



Surface water sodium (Na^+) concentration prediction using hybrid weighted exponential regression model with gradient-based optimization

Iman Ahmadianfar¹ · Seyedehelham Shirvani-Hosseini² · Arvin Samadi-Koucheksaraee¹ · Zaher Mundher Yaseen^{3,4,5} 

Received: 24 November 2021 / Accepted: 15 February 2022 / Published online: 14 March 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

Undeniably, there is a link between water resources and people's lives and, consequently, economic development, which makes them vital in health and the environment. Proper water quality forecasting time series has a crucial role in giving on-time warnings for water pollution and supporting the decision-making of water resource management. The principal aim of this study is to develop a novel and cutting-edge ensemble data intelligence model named the weighted exponential regression and hybridized by gradient-based optimization (WER-GBO). Indeed, this is to reach more meticulous sodium (Na^+) prediction monthly at Maroon River in the southwest of Iran. This developed model has advantages over other previous methodologies thanks to the following merits: (i) it can improve the performance and ability by mixing the outputs of four distinct data intelligence (DI) models, i.e., adaptive neuro-fuzzy inference system (ANFIS), least square support vector regression (LSSVM), Bayesian linear regression (BLR), and response surface regression (RSR); (ii) the proposed model can employ a Cauchy weighted function combined with an exponential-based regression model being optimized by GBO algorithm. To evaluate the performance of these models, diverse statistical indices and graphical assessment including error distributions, box plots, scatter-plots with confidence bounds and Taylor diagrams were conducted. According to obtained statistical metrics and verified validation procedures, the proposed WER-GBO resulted in promising accuracy compared to other models. Furthermore, the outcomes revealed the WER-GBO ($R = 0.9712$, $RMSE = 0.639$, and $KGE = 0.948$) reached more accurate and reliable results than other methods such as the ANFIS, LSSVM, BLR, and RSR for Na prediction in this study. Hence, the WER-GBO model can be considered a constructive technique to forecast the water quality parameters.

Keywords Water quality · Wavelet transform · Weighted exponential regression · Gradient-based optimization · Bayesian linear regression

Responsible Editor: Marcus Schulz

✉ Zaher Mundher Yaseen
zaheryaseen88@gmail.com

Iman Ahmadianfar
i.ahmadianfar@bkatu.ac.ir

Seyedehelham Shirvani-Hosseini
s.elham.shirvanihosseini@gmail.com

Arvin Samadi-Koucheksaraee
arvinsamadi.k@gmail.com

¹ Department of Civil Engineering, Behbahan Khatam Alanbia University of Technology, Behbahan, Iran

² Department of Chemical Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran

³ Adjunct Research Fellow, USQ's Advanced Data Analytics Research Group, School of Mathematics Physics and Computing, University of Southern Queensland, QLD 4350 Queensland, Australia

⁴ New Era and Development in Civil Engineering Research Group, Scientific Research Center, Al-Ayen University, Thi-Qar 64001, Iraq

⁵ Institute for Big Data Analytics and Artificial Intelligence (IBDAAI), Kompleks Al-Khawarizmi, Universiti Teknologi MARA, Shah Alam, Selangor 40450, Malaysia

Introduction

One of the serious problems attracting much attention recently is the issue of water environment pollution (Uddin et al. 2021). Owing to the importance of management and prediction of water quality (WQ) in a real-life water environment, researchers have developed numerous parameters (physical, biological, and chemical) that serve as the yardstick for the evaluation of water pollution (Vasistha and Ganguly 2020). The role of such parameters is to capture the balance between the rate of atmospheric exchange, oxygen consumption processes (such as nitrification, chemical oxidation, and aerobic respiration), and oxygen production processes (such as photosynthesis). Hence, there is a need to come up with a method for the accurate prediction of water quality (Haghiabi et al. 2018), because an accurate water quality prediction has a profound impact on the water management (Tao et al. 2019). Being that most of the available indicators of water quality normally vary with time, decision-makers can only rely on accurate water quality prediction in making accurate water environment management decisions that will ensure the maintenance of water quality values within an acceptable limit (Tiyasha et al. 2021a, b).

Data intelligence (DI) models have found increasing application in environmental engineering (Chen et al. 2008; Yetilmezsoy et al. 2011; Chakraborty et al. 2019; Hou et al. 2020; Al-Sulttani et al. 2021; Khan et al. 2021; Song et al. 2021). They have specifically been applied in DI-driven techniques in climate, morphology, hydrology, geology, and water chemical-related problems (Maier et al. 2014; Nourani et al. 2014; Yaseen et al. 2019; Li et al. 2021a). DI techniques have been successfully applied in the development of different learning models in the field of environmental monitoring, especially in the monitoring of hydrological parameters (Feng et al. 2021; Tiyasha et al. 2021a, b). In the study by (Fan et al. 2018), authors reported the real-time application of an intelligent water regimen monitoring system that could improve working efficiency. The performance of artificial neural network (ANN) integrated with wavelet transformation (WT) was tested for prediction the electrical conductivity (EC) by (Ravansalar and Rajaei 2015). A hybrid approach that relies on training an ANN model utilizing the multi-objective genetic algorithm (MOGA) was developed by Chatterjee et al. (2017) for improved WQ prediction accuracy (Chatterjee et al. 2017). The suggested model's performance was compared to that of the ANN-GA, the improved ANN model with particle swarm optimization method (ANN-PSO), and the SVM, and the findings revealed that the proposed model outperformed the benchmarking techniques. A hybrid model termed genetic algorithm (GA) and LSSVM (GA-LSSVM) was

developed by Bozorg-haddad et al. (2016) to predict water quality characteristics (Bozorg-Haddad et al. 2017). They conducted a comparison between the suggested model and genetic programming (GP). Na^+ , K^+ , Mg^+ , SO_4 and Cl dissolved solids (TDS) were modeled using a GA-LSSVR algorithm and a GP technique in the Sefidrood River in Iran. The findings show that the GA-LSSVR algorithm outperforms the GP algorithm when it comes to predicting water quality metrics. In another research, EC was predicted using coupled wavelet extreme learning machine (W-ELM) (Barzegar et al. 2018). Abobakr Yahya et al. (2019) revealed that an effective model for the estimation of WQ in the Ungauged River Catchment under Dual Scenarios had been developed (Abobakr Yahya et al. 2019). With modest prediction errors, their model could accurately forecast the WQ. Total dissolved solids were predicted using coupled multigene genetic programming (MGGP) with wavelet data pre-processing (Jamei et al. 2020). For the hybrid RNNs-DS model, Li et al. (2021a, b) merged recurrent neural networks (RNNs) with modified Dempster/Shافر (D-S) evidence theory to produce an RNNs-DS model that was both accurate and efficient (Li et al. 2021b). An array of ensemble machine learning models, comprising quantile regression forest (QRF), random forest (RF), radial support vector machine (SVM), stochastically gradient boosting (SGB), and gradient boosting machines (GBM), were used by Al-Sulttani et al. (2021) (Al-Sulttani et al. 2021) to estimate the amount of biochemical oxygen demand (BOD) in the Euphrates River, Iraq. Two separate feature extraction approaches, genetic algorithm (GA) and principal components analysis (PCA), were used to create integrative models. This investigation shows that the PCA-QRF model identified in research outperformed the solo methods and the GA-integrated models in this study. To estimate the BOD in the Haihe River Basin in China, Song et al. (2021) (Song et al. 2021) developed a novel hybrid model, the improved sparrow search algorithm (ISSA), which combined Cauchy mutation and opposition-based learning (OBL) with the long short-term memory (LSTM). The findings reveal that the addressing model outperforms the competitor models in terms of prediction accuracy.

One of the non-renewable resources affecting human life greatly is surface water owing to its importance to human existence (Asadollah et al. 2020; Tiyasha et al. 2021a, b). However, the rapid increase in economic development has negatively impacted the quality of surface water in some developing countries, and the sustained deterioration in surface water quality increases water stress and water demand, with consequential threats to human health in such regions (Tabari et al. 2011; Uprety et al. 2020; Wu et al. 2020; Dai et al. 2021; van Vliet et al. 2021). Hence, precise and timely

monitoring and prediction of water quality are greatly important to avert threats to human life. Although several review articles on the prediction of water quality parameters using DI models have been published (Chen et al. 2020; Rajaei et al. 2020; Giri 2021; Tiyasha et al. 2021a, b), some limitations of the existing models as reported in the literature gives rise to the development of new DI models using different water quality parameters to ensure prediction accuracy improvement. The environmental protection agencies must take certain rapid measures when alerted on impending deterioration in water quality. However, these agencies mostly rely on old models for water quality prediction, and the low precision of these old water quality prediction models may limit their future application; the availability of few training data in the literature may also limit the precision of these learning models even though there are still uncertainties regarding the relationship between the amount of training data and the performance of surface water quality prediction models.

The performance of DI models in water quality prediction is not usually only a function of the applied models and the amount of dataset, but also it is determined by the selected water parameters for the training of the learning models (Tiyasha et al. 2020; Abba et al. 2020). Hence, the most important water parameters for the training of the learning models must be identified and selected without sacrificing the predictive performance of the model; this could significantly improve the prediction efficiency and reduce the overall prediction cost (Yaseen et al. 2018; Tian et al. 2022). Nevertheless, there are still studies on the evaluation of the performance of learning models in surface water quality prediction to identify the most important water parameters from large training datasets.

The majority of DI models have achieved significant results as each model has its advantages and disadvantages. In order to deal with the concerns of each model or the advantages of all models, ensemble models are attaining more attention (Jamei et al. 2021b; Pandey et al. 2021). The main goal of this research is to develop ensemble models using artificial intelligence (AI) (Ahmadianfar et al. 2021, 2022) techniques to forecast sodium (Na) in the Tange-Takab station, in the south of Iran. The proposed model is based on a weighted exponential regression optimized by gradient-based optimization (WER-GBO). To the best of the authors' knowledge, there are no published studies demonstrating the development of the WER-GBO model in the field of modeling, estimating, and predicting. The goal was obtained in three main steps: (1) selection of the best combination of input variables through the best subset analysis, (2) using four DI models (i.e., adaptive neuro-fuzzy

inference system (ANFIS), least square support vector machine (LSSVM), Bayesian linear regression (BLR), and response surface regression (RSR)) to predict the Na and finally development of an ensemble model (i.e., WER-GBO) to boost the performance of the single DI models in forecasting the Na.

The current research objectives are (i) to develop a robust and accurate predictive model based on the hybridization weighted exponential regression model using gradient-based optimization for monthly records of sodium (Na) at Maroon River in the southwest of Iran and four DI models (i.e., ANFIS, LSSVM, RSR, and BLR), (ii) to introduce the data preprocessing and best input variables selection, (iii) to compare results by using the standalone DI models, wavelet-based DI models, and the proposed ensemble model, and (iv) finally, to describe the conclusion.

Methodology

Weighted exponential regression (WER)

WER is a weighted regression model based on the exponential function. The proposed WER uses a Cauchy function as a weighting function to increase the accuracy of the regression model's results. In addition, to show the ability of the WER model, predicting the Na is also modeled by a linear model called WLR. The WER and WLR models for the Na prediction are formulated as follows,

$$y_{WER} = w_1.X_1^{\beta_1} + w_2.X_1^{\beta_2} + w_3.X_1^{\beta_3} + w_4.X_1^{\beta_4} \quad (1)$$

$$y_{WLR} = w_1.X_1 + w_2.X_2 + w_3.X_3 + w_4.X_4 \quad (2)$$

where y_{WER} and y_{WLR} are the outputs of WER and WLR models. β_1 , β_2 , β_3 , and β_4 are the tuning parameters of WER model. w_1 , w_2 , w_3 , and w_4 are the weighting functions based on the Cauchy function, which is formulated as,

$$w = \frac{a}{\left(b + c.\|x_i - x_j\|^2\right)} \quad (3)$$

The gradient-based optimization algorithm is developed to extract optimal amounts of main unknown parameters of these two models based on measured data to predict the Na.

Accordingly, the following stages are implemented to derive optimal amounts of unknown coefficients of two models:

- (1) All datasets consisting of input and output parameters (i.e., lag-time of discharge (Q) and Na) are collected.

- (2) The collected datasets from 1980 to 2016 are utilized in the GBO to determine the best values of tuning parameters of two models. The main criterion to choose the best values for parameters of two models is the minimum objective functions, defined as,

$$\text{Minimize } g(x) = \sum_{i=1}^N (Na_{M,i} - Na_{DI,i})^2 \tag{4}$$

where $Na_{M,i}$ and $Na_{DI,i}$ are the measured and DI models, N denotes the number of the dataset.

- (3) Optimal amounts of the main parameters of each model are selected based on stage (2).

Gradient-based optimization (GBO)

GBO is a math-based optimization algorithm developed based on Newton’s method (Özban 2004) to solve complex optimization problems (Ahmadianfar et al. 2020a). The method consists of two operators: (1) gradient search rule (GSR) and (2) local escaping operator (LEO). The main stages of GBO are explained in the following parts.

Initialization

GBO is initialized by creating a set of random solutions (population) with the size of N and D dimension, which is expressed as,

$$x_m = LS + rand \times (US - LS) \tag{5}$$

where LS and US are the lower and upper limits of the problem space, and $rand$ has a random value in $[0, 1]$.

Gradient search rule (GSR)

GSR is a search mechanism in the GBO to find the promising regions in the solution space. The mechanism is acquired by Newton’s method (Özban 2004). The GSR is expressed as,

$$GSR = randn.R_1 \cdot \frac{2\Delta x \times x_m}{(dp_m - dq_m + \epsilon)} \tag{6}$$

in which

$$dp_m = rand \cdot \left(\frac{[u_{m+1} + x_m]}{2} + rand \cdot \Delta x \right) \tag{7}$$

$$dq_m = rand \cdot \left(\frac{[u_{m+1} + x_m]}{2} - rand \cdot \Delta x \right) \tag{8}$$

$$u_{m+1} = x_m - randn \cdot \frac{2\Delta x \cdot x_m}{(x_w - x_b + \epsilon)} + GD \tag{9}$$

where $randn$ has a random value with a normal distribution. ϵ has a small value in $[0, 0.1]$. x_w and x_b are the worst and best positions. R_1 denotes a weighting factor. GM means the gradient direction to improve the exploitation. The GD is expressed as,

$$GD = rand.R_2 \times (x_b - x_m^{it}) \tag{10}$$

In Eq. (6), Δx is a differential vector, which is expressed as,

$$\Delta x = rand \cdot |S| \tag{11}$$

$$S = \frac{(x_b - x_{a1}^{it}) + \mu}{2} \tag{11-1}$$

$$\mu = 2 \cdot rand \cdot \left[\left| \frac{x_{a1}^{it} + x_{a2}^{it} + x_{a3}^{it} + x_{a4}^{it}}{4} - x_m^{it} \right| \right] \tag{11-2}$$

where $a1, a2, a3,$ and $a4 (a1 \neq a2 \neq a3 \neq a4 \neq m)$ are the four different integers values, randomly chosen from $[1, N]$, and S is the step size.

Calculate the new solutions

To generate the new solution ($Z1_m^{it}$) in the GBO, the GSR and GD are used by the following equations:

$$Z1_m^{it} = x_m^{it} - GSR + GD \tag{12}$$

$$Z1_m^{it} = x_m^{it} - randn.R_1 \cdot \frac{2\Delta x \cdot x_m^{it}}{(dp_m - dq_m + \epsilon)} + GD \tag{13}$$

in which

$$GD = rand.R_2 \cdot (x_b - x_m^{it}) \tag{13-1}$$

where R_1 and R_2 are two adaptive weighting factors. The leading role of these factors is to make a proper equilibrium between the exploration and exploitation, which are formulated as,

$$R_1 = 2 \cdot rand \cdot \omega - \omega \tag{14}$$

$$R_2 = 2 \cdot rand \cdot \omega - \omega \tag{15}$$

$$\omega = \left| \alpha \times \sin\left(\frac{3\pi}{2}\right) + \sin\left(\alpha \times \frac{3\pi}{2}\right) \right| \tag{15-1}$$

$$\alpha = \alpha_{min} + (\alpha_{max} - \alpha_{min}) \cdot \left(1 - \left(\frac{it}{Maxit}\right)^3\right)^2 \quad (15-2)$$

where *it* denotes the number of iteration, *Maxit* is the maximum number of iterations, α_{min} and α_{max} are equal to 0.2 and 1.2, respectively.

To increase the exploitation in the GBO, another solution (16) is formulated as follows:

$$Z2_m^{it} = x_b - randn.R_1 \cdot \frac{2\Delta x \cdot x_m^{it}}{(dp_m^{it} - dq_m^{it} + \epsilon)} + rand.R_2 \cdot (x_{a1}^{it} - x_{a2}^{it}) \quad (16)$$

Finally, based on the solutions $Z1_m^{it}$ and $Z2_m^{it}$, the new solution (x_{new}) is calculated as,

$$x_{new} = r_1 \cdot (r_2 \cdot V1_l^k + (1 - r_2) \cdot V2_l^k) + (1 - r_1) \cdot Z3_m^{it} \quad (17)$$

if *rand* < 0.5

if *rand* < 0.5

$$x_{new} = x_m^{it} + \beta_1 \cdot (c_1 \cdot x_{bs} - c_2 \cdot x_{rand}^{it}) + \beta_2 \cdot R_1 \cdot (c_3 \cdot (Z2_m^{it} - Z1_m^{it})) + \beta_2 \cdot (x_{a1}^{it} - x_{a2}^{it}) / 2$$

else

$$x_{new} = x_{bs} + \beta_1 \cdot (c_1 \cdot x_{bs} - c_2 \cdot x_{rand}^{it}) + \beta_2 \cdot R_1 \cdot (c_3 \cdot (Z2_m^{it} - Z1_m^{it})) + \beta_2 \cdot (x_{a1}^{it} - x_{a2}^{it}) / 2$$

End

End

in which

$$Z3_m^{it} = x_m^{it} - w_1 \times (Z2_m^{it} - Z1_m^{it}) \quad (17-1)$$

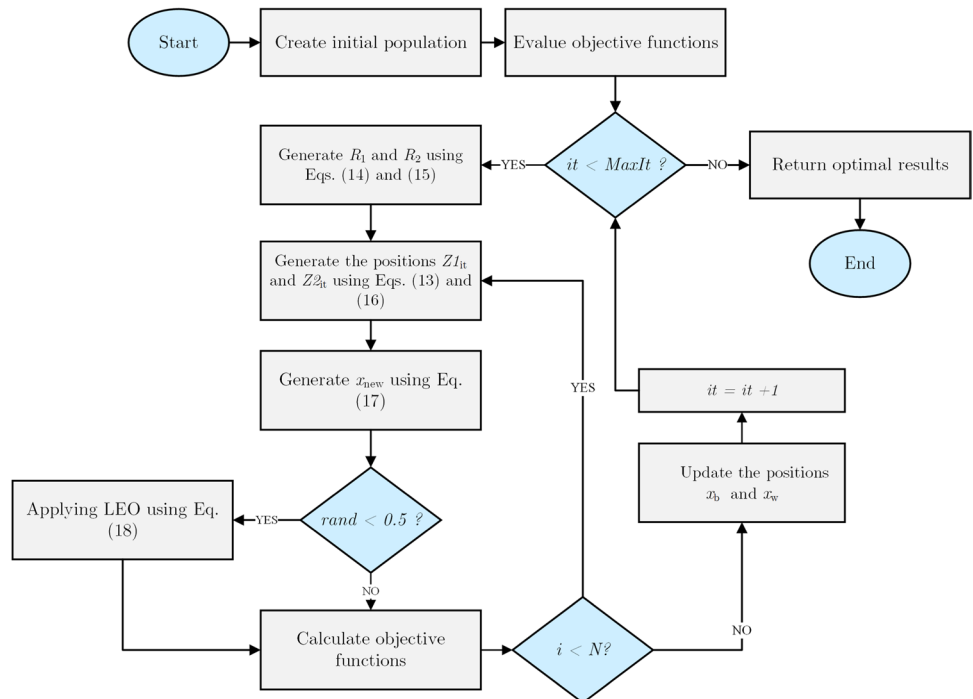
where r_1 and r_2 have two random values in [0, 1].

Local escaping operator (LEO)

The GBO uses an effective operator to avoid the local solutions called the LEO. The new solution (x_{new}) is created based on two random solutions (x_{a1}^{it} and x_{a2}^{it}) and the solutions $Z1_m^{it}$ and $Z2_m^{it}$, which is expressed as,

where β_1 and β_2 have two random values in [-1,1], and c_1 , c_2 , and c_3 have random values, which are defined as:

Fig. 1 Flowchart of GBO algorithm



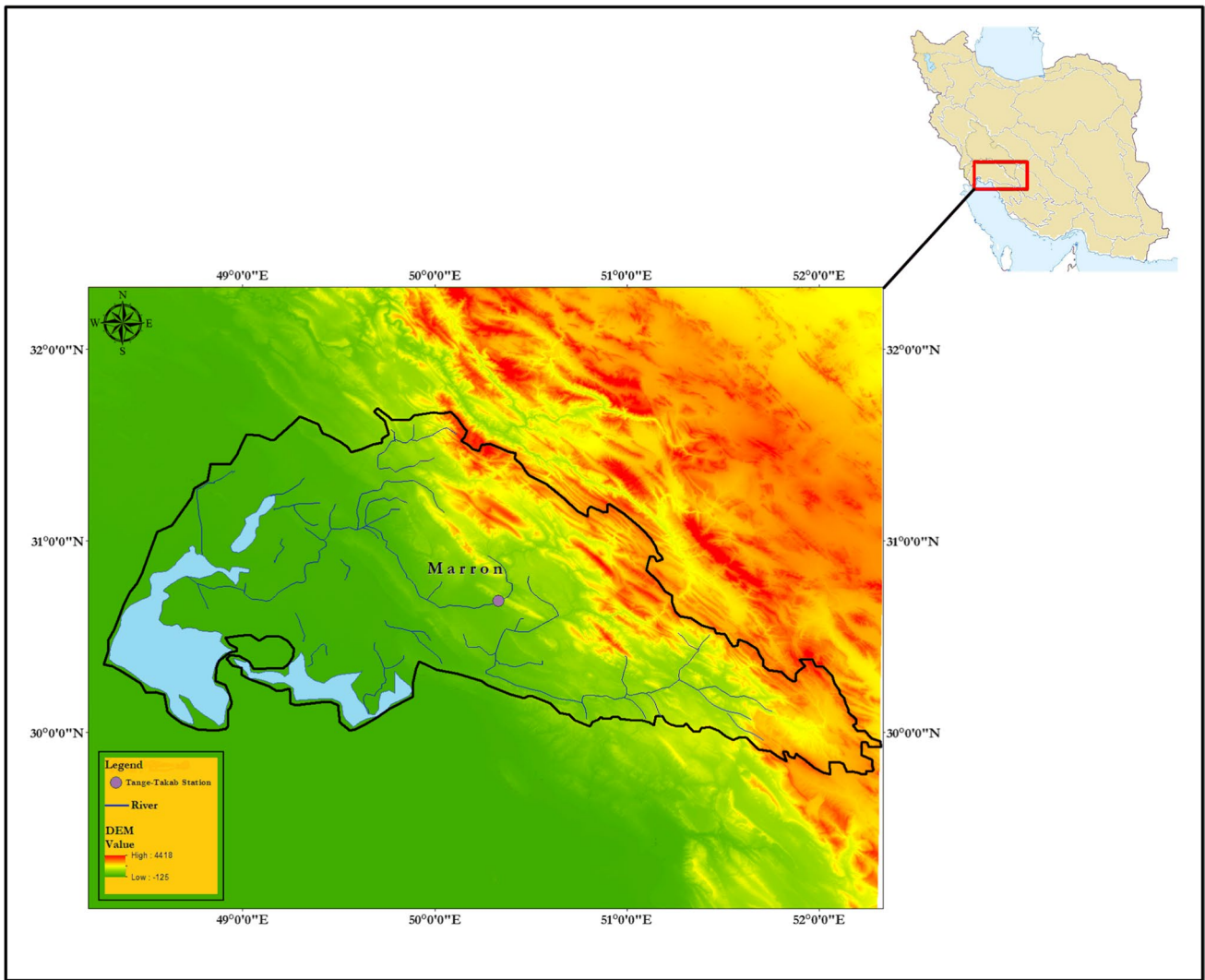


Fig. 2 Location of Tabge-Takab station

$$c_1 = \begin{cases} 2 \cdot rand & \text{if } \tau_1 < 0.5 \\ 1 & \text{otherwise} \end{cases} \quad (18-1)$$

$$c_2 = \begin{cases} rand & \text{if } \tau_1 < 0.5 \\ 1 & \text{otherwise} \end{cases} \quad (18-2)$$

$$c_3 = \begin{cases} rand & \text{if } \tau_1 < 0.5 \\ 1 & \text{otherwise} \end{cases} \quad (18-3)$$

where τ_1 has a random value in $[0, 1]$.

The solution x_{rand}^{it} in Eq. (18) is formulated as,

$$x_{rand}^{it} = \begin{cases} x_{rand} & \text{if } \tau_2 < 0.5 \\ x_{a5}^{it} & \text{otherwise} \end{cases} \quad (18-4)$$

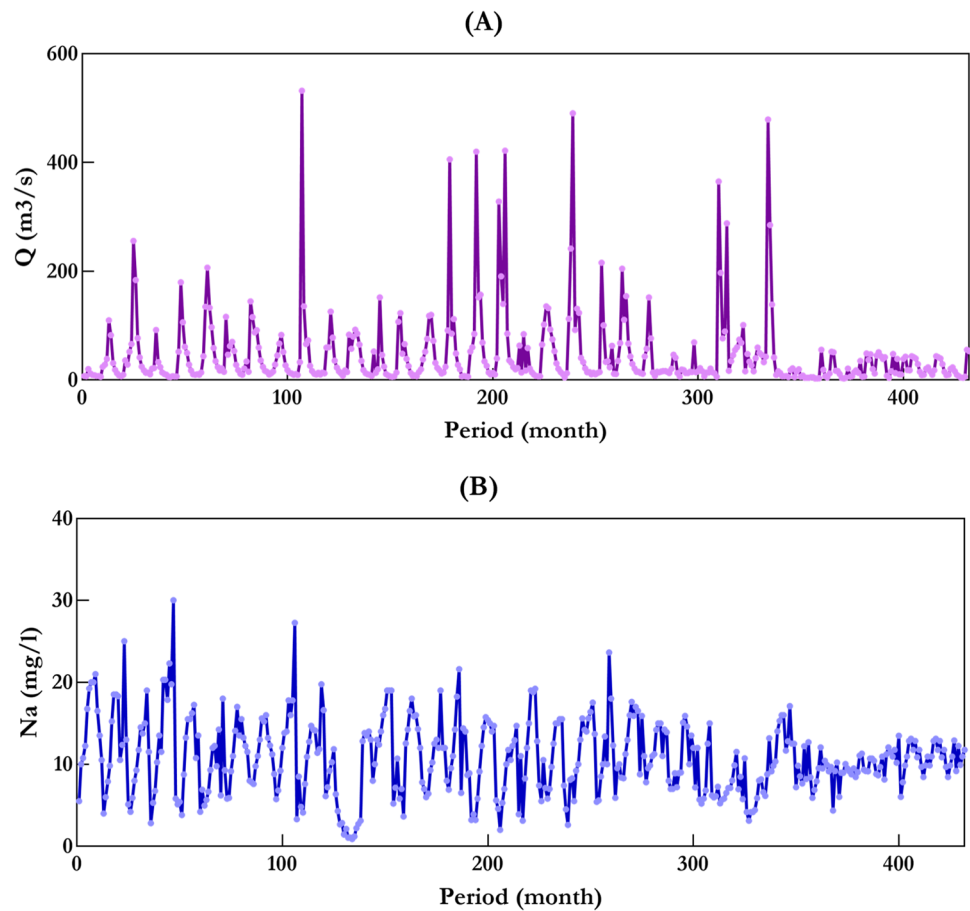
$$x_{rand} = LS + rand \cdot (US - LS) \quad (18-5)$$

where x_{rand} is a new random solution, x_{a5}^{it} is a random selected solution ($a5 \in [1, 2, \dots, N]$), and τ_2 has a random value in $[0, 1]$. The GBO's flowchart is displayed in Fig. 1.

ANFIS model

There are a large variety of supervised learning algorithms, among which the hybrid learning algorithm is regarded for prediction problems, primarily because the hybrid learning algorithm is widely utilized. ANFIS, a hybrid model, originates from integrating ANN and fuzzy systems, employing ANN learning method to gain fuzzy if-then rules with sufficient membership functions (Jang 1993). It can learn things through vague data and bring along the result. On the other hand, ANFIS uses memory and self-learning neural

Fig. 3 Time series of (A) Q (input) and (B) Na (target)



network's capabilities efficiently and creates a more constant training process (Huang et al. 2017).

Overall, ANFIS is, in fact, made by five distinct layers (Orouji et al. 2013). To put it simply, the input layer is a previous parameter and the other layers, namely rule-based along with three constant factors and a consequent factor, as well an output layer. Regarding the first layer, it can convert the input into a degree in the range of 0 and 1 or fuzzification, being named premise parameter. Indeed, it can be an activation function along with membership function including trapezoidal or generalized Bell, gauss and triangular (Mohammadi et al. 2016). Furthermore, the second layer can calculate the incoming signals to neurons through the product operator, while the third layer can bring the signals to a stable condition. However, the fourth layer is about fuzzification. The last layer can finally abstract the weighted output proportions.

Linear Regression (LR)

The linearity can be divided into two descriptions; linear to variable and linear to parameter. That is, modeling is considered a linear model on condition that the model is linear to parameter (Gujarati et al. 2012). Thus, the main

criterion for the linear model is linear model to the parameter even though the model is linear to the variable or not to the variable. Regarding multiple linear regression models, their objective is to assess the relationship of two or several independent variables with dependent ones, which is widely used. Making correlation or relationship between dependent and independent variables as well as the relationship between independent ones with each independent linear variable is essential before starting the process of linear regression modeling. Besides, it is important to test the possibility of whether the relationship between the independent and dependent variables could utilize Spearman or Pearson correlation test (Hauke and Kossowski 2011), although there are various testing procedures to examine the linearity test, namely Reset test, White test and Terasvirta test (Teräsvirta et al. 1993). Provided that one dependent variable (y) of m independent variables (x_1, x_2, \dots, x_m) is available, the models of multiple linear regression can be reformulated as:

$$y = a_0 + a_1x_1 + a_2x_2 + \dots + a_mx_m + \epsilon \quad (19)$$

$$y = aX + \epsilon \quad (20)$$

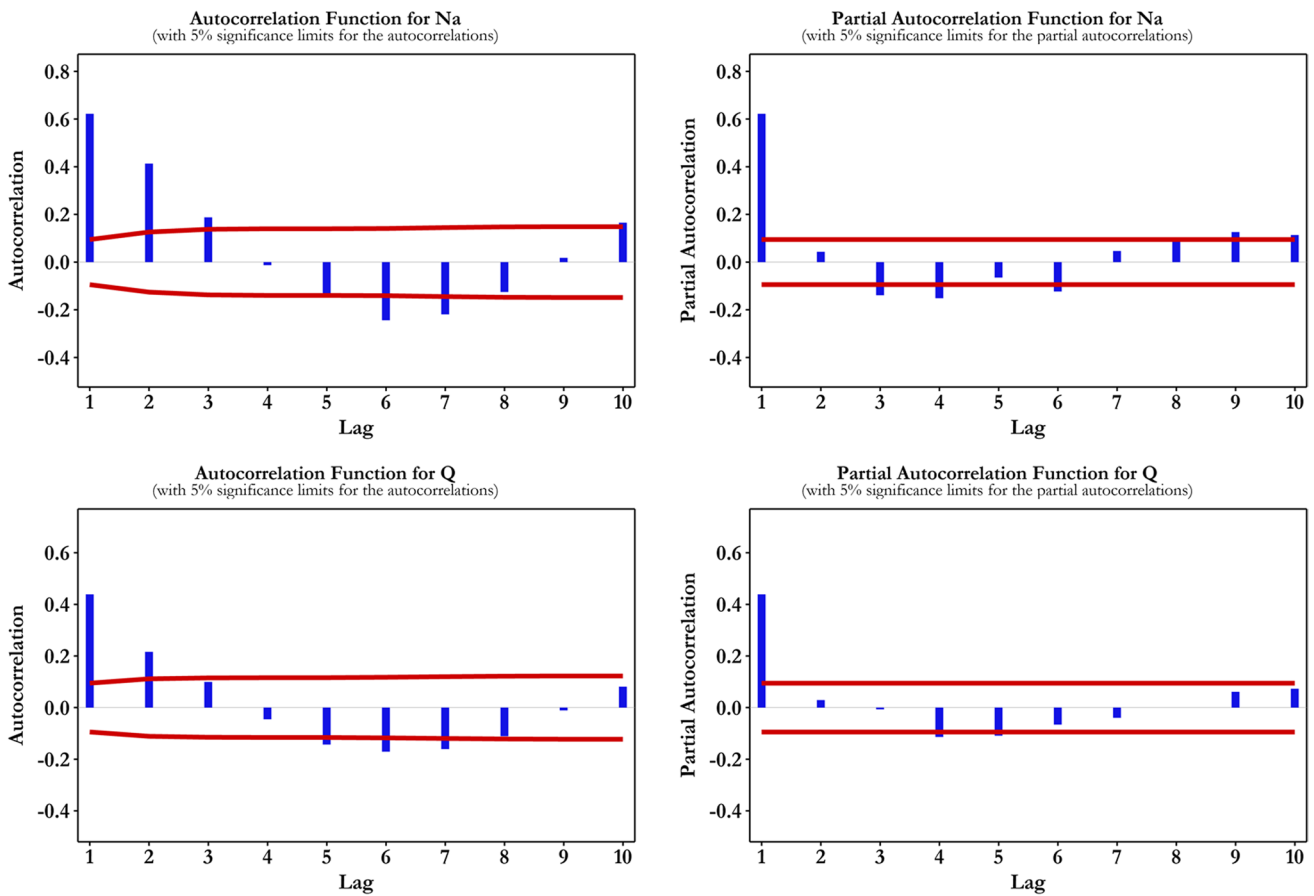


Fig. 4 ACF and PACF graphs of Q and Na datasets

where y denotes the dependent variable, a is the coefficients of the LR model, and ϵ denotes the error vector.

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} X = \begin{bmatrix} 1 & x_{11} & \dots & x_{1m} \\ 1 & x_{21} & \dots & x_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & \dots & x_{nm} \end{bmatrix} a = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{bmatrix} \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix} \quad (21)$$

To reach the regression model, the a parameter can be calculated by the following formula (Draper and Smith 1998):

$$\hat{a} = (X^T X)^{-1} X^T Y \quad (22)$$

Bayesian Linear Regression (BLR)

Regression modeling with parameter calculation technique is Bayesian linear regression (BLR), wherein the Bayesian method (Box and Tiao 2011) is used. This procedure lies on likelihood, prior and posterior distribution, in which parameters are calculated based on posterior distribution multiplying the prior one with likelihood distribution. Additionally, there is an OLS (i.e., ordinary least square) calculation

strategy for linear regression model, with it working by normal distributed error hypothesis, at $\epsilon \sim N(0, \delta^2)$. The variables $(Y|X, a, \delta^2)$ are highly likely to be distributed normally since they are similar to the error. Therefore, pdf or probability density function for variables $(Y|X, a, \delta^2) \sim N(Xa, \delta^2)$ can be defined as (Mutiarani et al. 2012):

$$prob(Y | X, a, \delta^2) = \frac{1}{\sqrt{2\pi\delta^2}} \exp \left[-\frac{1}{2\delta^2} (Y - Xa)^T (Y - Xa) \right] \quad (23)$$

According to the pdf, the likelihood of the mentioned variables is formulated in the following:

$$prob(Y | X, a, \delta^2) = \prod_{j=1}^m \frac{1}{\sqrt{2\pi\delta^2}} \exp \left[-\frac{1}{2\delta^2} (Y - Xa)^T (Y - Xa) \right] \quad (24)$$

$$prob(Y | X, a, \delta^2) = (\delta^2)^{-\frac{m}{2}} \exp \left[-\frac{1}{2\delta^2} (Y - Xa)^T (Y - Xa) \right] \quad (25)$$

Hence, many prior distributions are available to be employed in the Bayesian procedure of linear regression model, among which the distribution of prior conjugate

Table 1 Main information of dataset employed in train and test

Mode	Statistic	Q	Na
Training	No. of observations	302	302
	Minimum	4.900	0.900
	Maximum	532.000	30.000
	Median	23.550	11.450
	Mean	51.575	11.276
	SD	71.707	5.039
	Skewness (S)	3.646	0.262
	Kurtosis (K)	17.070	0.064
Testing	No. of observations	130	130
	Minimum	3.060	3.120
	Maximum	479.000	17.100
	Median	17.900	9.660
	Mean	37.099	9.556
	SD	2.733	64.282
	Skewness (S)	0.042	4.648
	Kurtosis (K)	-0.095	24.643
AC	R ₁	0.439	0.622
	R ₂	0.216	0.413
	R ₃	0.099	0.188
	R ₄	-0.045	-0.013

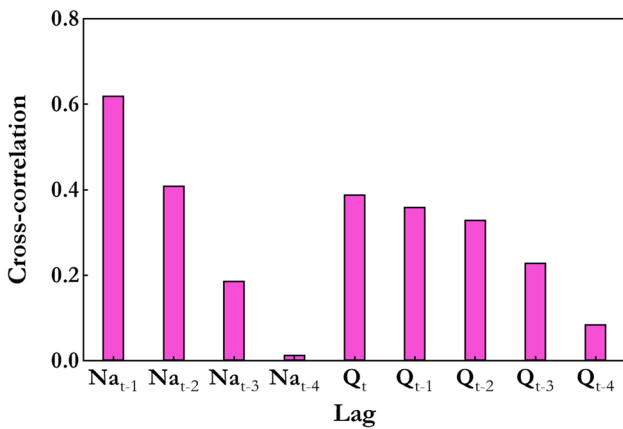


Fig. 5 Cross correlation values between input and target

Table 2 BRS to optimally select the input combination of Na for DI models

	Combinations	R ² %	R ² -Adj%	C _p	PC
1	Na _{t-1}	0.386	0.384	90.192	0.617
2	Q _t /Na _{t-1}	0.486	0.484	7.610	0.519
3	Q _t /Na _{t-2} /Na _{t-1}	0.494	0.490	3.053	0.513
4	Q _t /Na _{t-3} /Na _{t-2} /Na _{t-1}	0.496	0.492	2.991	0.513
5	Q _{t-2} /Q _t /Na _{t-3} /Na _{t-2} /Na _{t-1}	0.497	0.491	4.394	0.515
6	Q _{t-3} /Q _{t-2} /Q _t /Na _{t-3} /Na _{t-2} /Na _{t-1}	0.497	0.490	6.165	0.517
7	Q _{t-3} /Q _{t-2} /Q _{t-1} /Q _t /Na _{t-3} /Na _{t-2} /Na _{t-1}	0.497	0.489	8.000	0.519

(Rubio and Genton 2016) can be one of the probable methods. By considering the manner of iteration in marginal posterior, the calculation of regression model’s parameters through Bayesian procedure can be done. Also, posterior distribution can be estimated by multiplying the likelihood as well the prior distribution function (Mutiarani et al. 2012).

$$p(a, \sigma^2 | Y, X) \propto p(Y | X, a, \sigma^2) p(\sigma^2) p(a | \sigma^2) = (\delta^2)^{-\frac{m}{2}} \exp \left[-\frac{1}{2\delta^2} (Y - Xa)^T (Y - Xa) \right] \cdot (\delta^2)^{-\left(\frac{m}{2}+1\right)} \exp \left(-\frac{u\sigma^2}{2\delta^2} \right) \cdot (\sigma^2)^{-k/2} \exp \left[-\frac{1}{2a^2} (a - \mu)^T \Lambda (a - \mu) \right] \quad (26)$$

where k is the number of regression coefficients, $(\delta^2)^{-\left(\frac{m}{2}+1\right)} \cdot \exp \left(-\frac{u\sigma^2}{2\delta^2} \right)$ denotes the inverse $Gamma(a, b)$, where $a = u/2$ and $b = u\sigma^2$. μ is $[a_0, a_1, \dots, a_m]^T$.

MCMC or Markov Chain Monte Carlo (Geyer 1992) algorithm is employed to reach the calculation of the regression model’s parameters in the Bayesian procedure. Indeed, Gibbs sampling is one of the most well-known approaches in MCMC. In the following, the process of iteration for parameters calculation is done till the burn-in circumstances are completed.

LSSVM model

Vapnik (1995) presented initially the support vector machine or (SVM) algorithm based on statistical learning theory (Vapnik et al. 1995). This algorithm indeed focuses on formulating the training process in modeling by employing quadratic programming. Likewise, an improved SVM called the least squares SVM (LSSVM) was introduced by (Suykens and Vandewalle 1999) to minimize the computational time of the SVM algorithm process. Take into account a training set of M datasets and the input and output data x_m and y , respectively. The LSSVM-based models in feature space for regression matters are described as (Samui and Kothari 2011):

$$y = Q^T \varnothing(x) + c \quad (27)$$

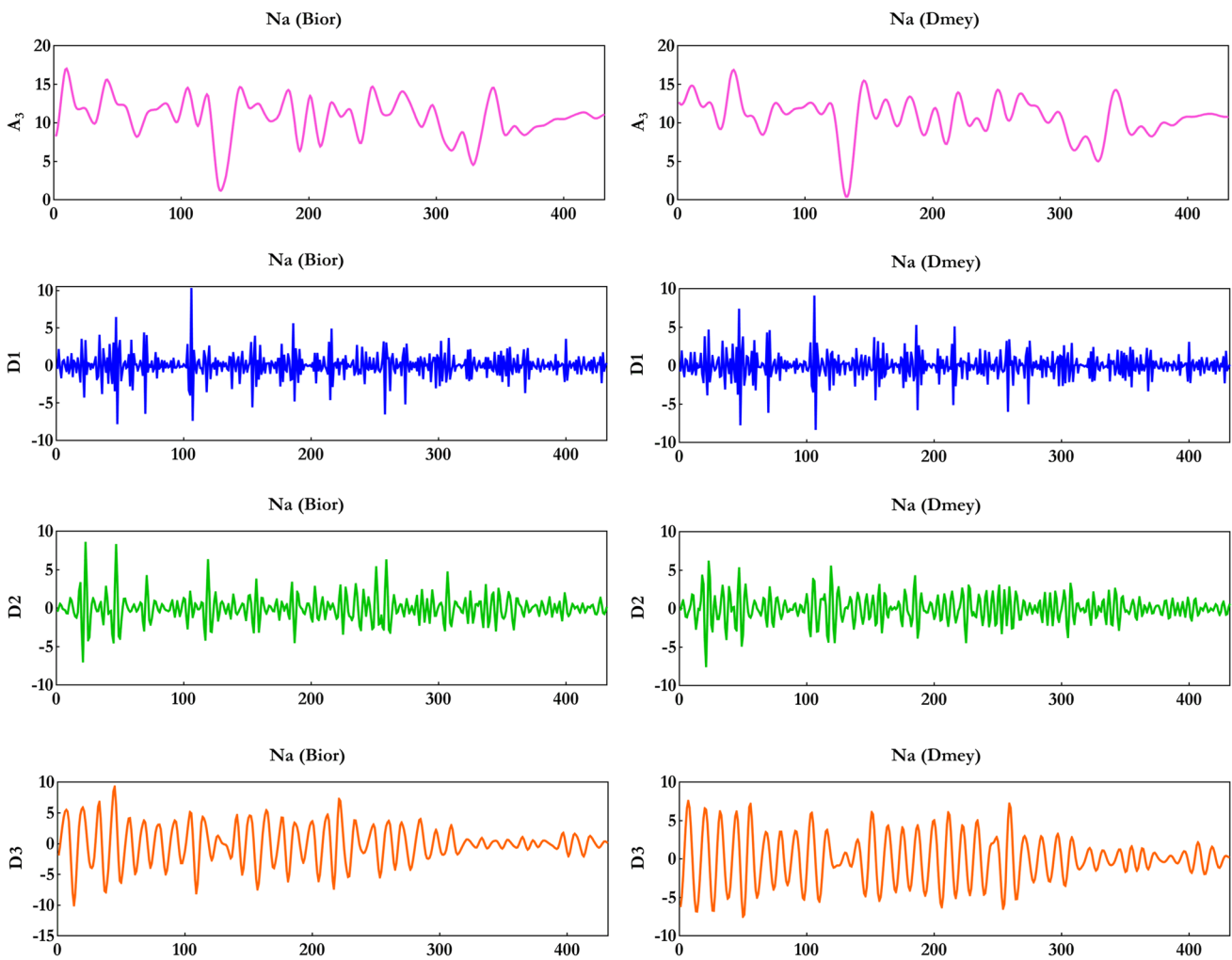


Fig. 6 Decomposition of Q and Na time series for Bior6.8 and Dmey

in which $\varnothing(x)$ is considered an adaptable weight vector, c denotes the scalar threshold regarded as bias. Q defines the mapping function moving the input data into a larger dimensional feature space. In the following, the given optimization problem is mathematically written for function calculation:

$$\begin{aligned} \text{Minimize : } & \frac{1}{2}Q^T Q + \frac{1}{2}\lambda \sum_{t=1}^M v_t^2 \text{ Subjected to : } y \\ & = Q^T \varnothing(x) + c + v_t, t = 1, \dots, M \end{aligned} \tag{28}$$

wherein v describes the error variable while λ being the regulative constant. Eventually, the LSSVM can be achieved by making a solution for the given optimization problems:

$$f(x) = \sum_{i=1}^M a_i Kr(x, x_i) + c \tag{29}$$

where $Kr(x, x_i)$ is considered the kernel function. In this research, radial basis function (RBF) as a kernel function is broadly utilized to support the previous research (Baesens

et al. 2000), in which diverse datasets compared the polynomial kernels and RBF. More significantly, the first kernel seems to be more appropriate than the second. RBF can be described by:

$$K(x, x_i) = \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right) \tag{30}$$

where σ denotes the tuning parameters of the kernel function.

Response surface regression (RSR)

In order to find a relationship between a number of input variables determined by (x_1, x_2, \dots, x_m) and a response Y , RSR (Gunst 1996; Bezerra et al. 2008) containing a set of approaches is employed. Overall, this relationship is estimated by the polynomial model as:

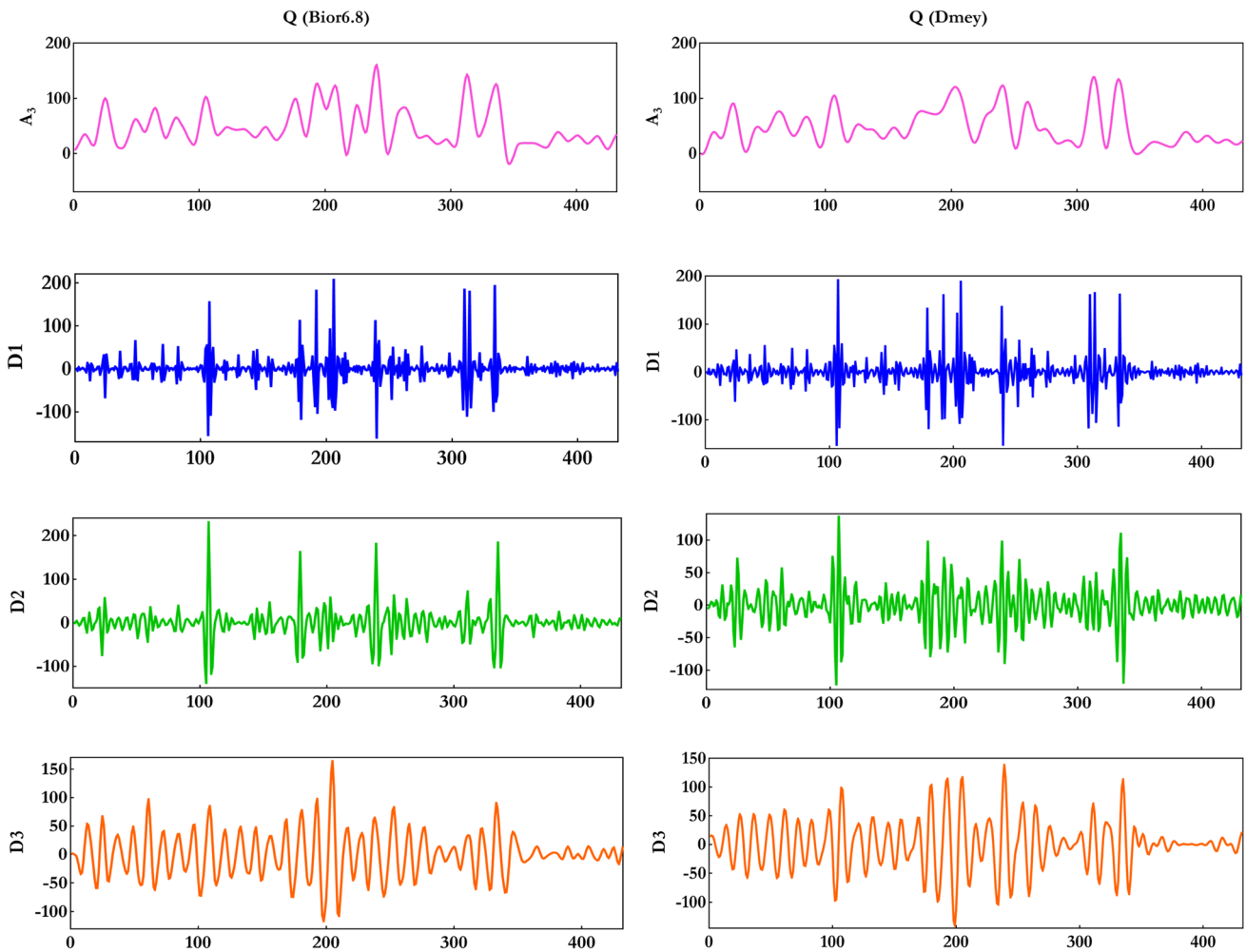


Fig. 6 (continued)

$$y = f'(X)a + \epsilon \tag{31}$$

in which $X = (x_1, x_2, \dots, x_m)$, $f(X)$ indicates the vector function of K containing the powers and cross-products of x_1, x_2, \dots, x_m to a distinct degree detected by $d(\geq 1)$. Besides, ϵ defines a random experimental error. a describes the vector of k stable parameters. In RSM, usually, two major conventional models are utilized, including the first-degree ($d = 1$) and second-degree model ($d = 2$). In this study, the second-degree model RSM is employed, wherein

$$y = a_0 + \sum_{j=1}^m a_j x_j + \sum_k \sum_{j>k} a_{kj} x_k x_j + \sum_{j=1}^m a_k x_j^2 + \epsilon \tag{32}$$

Evaluation criteria

In this study, seven well-known statistical metrics are used to assess the efficiency of DI models in forecasting the Na, including the correlation

coefficient (R), mean absolute percentage error (MAPE), root mean square error (RMSE), Willmott’s agreement Index (I_A), mean absolute error (MAE), relative absolute error (RAE), and Kling–Gupta efficiency (KGE) (Gupta et al.), which are formulated as,

$$RMSE = \left(\frac{1}{N} \sum_{i=1}^N (Na_{M,i} - Na_{DI,i})^2 \right)^{0.5} \tag{33}$$

$$R = \frac{\sum_{i=1}^N (Na_{DI,i} - \overline{Na_{DI}}) \cdot (Na_{M,i} - \overline{Na_M})}{\sqrt{\sum_{i=1}^N (Na_{DI,i} - \overline{Na_{DI}})^2 \sum_{k=1}^N (Na_{M,i} - \overline{Na_M})^2}} \tag{34}$$

$$RAE = \frac{\sum_{i=1}^N |Na_{DI,i} - Na_{M,i}|}{\sum_{i=1}^N |Na_{DI,i} - \overline{Na_M}|} \tag{35}$$

$$MAE = \left(\frac{1}{N}\right) \sum_{i=1}^N |Na_{DI,i} - Na_{M,i}| \tag{36}$$

$$MAPE = \left(\frac{100}{N}\right) \sum_{i=1}^N \left| \frac{Na_{DI,i} - Na_{M,i}}{Na_{M,i}} \right| \tag{37}$$

$$IA = 1 - \frac{\sum_{i=1}^N (Na_{DI,i} - Na_{M,i})^2}{\sum_{i=1}^N \left(\left| (Na_{DI,i} - Na_{DI}) \right| + \left| (Na_{M,i} - \overline{Na_M}) \right| \right)^2}, 0 < IA \leq 1 \tag{38}$$

$$KGE = 1 - \sqrt{(R - 1)^2 + (SD_{DI}/SD_M - 1)^2 + (\overline{R_{DI}}/\overline{R_M} - 1)^2} \tag{39}$$

where N is the total number of datasets, $Na_{M,i}$ denotes the measured dataset of Na , whilst $Na_{DI,i}$ expresses the predicted Na by DI models. $\overline{Na_{DI}}$ and $\overline{Na_M}$ denote the average amounts of the predicted Na and measured dataset correspondingly. SD_{DI} and SD_M are the standard deviation of the predicted and measured datasets.

In this paper, the performance index (PI), a multi-index criterion, was employed to convert the stated seven metrics into a single metric. This assists the user to facilitate decision making in selecting the best DI model. The proposed PI is formulated as,

$$PI = \frac{1}{7} \left(\frac{R_{min}}{R} + \frac{RMSE}{RMSE_{max}} + \frac{MAE}{MAE_{max}} + \frac{RAE}{RAE_{max}} + \frac{MAEP}{MAEP_{max}} + \frac{KGE_{min}}{KGE} + \frac{IA_{min}}{IA} \right) \tag{40}$$

where $MAEP_{max}$, MAE_{max} , $RMSE_{max}$ and MAE_{max} express the maximum amounts of $MAEP$, MAE , $RMSE$ and MAE computed by DI models, whereas IA_{min} , R_{min} and KGE_{min} indicate the minimum amounts of IA , R and KGE , which are calculated by DI models.

This study uses the Taylor diagram (Taylor 2001), a graphical metric, to clearly show the higher efficiency of the proposed model in comparison with other models. To implement this metric, three statistical metrics are used to plot this graph, such as standard deviation (SD), centered root mean square error ($cRMSE$), and (R). The performance of DI models is specified by geometrical distance to desired point (target point) in a polar space displayed in the graph (Taylor 2001). The $cRMSE$ is expressed as,

$$cRMSE^2 = SD_M^2 + SD_{DI}^2 - 2.R.SD_M.SD_{DI} \tag{41}$$

where SD_M and SD_{DI} denote the standard deviation of the measured and predicted amounts by DI models correspondingly.

Case study and data preprocessing

Study area

This research lies on water quality prediction based on Na prediction estimated at the Tange-Takab station in Maroun River, Khuzestan province, in Iran with Longitude $50^\circ 20' 02''$, Latitude $30^\circ 41' 09''$, and 280 heights from mean sea level. This study is conducted by considering the main water quality parameters including Sodium (Na) and discharge (Q) on a monthly basis. Since the area is comprised of the drainage area of $6,824 \text{ km}^2$ and roughly 310 km long, this river has an integral role in various purposes such as irrigation, drinking water as well recreational activities in Iran, in particular local dwellers in this province. Figure 2 illustrates the locality of this station (Fig. 3).

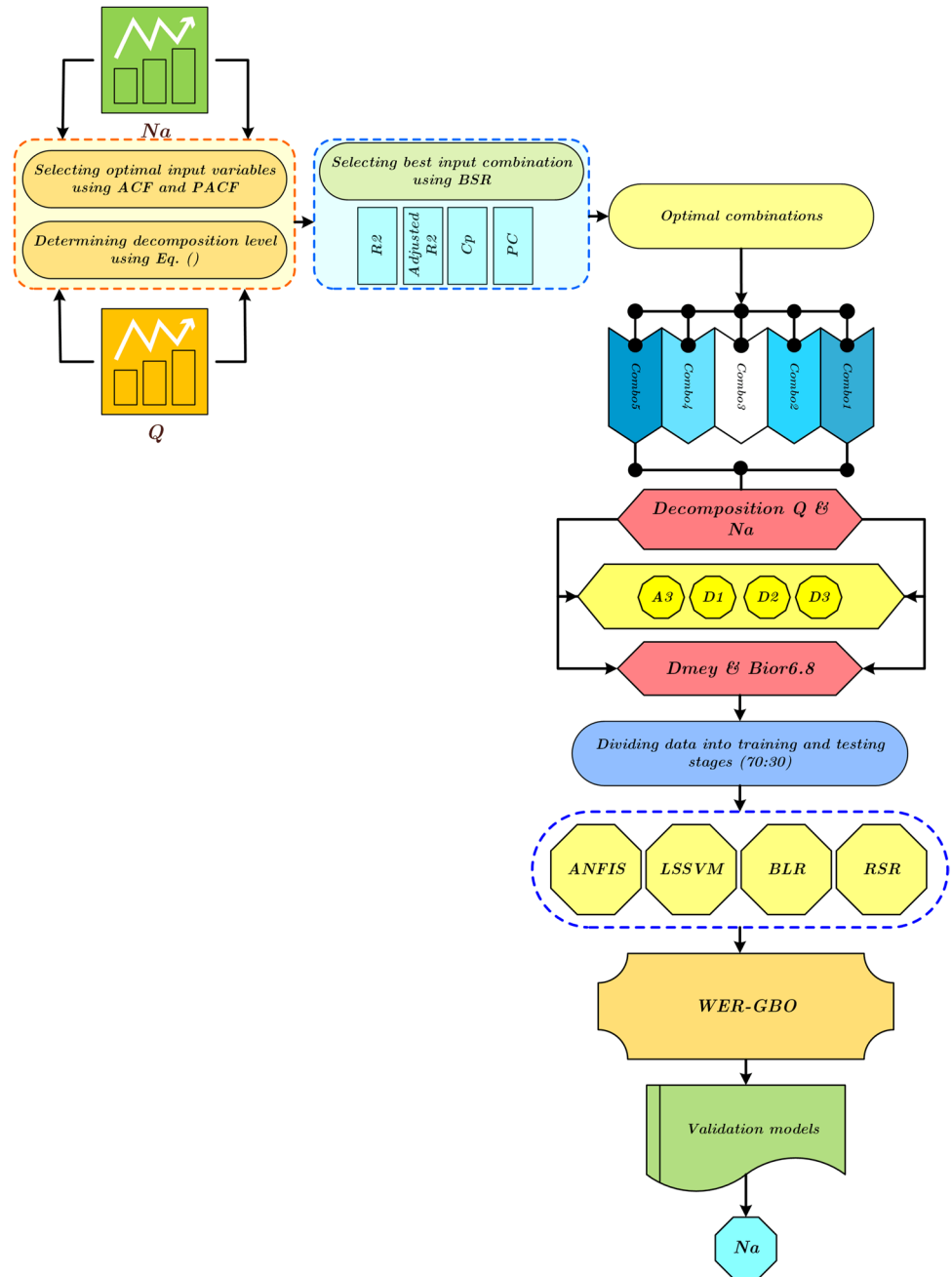
Pre-processing and selecting the best combination

In this part, the principal combinations of the gathered water quality datasets during the 36 years, from 1980 to 2016 on a monthly basis, are addressed in order to forecast the Na through DI models. These data consist of training and testing datasets, being 70 and 30 per cent or 302 and 130 months of the whole dataset correspondingly. The time series of Q as independent variables are displayed in Fig. 4A. To add to it, the Na time series as a target in testing and training intervals are shown in Fig. 4B. Table 1 provides the information on the factor’s categorizations, namely kurtosis (K), average (AVG), skewness (S), maximum (MAX), minimum (MIN), autocorrelation coefficients (AC), and standard deviation (SD) for testing, training and other data points. According to Table 1, the amounts of K and S for Na in the training and testing dataset stand at [0.064, 24.643] and [0.262, 4.648] ranges. Nonetheless, K range ([15.4, 48.81]) and S range ([3.469, 6.49]) for the Q indicate that the time series’ distribution has significant disparity from regular distribution.

One of the most significant steps in the prediction of Na parameter through DI models is to choose the most appropriate input variables’ combination. Simply put, the time series’ lags have a profound impact on this step (Malik et al. 2019; Jamei et al. 2020). Three statistical methods, namely cross correlation (CC), auto-correlation function (ACF) and partial auto correlation function (PACF), are employed to determine the input variables’ input combination (Barzegar et al. 2016).

As illustrated in Fig. 4, the calculation process of efficient input parameters is done by employing PACF and ACF. Apparently, the lagged input variables 1-month and 2-month’s AFC has a profound impact on Q_t , original input datasets. By contrast, the ACF of 1-month, 2-month, and 3-month lagged input variable Na_t is more efficient compared to the lagged times $Q_{t-3}, Q_{t-4}, \dots, ; Na_{t-4}, \dots$. The 1-month lagged signal can be set for Na_t and Q_t with regard to PCFA.

Fig. 7 Flowchart of DI models for predicting Na



As can be observed in Fig. 5, the input signal at the present time, $Q_t, Q_{t-1}, Q_{t-2}, Q_{t-3}$, and $Na_{t-1}, Na_{t-2}, Na_{t-3}$, time-lagged signals for Na_t have a large proportion of correlation, bringing along an influential predictive model in comparison with Na_{t-4} and Q_{t-4} . The comparison of cross-correlation and target signals (Na_t) indicates the Na_{t-1} and Na_{t-2} affect significantly Na_t parameter prediction in target thanks to high correlation coefficients, accounting for by 0.62 and 0.41 Na_{t-1}, Na_{t-2} correspondingly. Summing up, the assessment of ACF, CC, and PACF proves that there is a suitable range

for lags, as lagged t to $t - 4$ for Na_t and $t - 1$ to $t - 3$ months in the present month prediction of Na_t .

The best subset regression, BSR, was addressed in this research to specify the most appropriate input patterns from probable and current patterns. In this regard, a wide range of various factors such as Mallows (C_p) (Gilmour 1996), R^2 , Amemiya prediction criterion (PC) (Claeskens and Hjort 2008) and adjusted R are employed to detect optimal input pattern for water quality targets. The PC and C_p are formulated as (Kobayashi and Sakata 1990):

Table 3 Comparison of the statistic metrics achieved by four DI models to forecast the Na_t

Model	Criteria	Combination				
		Combo 1	Combo 2	Combo 3	Combo 4	Combo 5
ANFIS	R	0.610	0.595	0.578	0.564	0.471
	RMSE	3.113	2.990	2.981	2.991	3.272
	MAE	2.378	2.264	2.222	2.198	2.529
	RAE	1.106	1.053	1.033	1.022	1.176
	MAPE	30.459	29.007	29.005	28.141	31.665
	I_A	0.716	0.722	0.706	0.708	0.664
	KGE	0.559	0.559	0.542	0.535	0.442
	PI	0.898	0.875	0.879	0.877	1.000
LSSVM	R	0.567	0.568	0.578	0.567	0.574
	RMSE	2.973	2.926	2.997	2.935	2.897
	MAE	2.240	2.195	2.258	2.189	2.166
	RAE	1.041	1.020	1.050	1.018	1.007
	MAPE	29.236	28.720	29.572	28.468	28.154
	I_A	0.703	0.705	0.702	0.710	0.716
	KGE	0.531	0.531	0.536	0.535	0.544
	PI	0.995	0.984	0.996	0.980	0.969
BLR	R	0.540	0.543	0.536	0.535	0.535
	RMSE	2.799	2.781	2.849	2.854	2.851
	MAE	2.139	2.129	2.202	2.206	2.205
	RAE	0.995	0.990	1.024	1.026	1.025
	MAPE	29.183	29.114	30.099	30.154	30.135
	I_A	0.710	0.711	0.702	0.701	0.702
	KGE	0.526	0.526	0.516	0.515	0.516
	PI	0.978	0.975	0.999	1.000	0.999
RSR	R	0.572	0.577	0.568	0.559	0.570
	RMSE	2.958	2.704	2.837	3.082	2.889
	MAE	2.290	2.023	2.115	2.336	2.157
	RAE	1.065	0.941	0.983	1.086	1.003
	MAPE	30.034	27.086	28.095	30.645	28.725
	I_A	0.711	0.733	0.723	0.705	0.719
	KGE	0.542	0.557	0.546	0.525	0.545
	PI	0.977	0.909	0.938	1.000	0.949

Boldface indicates the best prediction results

$$C_p = \frac{RSS_k}{MSE_m} + 2k - N, m > k \tag{42}$$

$$PC = ((n + k)/(n - k))(1 - (R^2)) \tag{43}$$

wherein RSS_k indicates the squares' residual sum, MSE_m is considered mean squared error, N denotes the historical number of datasets and k defines the number of predictors. Table 2 depicts the classification of the outcomes of BSR analysis Na_t , and also the analysis of Na_t was assessed for five distinct suitable combinations. These combinations are the optimum input data in predictive models according to the most appropriate outcomes, including PC ([0.513, 0.517]), R^2 ([0.492, 0.497]), C_p ([2.991, 8.00]). Indeed, this method

cannot alone bring along the certainty of the most appropriate input combinations, which in turn a more authentic assessment in the most suitable and possible input combinations is necessary. That is, five input combinations are considered to improve the DI models being specified by grey highlight in Table 2.

Results and discussion

Model development

To predict the Na_t in the wavelet-complementary and standalone frameworks, several DI models, namely LSSVM,

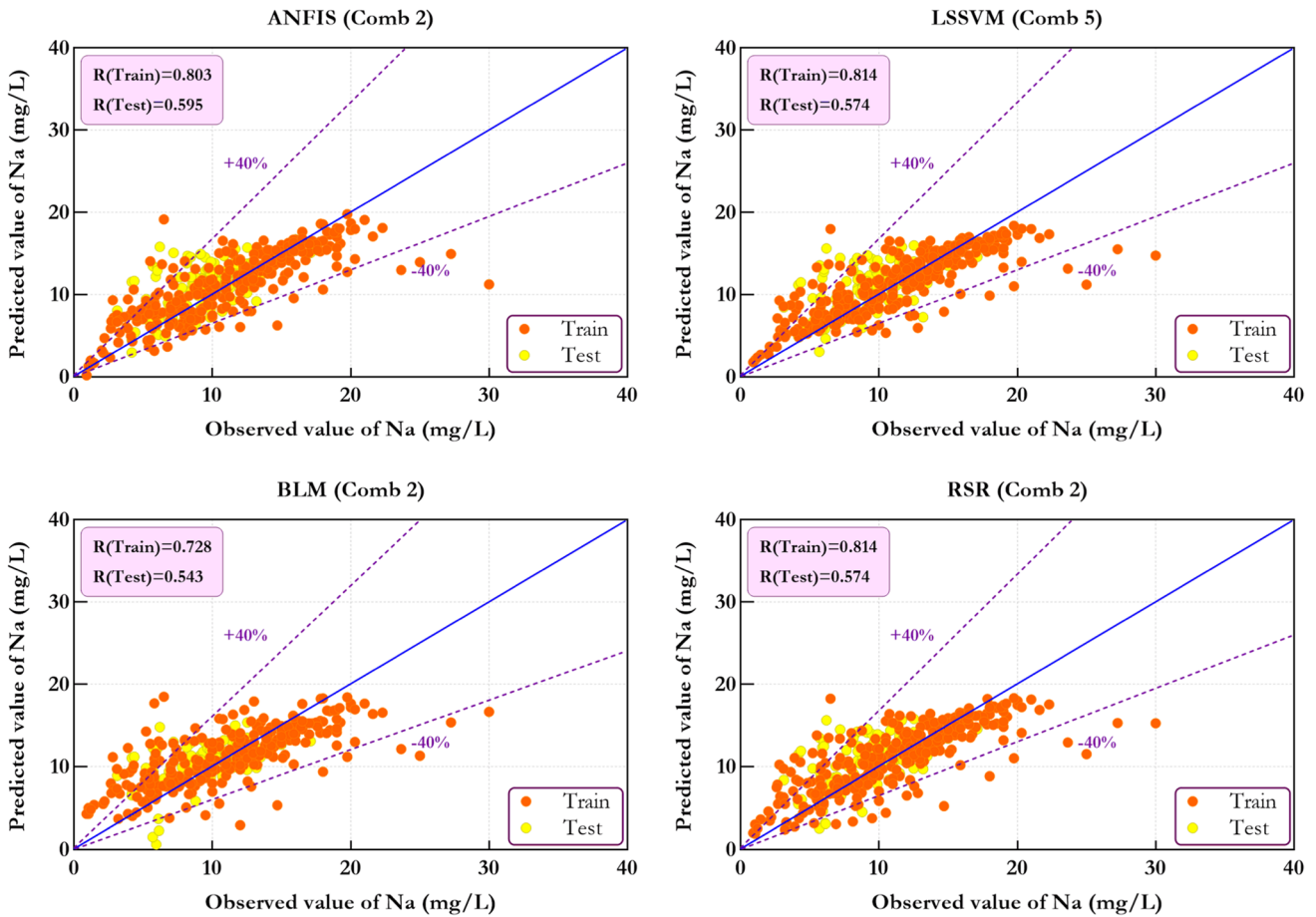


Fig. 8 Comparison of scatter graphs of predicted and measured amounts of Na applying the standalone DI models

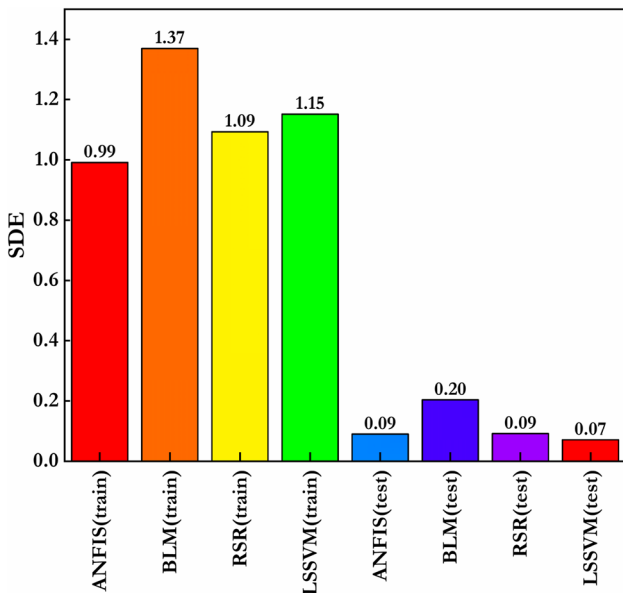


Fig. 9 SDE values of all models for training and testing stages

RSR, BLR, and ANFIS, are utilized in this stage, after which the obtained predicted amounts from these models are considered for inputs in a meticulous model, named weighted exponential regression. This model was in fact integrated with gradient-based optimization (WER-GBO) in order to predict the Na_t . This means it employs the principal benefit of these four models to reach more exact Na_t prediction. More significantly, the essential WER parameters in the proposed model are optimized by the algorithm of GBO. To add to it, using various cutting-edge machine learning methods, namely LSSVM, RSR, ANFIS, and BLR, played an integral role in proving the ability of WER-GBO in forecasting, which eventually brought about striking novelty in this study. The LSSVM acquired their setting parameters during the process of the trial-test procedure, in which the values of parameters are $\lambda = 8.80$ and $\gamma = 2.00$.

Wavelet-DI models

To boost the certainty and accuracy of DI models, wavelet can be a beneficial tool, mainly because wavelet decomposes

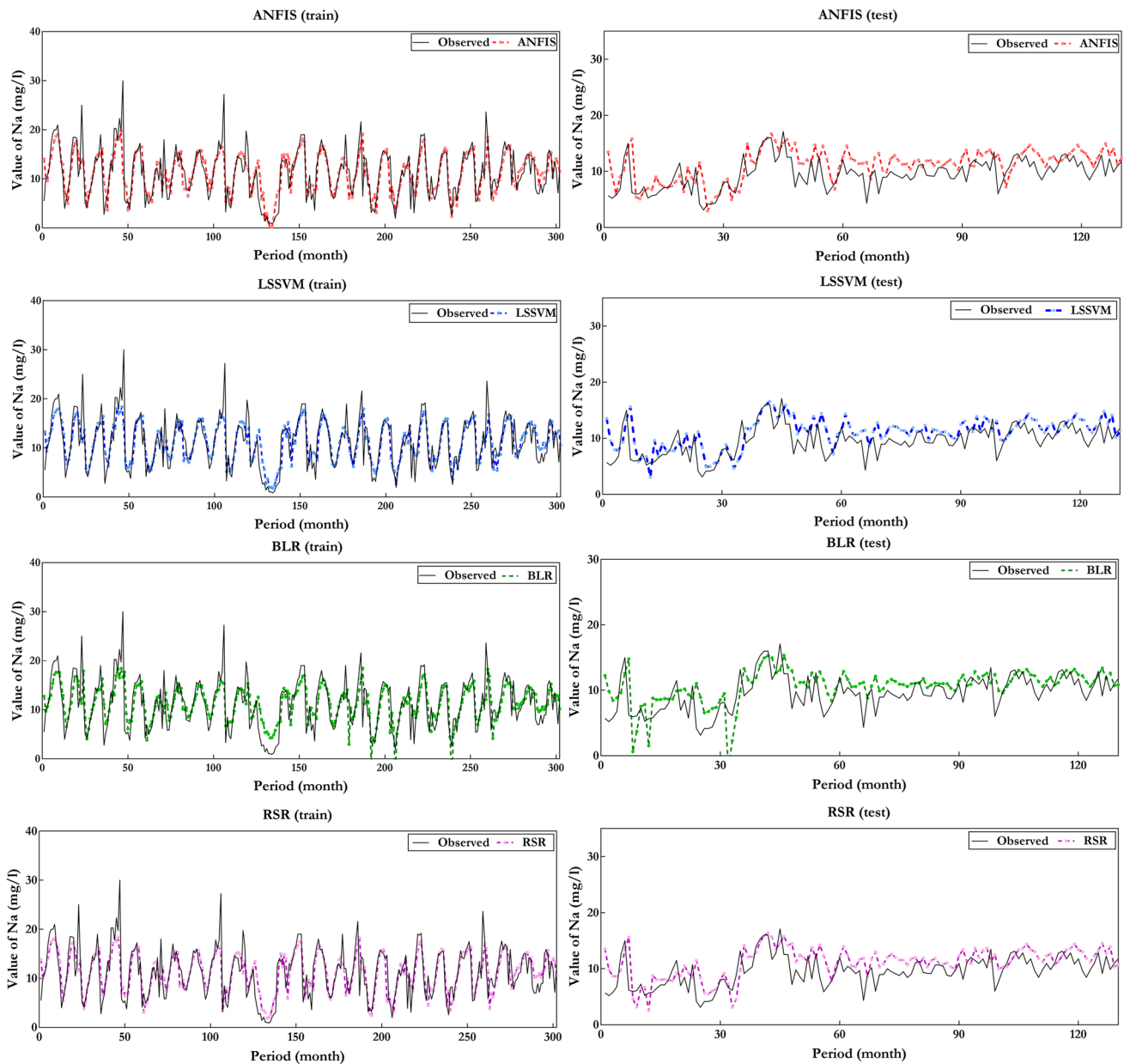


Fig. 10 Variations of the predicted and measured Na for the best performance of DI models

the inputs into two classifications, namely high-frequency (approximate) dataset and low-frequency (details) dataset. One of the most well-known wavelets in hydrological modeling, called discrete wavelet transform (DWT), is employed in this research. Two of most commonly used mother wavelets are discrete Meyer (Damey) and Biorthogonal 6.8 (Bior6.8) (Ahmadianfar et al. 2020b; Jamei et al. 2021a). In fact, these wavelets have been constructive ones in water quality predictive models. In other words, these mentioned wavelets support compressed form while being advantageous in creating time localization (Nourani et al. 2014; Freire et al. 2019; Ahmadianfar et al. 2020b; Jamei

et al. 2021a). The mother wavelet (i.e., Bior6.8 and Dmey) was utilized in this study to take apart in the time series. The optimal decomposition grade (1) of wavelet transform for the water quality time series was defined as below (Barzegar et al. 2018):

$$nW = \text{int}[\log(M)] \tag{44}$$

wherein M is considered the number of datasets, standing at 432. The proportion of disintegration grade will account for 3.

Table 4 Comparison of the statistic metrics achieved by ANFIS models for all combinations

Model	Criteria	Combination				
		Combo 1	Combo 2	Combo 3	Combo 4	Combo 5
Wavelet-Dmey	R	0.931	0.937	0.964	0.966	0.967
	RMSE	1.065	1.016	0.729	0.699	0.699
	MAE	0.833	0.745	0.561	0.530	0.526
	RAE	0.387	0.346	0.261	0.246	0.245
	MAPE	10.143	9.550	7.050	6.344	6.230
	I _A	0.963	0.966	0.981	0.983	0.982
	KGE	0.912	0.916	0.952	0.951	0.929
	PI	1.000	0.953	0.805	0.780	0.781
Wavelet-Bior	R	0.896	0.902	0.926	0.930	0.905
	RMSE	1.242	1.200	1.028	1.002	1.168
	MAE	0.904	0.816	0.767	0.770	0.867
	RAE	0.420	0.380	0.357	0.358	0.403
	MAPE	11.537	10.208	9.346	9.233	10.341
	I _A	0.938	0.947	0.960	0.962	0.946
	KGE	0.801	0.894	0.892	0.885	0.844
	PI	1.000	0.934	0.882	0.879	0.955

Boldface indicates the best prediction results

Table 5 Comparison of the statistic metrics achieved by BLR models for all combinations

Model	Criteria	Combination				
		Combo 1	Combo 2	Combo 3	Combo 4	Combo 5
Wavelet-Demy	R	0.945	0.966	0.966	0.966	0.968
	RMSE	0.899	0.705	0.703	0.704	0.682
	MAE	0.714	0.560	0.558	0.544	0.512
	RAE	0.332	0.260	0.259	0.253	0.238
	MAPE	8.712	7.153	7.058	6.597	6.015
	I _A	0.971	0.982	0.983	0.982	0.983
	KGE	0.924	0.955	0.955	0.947	0.931
	PI	1.000	0.872	0.870	0.858	0.834
Wavelet-Bior	R	0.913	0.931	0.929	0.925	0.932
	RMSE	1.117	0.996	1.008	1.035	0.990
	MAE	0.826	0.740	0.751	0.796	0.767
	RAE	0.384	0.344	0.349	0.370	0.356
	MAPE	9.890	9.024	9.010	9.478	8.999
	I _A	0.953	0.963	0.962	0.960	0.963
	KGE	0.885	0.904	0.898	0.890	0.879
	PI	0.999	0.934	0.940	0.968	0.946

Boldface indicates the best prediction results

In the following stage, disintegrated time series being constructive in this process were gathered (e.g., $Q = A_3 + \sum_{k=1}^3 D_k$) and set as inputs in order for complementary DI models with regard to the input combinations of N_a . Besides, the A_s , approximations, and D_s , details of signals of N_a and Q simulation, are illustrated in Fig. 6. In Fig. 7, DI models' flowchart for N_a prediction parameters is drawn.

Investigation of the standalone DI model abilities

In this step, four DI models, ANFIS, LSSVM, BLR, and RSR, were assessed in terms of performance and ability on the five diverse input parameters' combinations. The outcomes are given in Table. 3, in which the fundamental variety of DI models' forecasting ability in test data for combinations in N_a , prediction is provided. According to

Table 6 Comparison of the statistic metrics achieved by LSSVM models for all combinations

Model	Criteria	Combination				
		Combo 1	Combo 2	Combo 3	Combo 4	Combo 5
Wavelet-Demy	R	0.936	0.962	0.958	0.954	0.952
	RMSE	1.008	0.783	0.808	0.846	0.874
	MAE	0.781	0.604	0.620	0.646	0.660
	RAE	0.363	0.281	0.288	0.300	0.307
	MAPE	9.926	7.682	7.790	7.945	8.118
	I _A	0.960	0.977	0.976	0.973	0.971
	KGE	0.850	0.902	0.915	0.901	0.891
	PI	1.000	0.856	0.866	0.886	0.900
Wavelet-Bior	R	0.898	0.929	0.922	0.924	0.906
	RMSE	1.259	1.053	1.098	1.089	1.221
	MAE	0.960	0.799	0.823	0.827	0.931
	RAE	0.446	0.371	0.382	0.384	0.433
	MAPE	12.387	10.069	10.235	10.149	11.746
	I _A	0.937	0.957	0.953	0.954	0.940
	KGE	0.814	0.858	0.851	0.857	0.808
	PI	0.999	0.886	0.903	0.901	0.978

Boldface indicates the best prediction results

Table 7 Comparison of the statistic metrics achieved by RSR models for all combinations

Model	Criteria	Combination				
		Combo 1	Combo 2	Combo 3	Combo 4	Combo 5
Wavelet-Demy	R	0.940	0.961	0.952	0.948	0.932
	RMSE	0.970	0.749	0.847	0.866	1.023
	MAE	0.763	0.590	0.648	0.627	0.753
	RAE	0.355	0.274	0.301	0.292	0.350
	MAPE	9.251	7.595	8.064	7.595	9.210
	I _A	0.965	0.980	0.975	0.973	0.964
	KGE	0.901	0.952	0.946	0.941	0.928
	PI	0.991	0.857	0.903	0.892	0.991
Wavelet-Bior	R	0.902	0.921	0.927	0.814	0.878
	RMSE	1.177	1.069	1.027	1.705	1.331
	MAE	0.861	0.808	0.789	1.019	0.943
	RAE	0.400	0.376	0.367	0.474	0.438
	MAPE	10.208	10.173	9.669	12.778	11.333
	I _A	0.947	0.958	0.961	0.899	0.935
	KGE	0.863	0.904	0.902	0.808	0.870
	PI	0.852	0.818	0.802	1.000	0.905

Boldface indicates the best prediction results

Table 3, the performance and ability of the ANFIS model in forecasting the Na_t in the testing process proves that Combo 2 (R = 0.595, RMSE = 2.990, MAPE = 29.007, KGE = 0.559, and PI = 0.875) stands out compared to the other combinations. By considering five combinations, the best combination of inputs in BLR and RSR was detected. The most appropriate combinations for both are Combo 2 (RSR: (R = 0.577, RMSE = 2.704, MAPE = 27.086, KGE = 0.557,

and PI = 0.909), and BLR: (R = 0.543, RMSE = 2.781, MAPE = 29.114, KGE = 0.526, and PI = 0.975)). Turning to LSSVM model, the best combination is considered Combo 4 (R = 0.574, RMSE = 2.897, MAPE = 28.154, KGE = 0.544, and PI = 0.969)).

Different models, including ANFIS, LSSVM, BLR, and RSR, are utilized to determine the prediction outcomes of 6 water quality indicators demonstrated in

Fig. 11 Comparison of scatter graphs of predicted and measured amounts of Na applying the wavelet-based DI models

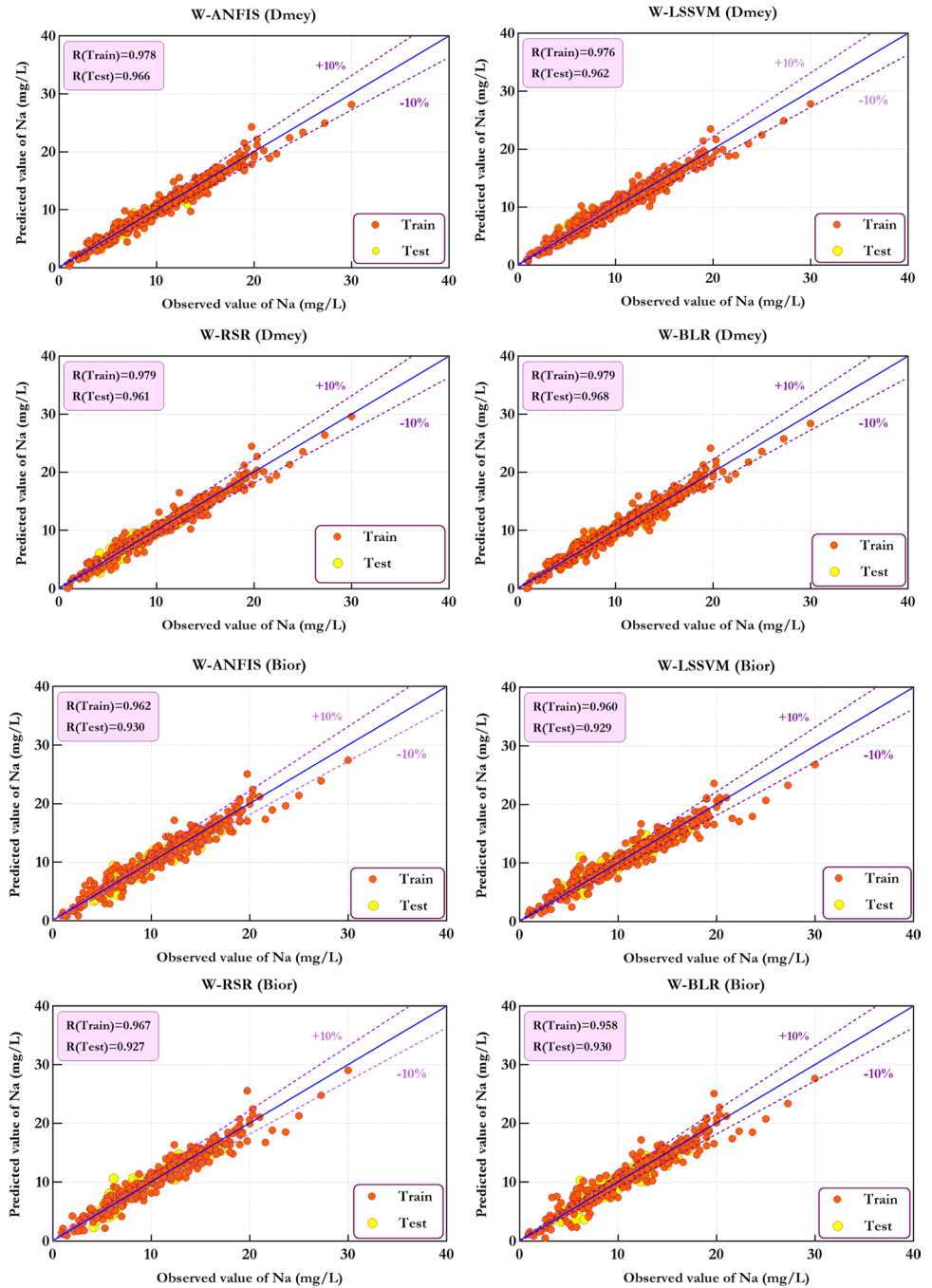


Fig. 8. The X-axis defines the measured value, and the Y-axis depicts the predicted value. More specifically, the ideal forecasting outcomes will be distributed on both sides of the line or $Y = X$. In turn, it is apparent that the error is conforming to the Gaussian distribution rule. In other words, the location of points based on the line has a direct relation with the amount of error. Thus, providing that the points are adjacent to the $Y = X$ line, the error is minor. On the other side, by considering the

forecasting of 4 indicators, the spots of DI models are out of the $Y = X$ line. As can be observed in Fig. 8, the distribution of prediction outcomes for many models is on the equal side of $Y = X$, which results in significant deviation and some under-fitting or over-fitting problems. Additionally, the error value stands at $\pm 40\%$ based on predicted proportions obtained by two distinct DI models. As a consequence, five standalone DI models are not able to predict the Na accurately.

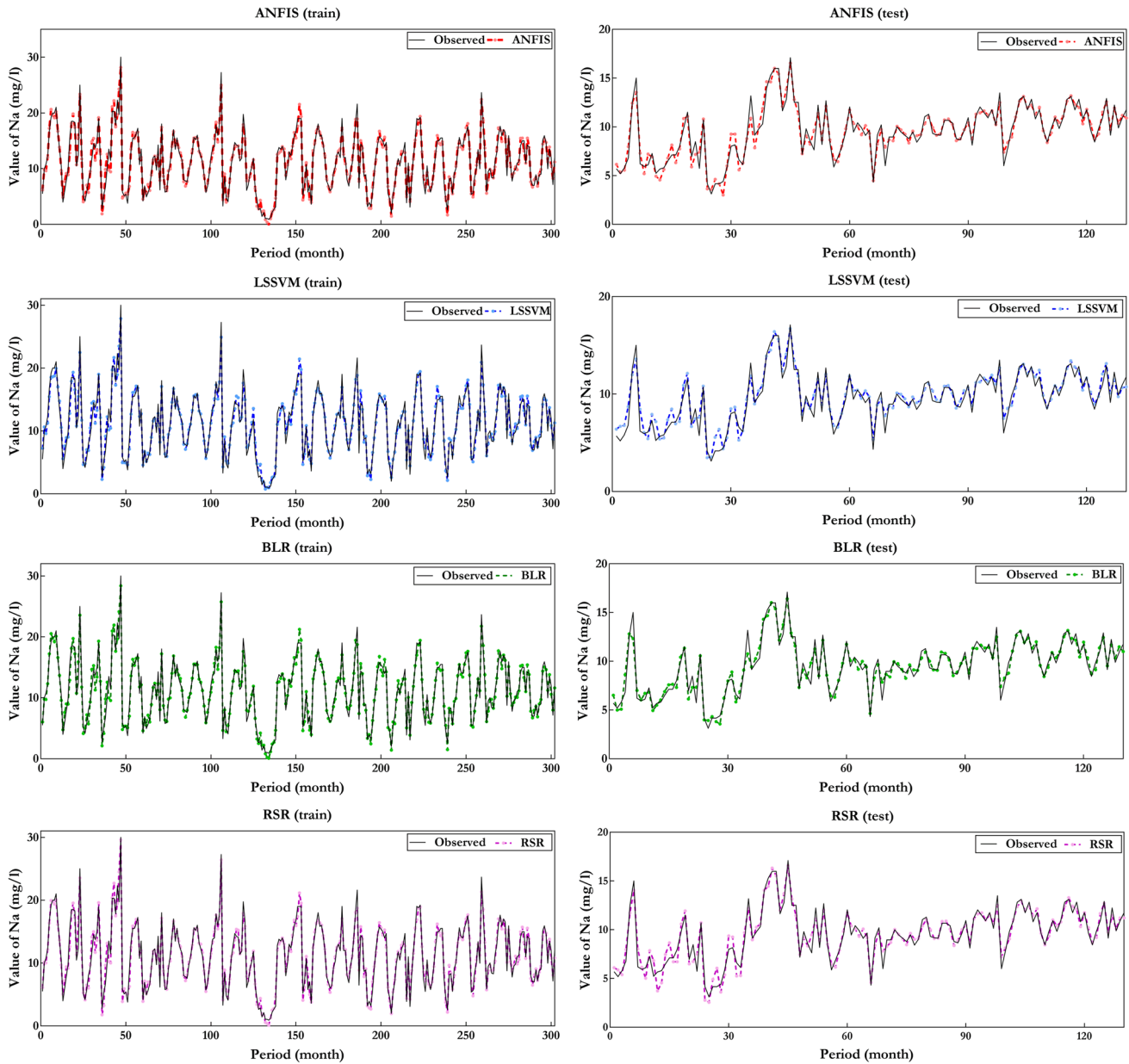


Fig. 12 Variations of the predicted and measured Na for the best performance of W-DI models

Since stability is a vital matter in prediction, the standard deviation error (SDE) can be an influential factor to investigate the models’ prediction stability. In Fig. 9, the SDEs of forecasting the different water quality indicators are drawn. More importantly, the SDEs of ANFIS reached the first and second-best place in training and testing stages, accounting for 0.99 and 0.09 correspondingly. Moreover, results reveal the SDEs of LSSVM, standing at 0.07, outperform in the testing step. The efficiency of ANFIS, LSSVM, BLR, and RSR models are addressed using time series diagrams

based on the measured and calculated proportion of Na within training and testing stages, which is provided in Fig. 10. As a result, all DI models cannot forecast the Na correctly.

Evaluate the performance of wavelet-based DI models

The W-ANFIS, W-LSSVM, W-BLR, and W-RSR models were improved with the purpose of enhancing the accuracy and certainty of standalone DI models (i.e., ANFIS, LSSVM,

Table 8 Comparison of statistical metrics of WER-GBO and WLR-GBO models

Mode	Criteria	WER-GBO	WLR-GBO
Train	R	0.9813	0.9811
	RMSE	0.9671	0.9741
	MAE	0.7028	0.7125
	RAE	0.1711	0.1735
	MAPE	7.7907	8.2642
	I_A	0.9905	0.9904
	KGE	0.9731	0.9730
	PI	0.987	1.000
Test	R	0.9721	0.9712
	RMSE	0.6399	0.6525
	MAE	0.4976	0.5030
	RAE	0.2314	0.2338
	MAPE	5.8858	6.0126
	I_A	0.9854	0.9847
	KGE	0.9486	0.9398
	PI	0.990	1.000

Boldface indicates the best prediction results

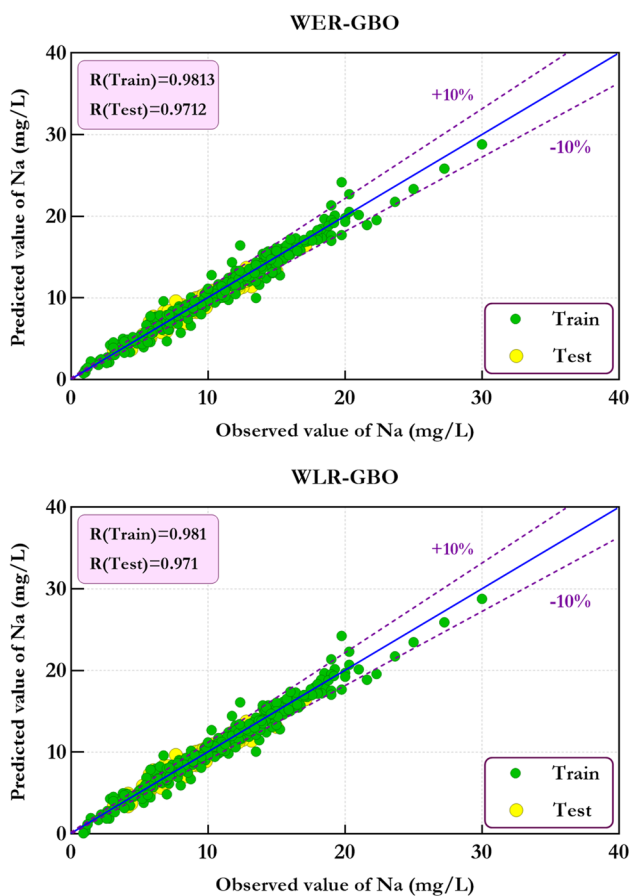


Fig. 13 Comparison of scatter graphs of predicted and measured amounts of Na applying the WER-GBO and WLR-GBO models

BLR, and RSR. Two mother wavelets (i.e., Bior6.8 and Dmey) are used in the decomposition process of time series for Na . The assessment of W-DI models’ ability with various mother wavelets is implemented by considering five distinct combinations of input variables. The parameter settings of LSSVM-Bior6.8 ($\lambda = 271.73$ and $\gamma = 34.24$) and LSSVM-Dmey ($\lambda = 111$ and $\gamma = 19.4$) are calculated based on trial-and-error.

Table 4 provides the data on the forecasting accuracy and certainty of W-ANFIS models considering all combinations and two mother wavelets. As shown, the most suitable combination for the W-ANFIS model with mother wavelets Dmey and Bior6.8 can be Comb 4 ($R=0.966$, $RMSE=0.699$, $MAPE=6.344$, $KGE=0.951$, and $PI=0.780$) and also for Na_t prediction in testing step can be Comb 4 ($R=0.930$, $RMSE=1.002$, $MAPE=9.233$, $KGE=0.885$, and $PI=0.879$). The ultimate results prove Dmey outperforms the Bior6.8 for W-ANFIS, which stems from its more accurate and exact performance than Bior6.8 mother wavelet (Table 5).

In the W-BLR model, Comb 4 is considered the best combination being equal to all mother wavelets, and the related data are provided in Table. Moreover, the final results of different mother wavelets for the most appropriate combination stand out at W-BRL-Dmey (Combo5: $R=0.968$, $RMSE=0.682$, $MAPE=6.015$, $KGE=0.931$, and $PI=0.834$), and W- BLR -Bior6.8 (Combo2: $R=0.931$, $RMSE=0.996$, $MAPE=9.024$, $KGE=0.904$, and $PI=0.934$). Accordingly, the outcomes confirm that the most suitable mother wavelet is admittedly Dmey in W-BLR, mainly because it reached higher certainty and accuracy than other models.

Regarding W-LSSVM, as can be seen in Table 6, Combo 2 is considered the winner for two mother wavelets. Besides, the estimated outcomes for Bior6.8 and Dmey mother wavelets for the most appropriate combination stand at: W-LSSVM -Dmey ($R=0.962$, $RMSE=783$, $MAPE=7.682$, $KGE=0.902$, and $PI=0.856$), and W-LSSVM-Bior6.8 ($R=0.929$, $RMSE=1.053$, $MAPE=10.069$, $KGE=0.858$, and $PI=0.886$). Eventually, the best mother wavelet for the W- LSSVM is considered Bior6.8, which results in the highest certainty in comparison with Dmey.

Table 7 reports the W-RSR model’s results with Bior6.8 and mother wavelets Dmey, with it indicating Comb 2 ($R=0.961$, $RMSE=0.749$, $MAPE=7.595$, $KGE=0.952$, and $PI=0.857$) and Comb 3 ($R=0.927$, $RMSE=1.027$, $MAPE=9.669$, $KGE=0.902$, and $PI=0.802$) as the best combinations for Dmey and Bior6.8, respectively. Therefore, Dmey considered the best mother wavelet that makes powerful performance and more accuracy than Bior6.8.

In Fig. 11, the scatter plot of forecasted values versus experimental ones is demonstrated. The main criterion to confirm the accuracy of the proposed ANFIS model in prediction ability can be the high density of points in the vicinity of the $Y=X$ line. According to this figure, there is a significant coincidence between experimental values and predicted ones. This means the proposed ANFIS model creates acceptable and reliable

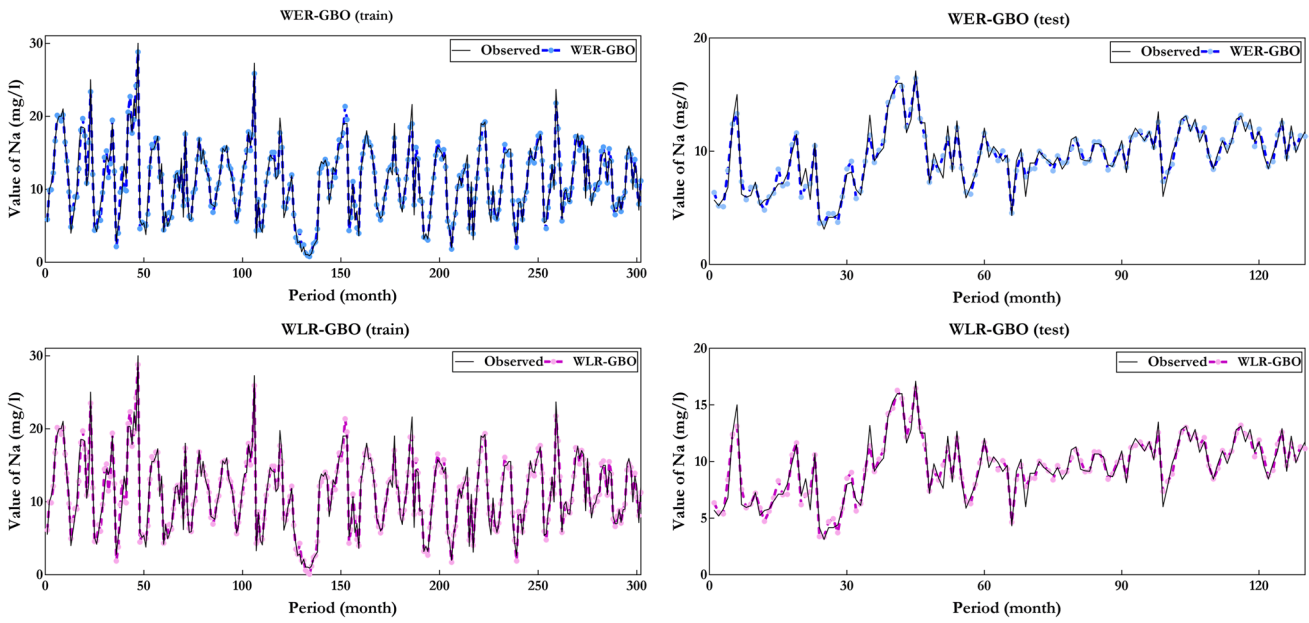


Fig. 14 Variations of the predicted and measured Na for the best performance of W-DI models

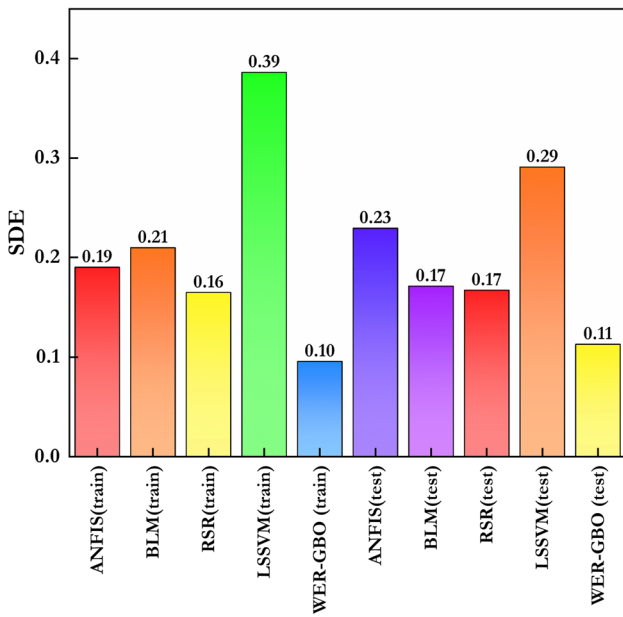


Fig. 15 SDE values of all models for training and testing stages

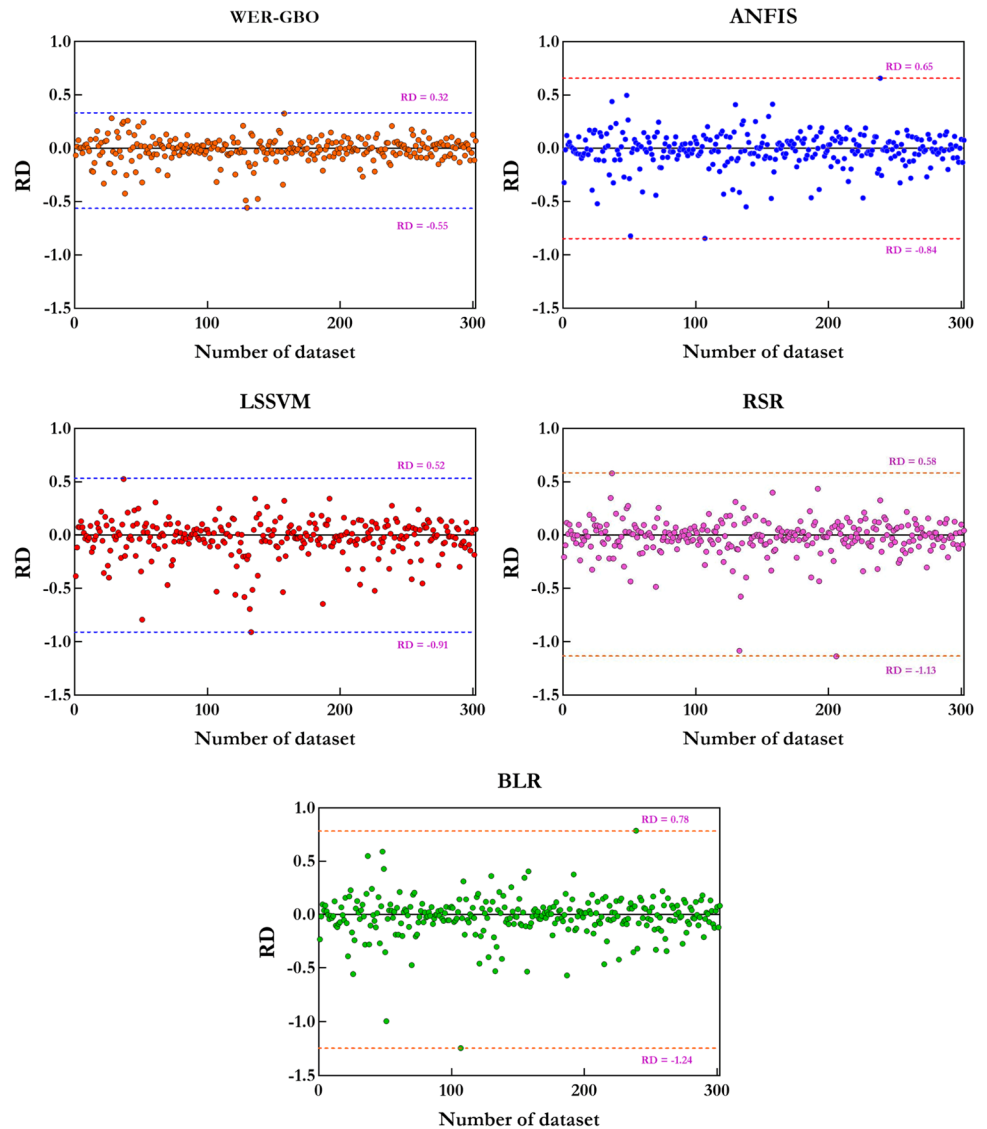
predictions. The time series plots, related to predicted and experienced Na in the testing and training phases, are depicted in Fig. 12. According to this figure, all predictive models reach considerable performance, and the predictive values are equivalent to the estimated Na.

Weighted exponential regression with gradient-based optimization

In this research, two new regression-based models, known as weighted linear regression (WLR) and weighted exponential regression (WER) models and optimized by the GBO algorithm for Na prediction, are introduced. These mentioned models are considered ensemble ones to utilize the most exact outcomes of ANFIS, LSSVM, BLR, and RSR models as inputs in order to forecast Na. To put it simply, the input parameters for WLR-GBO and WER-GBO models are in fact the predicted values gained by W-ANFIS-C4, W-LSSVM-C2, W-BLR-C5, and W-RSR-C2. Table 8 provides the data on statistical factors acquired by WLR-GBO and WER-GBO models for testing and training steps. Table 8 reports the WER-GBO outperforms according to the majority of important criteria for testing ($R=0.972$, $RMSE=0.639$, $MAPE=5.885$, $KGE=0.948$, and $PI=0.990$) and training ($R=0.9813$, $RMSE=0.9671$, $MAPE=7.7907$, $KGE=0.9731$, and $PI=0.987$) phases. Thus, it can be concluded that the WER-GBO model is able to reach more meticulous and accurate Na prediction compared to the WLR-GBO model. By drawing the scatter plots of these two models (Fig. 13), it reveals that they can predict Na meticulously because the calculated values in both models are found near the 45° straight line of the scatter plots.

In Fig. 14, the time series plots of predicted and observed values of the WER-GBO and WLR-GBO are illustrated, with the figure showing a close agreement between predicted and measured values. Overall, by comparing the computational intelligence strategies, it can be gained that both models, WER-GBO

Fig. 16 RD of all DI models in forecasting Na



and WLR-GBO, are concomitant with sufficient performance for both testing and training steps.

Comparison of the performance of all DI models

The analysis of the model is conducted to detect the most efficient model. That is, DI models, namely ANFIS, LSSVM, BLR, RSR, and WER-GBO, are employed for Na prediction. From what has been gained by DI models, the WER-GBO is able to bring about more accurate prediction amongst other models during the training and testing steps. In addition, Fig. 15 displays the prediction outcomes' SDEs for diverse water quality predictors. In this regard, the SDEs in the WER-GBO model for training and testing steps stand out at 0.10 and 0.11, gaining the first rank. Indeed, it can be ratification for suitable stability of the WER-GBO model in Na prediction.

Relative deviation (RD) analyzes the error to address the accuracy of the proposed model. The distribution of RD, $RD = \frac{Na_M - Na_{DI}}{Na_M}$, relative deviation, for predictive models during the testing stages was shown. In Fig. 16, it is apparent that the compression of error distribution in the WER-GBO model is far more than in other models. Besides, by constituting the relative deviation range of methods, it can be concluded WER-GBO model includes the minimum deviation range ($-0.55 \leq RD \leq 0.32$), which leads to better performance than ANFIS ($-0.84 \leq RD \leq 0.65$), LSSVM ($-0.91 \leq RD \leq 0.52$), RSR ($-1.13 \leq RD \leq 0.58$), and BLR ($-1.24 \leq Er \leq 0.78$) models.

The standard deviation and correlation coefficient (R), according to the Taylor diagram, are used to scrutinize the models' ability and performance, which elaborates the efficiency of models (Jamei et al. 2020). As can be seen in the diagram,

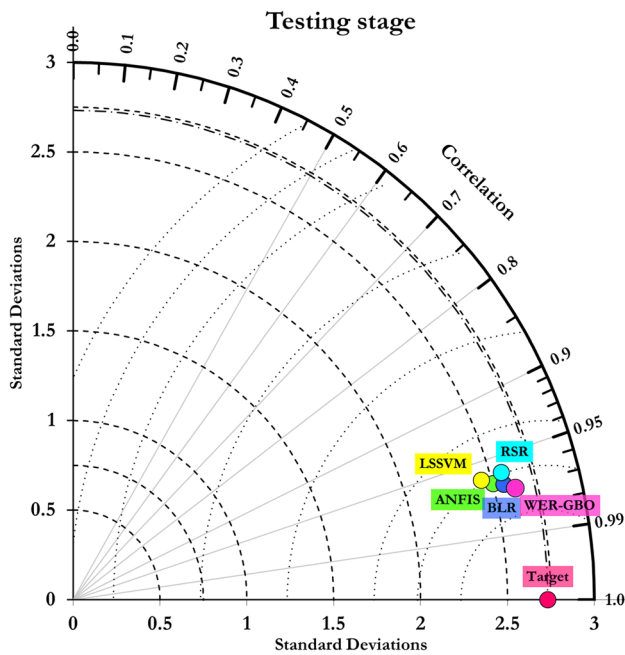


Fig. 17 Taylor diagram of all DI models for forecasting Na

by considering the standard deviation and R , there is a more compelling and tangible relationship between experienced and predicted Na. In Fig. 17, the Taylor diagram is depicted, while connecting the present monthly Na with all DI models. Hence, the best performance in the forecasting of Na is achieved by the WER-GBO model amongst other models while being the nearest model to the target.

Conclusion

Prediction of water quality parameters using authoritative and precise models has a profound impact on saving time and cost. The main objective of this research is to develop two state-of-the-art data intelligence methods based on the weighted exponential regression hybrid with GBO (WER-GBO) and the weighted linear regression hybrid with GBO (WLR-GBO) to forecast Na parameter. Indeed, these mentioned models utilize exponential and linear regression relationships with the Cauchy function as a weighted function to raise forecasting accuracy. Furthermore, the role of the GBO algorithm is to optimize the fundamental parameters in the WER model. In this regard, in the first step, in which the ANFIS, LSSVM, BLR, and RSR models are utilized, the proposed model is performed to reach the Na prediction. Consequently, the input variables in the WER-GBO model are the predicted values of Na gained by the mentioned DI models. In this research, the aim of developing a new ensemble model is to boost forecasting accuracy.

Thus, the ACF, PACF, and CC in this study are employed to identify the Na lags' number. Likewise, the measured time series of Na as well as Q on a monthly basis during the over 36 years in the Maroon River, located in southern Iran, are used in this study. The most appropriate subset regression employed in this process detects the input variable combinations' number. Because the relationship of Na and time series is nonlinear and sophisticated, the wavelet transform is utilized with three decomposition grades to maximize the prediction certainty in ANFIS, LSSVM, BLR, and RSR.

Monthly Na prediction is implemented in surface water by wavelet-DI models, using Dmey and Bior6.8 as two mother wavelets. Having integrated with ANFIS, LSSVM, BLR, and RSR models to forecast Na, Dmey confirmed the best and outstanding progress in the simulation accuracy grade, while Bior6.8 reached a suitable performance. The W-BLR-Dmey in Combo 5 gained striking ability in monthly Na prediction ($R=0.968$, $RMSE=0.682$, and $KGE=0.931$), followed by W-ANFIS-Dmey model in Combo 4 W-RSR-Dmey in Combo 2, and W-LSSVM-Dmey in Combo 2. Overall, the assessment of all DI-based models confirms that the supplementary model of W-BLR can accurately forecast the Na parameter.

The predicted values obtained by the mentioned DI models are applied for the input variables in WER-GBO and WLR-GBO models. From what has been gained as these models' outcomes, it can be concluded that WER-GBO outperforms WLR-GBO according to results of testing step as $R=0.9813$, $RMSE=0.967$, and $KGE=0.973$ for training stage and $R=0.9721$, $RMSE=0.9639$, and $KGE=0.948$. Moreover, by considering SDE and RD as statistical errors, the proposed WER-GBO has the advantage over ANFIS, LSSVM, BLR, and RSR models thanks to reaching more meticulous and accurate Na prediction. Based on the Taylor diagram, the predicted value of Na via WER-GBO has more similar features such as standard deviation and correlation to the measured Na than other models. Hence, the ensemble WER-GBO-based method, gathering the benefits of all complementary techniques, can be a breakthrough in predicting water quality parameters, particularly surface water.

Authors contributions Iman Ahmadianfar, Seyedehelham Shirvani-Hosseini, Arvin Samadi-Kouchehsaraee, and Zaher Mundher Yaseen contributed to data curation, formal analysis, methodology, investigation, visualization, writing—original draft, review and editing, draft preparation, resources, and software.

Data availability Data will be provided upon request from authors.

Declarations

Ethical approval The manuscript is conducted within the ethical manner advised by the Environmental Science and Pollution Research.

Consent to participate Not applicable

Consent to publish The research is scientifically consent to be published.

Competing interests The authors declare no conflict of interest.

References

- Abba SI, Hadi SJ, Sammen SS et al (2020) Evolutionary computational intelligence algorithm coupled with self-tuning predictive model for water quality index determination. *J Hydrol* 587:124974. <https://doi.org/10.1016/j.jhydrol.2020.124974>
- Abobakr Yahya AS, Ahmed AN, Binti Othman F, Ibrahim RK, Afan HA, El-Shafie A, Fai CM, Hossain MS, Ehteram M, Elshafie A (2019) Water quality prediction model based support vector machine model for ungauged river catchment under dual scenarios. *Water* 11:1231
- Ahmadianfar I, Bozorg-Haddad O, Chu X (2020a) Gradient-based optimizer: a new metaheuristic optimization algorithm. *Inf Sci* 540:131–159
- Ahmadianfar I, Heidari AA, Gandomi AH, Chu X, Chen H (2021) RUN beyond the metaphor: an efficient optimization algorithm based on Runge Kutta method. *Expert Syst Appl* 181:115079
- Ahmadianfar I, Heidari AA, Noshadian S, Chen H, Gandomi AH (2022) INFO: An efficient optimization algorithm based on weighted mean of vectors. *Expert Syst Appl* pp 116516
- Ahmadianfar I, Jamei M, Chu X (2020b) A novel hybrid wavelet-locally weighted linear regression (W-LWLR) model for electrical conductivity (EC) prediction in surface water. *J Contam Hydrol* 232:103641
- Al-Sulttani AO, Al-Mukhtar M, Roomi AB, Farooque AA, Khedher KM, Yaseen ZM (2021) Proposition of new ensemble data-intelligence models for surface water quality prediction. *IEEE Access* 9:108527–108541
- Asadollah SBHS, Sharafati A, Motta D, Yaseen ZM (2020) River water quality index prediction and uncertainty analysis: A comparative study of machine learning models. *J Environ Chem Eng*. <https://doi.org/10.1016/j.jece.2020.104599>
- Baesens B, Viaene S, Van Gestel T, Suykens JA, Dedene G, De Moor B, Vanthienen J (2000) An empirical assessment of kernel type performance for least squares support vector machine classifiers. In *KES'2000. Fourth International Conference on Knowledge-Based Intelligent Engineering Systems and Allied Technologies. Proceedings (Cat. No. 00TH8516) IEEE, (Vol. 1, pp. 313–316)*
- Barzegar R, Adamowski J, Moghaddam AA (2016) Application of wavelet-artificial intelligence hybrid models for water quality prediction: a case study in Aji-Chay River. *Iran Stoch Environ Res Risk Assess* 30:1797–1819
- Barzegar R, Asghari Moghaddam A, Adamowski J, Ozga-Zielinski B (2018) Multi-step water quality forecasting using a boosting ensemble multi-wavelet extreme learning machine model. *Stoch Env Res Risk Assess* 32:799–813
- Bezerra MA, Santelli RE, Oliveira EP, Villar LS, Escalera LA (2008) Response surface methodology (RSM) as a tool for optimization in analytical chemistry. *Talanta* 76:965–977
- Box GE, Tiao GC (2011) Bayesian inference in statistical analysis. John Wiley & Sons (Vol. 40), John Wiley & Sons
- Bozorg-Haddad O, Soleimani S, Loáiciga HA (2017) Modeling water-quality parameters using genetic algorithm–least squares support vector regression and genetic programming. *J Environ Eng* 143:04017021
- Chakraborty T, Chakraborty AK, Mansoor Z (2019) A hybrid regression model for water quality prediction. *Opsearch* 56:1167–1178
- Chatterjee S, Sarkar S, Dey N, Sen S, Goto T and Debnath NC (2017) Water quality prediction: multi objective genetic algorithm coupled artificial neural network based approach. Pages 963–968 in 2017 IEEE 15th International Conference on Industrial Informatics (INDIN)
- Chen SH, Jakeman AJ, Norton JP (2008) Artificial intelligence techniques: an introduction to their use for modelling environmental systems. *Math Comput Simul* 78:379–400
- Chen Y, Song L, Liu Y, Yang L, Li D (2020) A review of the artificial neural network models for water quality prediction. *Appl Sci* 10:5776
- Claeskens G, Hjort NL (2008) Model selection and model averaging. Cambridge Books
- Dai D, Brouwer R, Lei K (2021) Measuring the economic value of urban river restoration. *Ecol Econ* 190:107186
- Draper NR and Smith H (1998) Applied regression analysis. John Wiley & Sons
- Dunbabin M, Marques L (2012) Robots for environmental monitoring: significant advancements and applications. *IEEE Robot Autom Mag* 19:24–39
- Fan Y, Dong H, Jiang Y, Pan J, Fan S, Gui G (2018) An Intelligent Water Regimen Monitoring System. In *International Conference in Communications, Signal Processing, and Systems*. Springer, Singapore, pp 829–835
- Feng Z-K, Niu W-J, Tang Z-Y, Xu Y, Zhang H-R (2021) Evolutionary artificial intelligence model via cooperation search algorithm and extreme learning machine for multiple scales nonstationary hydrological time series prediction. *J Hydrol* 595:126062
- de Macedo Machado Freire PK, Santos CAG, da Silva GBL (2019) Analysis of the use of discrete wavelet transforms coupled with ANN for short-term streamflow forecasting. *Appl Soft Comput* 80:494–505
- Geyer CJ (1992) Practical markov chain monte carlo. *Statistical science* pp 473–483
- Gilmour SG (1996) The interpretation of Mallows's Cp-statistic. *J Royal Stat Soc: Ser D (The Statistician)* 45(1):49–56
- Giri S (2021) Water quality prospective in Twenty First Century: status of water quality in major river basins, contemporary strategies and impediments: a review. *Environ Pollut* 271:116332
- Gujarati DN, Porter DC, Gunasekar S (2012) Basic econometrics. Tata mcgraw-hill education
- Gunst RF (1996) Response surface methodology: process and product optimization using designed experiments. Taylor & Francis
- Gupta HV, Bastidas L, Sorooshian S, Shuttleworth W, Yang Z, Kling H, Yilmaz KK, Martinez GF (2009) Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. *J Hydrol* 377:80–91
- Haghiabi AH, Nasrolahi AH, Parsaie A (2018) Water quality prediction using machine learning methods. *Water Qual Res J* 53:3–13
- Hauke J, Kossowski T (2011) Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data. *Quaestiones Geographicae* 30:87–93
- Hou P, Jolliet O, Zhu J, Xu M (2020) Estimate ecotoxicity characterization factors for chemicals in life cycle assessment using machine learning models. *Environ Int* 135:105393
- Huang M, Zhang T, Ruan J, Chen X (2017) A new efficient hybrid intelligent model for biodegradation process of DMP with fuzzy wavelet neural networks. *Sci Rep* 7:1–9
- Jamei M, Ahmadianfar I, Chu X, Yaseen ZM (2020) Prediction of surface water total dissolved solids using hybridized wavelet-multigene genetic programming: new approach. *J Hydrol* 589:125335
- Jamei M, Ahmadianfar I, Karbasi M, Jawad AH, Farooque AA, Yaseen ZM (2021) The assessment of emerging data-intelligence technologies for modeling Mg+2 and SO4-2 surface water quality. *J Environ Manag* 300:113774
- Jamei M, Karbasi M, Olumegbon IA, Mosharaf-Dehkordi M, Ahmadianfar I, Asadi A (2021) Specific heat capacity of molten salt-based nanofluids in solar thermal applications: a paradigm of two modern ensemble machine learning methods. *J Mol Liq* 335:116434
- Jang J-S (1993) ANFIS: adaptive-network-based fuzzy inference system. *IEEE Trans Syst Man Cybern* 23:665–685

- Khan MSI, Islam N, Uddin J, Islam S, Nasir MK (2021) Water quality prediction and classification based on principal component regression and gradient boosting classifier approach. *J King Saud University-Comput Inf Sci*
- Kobayashi M, Sakata S (1990) Mallows' Cp criterion and unbiasedness of model selection. *J Econ* 45:385–395
- Li J, Garshick E, Hart JE, Li L, Shi L, Al-Hemoud A, Huang S, Koutarakis P (2021) Estimation of ambient PM_{2.5} in Iraq and Kuwait from 2001 to 2018 using machine learning and remote sensing. *Environ Int* 151:106445
- Li X, Xu Y, Li M, Ji R, Dolf R, Gu X (2021b) Water quality analysis of the Yangtze and the Rhine River: a comparative study based on monitoring data from 2007 to 2018. *Bull Environ Contam Toxicol* 106:825–831
- Maier H, Kapelan Z, Kasprzyk J, Kollat J, Matott L, Cunha M, Dandy G, Gibbs M, Keedwell E, Marchi A (2014) Environmental Modelling & Software Evolutionary algorithms and other metaheuristics in water resources: current status, research challenges and future directions*. *Environ Model Softw* 62:271–299
- Malik A, Kumar A, Singh RP (2019) Application of heuristic approaches for prediction of hydrological drought using multi-scalar streamflow drought index. *Water Resour Manage* 33:3985–4006
- Mohammadi K, Shamshirband S, Petković D, Khorasanizadeh H (2016) Determining the most important variables for diffuse solar radiation prediction using adaptive neuro-fuzzy methodology; case study: City of Kerman. *Iran Renew Sustain Energy Rev* 53:1570–1579
- Mutiarani V, Setiawan A and Parhusip HA (2012) Estimasi Parameter dan Interval Kredibel dengan Model Regresi Linier Berganda Bayesian.in Seminar Nasional Pendidikan Matematika Ahmad Dahlan 2012 (SENDIKMAD 2012) Universitas Ahmad Dahlan
- Nourani V, Baghanam AH, Adamowski J, Kisi O (2014) Applications of hybrid wavelet–artificial intelligence models in hydrology: a review. *J Hydrol* 514:358–377
- Orouji H, Bozorg Haddad O, Fallah-Mehdipour E, Mariño M (2013) Modeling of water quality parameters using data-driven models. *J Environ Eng* 139:947–957
- Özban AY (2004) Some new variants of Newton's method. *Appl Math Lett* 17:677–682
- Pandey M, Jamei M, Karbasi M, Ahmadianfar I, Chu X (2021) Prediction of maximum scour depth near spur dikes in uniform bed sediment using stacked generalization ensemble tree-based frameworks. *J Irrig Drain Eng* 147:04021050
- Rajaei T, Khani S, Ravansalar M (2020) Artificial intelligence-based single and hybrid models for prediction of water quality in rivers: a review. *Chemom Intell Lab Syst* 200:103978
- Ravansalar M, Rajaei T (2015) Evaluation of wavelet performance via an ANN-based electrical conductivity prediction model. *Environ Monit Assess* 187:1–16
- Rubio FJ, Genton MG (2016) Bayesian linear regression with skew-symmetric error distributions with applications to survival analysis. *Stat Med* 35:2441–2454
- Samui P, Kothari D (2011) Utilization of a least square support vector machine (LSSVM) for slope stability analysis. *Scientia Iranica* 18:53–58
- Song C, Yao L, Hua C, Ni Q (2021) A novel hybrid model for water quality prediction based on synchrosqueezed wavelet transform technique and improved long short-term memory. *J Hydrol* 603:126879
- Suykens JA, Vandewalle J (1999) Least squares support vector machine classifiers. *Neural Process Lett* 9:293–300
- Tabari H, Marofi S, Ahmadi M (2011) Long-term variations of water quality parameters in the Maroon River. *Iran Environ Monit Assess* 177:273–287
- Tao H, Bobaker AM, Ramal MM, Yaseen ZM, Hossain MS, Shahid S (2019) Determination of biochemical oxygen demand and dissolved oxygen for semi-arid river environment: application of soft computing models. *Environ Sci Pollut Res* 26:923–937
- Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram. *J Geophys Res Atmos* 106:7183–7192
- Teräsvirta T, Lin CF, Granger CW (1993) Power of the neural network linearity test. *J Time Ser Anal* 14:209–220
- Tian JW, Qi C, Peng K et al (2022) Improved permeability prediction of porous media by feature selection and machine learning methods comparison. *J Comput Civ Eng* 36:4021040
- Tiyasha T, Tung TM, Bhagat SK, Tan ML, Jawad AH, Mohtar WHMW, Yaseen ZM (2021a) Functionalization of remote sensing and on-site data for simulating surface water dissolved oxygen: Development of hybrid tree-based artificial intelligence models. *Marine Pollut Bull* 170:112639
- Tiyasha Tung TM, Yaseen ZM (2021b) Deep learning for prediction of water quality index classification: tropical catchment environmental assessment. *Nat Resour Res* 30(6):4235–4254
- Tiyasha TTM, Yaseen ZM (2020) A survey on river water quality modelling using artificial intelligence models: 2000–2020. *J Hydrol* 585:124670. <https://doi.org/10.1016/j.jhydrol.2020.124670>
- Uddin MG, Nash S, Olbert AI (2021) A review of water quality index models and their use for assessing surface water quality. *Ecol Indic* 122:107218
- Upriety S, Dangol B, Nakarmi P, Dhakal I, Sherchan SP, Shisler JL, Jutla A, Amarasiri M, Sano D, Nguyen TH (2020) Assessment of microbial risks by characterization of Escherichia coli presence to analyze the public health risks from poor water quality in Nepal. *Int J Hyg Environ Health* 226:113484
- van Vliet MT, Jones ER, Flörke M, Franssen WH, Hanasaki N, Wada Y, Yearsley JR (2021) Global water scarcity including surface water quality and expansions of clean water technologies. *Environ Res Lett* 16:024020
- Vapnik V, Guyon I, Hastie T (1995) Support vector machines. *Mach Learn* 20:273–297
- Vasistha P, Ganguly R (2020) Water quality assessment of natural lakes and its importance: an overview. *Mater Today: Proc* 32:544–552
- Wu H, Yang W, Yao R, Zhao Y, Zhao Y, Zhang Y, Yuan Q, Lin A (2020) Evaluating surface water quality using water quality index in Beiyun River, China. *Environ Sci Pollut Res* 27:35449–35458
- Yaseen ZM, Sulaiman SO, Deo RC, Chau K-W (2019) 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
- Yaseen ZM, Ramal MM, Diop L et al (2018) Hybrid adaptive neuro-fuzzy models for water quality index estimation. *Water Resour Manag*. <https://doi.org/10.1007/s11269-018-1915-7>
- Yetilmezsoy K, Ozkaya B, Cakmakci M (2011) Artificial intelligence-based prediction models for environmental engineering. *Neural Network World* 21:193

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