These results clearly show that precipitation variables are the dominant group of variables driving model predictions across all basins, accounting for approximately 61%.

Achieving high efficiency is often more challenging at a daily time step due to the stochastic nature of Q. However, the performance values for QP6 exceed the recommended satisfaction thresholds in hydrological simulations. The high predictive accuracy across all three rivers demonstrates that the modeling approach is effective for regional water resources management.

The superiority of the QP6 model confirms that Q is not only a response to sudden precipitations but is also influenced by basin's previous conditions. The model accurately captures the nonlinear temporal dynamics of the P–Q relationship. This aligns with the findings of Ma *et al* (2026), Dykstra and

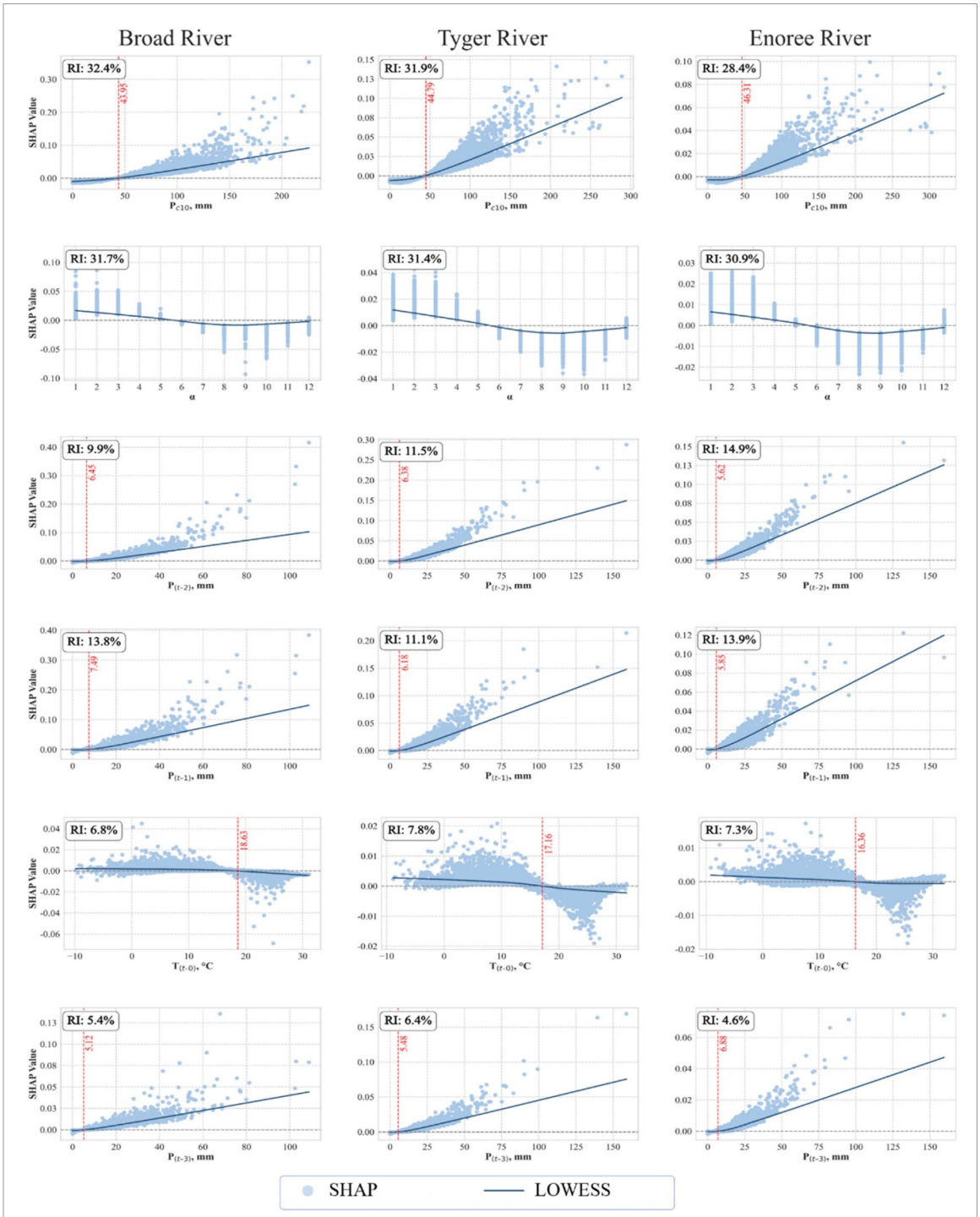


Figure 11. The SHAP values with LOWESS curves and relative importance of input variables for the SVR method in the QP6 model.

Dzwonkowski (2021), Kundu *et al* (2017), Rushworth *et al* (2013), and Zadeh *et al* (2010). All these studies emphasized the importance of lagged variables in enabling data-driven models to mimic a basin's memory.

Within the scope of this study, predicting daily Q with high success can be achieved using ML methods; however, the study still has certain limitations. The most significant one is the failure to utilize features such as soil moisture, evaporation, and snowmelt, which are generally used in Q prediction alongside P and T. Secondly, decomposition methods, which have been shown to improve prediction performance, have not been applied alongside ML methods. Thirdly, hybrid ML methods, which could enhance predictive performance of Q beyond classical ML methods, were not utilized in this study. Finally, recurrent neural network based methods, such as LSTM and GRU, are widely used in time series analysis, they could not be included in this study. This is because their sequence-based formulation is

not fully consistent with the predefined predictor structure adopted here. Precipitation was incorporated using selected lagged values, whereas P_{C10} represents cumulative precipitation and already contains antecedent precipitation information. Introducing these variables into a recurrent sequence would therefore result in overlapping representations of the same antecedent precipitation information and would require a different formulation of the predictor structure used in this study. Moreover, a consistent and fair comparison would require the same temporal input structure to be considered across the other ML models, which would alter the original modeling framework.

4. Conclusions

This study conducted a systematic comparison of eight ML methods for daily river discharge prediction using different combinations of precipitation, air temperature, cumulative precipitation, and temporal information across three gaging stations in the Broad River Basin. The results demonstrate that the model, which incorporated multiple lagged P inputs along with T, P_C, and α, provides the most effective representation of the precipitation–streamflow relationship across the three stations. Among the evaluated models, SVR showed the most consistent overall performance, while MLP and CNN also demonstrated highly competitive performance, particularly across low-flow regimes. The results further highlight the importance of cumulative and lagged precipitation information in capturing the temporal memory and nonlinear dynamics of the basin response. The SHAP analysis indicates that P_{C10} and α are the most influential predictors, while temperature and lagged precipitation also contribute substantially to daily Q prediction. Overall, the findings demonstrate the potential of appropriately configured ML models for reliable daily streamflow prediction across different hydrological conditions.

This study demonstrates that ML methods (except LR) provide reliable performance for daily Q prediction under the investigated case studies and can serve as useful tools for regional water resources assessment. Given the strategic importance of the Broad River Basin for hydroelectric and nuclear energy generation, the promising performance of the evaluated ML models highlights their potential applicability to streamflow prediction and hydrological planning. The ability to predict Q with low error margins can contribute to more informed reservoir operation and support energy generation planning under varying hydro-meteorological conditions.

Acknowledgments

For the tirelessly helping and supporting, the authors thank Professor Dr Murat Kankal, who is a top researcher, a good person, and a beloved friend, from Bursa Uludağ University, Türkiye.

Data availability statement

Daily discharge data used in this study are publicly available from the U.S. Geological Survey National Water Information System at https://waterdata.usgs.gov/nwis/uv?site_no=02156500&legacy=1, https://waterdata.usgs.gov/nwis/uv?site_no=02160105&legacy=1, and https://waterdata.usgs.gov/nwis/uv?site_no=02160700&legacy=1 in 2024. Meteorological data are publicly available from the National Weather Service at https://www.weather.gov/wrh/Climate?wfo=cae and. https://www.weather.gov/wrh/Climate?wfo=gsp in 2025.

Supplementary Data available at https://doi.org/10.1088/2515-7620/aeb18b/data1.

Availability of data and materials

The data supporting the findings of this study are partially or fully available from the corresponding author upon request.

Conflict of interest

No conflicts of interest, financial or personal, are declared by the authors in relation to the research described in this paper.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work, the authors used DeepL Translate and Claude AI to improve language and readability. After using this tool/service, the authors reviewed and edited the content as needed and took full responsibility for the published article.

Author contributions

Betul Mete 0000-0002-3689-6430
Conceptualization (equal), Data curation (equal), Formal analysis (equal), Investigation (equal), Methodology (equal), Resources (equal), Software (equal), Validation (equal), Visualization (equal), Writing – original draft (equal)

Emirhan Mustafa Anık 0000-0001-9342-2242
Formal analysis (equal), Investigation (equal), Methodology (equal), Resources (equal), Software (equal), Validation (equal), Visualization (equal), Writing – original draft (equal)

Sinan Nacar 0000-0003-2497-5032
Conceptualization (equal), Data curation (equal), Formal analysis (equal), Investigation (equal), Methodology (equal), Resources (equal), Supervision (equal), Validation (equal), Writing – original draft (equal), Writing – review & editing (equal)

Adem Bayram 0000-0003-4359-9183
Conceptualization (equal), Data curation (equal), Formal analysis (equal), Investigation (equal), Methodology (equal), Project administration (equal), Resources (equal), Supervision (equal), Validation (equal), Writing – original draft (equal), Writing – review & editing (equal)

References

Alipour M H 2025 Streamflow prediction in ungauged basins located within data-scarce areas using XGBoost: role of feature engineering and explainability *Int. J. River Basin Manag.* **23** 71–92

Anik E M, Togrul B, Akbas A and Kankal M 2026 Investigating the contribution of decomposition techniques to machine learning accuracy in SPEI-based drought forecasting for multiple Köppen-Geiger climates *Acta Geophys.* **74** 39

Arico C, Corato G, Tucciarelli T, Meftah M B, Petrillo A F and Mossa M 2010 Discharge estimation in open channels by means of water level hydrograph analysis *J. Hydraul. Res.* **48** 612–9

Asadollahi A, Magar B A, Poudel B, Sohrabifar A and Kalra A 2024 Application of machine learning models for improving discharge prediction in ungauged watershed: a case study in East DuPage, Illinois *Geographies* **4** 363–77

Bandara A, Hettiarachchi Y, Hettiarachchi K, Munasinghe S, Wijesinghe I and Thayasivam U 2020 *A generalized ensemble machine learning approach for landslide susceptibility modeling Data Management, Paper Presented at Analytics and Innovation: Proc. of ICDMAI* vol 2 pp 71–93

Bargam B, Boudhar A, Kinnard C, Bouamri H, Nifa K and Chehbouni A 2024 Evaluation of the support vector regression (SVR) and the random forest (RF) models accuracy for streamflow prediction under a data-scarce basin in Morocco *Discov. Appl. Sci.* **6** 306

Bhattarai S B and Shrestha R M 2023 Analysis of implications of a regulation on design discharge of run-of-river hydropower projects in Nepal *Sustain. Water Resour. Manag.* **9** 80

Borovykh A, Bohte S and Oosterlee C W 2017 Conditional time series forecasting with convolutional neural networks (arXiv:1703.04691)

Burnham K P and Anderson D R 2002 *Model Selection and Multimodel Inference: A Practical Information-theoretic Approach* (Springer)

Chen T and Guestrin C 2016 *Xgboost: a scalable tree boosting system Proc. 22nd Acm Sigkdd Int. Conf. on Knowledge Discovery and Data Mining* pp 785–94

Choo T H, Chae S K, Yoon H C and Choo Y M 2014 Discharge prediction using hydraulic characteristics of mean velocity equation *Environ. Earth Sci.* **71** 675–83

Cleveland W S 1979 Robust locally weighted regression and smoothing scatterplots *J. Am. Stat. Assoc.* **74** 829–36

Cortes C and Vapnik V 1995 Support-vector networks *Mach. Learn.* **20** 273–97

Dehghani A, Moazam H M Z H, Mortazavizadeh F, Ranjbar V, Mirzaei M, Mortezavi S and Dehghani A 2023 Comparative evaluation of LSTM, CNN, and ConvLSTM for hourly short-term streamflow forecasting using deep learning approaches *Ecol. Inform.* **75** 102119

Di Baldassarre G and Montanari A 2009 Uncertainty in river discharge observations: a quantitative analysis *Hydrol. Earth Syst. Sci.* **13** 913–21

Dykstra S L and Dzwonkowski B 2021 The role of intensifying precipitation on coastal river flooding and compound river-storm surge events, Northeast Gulf of Mexico *Water Resour. Res.* **57** e2020WR029363

Efron B 1979 Bootstrap methods: another look at the jackknife *Ann. Stat.* **7** 1–26

Efron B and Tibshirani R J 1993 *An Introduction to the Bootstrap* 1st edn (Chapman and Hall/CRC) (https://doi.org/10.1201/9780429246593)

Enung E, Kasyanto H, Sari R R and Lubis M F 2023 Application of multilayer perceptron (MLP) method for streamflow forecasting (case study: Upper Citarum River, Indonesia) *IOP Conf. Ser. Earth Environ. Sci.* **1203** 012032
Forkuor G, Hounkpatin O K, Welp G and Thiel M 2017 High resolution mapping of soil properties using remote sensing variables in south-western Burkina Faso: a comparison of machine learning and multiple linear regression models *PLoS One* **12** e0170478
Gardner M W and Dorling S R 1998 Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences *Atmos. Environ.* **32** 2627–36
Ghimire S, Yaseen Z M, Farooque A A, Deo R C, Zhang J and Tao X 2021 Streamflow prediction using an integrated methodology based on convolutional neural network and long short-term memory networks *Sci. Rep.* **11** 17497
Hastie T, Tibshirani R and Friedman J 2009 *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* 2nd edn (Springer) (https://doi.org/10.1007/978-0-387-84858-7)
Heberger M 2022a *Watershed Data Report* (available at: https://mghydro.com/app/report?lat=34.008&lng=−81.059&precision=high&simplify=true) (Accessed 1 August 2026)
Heberger M 2022b Global watershed (available at: https://mghydro.com/watersheds/) (Accessed 1 August 2026)
Hu J and Zheng K 2015 A novel support vector regression for data set with outliers *Appl. Soft Comput.* **31** 405–11
Jalbert J, Mathevet T and Favre A-C 2011 Temporal uncertainty estimation of discharges from rating curves using a variographic analysis *J. Hydrol.* **397** 83–92
Jimenez J I, Tripathi N K, Pandey A, Chao K C and Shrestha S 2025 Assessment of the spatiotemporal dynamics of river discharge and groundwater recharge to support sustainable water management for irrigated agriculture in the Padsan river watershed *Earth Sci. Inform.* **18** 203
Kaur S and Chavan S R 2025 Comparative analysis of deep learning and machine learning models for one-day-ahead streamflow forecasting in the Krishna River basin *J. Hydrol. Reg. Stud.* **60** 102549
Khan N, Shahid S, Juneng L, Ahmed K, Ismail T and Nawaz N 2019 Prediction of heat waves in Pakistan using quantile regression forests *Atmos. Res.* **221** 1–11
Khosravi K, Golkarian A and Tiefenbacher J P 2022 Using optimized deep learning to predict daily streamflow: a comparison to common machine learning algorithms *Water Resour. Manag.* **36** 699–716
Kisi O, Azamathulla H M, Cevat F, Kulls C, Kuhdaragh M and Fuladipanah M 2024 Enhancing river flow predictions: comparative analysis of machine learning approaches in modeling stage-discharge relationship *Results Eng.* **22** 102017
Kundu D, Vervoort R W and van Ogtrop F F 2017 Inferring catchment flow path responses using a data-driven model: an exploratory study based on a generalized additive model *Hydrol. Sci. J.* **62** 1965–79
Li B, Lu Y, Zhang S, Chi Y, Zhou H, Liu M and Guo Y 2025 Adapting newly constructed well depth to groundwater level changes *Water* **17** 2066
Li Z, Zhao Y, Hu X, Botta N, Ionescu C and Chen G H 2023 ECOD: unsupervised outlier detection using empirical cumulative distribution functions *IEEE Trans. Knowl. Data Eng.* **35** 12181–93
Liu C, Roy T, Tartakovsky D M and Dwivedi D 2025 Baseflow identification via explainable AI with Kolmogorov-Arnold networks *J. Geophys. Res. Mach. Learn. Comput.* **2** e2025JH000749
Liu Z *et al* 2025 *KAN: Kolmogorov-Arnold networks Paper Presented at Int. Conf. on Learning Representations* vol 2025 pp 70367–413
Lundberg S M and Lee S I 2017 A unified approach to interpreting model predictions *Proc. Paper Presented at Advances in Neural Information Processing Systems* vol 30 pp 4765–74
Ma C, Xu K, Li Y, Hao Y, Sun H, Gao S, Fan X and Wang X 2026 Heterogeneous spatiotemporal graph attention network for karst spring discharge prediction: advancing sustainable groundwater management under climate change *Sustainability* **18** 933
Marill K A 2004 Advanced statistics: linear regression, part II: multiple linear regression *Acad. Emerg. Med.* **11** 94–102
Mete B, Anik E M, Nacar S, Bayram A and Kankal M 2026 Predicting dissolved oxygen concentration using different neural network models *Ain Shams Eng. J.* **17** 103896
Mohammadi B 2021 A review on the applications of machine learning for runoff modeling *Sustain. Water Resour. Manag.* **7** 98
Moran G W 1984 Locally-weighted-regression scatter-plot smoothing (LOWESS): a graphical exploratory data analysis technique *PhD Thesis* (Naval Postgraduate School)
Moriasi D N, Arnold J G, Van Liew M W, Bingner R L, Harmel R D and Veith T L 2007 Model evaluation guidelines for systematic quantification of accuracy in 93 watershed simulations *Trans. ASABE* **50** 885–900
Moriasi D N, Gitau M W, Pai N and Daggupati P 2015 Hydrologic and water quality models: performance measures and evaluation criteria *Trans. ASABE* **58** 1763–85
Moridnejad A, Abdollahi H, Alavipanah S K, Samani J M V, Moridnejad O and Karimi N 2015 Applying artificial neural networks to estimate suspended sediment concentrations along the southern coast of the Caspian Sea using MODIS images *Arab. J. Geosci.* **8** 891–901
Muhebwa A, Gleason C J, Feng D and Taneja J 2024 Improving discharge predictions in ungauged basins: harnessing the power of disaggregated data modeling and machine learning *Water Resour. Res.* **60** e2024WR037122
Nacar S, Hınıs M A and Kankal M 2018 Forecasting daily streamflow discharges using various neural network models and training algorithms *KSCE J. Civ. Eng.* **22** 3676–85
Nacar S, Mete B and Bayram A 2019 *Estimation of daily dissolved oxygen concentration in the Broad River, South Carolina Paper Presented at Int. Civil Engineering and Architecture Conf. Trabzon, Türkiye* pp 879–85
Naghibi S A, Hashemi H, Berndtsson R and Lee S 2020 Application of extreme gradient boosting and parallel random forest algorithms for assessing groundwater spring potential using DEM-derived factors *J. Hydrol.* **589** 125197
Pfannerstill M, Guse B and Fohrer N 2014 Smart low flow signature metrics for an improved overall performance evaluation of hydrological models *J. Hydrol.* **510** 447–58
Rahimzad M, Moghaddam Nia A, Zolfonoon H, Soltani J, Danandeh Mehr A and Kwon H-H 2021 Performance comparison of an LSTM-based deep learning model versus conventional machine learning algorithms for streamflow forecasting *Water Resour. Manag.* **35** 4167–87
Rather S A, Patel M and Kapoor K 2025 AI-driven forecasting of river discharge: the case study of the Himalayan mountainous river *Earth Sci. Inform.* **18** 242
Rehamnia I, Anik E M, Nacar S and Kankal M 2026 Comparative evaluation of tree and neural network-based models for reservoir evaporation estimation *Pure Appl. Geophys.* **183** 1367–94
Rushworth A M, Bowman A W, Brewer M J and Langan S J 2013 Distributed lag models for hydrological data *Biometrics* **69** 537–44
Saleh M A, Rasel H M and Ray B 2024 From data to decisions: leveraging ML for improved river discharge forecasting in Bangladesh *Watershed Ecol. Environ.* **6** 209–26

San M, Nacar S, Bayram A and Kankal M 2026 Analysis of the changes in river water quality and quantity with innovative trend methods *Environ. Model. Assess.* **6** 209–26
Sanikhani H, Nikpour M R and Jamshidi F 2025 Advanced framework for predicting rainfall-runoff: comparative evaluation of AI models for enhanced forecasting accuracy *Water Resour. Manag.* **39** 3273–94
Shen K, Qin H, Zhou J and Liu G 2022 Runoff probability prediction model based on natural gradient boosting with tree-structured parzen estimator optimization *Water* **14** 545
Shibata K and Ito M 2014 Relationships of bankfull channel width and discharge parameters for modern fluvial systems in the Japanese Islands *Geomorphology* **214** 97–113
Shu X, Ding W, Peng Y, Wang Z, Wu J and Li M 2021 Monthly streamflow forecasting using convolutional neural network *Water Resour. Manag.* **35** 5089–104
Speir S L, Tamayo I G and Strauss A G 2025 Twenty-four year record reveals increasing discharge in tributaries to a key drinking water resource *Hydrol. Process.* **39** e70304
Sun W and Trevor B 2017 Combining k-nearest-neighbor models for annual peak breakup flow forecasting *Cold Reg. Sci. Technol.* **143** 59–69
Szczepanek R 2022 Daily streamflow forecasting in mountainous catchment using XGBoost, LightGBM and CatBoost *Hydrology* **9** 226
Talukdar S, Pal S, Naikoo M W, Parvez A and Rahman A 2023 Trend analysis and forecasting of streamflow using random forest in the Punarbhaba River basin *Environ. Monit. Assess.* **195** 153
Venkataramana V, Sarma V V S S, Gawade L and Reddy A M 2021 Magnitude of river discharge determines the food habit and community composition of zooplankton in the Indian estuaries during monsoon *Estuar. Coast. Shelf Sci.* **262** 107601
Wang S, Peng H, Hu Q and Jiang M 2022 Analysis of runoff generation driving factors based on hydrological model and interpretable machine learning method *J. Hydrol. Reg. Stud.* **42** 101139
Wang Y I, He B I N and Takase K 2009 Effects of temporal resolution on hydrological model parameters and its impact on prediction of river discharge *Hydrol. Sci. J.* **54** 886–98
Workneh H A and Jha M K 2025 Utilizing deep learning models to predict streamflow *Water* **17** 756
Xu K, Chen L and Wang S 2024 Kolmogorov-Arnold networks for time series: bridging predictive power and interpretability (arXiv:2406.02496)
Yaseen M, Dal Sasso S F, Izzaddin A, Margiotta M R, Onorati B, Sileo B and Iacobellis V 2026 Daily streamflow prediction using lightGBM and random forest in the Fiumarella di Corleto basin, Southern Italy *Model. Earth Syst. Environ.* **12** 188
Zadeh M R, Amin S, Khalili D and Singh V P 2010 Daily outflow prediction by multi layer perceptron with logistic sigmoid and tangent sigmoid activation functions *Water Resour. Manag.* **24** 2673–88
Zuo R, Xiong Y, Wang J and Carranza E J M 2019 Deep learning and its application in geochemical mapping *Earth-Sci. Rev.* **192** 1–14