



OPEN Development of a novel modeling framework based on weighted kernel extreme learning machine and ridge regression for streamflow forecasting

Arvin Samadi-Koucheksaraee & Xuefeng Chu✉

A precise streamflow forecast is crucial in hydrology for flood alerts, water quantity and quality management, and disaster preparedness. Machine learning (ML) techniques are commonly employed for hydrological prediction; however, they still face certain drawbacks, such as the need to optimize the appropriate predictors, the ability of the models to generalize across different time horizons, and the analysis of high-dimensional time series. This research aims to address these specific drawbacks by developing a novel framework for streamflow forecasting. Specifically, a hybrid ML model, WKELM-R, is developed to predict streamflow based on daily discharge and precipitation. The model combines ridge regression (RR), locally weighted linear regression (LWLR), and kernel extreme learning machine (KELM) to enhance multi-step-ahead predictions by accounting for both linear and nonlinear characteristics. In data preprocessing, this study applies multivariate variational mode decomposition (MVMD) for decomposition to handle non-stationarity and complexity, Boruta-XGBoost for feature selection to select the optimal inputs and decrease the dimension, and gradient-based optimizer (GBO) for adjustment of model parameters to overcome the need to optimize the appropriate predictors. To demonstrate the ability to handle real-world conditions and different time horizons, WKELM-R was applied to a watershed in North Dakota, USA to forecast discharge for three different time horizons. The results were compared with those from the existing standalone and hybrid models by multi-criteria decision-making (MCDM), demonstrating the efficacy and unique capabilities of the new hybrid model in streamflow forecasting (for the testing level at $t + 3$: $R = 0.992$, $RMSE = 0.426$, $NSE = 0.983$; at $t + 7$: $R = 0.997$, $RMSE = 0.249$, $NSE = 0.994$; at $t + 14$: $R = 0.996$, $RMSE = 0.304$, $NSE = 0.991$).

Keywords Streamflow forecasting, Multivariate variational mode decomposition (MVMD), Gradient-based optimization (GBO), Multi-criteria decision-making (MCDM), Hybrid WKELM-R model

Accurate forecasting of streamflow is essential for effective water resource management. Streamflow forecasting models are significantly influenced by the composite properties of severe nonlinearity, high levels of uncertainty, and spatiotemporal variability¹. Both physically-based and data-driven models have been widely used for streamflow simulations^{2,3}. Although the physical models provide detailed simulations of hydrological processes, the process of parameter adjustment in these models can be laborious⁴ because of limited data availability, inherent model complexities, and other factors. Therefore, under some circumstances, such physically-based models may not be applicable^{5–7}. In contrast, data-driven models are capable of estimating statistical connections among input variables and target variables without taking into account the mathematical equations that govern hydrological systems^{8–10}. In addition, data-driven models require notably less data compared to physically-based models. They can also effectively tackle the issue of missing data⁹.

Data-driven techniques for prediction can be broadly classified into two distinct categories, time-series (TS) methods and machine learning (ML) models¹¹. Some popular TS methods include least squares (LS), multiple linear regression (MLR), and stepwise cluster analysis (SCA). TS models in streamflow forecasting use a linear relationship among input and target variables with an assumption of stationarity for the recorded data¹¹. As

Department of Civil, Construction and Environmental Engineering (Dept 2470), North Dakota State University, PO Box 6050, Fargo, ND 58108-6050, USA. ✉email: xuefeng.chu@ndsu.edu

a consequence of these assumptions, the accuracy of both short-term and long-term forecasts is decreased¹². Given the issues with prediction accuracy and the need for improved results, there is an emphasis on ML models as more reliable options¹³. Unlike TS models, ML models are more adept at identifying intricate correlations, non-linearity, and non-stationarity among the elements of a phenomenon¹⁴. Some ML models commonly used for hydrological modeling include support vector machine (SVM)¹⁵, multi-layer perceptron (MLP)^{16,17}, extreme learning machine (ELM)¹⁸, artificial neural network (ANN)¹⁹, kernel extreme learning machine (KELM)²⁰, and several others.

Various ML models have been applied for streamflow forecasting in recent years. For example, El-Shafie and Nouredin²¹ performed multi-lead inflow forecasting by using autocorrelation and cross-correlation analyses to enhance input patterns and employing regularized neural network (RNN) and ensemble neural network (ENN) models to mitigate overfitting. Yaseen, et al.²² used the ELM for streamflow prediction and emphasized that a primary limitation of the ELM method was its reliance on a random starting process for the internal hidden layer weights. Yaseen, et al.⁹ further found that the integration of more reliable information as attributes could lead to improved efficiency of the ELM approach. In addition, Abozweita, et al.²³ assessed the predictive reliability of an optimized decision tree (ODT) model using historical inflow and rainfall data for the Terengganu River in Malaysia. Although there are some successful applications, ML models are often subject to their inherent “black box” characteristic, susceptibility to overfitting, and reliance on empirical methods for model construction²⁴. Therefore, there is room to improve their performance and efficacy.

Among various ML modeling efforts, development of hybrid ML models is one of the key alternatives for improving their prediction accuracy^{25–27}. Hybrid ML models can effectively exploit the respective advantages of different models and enhance their performances²⁸. For instance, Chang, et al.²⁹ demonstrated the advantages of employing a hybrid ML model that combined dynamic neural networks and a self-organizing map (SOM) to predict flooding using rainfall and previous flooding levels. Dariane and Azimi³⁰ used input selection methodologies based on singular spectrum analysis (SSA) and genetic algorithms (GA) to predict river flow and demonstrate improved accuracy. Moreover, Tikhamarine, et al.³¹ showed the capability of the grey wolf optimization method in enhancing the precision of streamflow forecasts. Ramaswamy and Saleh³² also combined ensemble-based and optimized algorithms to estimate reservoir outflow.

Recently, Ahmed, et al.³³ developed a novel modeling framework that combined multi-layer perceptron (MLP) with metaheuristic algorithms (MHAs) for estimating monthly streamflow and showed that the MLP using nuclear reaction optimization (NRO) achieved the highest predicting accuracy. Moreover, Adnan, et al.³⁴ presented a new hybrid model named ELM-IRSA to improve river flow modeling. Kilinc, et al.³⁵ developed a hybrid ML model that combined CatBoost and genetic algorithm (GA) to forecast river flow on a daily basis. Momeneh and Nourani³⁶ examined the efficacy of hybrid models for streamflow forecasting by combining an ANN model with several ML models. In addition, Jamei, et al.³⁷ created a hybrid model, MVMD-CNN-BiGRU, to predict daily streamflow with one and three days in advance, demonstrating its superior accuracy. In another study, Adnan, et al.³⁸ examined the effectiveness of a relevance vector machine mixed with a dwarf mongoose optimization algorithm in monthly streamflow modeling.

In spite of the attempts to develop hybrid ML models, they still face several major drawbacks, such as the need to optimize the appropriate predictors, the ability to generalize across different time horizons, and the analysis of high-dimensional time series. Therefore, more efforts are needed to develop effective methodologies for deriving optimum solutions in hydrologic forecasting³¹. One of these efforts involves the integration of hybrid ML models with data preprocessing techniques³¹. Such combinations can help models overcome the mentioned drawbacks. One of the preprocessing techniques is time series decomposition, which has been extensively investigated in hydrologic modeling³⁹. Among the choices for time series decomposition, MVMD has been found to be reliable. In contrast to variational mode decomposition (VMD), MVMD can concurrently break down all input characteristics with adequate precision and hence reduce computing cost and time⁴⁰. In addition, decomposing the inputs into different frequency components helps in isolating and examining each portion individually, making it easier to identify patterns and correlations. Furthermore, the modeling accuracy can be potentially enhanced by using a combination of two decomposition methods that reveal the intricate correlations among the input features^{41,42}. So, feature selection strategies can effectively improve predictive performance by reducing network noise and optimizing network processing learning. In addition, feature selection strategies with selecting the optimal inputs can help decrease dimension of the problems. Furthermore, adjusting the scaling factors associated with hybrid models is amongst the most critical phases to obtain the greatest potential efficiency and accuracy. In this regard, metaheuristic optimization algorithms, random search, grid search strategy, and manual search (trial and error strategy) are able to improve the parameters of ML models.

In this study, a new hybrid ML model, B-MVMD-WKELM-R-GBO, is developed for streamflow prediction with combination of linear and nonlinear ML models and data preprocessing strategies. This study uses multi-stage MVMD for data preprocessing to handle non-stationarity and complexity, and XGBoost feature selection for choosing the most related inputs and decreasing dimension of the problem. In addition, this study integrates RR and LWLR with KELM to effectively analyze high-dimensional time series and utilizes GBO to optimize the hyperparameters. To demonstrate the improvements in computational efficiency and accuracy, B-MVMD-WKELM-R-GBO is compared with several other robust models. In addition, three multi-criteria decision-making (MCDM) techniques (MARCOS, ARAS, and COPRAS) are utilized to determine the most precise forecasting framework for streamflow, while considering six metrics including R, RMSE, MAPE, NSE, IA, and U95%. More importantly, the new approach is tested through a real application in the Upper Turtle River watershed in North Dakota, USA to highlight its broad implications and the ability to generalize across different time horizons.

Materials and methods

A hybrid ML model, B-MVMD-WKELM-R-GBO, was developed for understanding the dynamic interconnections between precipitation and streamflow in this study. The B-MVMD-WKELM-R-GBO model was enhanced in the MATLAB R2022a and Python 3.9 environments. Therefore, the proposed ML model developing process was carried out in two distinct steps. Firstly, data pre-processing was conducted by using XGBoost feature selection and MVMD. Secondly, a hybrid model was created by coupling RR, LWLR, KELM, and GBO. The model was executed on a Windows 10 device with an Intel Core i7 CPU running at a speed of 3.20 GHz and 16 GB of RAM. The hybrid B-MVMD-WKELM-R-GBO model was then applied to forecast the discharge of the Upper Turtle River for 3 days ahead ($t+3$), 7 days ahead ($t+7$), and 14 days ahead ($t+14$). Figure 1 shows the overall flowchart of this hybrid model. The specific modeling procedures and steps are detailed in the following subsections.

Data pre-processing and preparing

One of the difficult problems in constructing ML models that rely on a large number of input characteristics is selecting the appropriate and most effective mix of input features to be used as inputs of the models. In the last decades, the introduction of AI-based techniques and the progress of data mining technologies have led to the emergence of feature selection techniques. These techniques facilitate effective identification of the characteristics that are important to selecting an objective variable from current variables based on a certain logic. The process of selecting features is crucial in constructing a prediction model because it directly affects the correctness of the model. Therefore, the model's efficacy and cost can be improved by selecting the most suitable features.

In the proposed approach, both the lags of the observed data and the observed dataset itself are listed as input data. Then, the selection of optimum input parameters is carried out through two distinct phases. The preliminary phase occurs before the execution of decomposition. This phase is primarily focused on finding the relevant delays. In the subsequent phase, the feature selection technique identifies the optimal input data

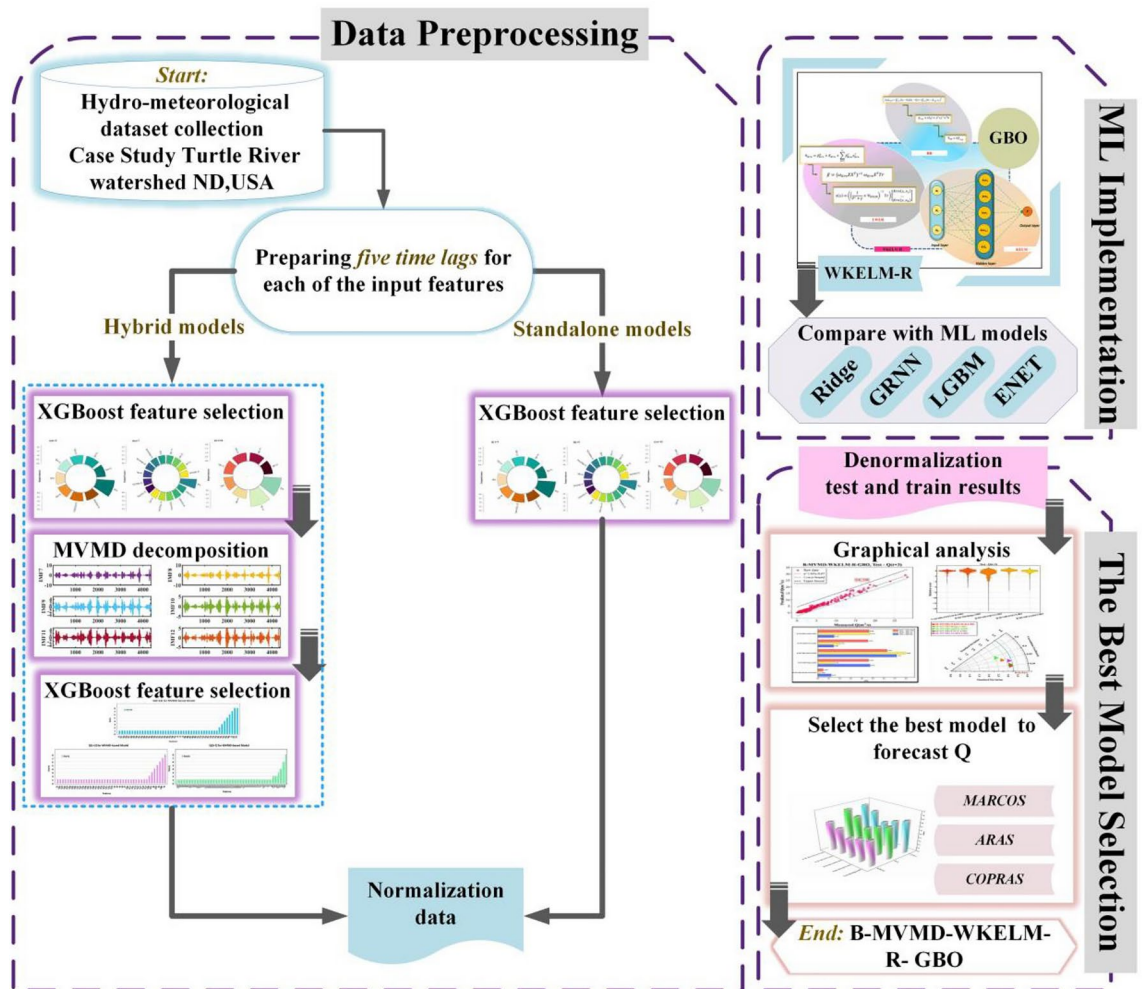


Fig. 1. Schematic flowchart depicting the development procedures of the hybrid model (this figure was generated using Microsoft Visio Professional 2019, version 1808: <https://www.microsoft.com/en/microsoft-365/visio/flowchart-software>).

decomposed by the MVMD approach. The second phase is to tackle the extensive number of decomposed inputs and determine the inputs with the greatest influence.

First stage of feature selection

The XGBoost algorithm (Sánchez-Marño et al.⁴³; Kursa and Rudnicki⁴⁴) was used to detect the lags of the features that have statistical significance. An analysis was conducted on each time horizon to identify the important factors linked to the five distinct lags of the $Q(t+3)$, $Q(t+7)$, and $Q(t+14)$ indices in this study. Therefore, 9, 16, and 8 characteristics were selected (Fig. 2) based on their ranks to forecast $Q(t+3)$, $Q(t+7)$, and $Q(t+14)$, respectively. The optimal time delays selected in the present step were necessary for the subsequent stages.

Signals decomposition

Signals decomposition is also vital because the success of an ML model is completely contingent on the information acquired from the input dataset. In fact, the more reliable the input dataset is, the more reliable the model's prediction is. MVMD, created by Ur Rehman and Aftab⁴⁰, represents a generalized variant of the variational mode decomposition (VMD) process. It is specifically designed to handle multivariate collections of data. Furthermore, in contrast to the univariate mode decomposition approach, MVMD exhibits a more precise mathematical foundation and a superior ability to identify the shared frequency patterns across signals⁴⁵. Therefore, the MVMD approach concurrently decomposed the selected input characteristics identified in the previous stage. This process entailed decomposing the independent variables into many sub-series, known as Intrinsic Mode Functions (IMFs), along with a residual component. The goal of this process was to reduce the complexity of the main signals. Thus, the optimal values of K , representing the total number of modes, were 5 for $Q(t+3)$, 10 for $Q(t+7)$, and 12 for $Q(t+14)$ in this study. In addition, these numbers were obtained by a series

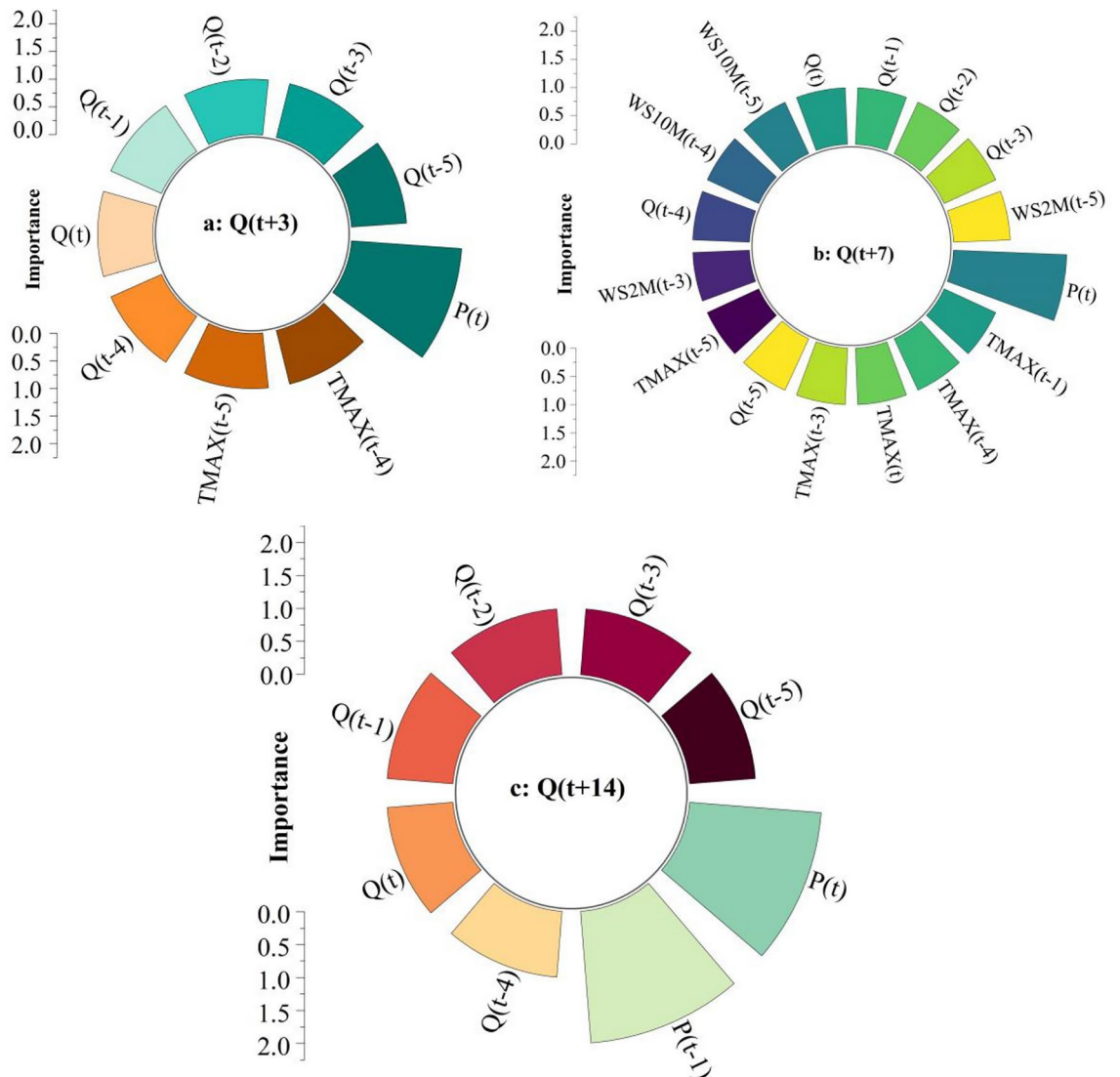


Fig. 2. Scores of input modes achieved by XGBoost (a) for $Q(t+3)$; (b) for $Q(t+7)$; and (c) for $Q(t+14)$.

of trial and error processes. Note that choosing a correct K value is essential. If K is too high, it can lead to mode aliasing, while a too low K value results in insufficient feature extraction and poor decomposition.

Second stage of feature selection

Following the breakdown of signals, the input dataset included 45 modes, represented as 5×9 IMFs for $Q(t+3)$. For $Q(t+7)$, there were 160 modes, stated as 10×16 IMFs, while there were 96 modes for $Q(t+14)$, represented as 12×8 IMFs. As a result of the increase in variables after decomposition, the XGBoost method was used again to determine the effective time delays of the features and remove ineffective variables from the input data acquired from the decomposition process to improve computational efficiency. During this stage, the variables with the highest rank (rank 1) were chosen as the input dataset for the ML models, instead of picking ranks 1 and 2 as performed in the first feature selection step. Following the implementation of the XGBoost technique for feature selection, 35 out of the 45 input modes for $Q(t+3)$, 68 out of the 160 input modes for $Q(t+7)$, and 39 out of the 96 input modes for $Q(t+14)$ were chosen as feed for the ML models.

Development of the hybrid WKELM-R-GBO model

Kernel extreme learning machine

Huang et al.⁴⁶ improved the ELM through the utilization of the kernel function (K), which altered the feature mapping $g(x)$ associated with the hidden layer⁴⁷. Not only does the kernel function $Krn(x, x_1)$ reduce the number of internal variables in the KELM, but it also evolves convergence and more powerful performance in generalization⁴⁸. In addition, the concept of ELM as a single hidden layer feed-forward neural network (SLFN) was considered by Huang et al.⁴⁹. The equation representing the generalized SLFN can be expressed as^{50–52}:

$$Z(x) = G(x)\omega = Tr \quad (1)$$

in which x and G respectively represent the given input variable and matrix, which reflects the feature mapping of the hidden layer; Tr denotes the data extracted from the training set (i.e., the output matrix from the hidden layer); and ω is the weight associated with the comprehensive feature transformation, which is calculated by finding the pseudoinverse of matrix G ^{46,47,50}:

$$\omega = G^T \left(\frac{I}{\gamma} + GG^T \right)^{-1} Tr \quad (2)$$

in which I and γ denote the identity matrix and the regularization coefficient, respectively. Thus, the function for ELM can be formulated as^{46,49}:

$$Z(x) = G(x)\omega = G(x) \left(G^T \left(\frac{I}{\gamma} + GG^T \right)^{-1} Tr \right) \quad (3)$$

In the subsequent phase, the substitution of the feature mapping in the hidden layer with a kernel function (K) is undertaken to augment both stability and generalization, resulting in the creation of a KELM. The function for KELM is given by⁴⁶:

$$Z(x) = G(x) \left(G^T \left(\frac{I}{\gamma} + GG^T \right)^{-1} Tr \right) = \begin{bmatrix} Krn(x, x_1) \\ \vdots \\ Krn(x, x_N) \end{bmatrix}^T \left(\left(\frac{I}{\gamma} + \Psi_{KELM} \right)^{-1} Tr \right) \quad (4)$$

in which

$$\Psi_{KELM} = GG^T = G(x_i) \cdot G(x_j) = Krn(x_i, x_j) \quad (5)$$

Because the wavelet kernel function (Krn_w) is capable of identifying both global and local characteristics in data, it is often chosen in a variety of ML applications. Several Krn_w exist, including the Gaussian, linear, exponential, and polynomial kernel⁵³. In this research, we used the following Krn_w because of its effectiveness in both training and testing^{54,55}:

$$Krn_w(x_i - x_i^T) = \left[\cos \left(a \cdot \frac{(x_i - x_i^T)}{b} \right) \cdot \exp \left(-\frac{\|x_i - x_i^T\|^2}{c} \right) \right] \quad (6)$$

in which a , b , and c are the tuning parameters. Since these parameters have significant impacts on the efficacy⁵⁶, it is crucial to determine their optimal values. In this study, the optimal a , b , and c were determined by using the GBO algorithm.

Thus, using the KELM learning framework can eliminate the variations in optimization constraints in the Support Vector Machine (SVM) and Least Squares Support Vector Machine (LSSVM). Because of its optimization restrictions, KELM exhibits an improved generalization performance compared with SVM and LSSVM⁴⁶. It

can be inferred that the KELM method has a comparable performance in terms of its ability to generalize from training data to novel, unknown data, making it a viable technique in the field of machine learning.

Locally weighted kernel extreme learning machine

This study combines the Locally Weighted Linear Regression (LWLR) method (Atkeson, et al.⁵⁷) with the KELM model to enhance its predictive capability. Particularly, we utilized the LWLR method employing the Multiple Linear Regression (MLR) approach, commonly known as the locally weighted KELM (WKELM)⁵⁸. With the resulting method, the dependent variable Υ_s is determined by a linear function of independent x ^{59,60}:

$$\Upsilon_s = \beta_s^0 + \sum_{j=1}^L \beta_s^j x_s^j + E_s \quad (7)$$

in which E_s is the random error; $\beta_s^0, \beta_s^1, \dots, \beta_s^L$ are the regression coefficients obtained by the Least Square (LS) method. Furthermore, a function of fitness is used to choose the path that provides the highest level of agreement with the observed data (Υ_O) by applying the multiple regression approach, as follows⁵⁷:

$$\text{Minimize } F = \frac{0.5}{N_{in}} \sum_{k=1}^{N_{in}} \left(\beta_s^0 + \sum_{j=1}^L \beta_s^j x_s^j + E_s - \Upsilon_O \right)^2 \quad (8)$$

in which N_{in} is the total number of input variables.

The matrix form of Eq. 8 can be expressed as $(X\beta - \Upsilon)^T (X\beta - \Upsilon)$. $(X)^T (X\beta - \Upsilon)$ is provided by the derivative of $(X\beta - \Upsilon)^T (X\beta - \Upsilon)$ with respect to β . The resultant matrix is considered to be equal to 0 in order to derive the LS as shown in Eq. (9)^{57,61}:

$$\beta = (X^T X)^{-1} X^T \Upsilon \quad (9)$$

in which X and Υ indicate the input and output dataset matrix, respectively.

The LWLR utilizes a weight function to characterize the relationship between the training dataset and the prediction^{57,59}:

$$\text{Minimize } F = \frac{0.5}{N_{in}} \sum_{k=1}^{N_{in}} W_{LWLR} \left(\beta_s^0 + \sum_{j=1}^L \beta_s^j x_s^j + E_s - \Upsilon_O \right)^2 \quad (10)$$

in which W_{LWLR} represents the weight function matrix.

The matrix form of the aforementioned equation can be expressed as $(X\beta - \Upsilon)^T W (X\beta - \Upsilon)$. To get optimum outcomes from Eq. (10), it is necessary to let the derivative of the fitness function (F) with respect to β be 0^{57,61}:

$$\frac{\partial F(\beta)}{\partial \beta} = X^T W X \beta - X^T W \Upsilon = 0 \quad (11)$$

$$\beta = (X^T W X)^{-1} X^T W \Upsilon \quad (12)$$

The WKELM, based on Eq. (12), is then integrated into the kernel function to enhance the predictive efficiency. Hence, Eqs. (4) and (5) can be rewritten as:

$$\Psi_{KELM} = G(x_i) \cdot G(x_l) = \text{Krn}(\beta x_i, \beta x_l) \quad (13)$$

$$Z(x) = \begin{bmatrix} \text{Krn}(x, x_1) \\ \vdots \\ \text{Krn}(x, x_N) \end{bmatrix}^T \left(\left(\frac{I}{P^* \times \gamma} + \Psi_{KELM} \right)^{-1} Tr \right) \quad (14)$$

in which the penalty variable P^* is given by:

$$P^* = \frac{N_t}{\|Tr\|} \quad (15)$$

where N_t is the total size of the dataset.

To enhance the efficiency and accuracy of the model, the WKELM model was combined with the RR model (Hoerl and Kennard⁶²) in this study, forming a hybrid model referred to as WKELM-R. Figure 3 shows its conceptual modeling framework. The resulting hybrid WKELM-R model can be expressed as:

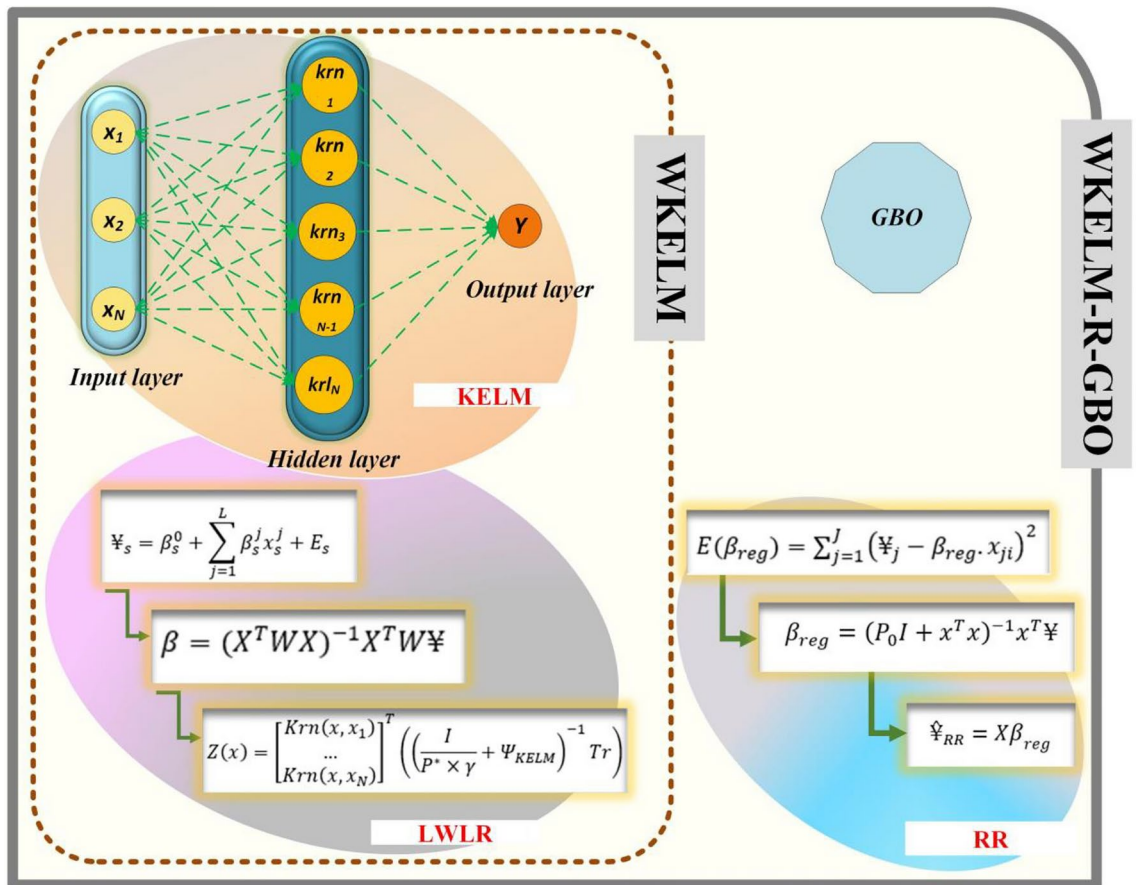


Fig. 3. Conceptual framework of the WKELM-R model (this figure was generated using Microsoft Visio Professional 2019, version 1808: <https://www.microsoft.com/en/microsoft-365/visio/flowchart-software>).

$$\hat{y}_{WKELM-R} = \eta_1 \cdot \hat{y}_{WKELM} + \eta_2 \cdot \hat{y}_{Ridge} \tag{16}$$

where η_1 and η_2 are positive real numbers within $[0, 1]$, which are determined by optimization using GBO.

Gradient-based optimizer algorithm

The GBO (Ahmadianfar et al.⁶³) centers on the utilization of two discrete mechanisms, Local Escaping Operator (LEO) and Gradient Search Rule (GSR). The primary search operator prioritizes the identification of local optima, emphasizing exploitation. Conversely, the secondary search mechanism is designed to seek global optima, emphasizing exploration. Therefore, the GBO algorithm integrates the advantages of gradient-based and population-based techniques, resulting in a very efficient and robust search mechanism^{64–66}.

Requirements for ML models

The partitioning technique for input data is crucial and complex for building ML forecasting models. Implementing an appropriate method based on the length of the time series for predicting difficulties may effectively avoid or reduce the incidence of overfitting. Therefore, overfitting can be prevented as long as the dataset is long enough (in this study, the dataset ranges from January 1, 2010 to December 31, 2021). During this step, the dataset was divided into training and testing subsets, with 70% of the data for model training and the remaining 30% for model testing. Furthermore, all the inputs and goals were scaled to a range of 0–1 in order to optimize the convergence and effectiveness of the ML models. In addition, standalone ML models were trained and tested with the datasets before the decomposition phase.

It is necessary to carefully adjust the controlling factors in order to optimize the performance of ML-based forecasting methods. This study employed the GBO optimization technique (with 100 iterations and 50 population sizes) to get the best tuning settings, using the root-mean-squared error (RMSE) as the convergence criterion to achieve the highest feasible accuracy. That is, RMSE was utilized as an evaluation criterion inside the GBO framework, successfully assessing the sensitivity of the tuning parameters of the models. It was notable that even minor changes to the parameters could result in significant variations in RMSE, indicating their influence

on the model efficacy. The control factors of the hybrid ML models for $Q(t+3)$, $Q(t+7)$, and $Q(t+14)$ are summarized in Table 1.

Study area and datasets

The Upper Turtle River (UTR) watershed in North Dakota (ND), USA (Fig. 4) was selected for this study because streamflow prediction is of the utmost importance for agriculturally dominant areas. The area of the UTR watershed is 664.24 km², covering parts of Grand Forks and Nelson counties in ND. The lowest elevation point (295 m) in the watershed is located at the southeast boundary, while the highest elevation point (481 m) is in the west. According to the 2019 National Land Cover Database, agricultural lands account for 67.03% of the watershed area. In addition, the areas with poor soil conditions such as riparian zones and salt marshes are used for livestock farming. The watershed drains to its outlet located at the USGS gaging station 05082625 at Turtle River State Park near Arvilla, ND (47° 55' 55" N, 97° 30' 51" W). On average, 44.7-cm precipitation falls in this watershed each year, which is characterized by a climate mostly defined by dynamic continental circumstances. In this study, a 12-year dataset (January 1, 2010–December 31, 2021) was used for prediction of discharge of the Upper Turtle River. Table 2 shows the basic hydroclimatic data and Table 3 presents their statistical details.

Table 3 shows different trends of the hydrometeorological data. The two major hydrological variables (discharge Q and precipitation P) are highly variable as shown by the high kurtosis (44.15 for Q and 40.21 for P) and skewness (5.26 for Q and 5.34 for P). The statistics indicate strongly peaked distributions and numerous low values. It is hence challenging to accurately forecast streamflow in the watershed. These complicated and changing features of this watershed make it a good site to test the capability of the new ML model predicting streamflow under a range of real situations.

Results and discussion

Analyses of results

Tables 4 and 5 provide a comprehensive overview of the statistical results (including R , RMSE, MAPE, NSE, I_A , and $U_{95\%}$) for the standalone and hybrid models in prediction of $Q(t+3)$, $Q(t+7)$, and $Q(t+14)$ under varying situations (such as variable K values) throughout both the training and testing phases. MAPE measures the percentage error and RMSE measures the magnitude of the prediction error. Thus, these two metrics complement one another and provide useful information. In addition, $U_{95\%}$ quantifies the uncertainty surrounding predictions; NSE measures how well the model predicts outcomes in comparison to the mean of observed values; and R quantifies the proportion of variance in the dependent variable that can be explained by the independent variables in the model. Furthermore, I_A offers a thorough evaluation of model correctness by considering both the amount and direction of errors.

Beginning with $Q(t+3)$ in Table 4, the B-MVMD-WKELM-R-GBO model provided reliable results in both training and testing phases when it is juxtaposed with other hybrid models. Throughout the training process, it achieved remarkable values of R (0.996) and NSE (0.993), and very low $U_{95\%}$ (0.805) and RMSE (0.291), indicating its precise and reliable forecasts. During the testing phase, the values of R and NSE remained high (0.992 and 0.996, respectively), while the RMSE, MAPE, and $U_{95\%}$ experienced marginal increases. On the contrary, the efficacy of all other models exhibited a decline, indicating the possibility of overfitting and difficulties in extrapolating findings throughout the testing and training phases. For example, the B-MVMD-

Time horizon	Models	Tuning parameters
t+3	B-WKELM-R-GBO	$a = 1.90E + 03, b = 2.00E + 02, c = 2.00E + 03, \gamma = 1985.56, P^* = 0.48$
	B-Ridge-GBO	$P_0 = 0.45$
	B-GRNN-GBO	$spreadparameter = 3$
	B-ENET-GBO	$\alpha = 0.01, L1_ratio = 1$
	B-LGBM-GBO	$r = 0.04, ms = 43.25, mv = 0.28, ra = 5.99, s = 7.19, c = 0.93$
	MVMD	$K = 5$
t+7	B-WKELM-R-GBO	$a = 1.64E + 03, b = 1.76E + 03, c = 1.25E + 03, \gamma = 200.99, P^* = 6.99E - 05$
	B-Ridge-GBO	$P_0 = 7.05E - 5$
	B-GRNN-GBO	$spreadparameter = 5$
	B-ENET-GBO	$\alpha = 0.01, L1_ratio = 1$
	B-LGBM-GBO	$r = 0.15, ms = 66.30, mv = 0.57, ra = 3.06, s = 7.88, c = 0.01$
	MVMD	$K = 10$
t+14	B-WKELM-R-GBO	$a = 1.70E + 03, b = 1.89E + 03, c = 2.09E + 02, \gamma = 33.67, P^* = 4.13E - 05$
	B-Ridge-GBO	$P_0 = 2.41E - 05$
	B-GRNN-GBO	$spreadparameter = 2.5$
	B-ENET-GBO	$\alpha = 0.0001, L1_ratio = 1$
	B-LGBM-GBO	$r = 0.39, ms = 66.78, mv = 0.25, ra = 1.02, s = 3.44, c = 1E - 3$
	MVMD	$K = 12$

Table 1. Control parameters.

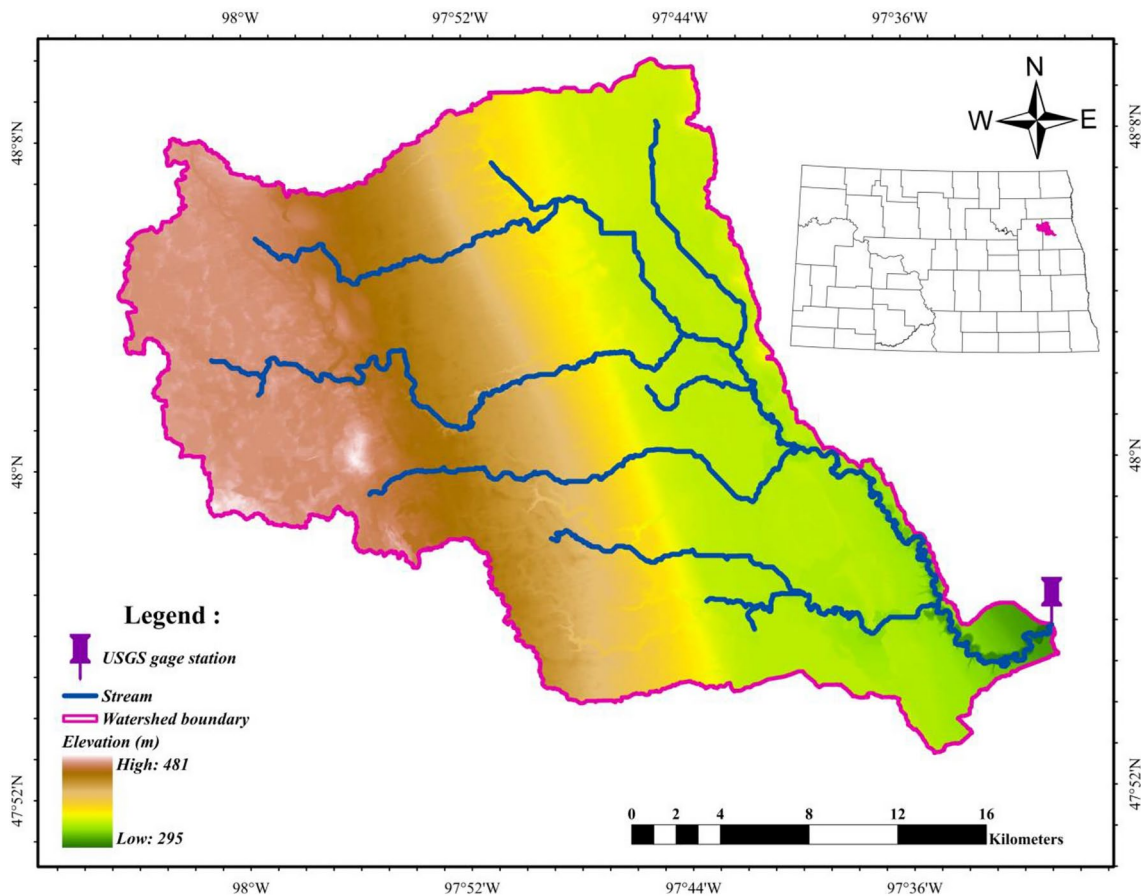


Fig. 4. Location of the Upper Turtle River watershed (this map was generated using ArcGIS Pro, version 3.1.2: <https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview>).

Data	Sources
Discharge (m ³ /s)	United states geological survey (USGS)
USGS 05082625 at Turtle River State Park near Arvilla, ND	
Precipitation (mm)	National Oceanic and Atmospheric Administration (NOAA)
Maximum temperature (°C)	National Oceanic and Atmospheric Administration (NOAA)
Minimum temperature (°C)	National Oceanic and Atmospheric Administration (NOAA)
Petersburg 2 N, ND US, USC00327027, 48.0356°, -98.0096°	
Wind speed at 2 m (m/s)	POWER (NASA prediction of worldwide energy resources)
Wind direction at 2 m (degrees)	
Wind speed at 10 m (m/s)	
Wind direction at 10 m (degrees)	
Location: Latitude 48.0356°, Longitude -98.0096°	

Table 2. Description of the hydrometeorological data and their sources.

GRNN-GBO model showed difficulties during testing, suggesting constraints in making forecasts for $Q(t+3)$. In contrast, the B-MVMD-ENET-GBO and B-MVMD-LGBM-GBO models provided better performances in both training and testing phases, highlighting their dependability as the second and third models.

For the medium-term forecasting horizon $Q(t+7)$, as shown in Table 4, the B-MVMD-WKELM-R-GBO model consistently yielded better results in both testing and training phases, demonstrating its capability generating reliable forecasts in the medium future. The results of the B-MVMD-Ridge-GBO model demonstrated its dependability in testing, despite a small reduction in R during the training phase. The B-MVMD-GRNN-GBO model had difficulties during testing, indicating possible restrictions for this period. The B-MVMD-ENET-GBO and B-MVMD-LGBM-GBO models continued to provide competitive performances.

This modeling study revealed that when the prediction time expanded to $(t+14)$, all models encountered more difficulties in generalization. However, the B-MVMD-WKELM-R-GBO model remained successful, with higher R and NSE values (0.996 and 0.991) and lower RMSE and MAPE values (0.304 and 16.694) throughout

Statistics	Q (m ³ /s)	P (mm)	TMax (°C)	TMin (°C)	WS2M (m/s)	WD2M (degrees)	WS10M (m/s)	WD10M (degrees)
Kur	44.15	40.21	1.99	2.23	4.26	2.38	4.12	2.38
Max	53.66	61.00	39.40	22.20	13.51	351.81	17.70	351.81
Min	0.05	0.00	- 27.80	- 37.80	0.52	5.69	0.75	5.69
Ave	1.64	1.66	9.94	- 1.34	3.63	210.19	5.19	210.24
Med	0.31	0.00	23.90	13.90	3.30	188.00	5.05	202.50
SD	3.34	5.05	14.61	13.31	1.45	77.02	1.93	77.03
Skew	5.26	5.34	- 0.29	- 0.42	0.82	- 0.46	0.76	- 0.46

Table 3. Descriptive statistics of all historical hydrometeorological data. *Kur* kurtosis, *Max* maximum, *Min* minimum, *Ave* average, *SD* standard deviation, *Skew* skewness), *Med* median, *Q* discharge, *P* precipitation, *TMax* maximum temperature, *TMin* minimum temperature, *WS2M* wind speed at 2 m, *WD2M* wind direction at 2 m, *WS10M* wind speed at 10 m, *WD10M* wind direction at 10 m.

	Hybrid model		R	RMSE	MAPE	NSE	I _A	U _{95%}
Q(t+3) K=5	B-MVMD-WKELM-R-GBO	Train	0.996	0.291	13.003	0.993	0.998	0.805
		Test	0.992	0.426	22.548	0.983	0.996	1.181
	B-MVMD-Ridge-GBO	Train	0.943	1.168	46.780	0.880	0.964	3.238
		Test	0.954	1.031	62.880	0.901	0.971	2.857
	B-MVMD-GRNN-GBO	Train	0.957	1.035	49.480	0.906	0.973	2.852
		Test	0.920	1.360	98.449	0.827	0.943	3.769
	B-MVMD-ENET-GBO	Train	0.977	0.720	37.757	0.954	0.988	1.480
		Test	0.987	0.534	43.882	0.973	0.993	1.480
	B-MVMD-LGBM-GBO	Train	0.984	0.604	11.361	0.968	0.992	1.674
		Test	0.968	0.826	33.167	0.936	0.983	2.289
Q(t+7) K=10	B-MVMD-WKELM-R-GBO	Train	0.991	0.459	16.993	0.982	0.995	1.270
		Test	0.997	0.249	21.267	0.994	0.999	0.690
	B-MVMD-Ridge-GBO	Train	0.977	0.717	41.200	0.955	0.988	1.987
		Test	0.988	0.509	46.479	0.976	0.994	1.410
	B-MVMD-GRNN-GBO	Train	0.958	1.024	47.767	0.908	0.973	2.830
		Test	0.866	1.674	109.882	0.738	0.909	4.640
	B-MVMD-ENET-GBO	Train	0.962	0.926	46.902	0.925	0.980	2.567
		Test	0.978	0.687	56.400	0.956	0.988	1.904
	B-MVMD-LGBM-GBO	Train	0.987	0.553	9.874	0.973	0.993	1.532
		Test	0.948	1.083	46.010	0.890	0.967	3.000
Q(t+14) K=12	B-MVMD-WKELM-R-GBO	Train	0.985	0.589	14.543	0.970	0.992	1.634
		Test	0.996	0.304	16.694	0.991	0.998	0.842
	B-MVMD-Ridge-GBO	Train	0.938	1.176	63.791	0.879	0.966	3.260
		Test	0.957	1.001	73.758	0.907	0.973	2.774
	B-MVMD-GRNN-GBO	Train	0.959	1.000	43.152	0.913	0.975	2.765
		Test	0.820	1.875	97.717	0.672	0.891	5.198
	B-MVMD-ENET-GBO	Train	0.939	1.166	67.669	0.881	0.967	3.234
		Test	0.956	1.004	78.862	0.906	0.973	2.783
	B-MVMD-LGBM-GBO	Train	0.977	0.729	40.543	0.953	0.988	2.022
		Test	0.902	1.422	83.286	0.811	0.943	3.942

Table 4. Statistical metrics of the B-MVMD-WKELM-R-GBO model and four other hybrid ML models. Significant values are given in bold.

testing. The B-MVMD-Ridge-GBO and B-MVMD-ENET-GBO models exhibited comparable performances during the testing phase. Nevertheless, the results of the B-MVMD-GRNN-GBO model showed a significant disparity in performance between the training and testing phases.

The comparisons of the hybrid and standalone ML models demonstrate the persistent advantages of the hybrid ML models in predicting streamflow across different horizons. For the short-term (Q(t+3)), the B-MVMD-WKELM-R-GBO model stood out with the best performance and showed much better R and NSE values, lower RMSE, U95%, and MAPE in comparison with the standalone models. These findings indicated that combining MVMD and WKELM-R with a robust optimizer GBO improved the precision of forecasts. The hybrid models,

	Standalone model		R	RMSE	MAPE	NSE	I_A	$U_{95\%}$	
Q(t + 3) K = 0	B-WKELM-R-GBO	Train	0.817	1.979	45.309	0.657	0.872	5.485	
		Test	0.897	1.504	69.562	0.788	0.929	4.170	
	B-Ridge-GBO	Train	0.729	2.310	77.638	0.532	0.819	6.404	
		Test	0.873	1.674	110.511	0.738	0.906	4.640	
	B-GRNN-GBO	Train	0.724	2.353	94.431	0.515	0.795	6.515	
		Test	0.818	1.916	169.945	0.657	0.873	5.308	
	B-ENET-GBO	Train	0.730	2.309	77.123	0.533	0.820	6.401	
		Test	0.874	1.667	109.984	0.740	0.907	4.621	
	B-LGBM-GBO	Train	0.797	2.043	66.386	0.634	0.869	5.664	
		Test	0.872	1.608	95.906	0.758	0.923	4.457	
	Q(t + 7) K = 0	B-WKELM-R-GBO	Train	0.598	2.801	89.961	0.313	0.589	7.745
			Test	0.727	2.397	140.414	0.463	0.726	6.639
B-Ridge-GBO		Train	0.442	3.031	150.187	0.195	0.493	8.403	
		Test	0.681	2.532	221.344	0.401	0.668	7.016	
B-GRNN-GBO		Train	0.558	2.840	159.996	0.293	0.581	7.872	
		Test	0.613	2.657	282.758	0.340	0.630	7.355	
B-ENET-GBO		Train	0.448	3.021	145.311	0.200	0.511	8.375	
		Test	0.684	2.504	213.889	0.414	0.684	6.939	
B-LGBM-GBO		Train	0.653	2.922	181.653	0.252	0.456	8.099	
		Test	0.525	2.873	282.853	0.229	0.485	7.950	
Q(t + 14) K = 0		B-WKELM-R-GBO	Train	0.359	3.192	115.838	0.109	0.384	8.807
			Test	0.436	2.963	190.335	0.181	0.462	8.206
	B-Ridge-GBO	Train	0.300	3.226	191.862	0.090	0.317	8.941	
		Test	0.413	2.999	274.214	0.161	0.402	8.312	
	B-GRNN-GBO	Train	0.333	3.195	193.917	0.107	0.332	8.855	
		Test	0.348	3.077	289.754	0.117	0.330	8.529	
	B-ENET-GBO	Train	0.300	3.226	191.906	0.090	0.317	8.941	
		Test	0.413	3.000	274.280	0.161	0.402	8.312	
	B-LGBM-GBO	Train	0.370	3.146	137.049	0.134	0.382	8.720	
		Test	0.399	3.007	189.485	0.156	0.446	8.338	

Table 5. Statistical metrics of the B-WKELM-R-GBO model and four other standalone ML models. Significant values are given in bold.

particularly B-MVMD-WKELM-R-GBO, continued to outperform the standalone models for both medium-term Q(t + 7) and long-term Q(t + 14) projections. This study also emphasized the susceptibility of standalone models to forecast horizons, resulting in a decrease in accuracy over extended timeframes. On the other hand, the hybrid models exhibited robustness and resilience, making them especially well-suited for prediction that needs the flexibility of adjusting to different temporal patterns. Thus, the B-MVMD-WKELM-R-GBO model provided an attractive option for precise and adaptable time series forecasting over various timeframes.

Figures 5, 6 and 7 show the scatter plots of the observed and predicted discharges of the Upper Turtle River at (t + 3), (t + 7), and (t + 14) during the testing period. The evaluation of the forecast results for these three time horizons shows that the B-MVMD-WKELM-R-GBO model outperformed the B-MVMD-Ridge-GBO, B-MVMD-GRNN-GBO, B-MVMD-ENET-GBO, and B-MVMD-LGBM-GBO methods with the highest R^2 . Although the B-MVMD-ENET-GBO model's accuracy dropped at (t + 7) and (t + 14), it still showed an overall good performance, especially for Q(t + 3) ($R^2 = 0.9745$). On the other hand, the B-MVMD-Ridge-GBO model produced middling results with R^2 ranging from 0.9117 to 0.9186, indicating an acceptable match but lacking robustness. The B-MVMD-GRNN-GBO model's difficulties in properly representing discharge dynamics were most apparent with R^2 values below 0.85. However, the B-MVMD-LGBM-GBO model showed steady but lower results overall. Therefore, the results reinforced the effectiveness of the B-MVMD-WKELM-R-GBO model in delivering precise and dependable forecasts, highlighting its ability to improve the accuracy of streamflow predictions.

The violin plots in Figs. 8, 9 and 10 respectively illustrate the relative errors of the discharges forecasted by the hybrid ML models at (t + 3), (t + 7), and (t + 14) throughout the training phase. It is evident that the proposed B-MVMD-WKELM-R-GBO model exhibited a more precise and consistent distribution of violins, with smaller relative errors. Specifically, the relative errors ranged from - 4.743 to +0.632 for (t + 3), - 3.009 to +0.963 for (t + 7), and - 2.713 to +0.838 for (t + 14) when predicting Q of the Upper Turtle River. The proposed model outperformed all other hybrid ML models in terms of accuracy. Therefore, the B-MVMD-WKELM-R-GBO model demonstrated its superior performances in predicting streamflow for (t + 3), (t + 7), and (t + 14).

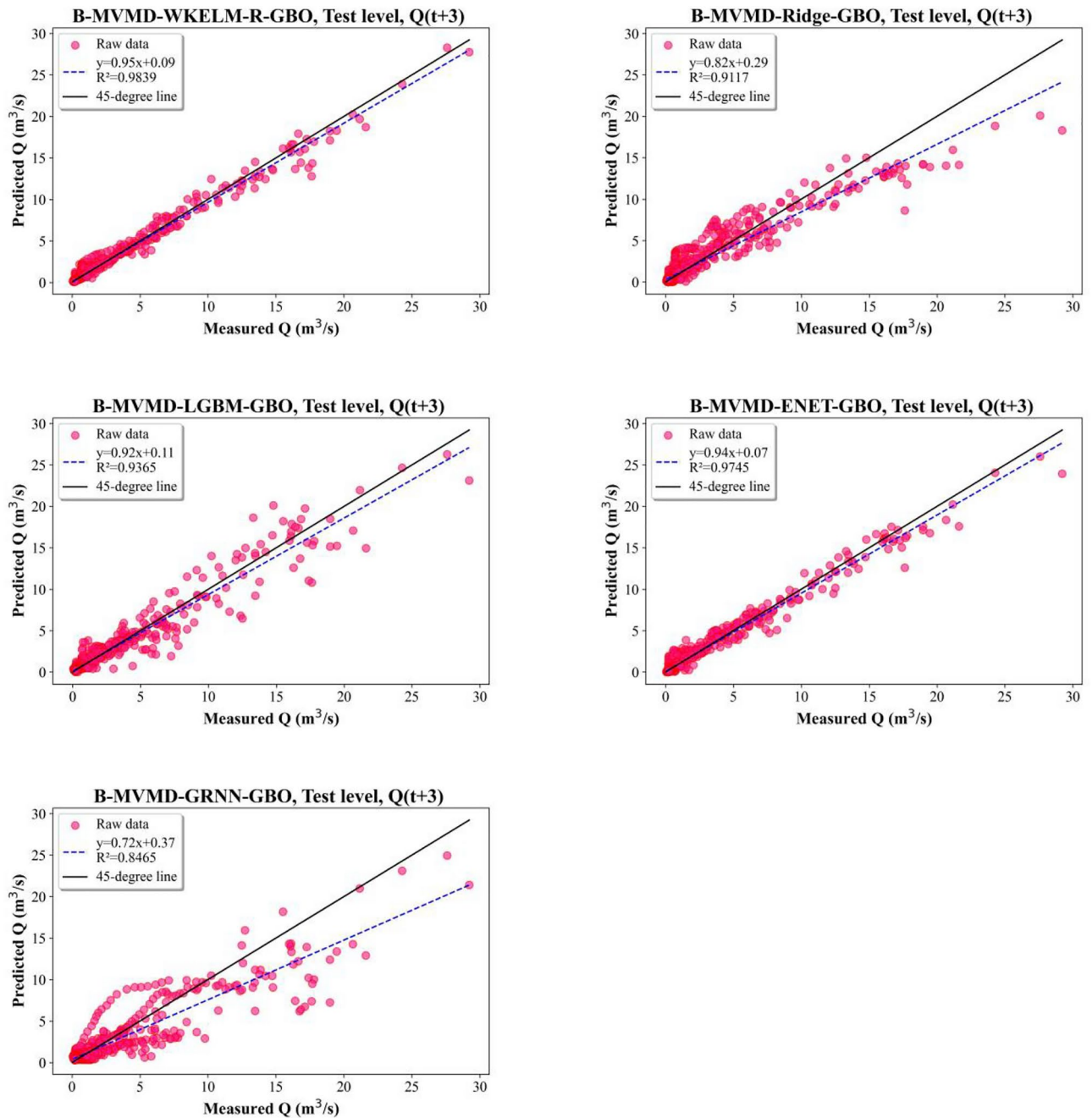


Fig. 5. Comparisons of the predicted and measured discharges (Q) of the Upper Turtle River at $(t+3)$ during the testing period for five hybrid ML models.

To compare the consistency of predictions from different models, it is essential to evaluate the standard deviation errors (SDEs) since stability is very important in forecasting. Figure 11 shows the SDEs of Q predicted by the five ML models in the testing stage. For $Q(t+3)$, the B-MVMD-WKELM-R-GBO model provided reliable results in testing, with the lowest SDEs (0.122), suggesting that it successfully captured fundamental patterns and demonstrated consistency throughout the testing phase. The proposed model continued to demonstrate outstanding performances for $Q(t+7)$ and $Q(t+14)$, with the lowest SDEs (0.026 for $Q(t+7)$ and 0.049 for $Q(t+14)$). However, the B-MVMD-Ridge-GBO, B-MVMD-ENET-GBO, and B-MVMD-LGBM-GBO models exhibited poor performances. In summary, this study showed that the B-MVMD-WKELM-R-GBO model consistently demonstrated robust performances in all circumstances, making it a viable and appropriate method for streamflow forecasting.

The MCDM approaches were used to compare the rankings of the hybrid MVMD-based models. Figures 12, 13 and 14 display the rankings of the five different hybrid MVMD-based models based on the MARCOS, ARAS, and COPRAS scores, which range from 0 to 1. As shown in Fig. 12 for $Q(t+3)$, the B-MVMD-WKELM-

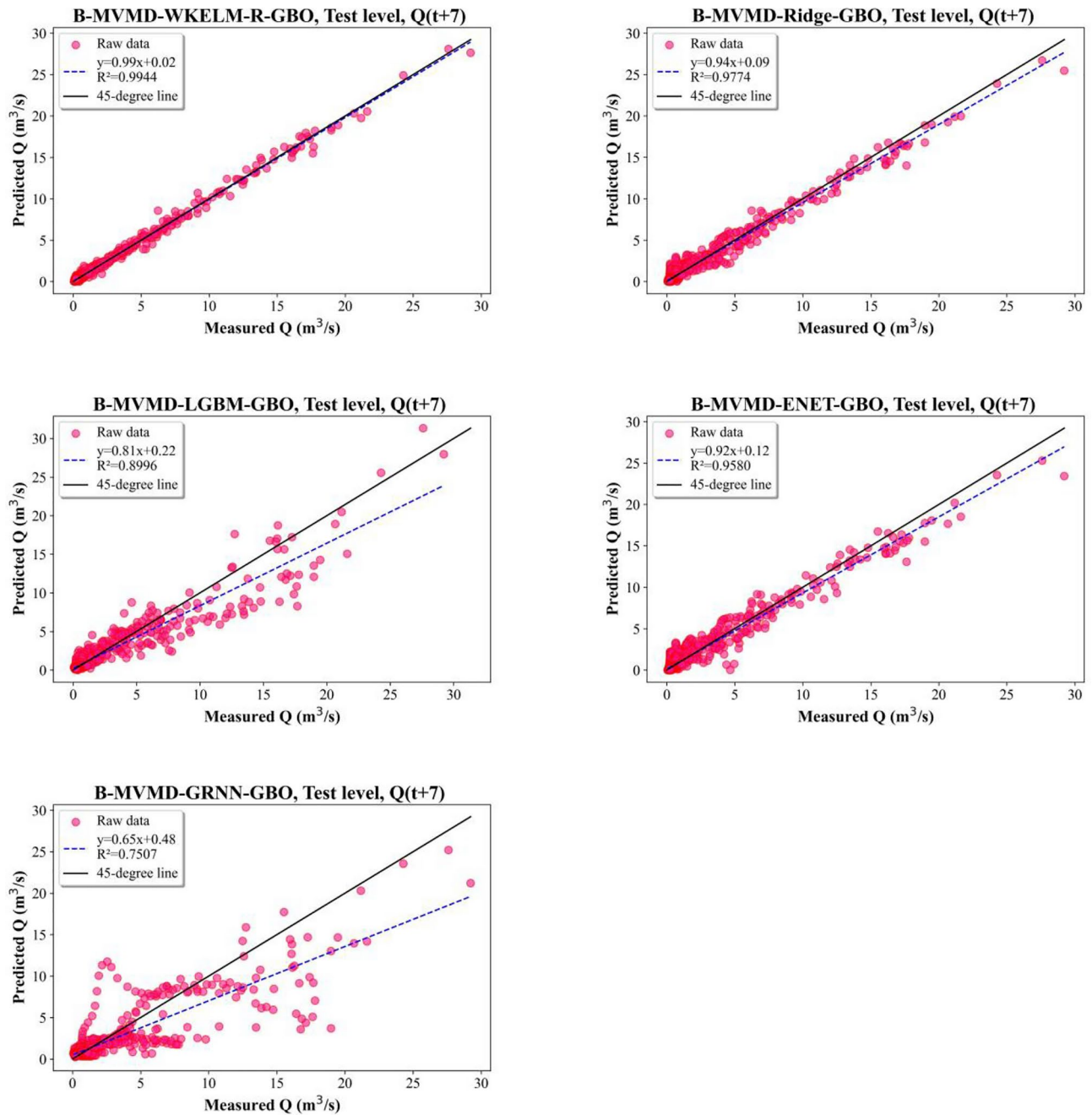


Fig. 6. Comparisons of the predicted and measured discharges (Q) of the Upper Turtle River at $(t+7)$ during the testing period for five hybrid ML models.

R-GBO model stood out as the leading choice, with a MARCOS score of 0.828, an ARAS score of 1, and a COPRAS score of 1. This again demonstrated its exceptional precision of operation. The B-MVMD-GRNN-GBO and B-MVMD-Ridge-GBO models were placed at the bottom of the list, showing that there was room for development. The B-MVMD-ENET-GBO model showed competitive performances, but it was inferior to the new proposed model. The B-MVMD-WKELM-R-GBO model maintained its dominance for $Q(t+7)$ (Fig. 13), further confirming its reliable performance. The B-MVMD-GRNN-GBO and B-MVMD-Ridge-GBO models exhibited better performances, indicating their capacities to adapt to the prediction patterns. Therefore, the highest MARCOS, ARAS, and COPRAS scores confirmed that the hybrid MVMD-based WKELM-R model had a higher level of accuracy in forecasting Q for $(t+3)$, $(t+7)$, and $(t+14)$ than other models.

Figure 15 displays the Taylor plots of the simulation results from the five hybrid ML models (B-MVMD-WKELM-R-GBO, B-MVMD-Ridge-GBO, B-MVMD-GRNN-GBO, B-MVMD-ENET-GBO, and B-MVMD-LGBM-GBO) for $(t+3)$, $(t+7)$, and $(t+14)$. The Taylor diagrams provide a detailed comparison of the observed and forecasted discharges using standard deviation and correlation coefficient. The proposed model showed a strong correlation with the reference Q , ranging from 0.97 to 0.995 at $t+3$, 0.99 to 0.999 at $t+7$, and 0.99 to

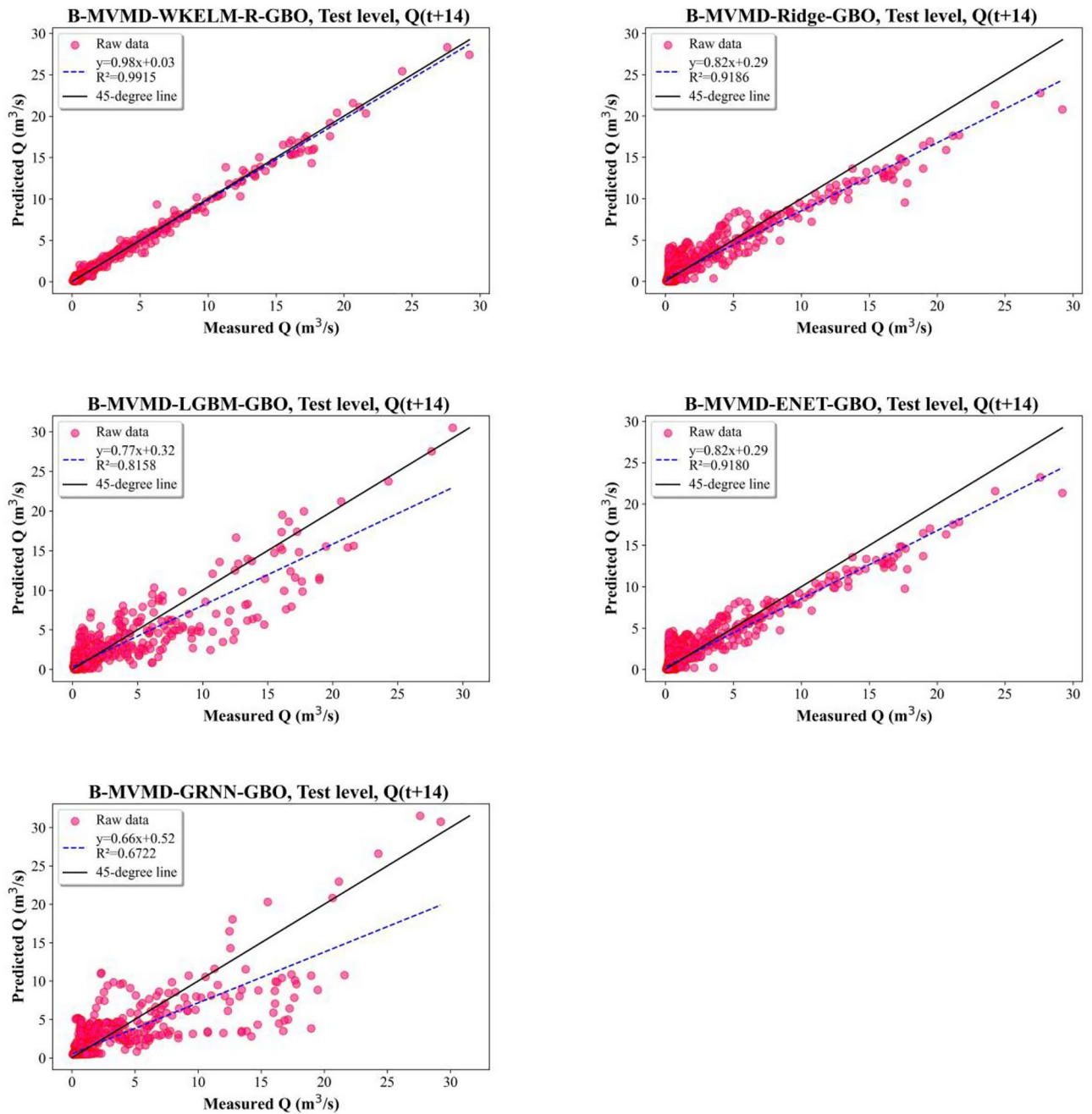


Fig. 7. Comparisons of the predicted and measured discharge (Q) of the Upper Turtle River at (t + 14) during the testing period for five hybrid ML models.

0.997 at t + 14. The standard deviation values for these correlations were between 0.30 and 0.35. The B-MVMD-Ridge-GBO, B-MVMD-GRNN-GBO, B-MVMD-ENET-GBO, and B-MVMD-LGBM-GBO models, while reasonably satisfactory, did not surpass the performance of the B-MVMD-WKELM-R-GBO model. Therefore, the predictions from the B-MVMD-WKELM-R-GBO model are more accurate than the other hybrid models for all time horizons.

Figures 16, 17 and 18 depict the variations in Q values predicted by the five ML models over the evaluation period. The B-MVMD-WKELM-R-GBO model exhibits considerable proficiency in forecasting daily discharge Q, with patterns closely aligned with the measured Q at (t + 3), (t + 7), and (t + 14). Therefore, the B-MVMD-WKELM-R-GBO model shows enhanced precision and reliability in characterizing the intricate temporal dynamics associated with daily discharge across all evaluated scenarios.

Discussion

The novel hybrid B-MVMD-WKELM-R-GBO model was developed and applied for forecasting daily streamflow of the Upper Turtle River in ND, USA based on hydrometeorological factors at time intervals of (t + 3), (t + 7),

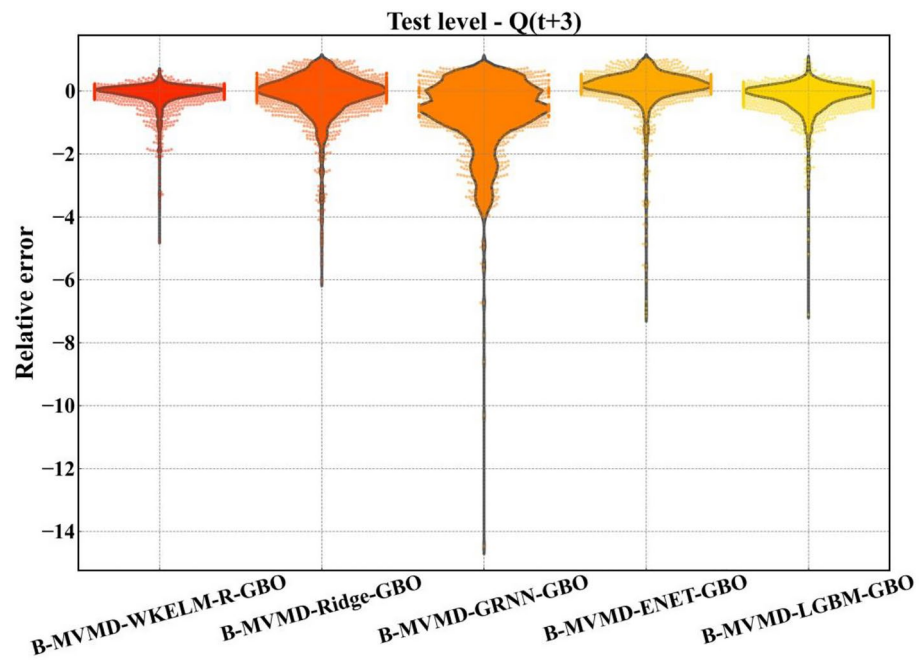


Fig. 8. Relative errors of discharges of the Upper Turtle River forecasted by five hybrid ML models for (t + 3) during the testing period.

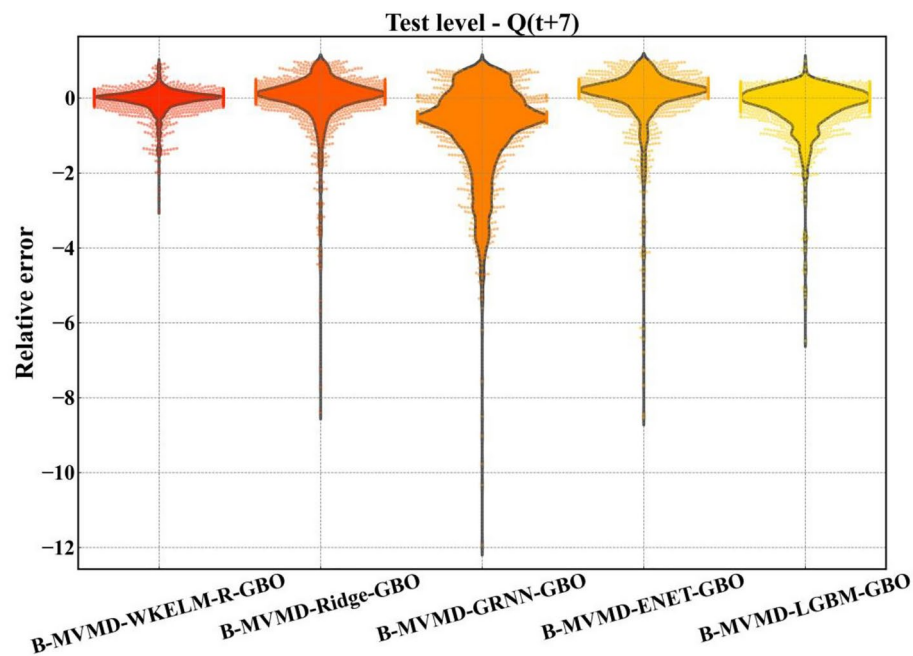


Fig. 9. Relative errors of discharges of the Upper Turtle River forecasted by five hybrid ML models for (t + 7) during the testing period.

and (t + 14). The evaluations revealed that all hybrid models outperformed the standalone models. Several other studies (e.g., Adnan et al.⁶⁷; Mostafa et al.⁶⁸; Adnan et al.⁶⁹; and Panahi et al.⁷⁰) also indicated that hybrid ML techniques can efficiently use the distinct benefits of multiple models to improve forecasting of high-dimensional time series. Furthermore, the comparative evaluation of the performances for the B-MVMD-WKELM-R-GBO model against both standalone ML models and other hybrid ML models (Tables 4 and 5) demonstrated its efficiency in streamflow forecasting.

This study employed different approaches to improve the precision of streamflow prediction models. The MVMD method can be a feasible solution for dealing with the multivariate oscillatory characteristics of input

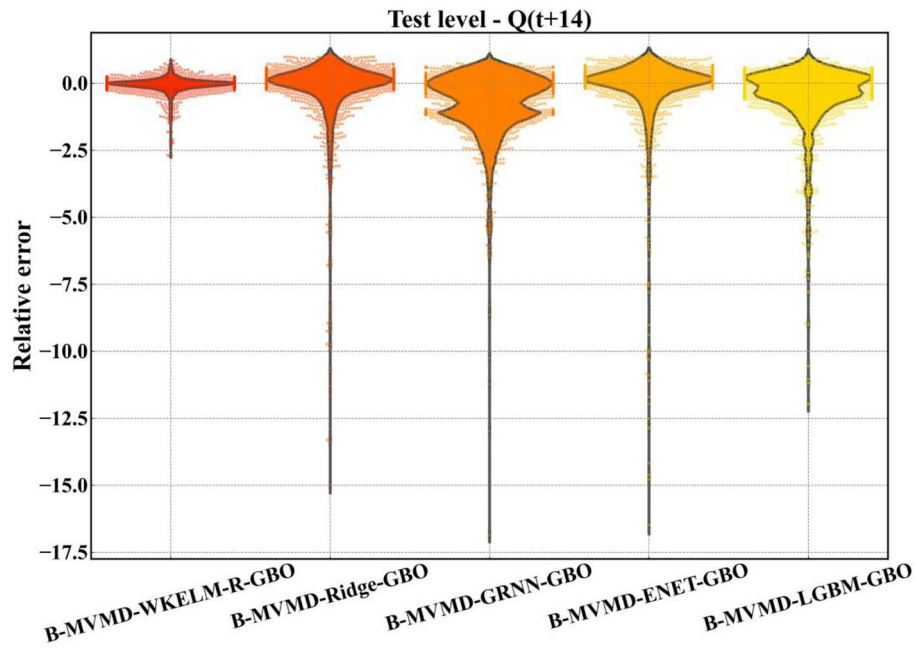


Fig. 10. Relative errors of discharges of the Upper Turtle River forecasted by five hybrid ML models for (t + 14) during the testing period.

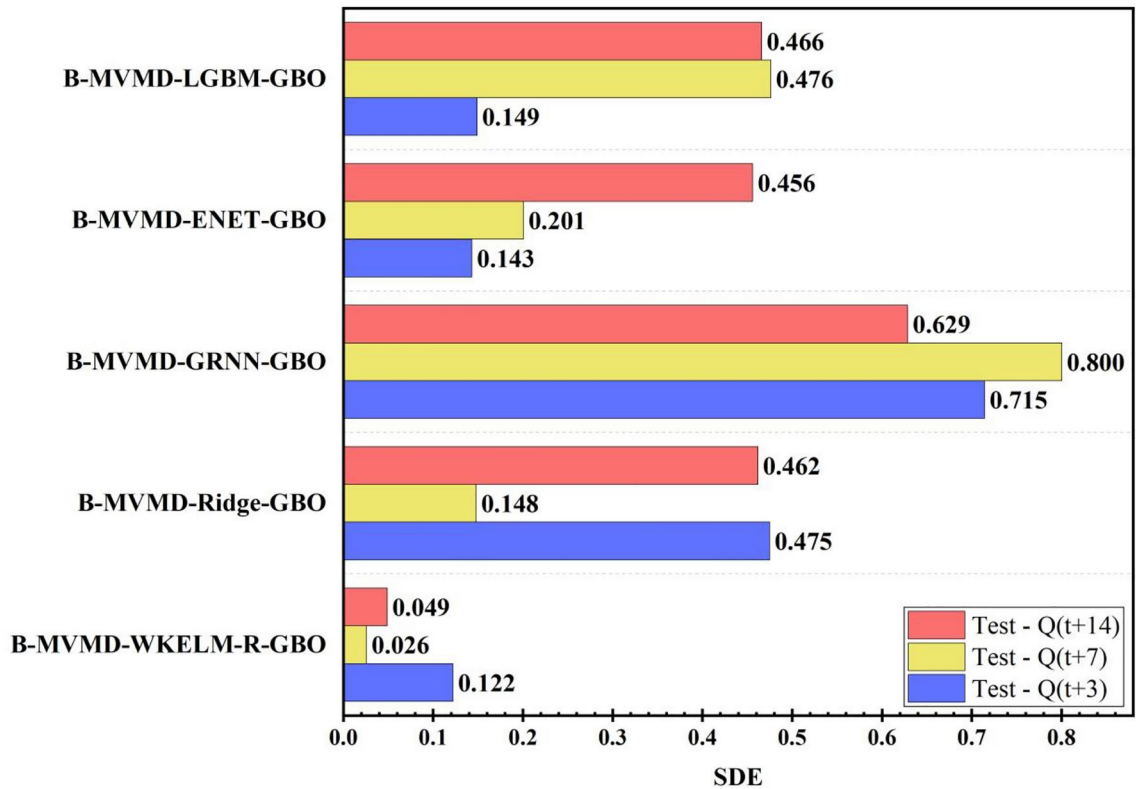


Fig. 11. Standard deviation errors (SDE) of the predictions from the five hybrid ML models during the testing period.

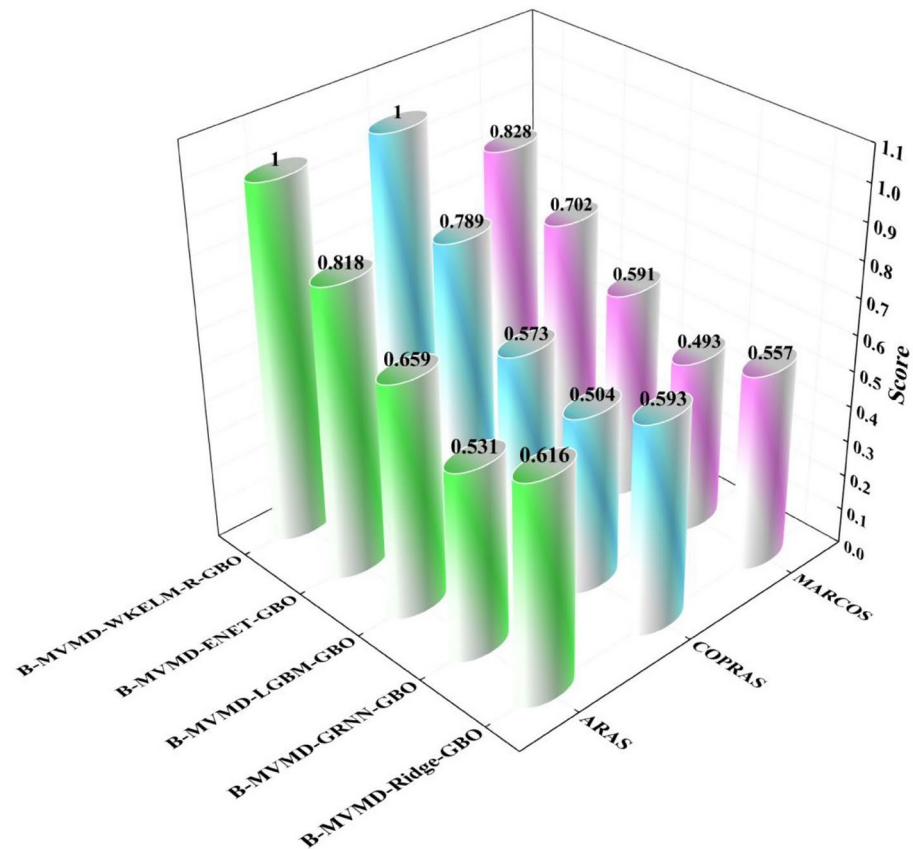


Fig. 12. Performance evaluation of five hybrid ML models for forecasting Q of the Upper Turtle River using MARCOS, ARAS, and COPRAS scores for $(t+3)$.

datasets as pointed out by Tao et al.⁷¹. It provides a distinct way to select adaptive mode parameters via scale segmentation. It improves the precision of the XGBoost technique by using multivariate modulated oscillations that share common frequency components across all of the data input channels. Several studies confirmed MVMD effectively and simultaneously captured both non-stationarity and non-linearity in multivariate data, allowing it to overcome mode mixing challenges. For instance, the research conducted by Fang, et al.⁷² on daily streamflow prediction using the MVMD-ensembled transformer model revealed that the use of MVMD for decomposing both input and target variables markedly improved model efficacy. This technique accurately reflects the complicated interdependence, hence enhancing the accuracy of streamflow forecasting.

Furthermore, using two decomposition approaches together can help improve the model performance by quantifying the complex relationships between the input characteristics, as indicated by Jamei et al.⁴¹ and Prasad et al.⁴². In addition, according to the results from this study, employing MVMD to decompose the data into various frequency components increased the dimensionality of the problem in relation to the number K . Therefore, feature selection approaches such as XGBoost can play a crucial role in improving prediction precision by selecting the most effective signals decomposed by MVMD and decreasing the mentioned dimensionality. In this study, XGBoost accurately assessed the significance of individual time-lagged data components, making it easier to choose a subset of the most important inputs for streamflow prediction.

At the first phase of model development, the framework enhanced the precision of the KELM model combined with LWLR. Then, WKELM was further integrated with RR using a linear correlation (Eq. 16) that reduced computational time and cost and prevented overfitting. In addition, the GBO method was applied to find optimal hyperparameters. Hence, the model's accuracy was improved by optimizing the ML model parameters to adapt iteratively to the unique features of datasets, resulting in a substantial enhancement of the model's predictive capabilities. Kadkhodazadeh and Farzin⁷³, Khozani et al.⁷⁴, and Adnan et al.⁷⁵ came to the same conclusion.

Therefore, the B-MVMD-WKELM-R-GBO model was identified as one of the most suitable options for streamflow forecasting after conducting a thorough investigation of statistical outcomes, comparative assessments, and model performances across several forecasting timeframes. For short-term forecasting ($Q(t+3)$), this model showed outstanding accuracy in both training and testing stages with high R and NSE and low $RMSE$ and $MAPE$. Although there were slight changes in $RMSE$, $MAPE$, and $U95\%$ throughout the testing, the B-MVMD-WKELM-R-GBO model outperformed other ML models, demonstrating its resilience and dependability. Furthermore, the model's efficacy was also demonstrated for medium-term ($t+7$) and long-term ($t+14$) forecasts compared to other simple and hybrid ML models. The scatter plots, violin plots, Taylor

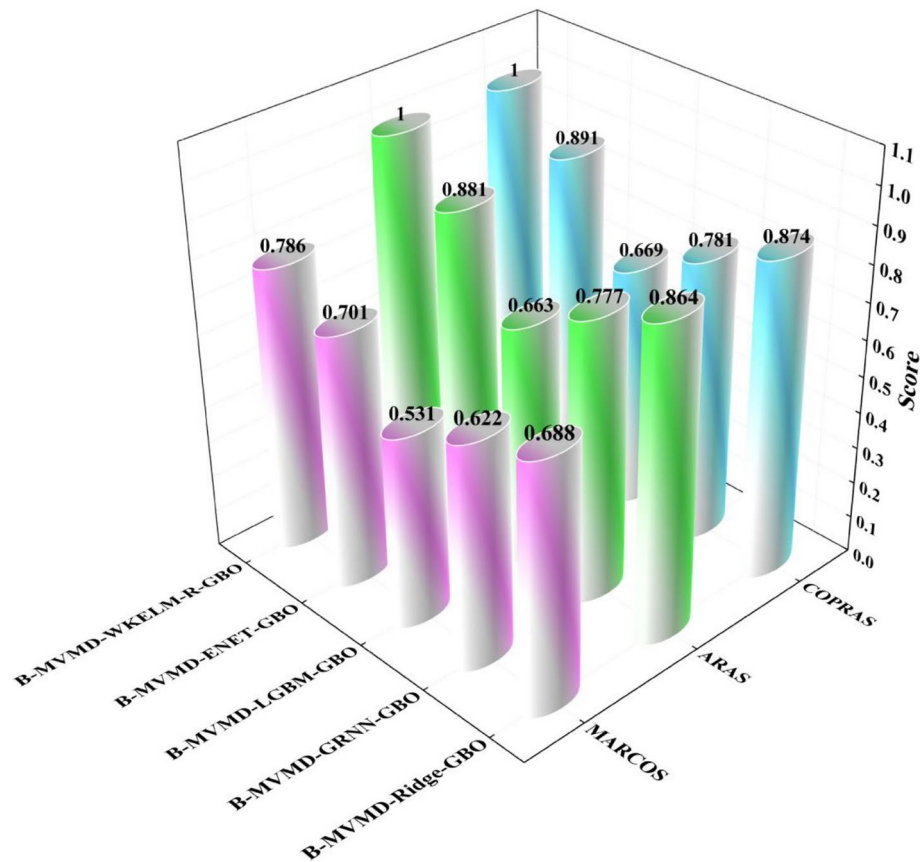


Fig. 13. Performance evaluation of five hybrid ML models for forecasting Q of the Upper Turtle River using MARCOS, ARAS, and COPRAS scores for $(t+7)$.

diagrams, as well as MCDM evaluation and standard deviation errors provided further evidence of the constant precision of the B-MVMD-WKELM-R-GBO model at various time intervals.

Conclusions

Streamflow forecasting is essential in hydrology for comprehending water availability and managing water resources. A novel hybrid B-MVMD-WKELM-R-GBO model, which incorporated MVMD, XGBoost, WKELM, RR model, and the GBO method, was developed and applied for predicting daily discharge of the Upper Turtle River watershed in North Dakota, USA. This study showcased the importance of feature selection in building precise forecasting models using XGBoost to find pertinent features and improve the input values. Particular efforts were made in this study to tackle the issues of non-stationarity, intricacy, and noisy input datasets by using the MVMD method to enhance the overall modeling accuracy. The findings demonstrated the dominance of the hybrid models, especially the new B-MVMD-WKELM-R-GBO model, in terms of precision, dependability, and resilience over several forecasting timeframes. The thorough statistical assessments confirmed the efficacy and reliability of the B-MVMD-WKELM-R-GBO model in capturing the underlying patterns and assuring accuracy in streamflow forecasts. The B-MVMD-WKELM-R-GBO modeling results demonstrated its superior precision based on various statistical metrics. In addition, the MCDM provided more understanding of the model performances, and the B-MVMD-WKELM-R-GBO model constantly received top rankings in terms of MARCOS, ARAS, and COPRAS scores. It is also demonstrated that MVMD efficiently handled non-stationarity and complexity and XGBoost notably decreased the dimension of the problem by selecting the optimal inputs. The results highlighted that the WKELM-R-GBO accurately analyzed the high-dimensional time series through the precise parameter optimization provided by GBO and combination of linear and nonlinear ML models.

While B-MVMD-WKELM-R-GBO showed impressive efficacy in streamflow prediction, its structure presented constraints in comprehending and validating the intricate connections between predictors and targets. Further research is needed for potential enhancements, such as combining ML models with numerical models. In addition, the B-MVMD-WKELM-R-GBO model has the potential to be diversified and can be improved by using approaches such as Bayesian model averaging⁷⁶ and bootstrapping models⁷⁷. Evaluating the new hybrid ML framework in various applications can potentially advance the related knowledge and modeling methodologies and also facilitate the development of more effective hydrological and agricultural management strategies under climate change scenarios.

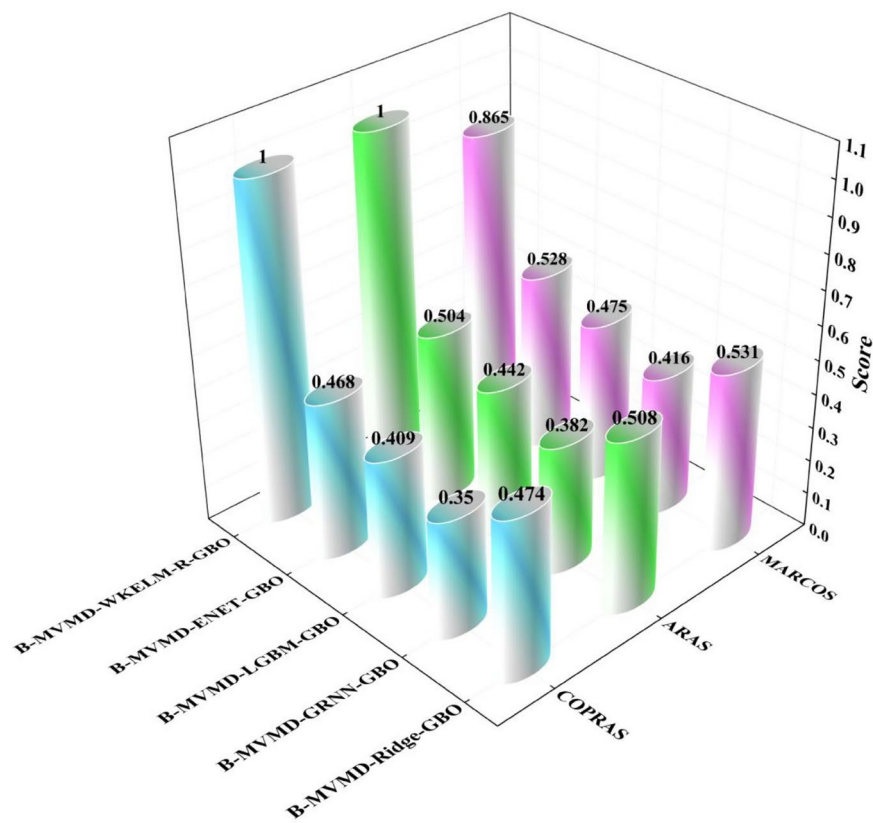


Fig. 14. Performance evaluation of five hybrid ML models for forecasting Q of the Upper Turtle River using MARCOS, ARAS, and COPRAS scores for (t+14).

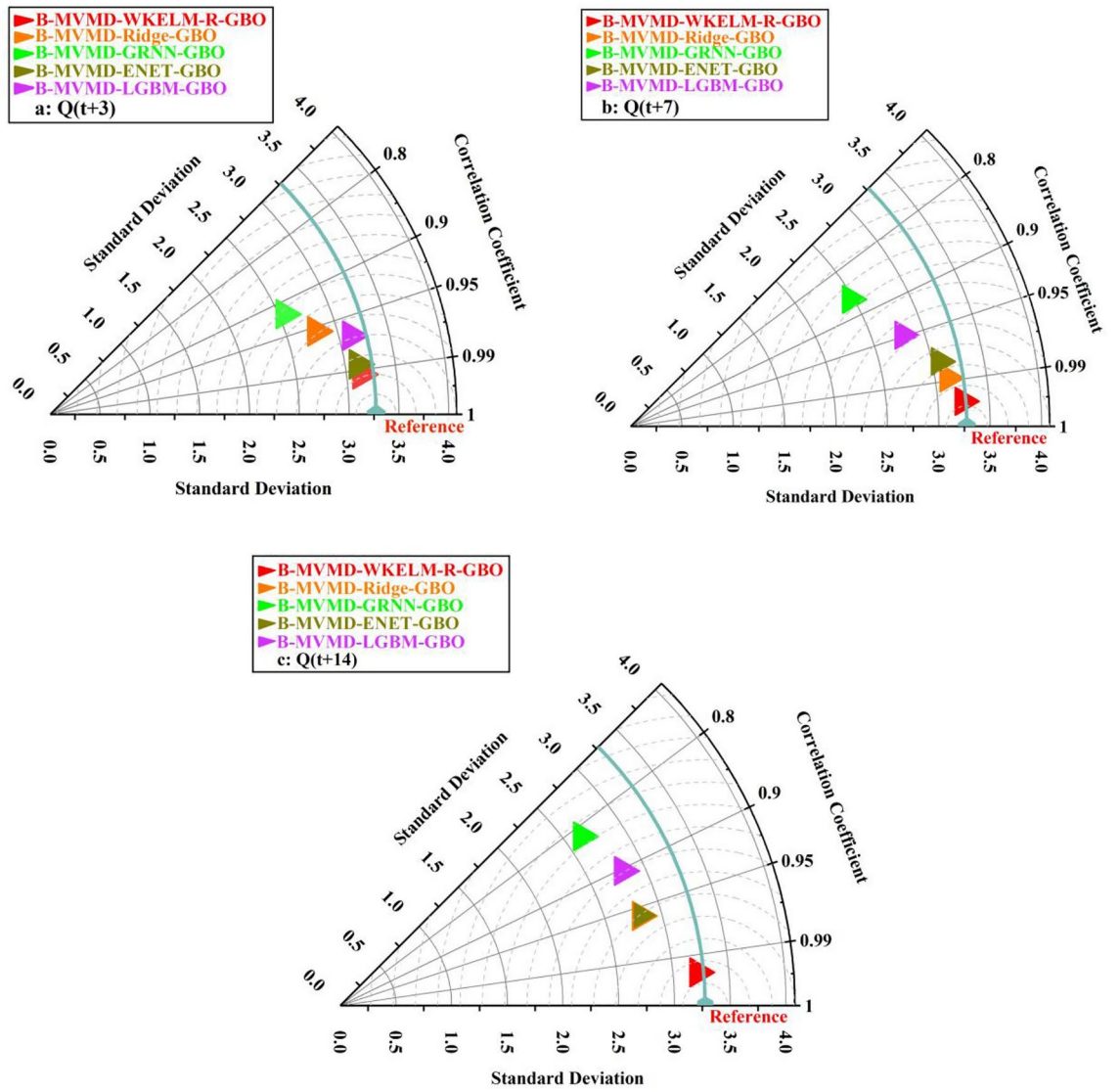


Fig. 15. Taylor diagrams of the simulations from five hybrid ML models for (a) Q (t + 3); (b) Q (t + 7); (c) Q (t + 14).

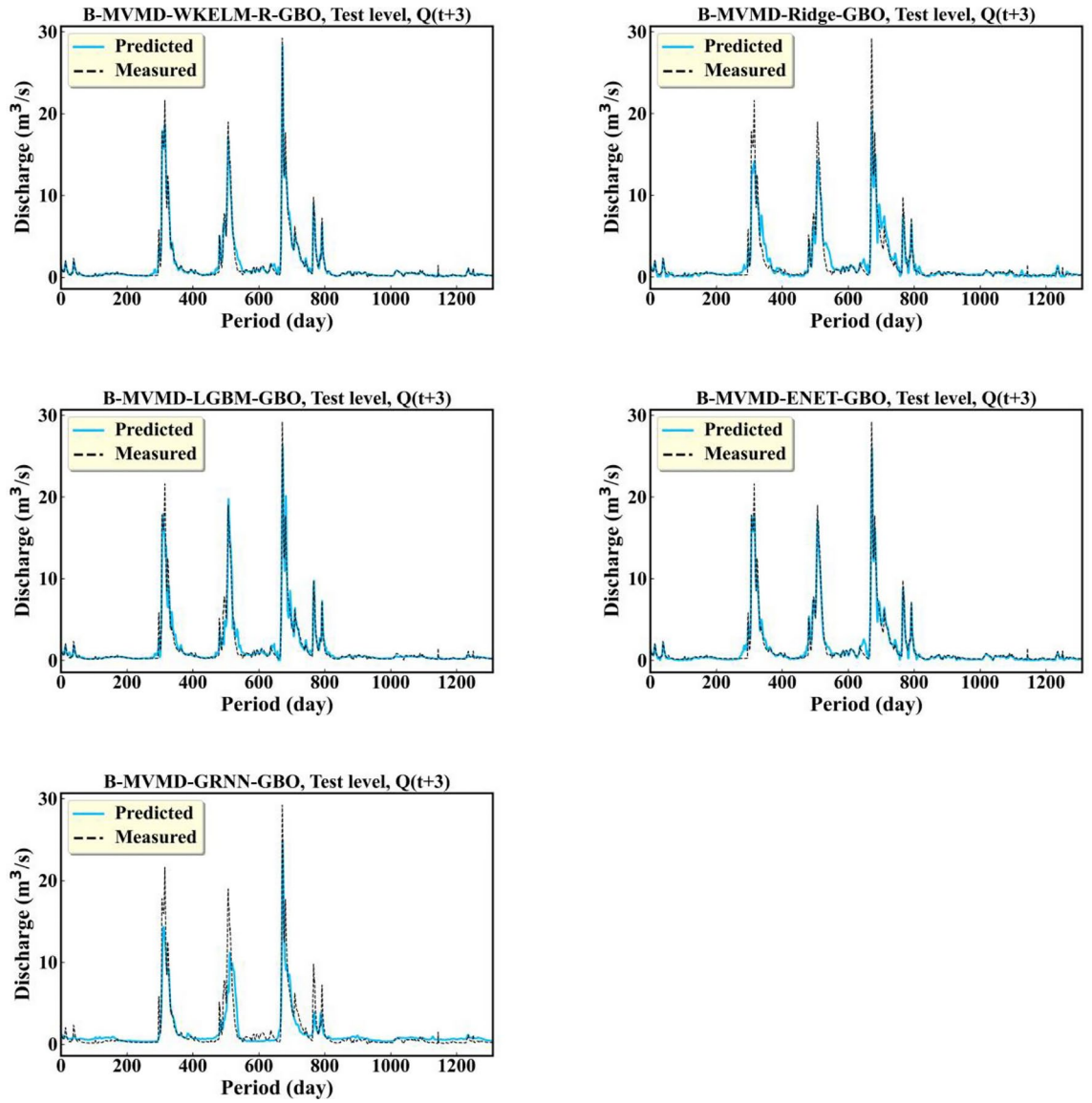


Fig. 16. Comparisons of the predicted and measured discharges (Q) of the Upper Turtle River at $(t+3)$ during the testing period for five hybrid ML models.

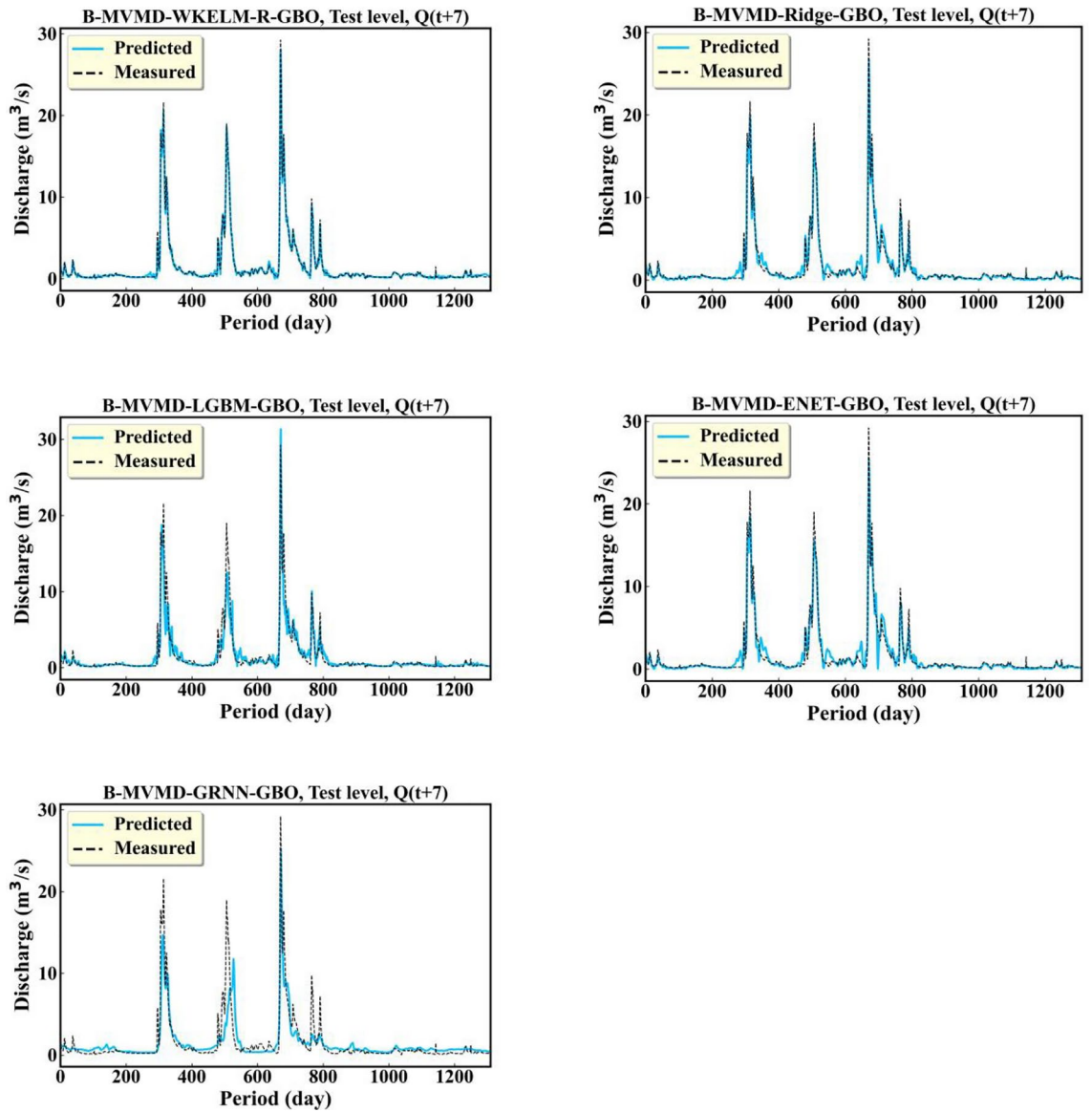


Fig. 17. Comparisons of the predicted and measured discharges (Q) of the Upper Turtle River at (t+7) during the testing period for five hybrid ML models.

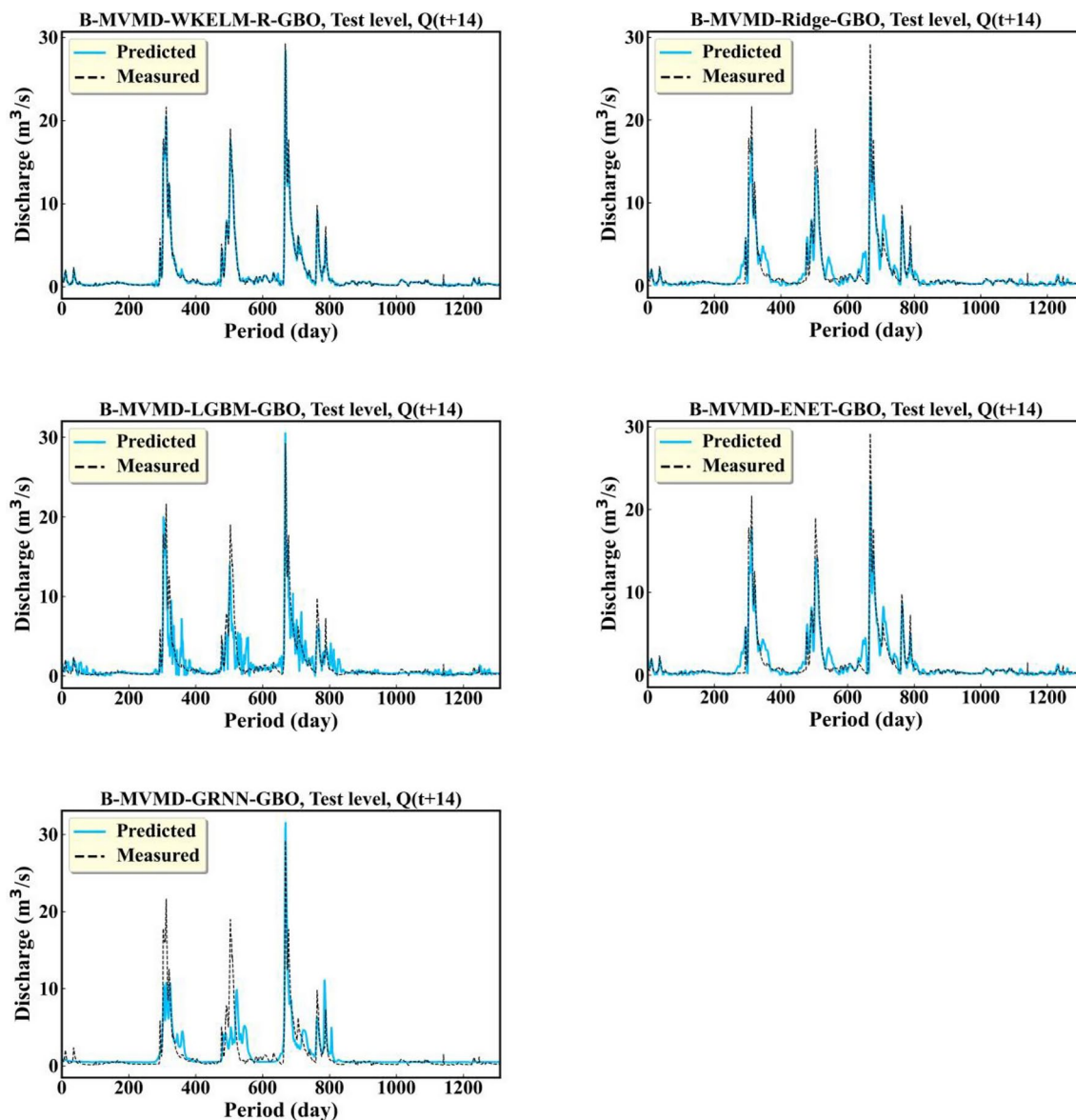


Fig. 18. Comparisons of the predicted and measured discharges (Q) of the Upper Turtle River at ($t+14$) during the testing period for five hybrid ML models.

Data availability

The datasets used in this study are available at the Parameter-elevation Regressions on Independent Slopes Model (PRISM; <https://prism.oregonstate.edu>), the Prediction of Worldwide Energy Resources (POWER) Data Access Viewer (<https://power.larc.nasa.gov/>), and the National Oceanic and Atmospheric Administration (NOAA; <https://www.noaa.gov/>). The discharge data are available at the USGS National Water Information System (<http://waterdata.usgs.gov/nwis/>). The modeling data generated in this study are available from the authors.

Received: 11 September 2024; Accepted: 27 November 2024

Published online: 28 December 2024

References

1. Bayazit, M. Nonstationarity of hydrological records and recent trends in trend analysis: A state-of-the-art review. *Environ. Process.* **2**, 527–542 (2015).
2. Ng, K. *et al.* A review of hybrid deep learning applications for streamflow forecasting. *J. Hydrol.* 130141 (2023).
3. Adnan, R. M. *et al.* Daily streamflow prediction using optimally pruned extreme learning machine. *J. Hydrol.* **577**, 123981 (2019).
4. Pandhiani, S. M., Sihag, P., Shabri, A. B., Singh, B. & Pham, Q. B. Time-series prediction of streamflows of Malaysian rivers using data-driven techniques. *J. Irrig. Drain. Eng.* **146**, 04020013 (2020).
5. Cirilo, J. A. *et al.* Development and application of a rainfall-runoff model for semi-arid regions. *Rbrh* **25** (2020).
6. Okkan, U. & Serbes, Z. A. Rainfall-runoff modeling using least squares support vector machines. *Environmetrics* **23**, 549–564 (2012).

7. Zhang, D. et al. Modeling and simulating of reservoir operation using the artificial neural network, support vector regression, deep learning algorithm. *J. Hydrol.* **565**, 720–736 (2018).
8. Liu, Z., Zhou, P., Chen, X. & Guan, Y. A multivariate conditional model for streamflow prediction and spatial precipitation refinement. *J. Geophys. Res. Atmos.* **120**, 10116–110129 (2015).
9. Yaseen, Z. M., Sulaiman, S. O., Deo, R. C. & Chau, K.-W. An enhanced extreme learning machine model for river flow forecasting: State-of-the-art, practical applications in water resource engineering area and future research direction. *J. Hydrol.* **569**, 387–408 (2019).
10. Jahangir, M. S., You, J. & Quilty, J. A quantile-based encoder-decoder framework for multi-step ahead runoff forecasting. *J. Hydrol.* **619**, 129269 (2023).
11. Ahmadi, F., Tohidi, M. & Sadrianzade, M. Streamflow prediction using a hybrid methodology based on variational mode decomposition (VMD) and machine learning approaches. *Appl. Water Sci.* **13**, 135 (2023).
12. Ibrahim, K. S. M. H., Huang, Y. F., Ahmed, A. N., Koo, C. H. & El-Shafie, A. A review of the hybrid artificial intelligence and optimization modelling of hydrological streamflow forecasting. *Alex. Eng. J.* **61**, 279–303 (2022).
13. Ghimire, S. et al. Streamflow prediction using an integrated methodology based on convolutional neural network and long short-term memory networks. *Sci. Rep.* **11**, 17497 (2021).
14. Meng, E. et al. A hybrid VMD-SVM model for practical streamflow prediction using an innovative input selection framework. *Water Resour. Manage.* **35**, 1321–1337 (2021).
15. Feng, Z.-K. et al. Monthly runoff time series prediction by variational mode decomposition and support vector machine based on quantum-behaved particle swarm optimization. *J. Hydrol.* **583**, 124627 (2020).
16. Asadi, S., Shahrabi, J., Abbaszadeh, P. & Tabanmehr, S. A new hybrid artificial neural networks for rainfall-runoff process modeling. *Neurocomputing* **121**, 470–480 (2013).
17. Li, X.-L., Lü, H., Horton, R., An, T. & Yu, Z. Real-time flood forecast using the coupling support vector machine and data assimilation method. *J. Hydroinf.* **16**, 973–988 (2014).
18. Feng, Z.-K., Niu, W.-J., Tang, Z.-Y., Xu, Y. & Zhang, H.-R. Evolutionary artificial intelligence model via cooperation search algorithm and extreme learning machine for multiple scales nonstationary hydrological time series prediction. *J. Hydrol.* **595**, 126062 (2021).
19. Zhang, Z. & Zhang, Z. Artificial neural network. In *Multivariate time series analysis in climate and environmental research*, 1–35 (2018).
20. Sebbar, A., Heddam, S. & Djemili, L. Kernel extreme learning machines (KELM): A new approach for modeling monthly evaporation (EP) from dams reservoirs. *Phys. Geogr.* **42**, 351–373 (2021).
21. El-Shafie, A. & Noureldin, A. Generalized versus non-generalized neural network model for multi-lead inflow forecasting at Aswan High Dam. *Hydrol. Earth Syst. Sci.* **15**, 841–858 (2011).
22. Yaseen, Z. M., Awadh, S. M., Sharafati, A. & Shahid, S. Complementary data-intelligence model for river flow simulation. *J. Hydrol.* **567**, 180–190 (2018).
23. Abozweita, O. A. et al. Enhancing hydrological predictions: optimised decision tree modelling for improved monthly inflow forecasting. *J. Hydroinf.* **2024**, 205 (2024).
24. Bai, X. et al. Explainable deep learning for efficient and robust pattern recognition: A survey of recent developments. *Pattern Recogn.* **120**, 108102 (2021).
25. Abbasi, M., Farokhnia, A., Bahreinimotlagh, M. & Roozbahani, R. A hybrid of Random Forest and Deep Auto-Encoder with support vector regression methods for accuracy improvement and uncertainty reduction of long-term streamflow prediction. *J. Hydrol.* **597**, 125717 (2021).
26. Xie, Y. et al. Stacking ensemble learning models for daily runoff prediction using 1D and 2D CNNs. *Expert Syst. Appl.* **217**, 119469 (2023).
27. Adnan, R. M., Keshtegar, B., Abusurrah, M., Kisi, O. & Alkabaa, A. S. Enhancing solar radiation prediction accuracy: A hybrid machine learning approach integrating response surface method and support vector regression. *Ain Shams Eng. J.* 103034 (2024).
28. Yue, Z., Ai, P., Yuan, D. & Xiong, C. Ensemble approach for mid-long term runoff forecasting using hybrid algorithms. *J. Ambient Intell. Hum. Comput.* **13**, 5103–5122 (2022).
29. Chang, L.-C., Shen, H.-Y. & Chang, F.-J. Regional flood inundation nowcast using hybrid SOM and dynamic neural networks. *J. Hydrol.* **519**, 476–489 (2014).
30. Dariane, A. & Azimi, S. Streamflow forecasting by combining neural networks and fuzzy models using advanced methods of input variable selection. *J. Hydroinf.* **20**, 520–532 (2018).
31. Tikhmarine, Y., Souag-Gamane, D., Ahmed, A. N., Kisi, O. & El-Shafie, A. Improving artificial intelligence models accuracy for monthly streamflow forecasting using grey Wolf optimization (GWO) algorithm. *J. Hydrol.* **582**, 124435 (2020).
32. Ramaswamy, V. & Saleh, F. Ensemble based forecasting and optimization framework to optimize releases from water supply reservoirs for flood control. *Water Resour. Manage.* **34**, 989–1004 (2020).
33. Ahmed, A. N. et al. A comprehensive comparison of recent developed meta-heuristic algorithms for streamflow time series forecasting problem. *Appl. Soft Comput.* **105**, 107282 (2021).
34. Adnan, R. M. et al. Enhancing accuracy of extreme learning machine in predicting river flow using improved reptile search algorithm. *Stoch. Environ. Res. Risk Assess.* **37**, 3063–3083 (2023).
35. Kilinc, H. C. et al. Daily scale river flow forecasting using hybrid gradient boosting model with genetic algorithm optimization. *Water Resour. Manage.* 1–16 (2023).
36. Momeneh, S. & Nourani, V. *Performance evaluation of artificial neural network model in hybrids with various preprocessors for river streamflow forecasting* (Ecosystems and Society, 2023).
37. Jamei, M. et al. Quantitative improvement of streamflow forecasting accuracy in the Atlantic zones of Canada based on hydro-meteorological signals: A multi-level advanced intelligent expert framework. *Ecol. Inf.* **80**, 102455 (2024).
38. Adnan, R. M. et al. Comparison of improved relevance vector machines for streamflow predictions. *J. Forecast.* **43**, 159–181 (2024).
39. Wang, M., Rezaie-Balf, M., Naganna, S. R. & Yaseen, Z. M. Sourcing CHIRPS precipitation data for streamflow forecasting using intrinsic time-scale decomposition based machine learning models. *Hydrol. Sci. J.* **66**, 1437–1456 (2021).
40. Ur Rehman, N. & Aftab, H. Multivariate variational mode decomposition. *IEEE Trans. Signal Process.* **67**, 6039–6052 (2019).
41. Jamei, M. et al. Forecasting daily flood water level using hybrid advanced machine learning based time-varying filtered empirical mode decomposition approach. *Water Resour. Manage.* **36**, 4637–4676 (2022).
42. Prasad, R., Ali, M., Xiang, Y. & Khan, H. A double decomposition-based modelling approach to forecast weekly solar radiation. *Renew. Energy* **152**, 9–22 (2020).
43. Sánchez-Marroño, N., Alonso-Betanzos, A. & Calvo-Estévez, R. M. In *International work-conference on artificial neural networks*. 456–463 (Springer).
44. Kursu, M. B. & Rudnicki, W. R. Feature selection with the Boruta package. *J. Stat. Softw.* **36**, 1–13 (2010).
45. Cao, P., Wang, H. & Zhou, K. Multichannel signal denoising using multivariate variational mode decomposition with subspace projection. *IEEE Access* **8**, 74039–74047 (2020).
46. Huang, G.-B., Zhou, H., Ding, X. & Zhang, R. Extreme learning machine for regression and multiclass classification. *IEEE Trans. Syst. Man Cybern. Part B (Cybernetics)* **42**, 513–529 (2011).
47. Gaspar, A., Oliva, D., Hinojosa, S., Aranguren, I. & Zaldivar, D. An optimized Kernel Extreme Learning Machine for the classification of the autism spectrum disorder by using gaze tracking images. *Appl. Soft Comput.* **120**, 108654 (2022).

48. Gan, L., Zhao, X., Wu, H. & Zhong, Z. Estimation of remaining fatigue life under two-step loading based on kernel-extreme learning machine. *Int. J. Fat.* **148**, 106190 (2021).
49. Huang, G.-B., Wang, D. H. & Lan, Y. Extreme learning machines: A survey. *Int. J. Mach. Learn. Cybern.* **2**, 107–122 (2011).
50. Huang, G.-B., Zhu, Q.-Y. & Siew, C.-K. Extreme learning machine: theory and applications. *Neurocomputing* **70**, 489–501 (2006).
51. Yan, Z., Huang, J. & Xiang, K. Kernel extreme learning machine optimized by the sparrow search algorithm for hyperspectral image classification. arXiv preprint. [arXiv:2204.00973](https://arxiv.org/abs/2204.00973) (2022).
52. Zhou, Y., Peng, J. & Chen, C. P. Extreme learning machine with composite kernels for hyperspectral image classification. *IEEE J. Select. Top. Appl. Earth Observ. Remote Sens.* **8**, 2351–2360 (2014).
53. Ding, S., Zhang, Y., Xu, X. & Bao, L. A novel extreme learning machine based on hybrid kernel function. *J. Comput.* **8**, 2110–2117 (2013).
54. Avci, D. & Dogantekin, A. An expert diagnosis system for parkinson disease based on genetic algorithm-wavelet kernel-extreme learning machine. *Parkinson's Dis.* **2016**, 5264743 (2016).
55. Chen, H., Ahmadianfar, I., Liang, G. & Heidari, A. A. Robust kernel extreme learning machines with weighted mean of vectors and variational mode decomposition for forecasting total dissolved solids. *Eng. Appl. Artif. Intell.* **133**, 108587 (2024).
56. Cai, Z. et al. Evolving an optimal kernel extreme learning machine by using an enhanced grey wolf optimization strategy. *Expert Syst. Appl.* **138**, 112814 (2019).
57. Atkeson, C. G., Moore, A. W. & Schaal, S. Locally weighted learning for control. *Lazy Learn.* 75–113 (1997).
58. Kisi, O. & Ozkan, C. A new approach for modeling sediment-discharge relationship: Local weighted linear regression. *Water Resour. Manag.* **31**, 1–23 (2017).
59. Zhang, X., Deng, X. & Wang, P. Double-level locally weighted extreme learning machine for soft sensor modeling of complex nonlinear industrial processes. *IEEE Sensors J.* **21**, 1897–1905 (2020).
60. Rencher, A. C. & Schaalje, G. B. *Linear models in statistics.* (John Wiley & Sons, 2008).
61. Ahmadianfar, I., Jamei, M. & Chu, X. A novel hybrid wavelet-locally weighted linear regression (W-LWLR) model for electrical conductivity (EC) prediction in surface water. *J. Contam. Hydrol.* **232**, 103641 (2020).
62. Hoerl, A. E. & Kennard, R. W. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **12**, 55–67 (1970).
63. Ahmadianfar, I., Bozorg-Haddad, O. & Chu, X. Gradient-based optimizer: A new metaheuristic optimization algorithm. *Inf. Sci.* **540**, 131–159 (2020).
64. Premkumar, M., Jangir, P. & Sowmya, R. MOGBO: A new multiobjective gradient-based optimizer for real-world structural optimization problems. *Knowl. Based Syst.* **218**, 106856 (2021).
65. Rezk, H. et al. Optimal parameter estimation strategy of PEM fuel cell using gradient-based optimizer. *Energy* **239**, 122096 (2022).
66. Li, L.-L. et al. Optimization and performance assessment of solar-assisted combined cooling, heating and power system systems: Multi-objective gradient-based optimizer. *Energy* **289**, 129784 (2024).
67. Adnan, R. M. et al. Pan evaporation estimation by relevance vector machine tuned with new metaheuristic algorithms using limited climatic data. *Eng. Appl. Comput. Fluid Mech.* **17**, 2192258 (2023).
68. Mostafa, R. R., Kisi, O., Adnan, R. M., Sadeghifar, T. & Kuriqi, A. Modeling potential evapotranspiration by improved machine learning methods using limited climatic data. *Water* **15**, 486 (2023).
69. Adnan, R. M. et al. Estimating reference evapotranspiration using hybrid adaptive fuzzy inferencing coupled with heuristic algorithms. *Comput. Electron. Agric.* **191**, 106541 (2021).
70. Panahi, F. et al. Streamflow prediction with large climate indices using several hybrid multilayer perceptrons and copula Bayesian model averaging. *Ecol. Indic.* **133**, 108285 (2021).
71. Tao, H. et al. PM2.5 concentration forecasting: Development of integrated multivariate variational mode decomposition with kernel Ridge regression and weighted mean of vectors optimization. *Atmos. Pollut. Res.* **15**, 102125 (2024).
72. Fang, J. et al. Ensemble learning using multivariate variational mode decomposition based on the Transformer for multi-step-ahead streamflow forecasting. *J. Hydrol.* **636**, 131275 (2024).
73. Kadhodazadeh, M. & Farzin, S. A novel hybrid framework based on the ANFIS, discrete wavelet transform, and optimization algorithm for the estimation of water quality parameters. *J. Water Clim. Change* **13**, 2940–2961 (2022).
74. Khozani, Z. S., Banadkooki, F. B., Ehteram, M., Ahmed, A. N. & El-Shafie, A. Combining autoregressive integrated moving average with Long Short-Term Memory neural network and optimisation algorithms for predicting ground water level. *J. Clean. Prod.* **348**, 131224 (2022).
75. Adnan, R. M. et al. Development of new machine learning model for streamflow prediction: Case studies in Pakistan. *Stoch. Environ. Res. Risk Assess.* 1–35 (2022).
76. Raftery, A. E., Madigan, D. & Hoeting, J. A. Bayesian model averaging for linear regression models. *J. Am. Stat. Assoc.* **92**, 179–191 (1997).
77. Freedman, D. A. Bootstrapping regression models. *Ann. Stat.* **9**, 1218–1228 (1981).

Acknowledgements

The North Dakota Water Resources Research Institute provided partial financial support in the form of a graduate fellowship for the first author. In addition, this research is based upon work supported by the United States Environmental Protection Agency under Grant No. CD-95811400.

Author contributions

Arvin Samadi-Koucheksaraee: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Visualization. Xuefeng Chu: Conceptualization, Methodology, Investigation, Writing – review & editing, Data curation, Supervision, Project administration, Funding acquisition.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to X.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024