



Study of water resources parameters using artificial intelligence techniques and learning algorithms: a survey

Mojtaba Poursaeid¹ · Amir Hossein Poursaeed² · Saeid Shabanlou³

Received: 6 October 2021 / Accepted: 8 April 2022 / Published online: 13 May 2022
© The Author(s) 2022

Abstract

Qualitative analysis of water resources is one of the most widely used topics in water resources research today. Researchers use various analysis methods of water parameters to achieve the desired goals in this field. This research uses artificial intelligence (AI), learning machine (LM), data mining, and mathematical techniques to simulate water behavior and estimate its parametric changes. The proposed model used in this study was a Self-adaptive Extreme learning machine (SAELM) to estimate hydrogeological parameters of the Meghan wetland located in Markazi province in Iran. In addition, SAELM simulation results were compared to Least square support vector machine (LSSVM), Multiple linear regression (MLR), and Adaptive Neuro-fuzzy inference system (ANFIS) models. The simulated parameters were Electrical Conductivity (EC), Total Dissolved Solids (TDS), Groundwater Level (GWL), and salinity. This information was related to sampling for 175 months in the study area. Finally, after simulation operation, four models were introduced as superior models. Mentioned exceptional models were SAELM in GWL modeling, SAELM in modeling the EC, MLR in salinity simulation, and LSSVM in the simulation of TDS parameters. Moreover, by five approaches, the models' performance was evaluated. Suggested strategies were performance evaluation by statistical indicators, Wilson score method uncertainty analysis (WSMUA), response & correlation plots, discrepancy ratio charts, and distribution error diagrams. Based on statistical indicators, the SAELM_{GWL} model was the most accurate model with RMSE, MAPE, and R^2 indices equal to 0.1496, 0.0043, and 0.9933, respectively. The ANFIS model had the worst results in simulation.

Keyword Data mining · LSSVM · MLR · SAELM · ANFIS · Water parameters

Introduction

Basic statement

Needs assessment proved that one of the vital human needs is access to drinking water. For this reason, human beings have started various works such as digging wells, building dams, aqueducts, and so on. But today, there are many concerns. For example, we have recently seen global warming climate change. The increase in the human population is

another cause for concern. But with the advancement of science, the use of new scientific methods to achieve the goal has also increased. Modeling the water quality parameters is one of the fundamental challenges investigated by several studies (Parmar and Bhardwaj 2014; Zhou 2020). One of the crucial human issues is access to drinking water. In this case, AI and LMs, artificial neural networks (ANN), and engineering sciences were tools to help human beings. Due to water scarcity in countries around the world, the study of water resources in different scientific ways has been considered by researchers (Poursaeid et al. 2020; Chang et al. 2021). The following will be reviewed related studies to water resources and their quantitative and qualitative simulations.

Related studies

Today, artificial intelligence methods have been developed for various purposes. In previous research, artificial intelligence techniques have been used for scientific challenges.

✉ Mojtaba Poursaeid
mojtaba.poursaeid.ac@gmail.com

¹ MPO-Plan and Budget Organization, Deputy of Technical and Engineering, Khorramabad, Lorestan, Iran

² Department of Electrical Engineering, Lorestan University, Khorramabad, Lorestan, Iran

³ Department of Water Engineering, Kermanshah Branch, Islamic Azad University, Kermanshah, Iran

In addition, several studies have been conducted in the field of water resources and related modeling. In these studies, numerical modeling, analytical modeling, artificial intelligence simulations, etc., have been used. But in this research, we review the articles related to water resources modeling and quantitative and qualitative simulations that have used artificial intelligence.

On the one hand, some studies used ANNs and LMs to model water resources separately. ANNs were applied to simulate water parameters. ANNs such as multi-layer perceptron neural network, backpropagation neural network, Radial basis function neural network, and LMs such as support vector machine, extreme learning machine, and reinforcement learning machine. In mentioned works, the accuracy of ANNs and LMs was compared to similar articles. Finally, the superiority of these models over previous works was confirmed (Gholami et al. 2011; Wu et al. 2014; Yang et al. 2014; Kalteh 2014, 2015; Kheradpisheh et al. 2015; Shahid and Ehteshami 2015; Nema et al. 2017; Manu and Thalla 2017; Wang et al. 2019; Qu et al. 2020; Vijay and Kamaraj 2021; Sada and Ikpeseni 2021; Che Nordin et al. 2021; Sarkar et al. 2021). On the other hand, in some other papers, Optimizer & Heuristic algorithms (OHA) simulated quantitative and qualitative water parameters. OHAs such as genetic algorithm, differential evolution, and particle swarm optimization algorithm implemented to model the water parameters. Specifically, the OHAs proportionally improved the modeling's speed and accuracy in these studies. (Parmar and Bhardwaj 2014; Walker et al. 2015; Guneshwor et al. 2018; Vaheddoost and Aksoy 2018; Elkiran et al. 2019). Besides, other articles combined ANNs & LMs with OHAs to model water parameters. The results showed that hybrid models' simulation, speed, accuracy, and ability were often higher than the original models (Heddami et al. 2019; Zhang et al. 2019, 2021; Majumder and Eldho 2020; Azimi and Azhdary Moghaddam 2020; Poursaeid et al. 2020, 2021).

Contributions

In this paper, quantitative and qualitative groundwater parameters were predicted using LM techniques and mathematical methods. However, there are many articles in this field, but in this study, for the first time in the study area, four AI and mathematical models have been used simultaneously to quantitatively and qualitatively simulate the water of Meghan Wetland located in the Arak plain, Markazi province of Iran. Three models of artificial intelligence SAELM LSSVM, ANFIS, and MLR as a mathematical model, simulated the EC, TDS, GWL, and salinity parameters. It should be noted that there are several parameters in the water resources quality management, including Cl^- , EC, TDS, SO_4^{2+} , etc. Many researchers have used the most

widely used parameters in this field, such as above-mentioned parameters.

The structure of this work in the following sections is as follows: Defining the water quality parameters discussed in the second part. The third part of the article describes materials and methods, including AI models and mathematical models. Moreover, In mentioned section, the characteristics of the study area and the data collection steps are presented. Also, in Sect. 5, the results of the research are discussed. Finally, the general conclusion is summarized in the last section.

Problem statement

In this part, water quality parameters are introduced as follows.

Water quality parameters

This section explains the water quality parameters. Although there are several parameters in this field (Solanki et al. 2015), we have tried to describe the most widely used parameters in this work.

Total dissolved solids (TDS)

Total dissolved solids (TDS) means the numerical sum of solids soluble in water (Jamei et al. 2020). This parameter is measured in milligrams per liter (mg l^{-1}). This parameter includes different types of mineral salts such as soluble bicarbonates (HCO_3^-), organic matter, magnesium (Mg^{+2}), calcium (Ca^{+2}), sodium (Na^+), potassium (K^+), chloride (Cl^-), sulfate (SO_4^{-2}), nitrates (NO_3^-) (Ahmadianfar et al. 2020).

Salinity

Water salinity is one of the water quality parameters. This parameter is equivalent to salt concentration in water according to the definition. However, some research has defined salinity as the concentration of water-soluble mineral salt in a specific volume or weight per square meter (Sparks 2003). It should be noted that some factors such as increased consumption or high evaporation rate lead to increased salinity (Harris 2009). It should be noted that this parameter is measured in milligrams per liter (mg l^{-1}).

Electrical conductivity (EC)

Electrical conductivity (EC) is one of the quality parameters of water sources and is measured in micro Siemens per centimeter (μScm^{-1}). EC is considered equivalent to the salinity

of water (Serrano-Finetti et al. 2019). Because this parameter indicates the number of salts in the water and is equal to the amount of electrical transfer of water, this parameter is an essential factor in drinking water quality and agriculture. However, the amount of ionic salt in water reduces its quality for drinking (Ahmadianfar et al. 2020).

Materials and methods

Study area

The study area of this research is Meghan Wetland, located in Markazi province in Iran as Fig. 1. According to the statistical results of synoptic stations, the maximum and

minimum precipitation varies from 461 mm in the northeast to 208 mm in the center of Arak plain.

AI models

In the following sections, the AI and MLR models will be explained.

Self Adaptive Extreme Learning Machine (SAELM)

The proposed method in this study is SAELM model. Extreme learning machine was proposed in 2004 by Huang et al. (2004). This model is one of the learning



Fig. 1 Meghan wetland. Source: Wikimedia and GoogleMap

machine types, and in various research, its superiority over other methods (including neural networks and learning machines) has been proved (Huang et al. 2006, 2012). If we have n neurons in the hidden layer, then we can define the single-layer feedforward network of the learning machine based on mathematical relations as follows (Liang et al. 2006; Poursaeid et al. 2020):

$$y_i = \sum_{j=1}^k \beta_j g(a_j x_i + c_j) = u_i, \quad i = 1, 2, 3, \dots, N \quad (1)$$

So that g , c_i and β_i are the transfer function between input and output layers, respectively. The above relation can be rewritten in the form of the following:

$$\begin{aligned} c_j &= [c_{j1}, c_{j2}, \dots, c_{jn}]^T, c_j \in R, \quad j = 1, 2, 3, \dots, k \\ a_j &= [a_{j1}, a_{j2}, \dots, a_{jn}]^T, a_j \in R^t, \quad j = 1, 2, 3, \dots, k \\ \beta_j &= [\beta_{j1}, \beta_{j2}, \dots, \beta_{jn}]^T, \beta_j \in R^L, \quad j = 1, 2, 3, \dots, k \end{aligned} \quad (2)$$

$$G\beta = Y$$

$$G = \begin{bmatrix} g(x_1; c_1, w_1) & \dots & g(x_1; c_k, w_k) \\ \vdots & \ddots & \vdots \\ g(x_n; c_1, w_1) & \dots & g(x_n; c_k, w_k) \end{bmatrix}_{n \times k} \quad (3)$$

$$\beta = (\beta_1^T, \beta_2^T, \dots, \beta_h^T)_{h \times m}^T \quad (4)$$

Finally, the output weights can be calculated using the Moore–Penrose generalized inverse matrix method.

$$\beta = G^+ Y \quad (5)$$

Least square support vector machine (LSSVM)

LSSVM model is a type of SVMs that can adjust the constant factors of the support vector model with least-square solutions and self-adapting changes. The support vector model was developed by Vapnik (Sapankevych and Sankar 2009). These learning machines operate based on *Structural risk minimization*. Meanwhile, some other AI methods use the *Empirical risk minimization* method. (Cristianini and Shawe-Taylor 2000; Dibike et al. 2001). SVM is used in classification problems. In short, an equation is obtained in a *Quadratic programming* problem in this theory. The fixed parameters of the model are determined, and then with optimization algorithms such as genetic algorithm (GA) or other methods, we can get the optimal values for this equation. The SVM can also be used for regression problems. The mathematical definition of LSSVM is that if x_i and y_i are the inputs and outputs, then

the nonlinear regression function is as follows (Valyon and Horvath 2007):

$$f(x) = \sum_{i=1}^n w_i \varphi_i(x) + b = w^T \cdot \varphi(x_i) + b \quad (6)$$

where w is the weight vector, and b is the bias, and φ are nonlinear functions for mapping data into large feature spaces, so:

$$w = \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}, \quad \varphi = \begin{bmatrix} \varphi_1 \\ \vdots \\ \varphi_n \end{bmatrix} \quad (7)$$

The nonlinear regression problem can be solved by minimizing the following Quadratic programming problem:

$$\text{Min}_{w,e,b} \left\{ \lambda(w, e) = \frac{1}{2} w^T w + \frac{1}{2} C \sum_{i=1}^m e_i^2 \right\} \quad (8)$$

where C is the tradeoff variable between two terms of the equation, so:

$$Y_i = w^T \cdot \varphi(x_i) + b + e_i, \quad Y_i = y_i + \delta_i \quad (9)$$

It should be noted that δ_i is defined as network noise. Then for each x_i , the output is a weighted set of n kernel functions, in which the central variable of the kernel functions is determined using the x_i as inputs data. We will have the Lagrangian form of the equation as follows:

$$\psi(w, b, e, \alpha) = \lambda(w, e) - \sum_{i=1}^n a_i (w^T \cdot \varphi(x_i) + b + e_i - Y_i) \quad (10)$$

In Eq. 10, a_i are lagrangian multipliers. In the following solution, we solve the problem with a constrained optimization problem. Then we have the optimization with the following conditions as follows:

$$\begin{aligned} \frac{\partial \psi}{\partial w} = 0 &\rightarrow w = \sum_{i=1}^m a_i \varphi(x_i) \\ \frac{\partial \psi}{\partial b} = 0 &\rightarrow \sum_{i=1}^m a_i = 0 \\ \frac{\partial \psi}{\partial e_i} = 0 &\rightarrow a_i = C e_i, \quad i = 1, 2, \dots, m \\ \frac{\partial \psi}{\partial a_i} = 0 &\rightarrow w^T \varphi(x_i) + b + e_i - Y_i = 0, \quad i = 1, 2, \dots, m \end{aligned} \quad (11)$$

And the final solution to the problem is as follows:

$$\begin{bmatrix} 0 & 1 & 1 & \dots & 1 \\ 1 & \Phi_{11} + \frac{1}{C} & \Phi_{12} & \dots & \Phi_{1m} \\ 1 & \Phi_{21} & \Phi_{12} + \frac{1}{C} & \dots & \Phi_{2m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \Phi_{(m-1)1} & \Phi_{(m-1)2} & \dots & \Phi_{(m-1)m} \\ 1 & \Phi_{m1} & \Phi_{m2} & \dots & \Phi_{mm} + \frac{1}{C} \end{bmatrix} \begin{bmatrix} b \\ a_1 \\ a_1 \\ \vdots \\ a_m \end{bmatrix} = \begin{bmatrix} 0 \\ Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{bmatrix}$$

$Y = [Y_1, \dots, Y_m], a = [a_1, \dots, a_m], 1 = [1, \dots, 1]$

(12)

And in Eq. 12, the $\Phi_{i,j}$ is the kernel matrix. Also, $\phi(x_i, x_j)$ is the kernel functions:

$$\Phi_{ij} = \phi(x_i, x_j) \quad , \quad \Phi_{ij} = \varphi^T(x_i) \cdot \varphi(x_j) \quad , \quad i, j = 1, \dots, m$$

(13)

Adaptive neuro-fuzzy inference system (ANFIS)

Adaptive Neuro-Fuzzy Inference System, or ANFIS for short, is a feed neural network that simulates based on fuzzy logic (El-Shafie et al. 2006). In this type of network, two types of inferential systems based on FIS fuzzy logic are used (Tokachichu and Gaddam 2021; Arora and Keshari 2021):

- A fuzzy inference system-based network, called Mamdani, is known as *M-FIS* for short.
- Takagi–Sugeno fuzzy inference system-based network, known as *TS-FIS* for short.

In these networks, there are at least two inputs I_1 and I_2 , for the network based on TS-FIS fuzzy inference system and two *if-then conditional principles* for each output O_i and the conditional rules of these fuzzy networks are as follows:

Rule (1): If x is input I_1 and output O_1 , then we have:

$$f_1 = a_1x_1 + b_1y_1 + \alpha_1$$

Rule (2): If x is input I_2 and output O_2 , then we have:

$$f_2 = a_2x_2 + b_2y_2 + \alpha_2$$

Neuro-fuzzy networks are organized of one input layer and the other five layers, which can be a type of multi-layered neural network.

- **Layer 0:** Input layer with n Input nodes
- **Layer 1:** This layer provides a membership function for points using gaussian rules by fuzzifying each node.

$$\mu_{D_i}(x) = \exp \left\{ - \left[\left(\frac{x - h_i}{\varphi_i} \right)^2 \right]^{t_i} \right\}$$

(14)

where φ_i , t_i , and h_i are adaptive functions in a fuzzy network.

- **Layer 2:** all of fuzzified data are passed into operators. Also, I_i and O_i are the membership parameters of the antecedent parameters of rule (1).

$$w_i = \mu_{D_i}(x) \times \mu_{O_i}(x)$$

(15)

- **Layer 3:** in this layer, All of the nodes are normalized as below:

$$\bar{w}_i = \frac{w_i}{\sum_{j=1}^m w_j}$$

(16)

where the \bar{w}_i , in 2nd layer, is the Sum of Operator in the i^{th} order.

- **Layer 4:** The corresponding linear function is calculated for each node in this layer. Then the coefficients of the functions are computed using the BNN-error (backpropagation neural network error).

$$\bar{w}_i f_i = \bar{w}_i (a_0x_0 + a_1x_1 + a_2)$$

(17)

where a_i is the parameter for the input I , also \bar{w}_i as the output parameter of Layer 3.

- **Layer 5:** This layer is the sum of the outputs of each node from the 4th layer, which is calculated as follows in Eq. 18:

$$\sum \bar{w}_i f_i = \frac{\sum w_i f_i}{\sum w_i}$$

(18)

Multiple Linear Regression (MLR)

Multiple linear regression methods are based on statistical and mathematical calculations. These methods can be used to study the relationships between input variables and multi-target variables. The mathematical definition of this model is as follows:

$$f(x_i) = a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n + \epsilon$$

(19)

where the $f(x_i)$ is a secondary variable, x_i 's are multiple Primary variables, a_i are regression multipliers, and ϵ is a random error in the equation (Çamdevýren et al. 2005; Asadi et al. 2014).

Data collection and performance indicators

Data analysis

The Input dataset was sampled and collected for 175 months in the study area. This work used the primary dataset of the TDS, salinity, GWL, and EC parameters. After collecting data, the K-Fold cross-validation method was used to

randomize the data to improve the accuracy of the models. 70% of the data were used in the training phase, and 30% were assigned for the testing phase.

Performance indicators

In the present study, to evaluate the accuracy of models, the statistical indices: Absolute mean error percent (MAPE), root means square error (RMSE), and coefficient of determination (R^2) are used as Eq. 20–22:

$$R^2 = \frac{(n \sum_{i=1}^n O_i I_i - \sum_{i=1}^n O_i \sum_{i=1}^n I_i)^2}{(n \sum_{i=1}^n (O_i)^2 - \sum_{i=1}^n (O_i))^2 (n \sum_{i=1}^n (I_i)^2 - \sum_{i=1}^n (I_i))^2} \tag{20}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (O_i - I_i)^2} \tag{21}$$

$$MAPE = \frac{\sum_{i=1}^n |O_i - I_i|}{n} \times 100 \tag{22}$$

where I_i the input values, O_i output values, \bar{I} the mean of observational values, and n the number of observational values.

Results and discussions

First, input values were entered for all models. Then EC, TDS, GWL, and Salinity parameters were considered output variables. Moreover, the superior model was selected in each simulation phase for four output variables. Finally, after the simulation, the performance of each model is

examined. In this study, models' performance was evaluated by four different methods: performance evaluation with statistical indicators, performance evaluation by uncertainty analysis by Wilson Score Method (WSMUA), performance evaluation with response figures, performance evaluation with discrepancy ratio charts, and error distribution diagrams (Figs. 2 and 3).

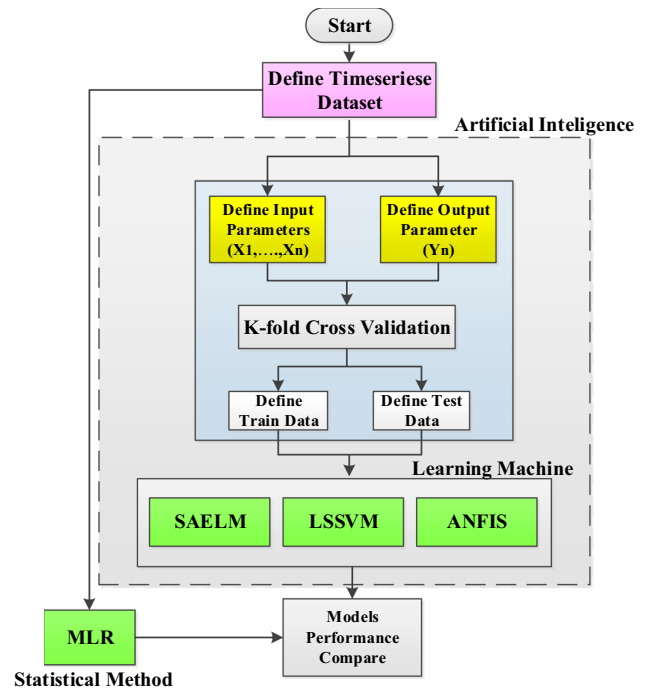


Fig. 3 Research flowchart

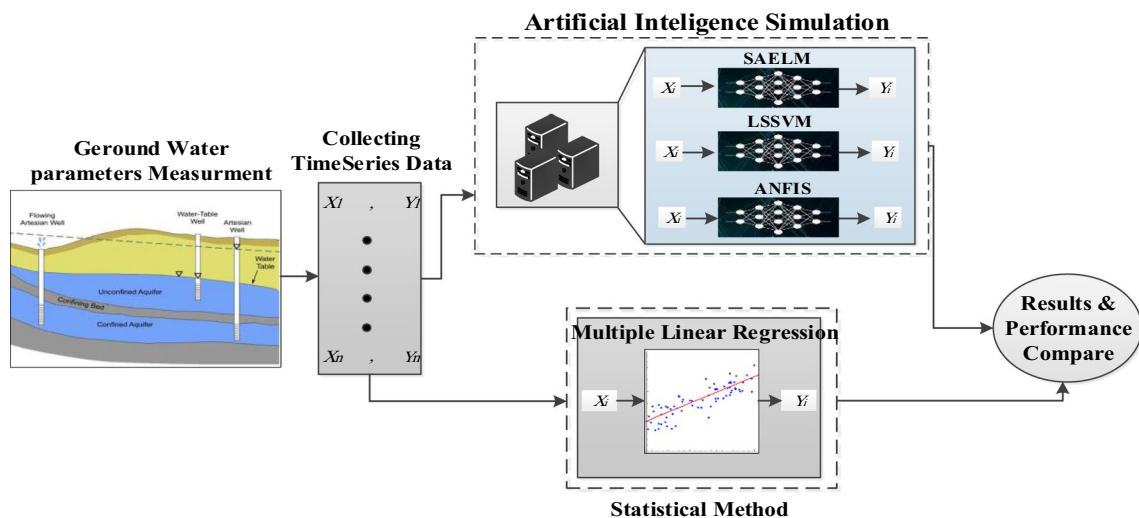


Fig. 2 Graphical abstract

Evaluation indicators

This section investigates the values of evaluation indicators for models performance in the output parameters simulation. In this study, statistical indicators MAPE, RMSE, and R² have been used for evaluation.

According to Fig. 4 and Table 1, the best model in quantitative and qualitative water simulation were known, and bold fonts show the superior model in each output's simulation. The SAELM in GWL and EC modeling, the LSSVM model in TDS simulation, and the MLR in salinity simulation were determined as the best models and presented the most accurate results.

Wilson score method uncertainty analysis (WSMUA)

After simulation, to investigate the prediction error, the amount of uncertainty of models can be calculated, and the performance of models can be evaluated in estimating the targets. The present study performed uncertainty analysis by the WSMUA (Poursaeid et al. 2020; Bonakdari et al. 2020).

Computational parameters in this analysis are prediction error (Δ_i), mean error (Mean), and standard deviation of error values (Std), each of which is calculated according to Eq. 23–25. The results of the uncertainty analysis with a *width of uncertainty band* that denoted by the (WUB), with 95% confidence bound (PEI 95%) was calculated. Also, based on the (Mean) values, if its sign is positive, the model will have an

Table 1 Models evaluation indices

Parameter	Model	RMSE	R ²	MAPE
Salinity	SAELM	24.3524	0.99980001	0.0811
	LSSVM	0.0308	0.99980001	1.36E-04
	ANFIS	50.7474	0.99980001	0.1282
	MLR	5.00E-12	0.80464488	1.28E-14
EC	SAELM	70.9114	0.99798102	1.4424
	LSSVM	96.0246	0.99199608	1.5016
	ANFIS	190.6713	0.969063048	2.6701
	MLR	190.4244	0.967842764	2.86E+00
TDS	SAELM	418.1382	0.917917286	12.7324
	LSSVM	66.473	0.9900648	1.6597
	ANFIS	95.65	0.97495876	2.6226
	MLR	97.6525	0.97223544	2.821
GWL	SAELM	0.1496	0.98664489	0.0043
	LSSVM	0.3775	0.91356752	0.0178
	ANFIS	14.2444	0.67895456	8.392
	MLR	0.6161	0.8156451	0.032

*The bold fonts show the superior models

overestimation performance. If its mean error is negative, the model will have an *underestimation* performance.

$$\Delta_i = O_i - I_i \tag{23}$$

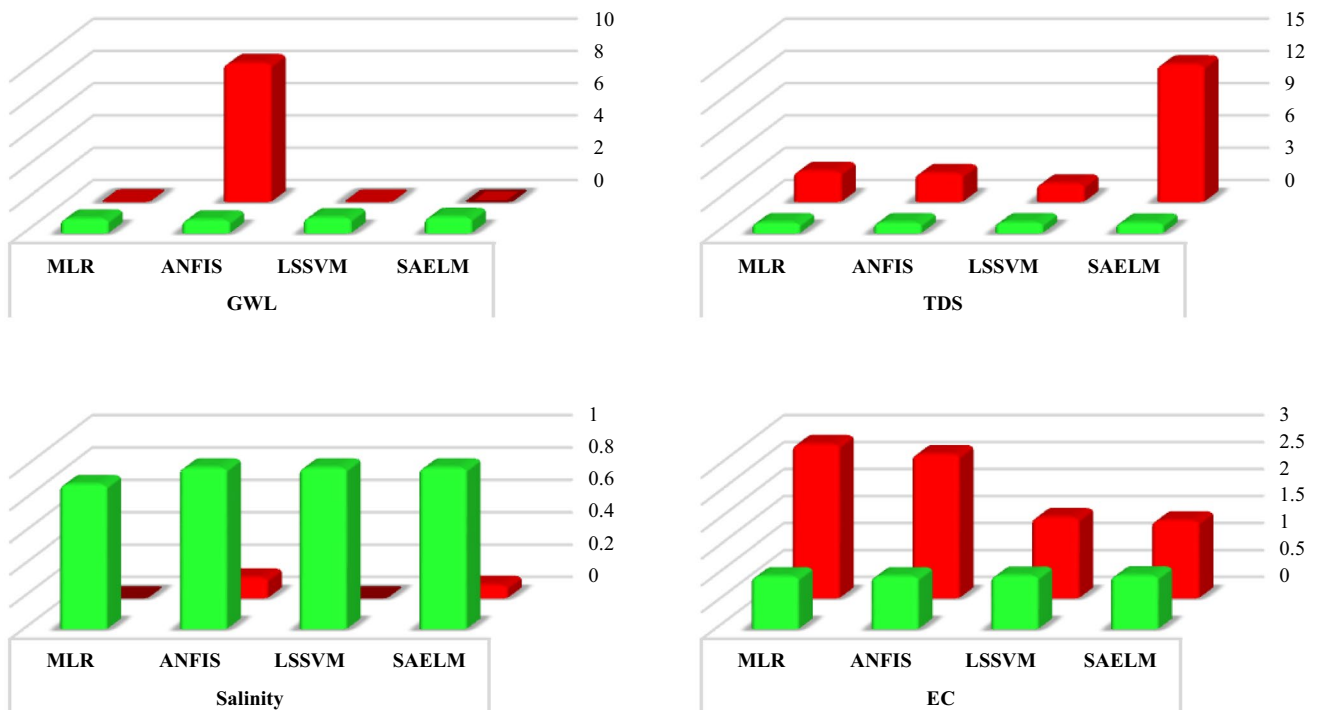


Fig. 4 Models performance indices: MAPE R²

$$\text{Mean} = \frac{1}{n} \sum_{i=1}^n \Delta_i \tag{24}$$

$$\text{Std} = \sqrt{\sum_{i=1}^n (\Delta_i - \text{Mean})^2 / (n - 1)} \tag{25}$$

In the above equations, I_i is the input values, O_i is the output values, and n is the number of observational samples. The results of the uncertainty analysis are shown in Table 2.

According to Table 2 results, the SAELM is the superior model in GWL simulation and had the *Underestimation* performance. The performance of other models was *Overestimation*, and the SAELM model was the most accurate with the minimum of average error equal to 0.0128.

Response and correlation figures

The models' output was illustrated and presented as response plots in this section. In addition, correlation plots also are displayed.

The proposed model had excellent performance in GWL simulation. Moreover, this model was more accurate than over models for simulating the EC parameters. The SAELM had the most accurate plots with the best correlation between observed_predicted values based on Fig. 5 diagrams (Figs. 6, 7, 8 and 9).

Discrepancy ratio and error distribution

According to the mathematical definition of Discrepancy Ratio (DR) written in Eq. 26, according to this definition, the closeness of DR values to a horizontal line (DR = 1) shows the high accuracy of the simulation.

$$\text{DR} = \left[\frac{\text{Value}_{\text{predicted}}}{\text{Value}_{\text{observed}}} \right] \tag{26}$$

The diagram of the DR shows the high accuracy of the SAELM and the MLR models. The closeness of DR values for SAELM and MLR models to the DR line (DR = 1) shows the high accuracy of the mentioned models in GWL and salinity simulation. Additionally, Error percent is defined as Eq. 27.

$$\text{Error Percent} = \left| \frac{\text{predicted} - \text{observed}}{\text{observed}} \right| \times 100 \tag{27}$$

According to the prediction results error in Fig. 10, The SAELM and MLR model has the most Error-percent (100%) in the range of less than 1%. Meanwhile, the LSSVM has the best accuracy in prediction in the next rank. The SAELM in EC simulation had the worst performance with the most error percent in the range of greater than 2%.

Conclusions

This study collected sampling data related to 175 months for groundwater of Meghan Wetland located in Arak plain in Markazi province in Iran. The primary parameters of this study were sampling (t), (TDS), (EC), (Cl), (Salinity), and (GWL). Then, using three AI models and a mathematical model, water parameters modeling was performed. The proposed method in this article was the SAELM model. Other models are LSSVM, ANFIS, and the mathematical model is MLR. After analyzing the results, the performance of the models was evaluated with five approaches: Based on statistical indicators, the best results were recorded for models SAELM_{GWL} simulation and MLR_{salinity}, with the lowest value of RMSE and MAPE. Additionally, the mentioned models had the closest R² values to 1.

Based on the response & correlation plots, the best performance was assigned to the SAELM_{GWL} model with better mapping of simulated values than observed values. Based on the results of the WSMUA, the SAELM_{GWL} model with a minimum mean error value equal to 0.0128 was the best and most accurate model. According to the DR diagram, the SAELM_{GWL} and MLR_{salinity} models had the highest concentration of output points near the DR = 1 line. Also, based on the error distribution percentage diagram, the best forecast accuracy was assigned to SAELM_{GWL} and MLR_{salinity} models with the lowest error percent in the range of less than 1%.

Table 2 Models uncertainty analysis

Model	WUB	Average (Δ_i)	Std	95% PEI		Performance of models
				Lower bound	Upper bound	
SAELM(GWL)	0.033	-0.0128	0.11535	-0.04665	0.02021	Underestimated
SAELM(EC)	9.0267	9311.883	9527.88	9302.8563	9320.9097	Overestimated
LSSVM(TDS)	15.0058	5248.172	5346.963	5233.1662	5263.1778	Overestimated
MLR(Salinity)	1.29E-12	53,897.167	57,759/64	53,897.167	53,897.167	Overestimated

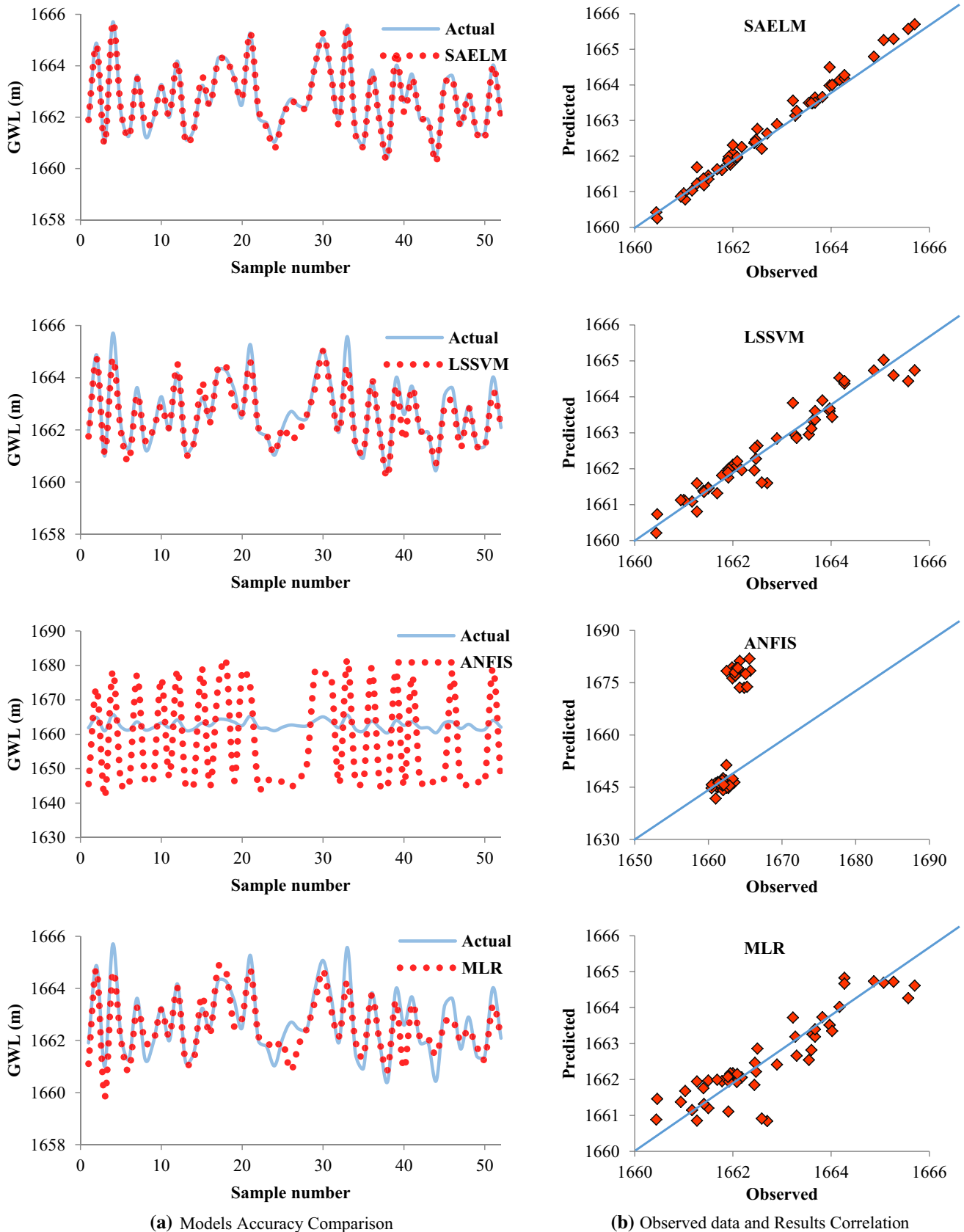


Fig. 5 Model performance in GWL simulation

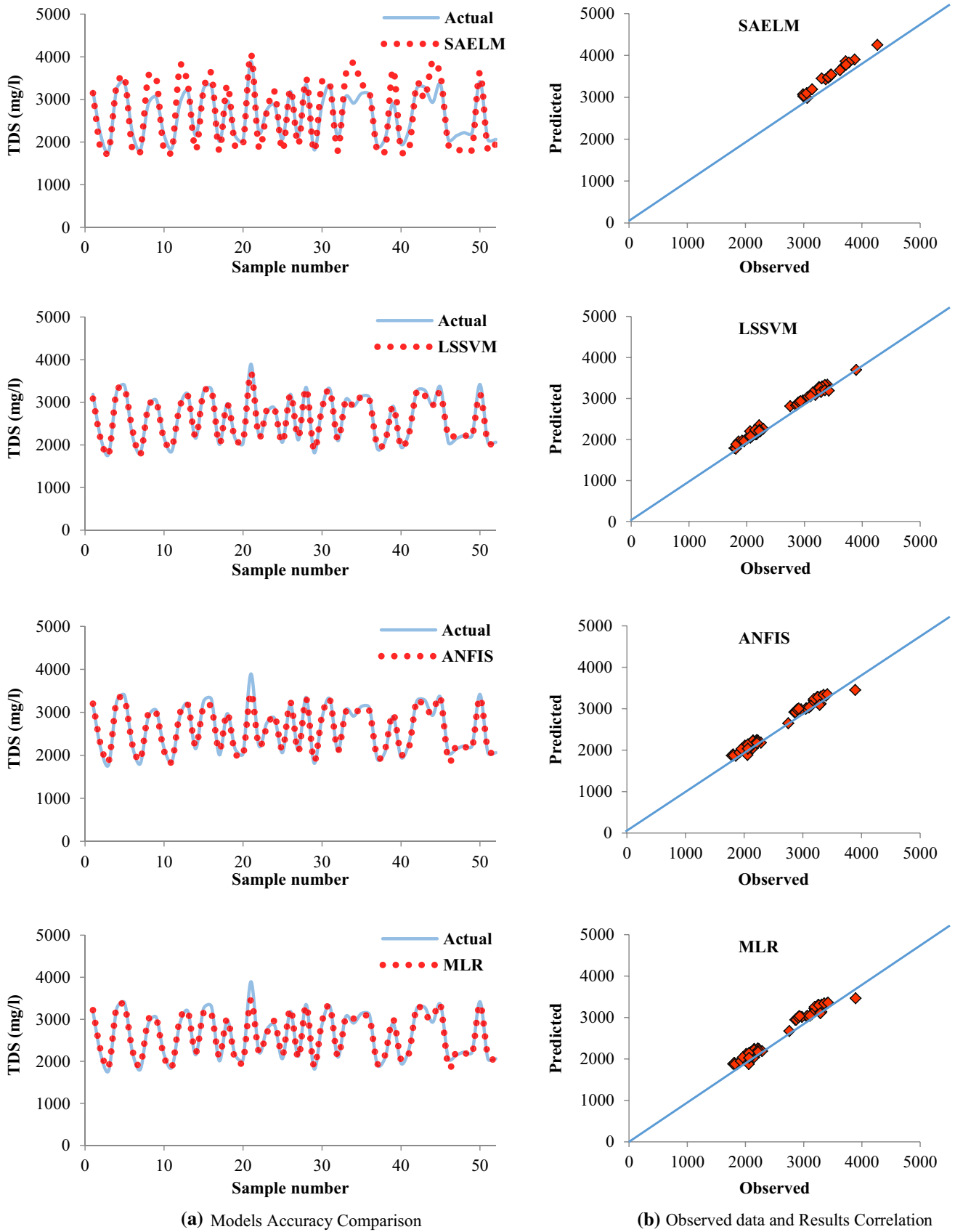


Fig. 6 Model performance in TDS simulation

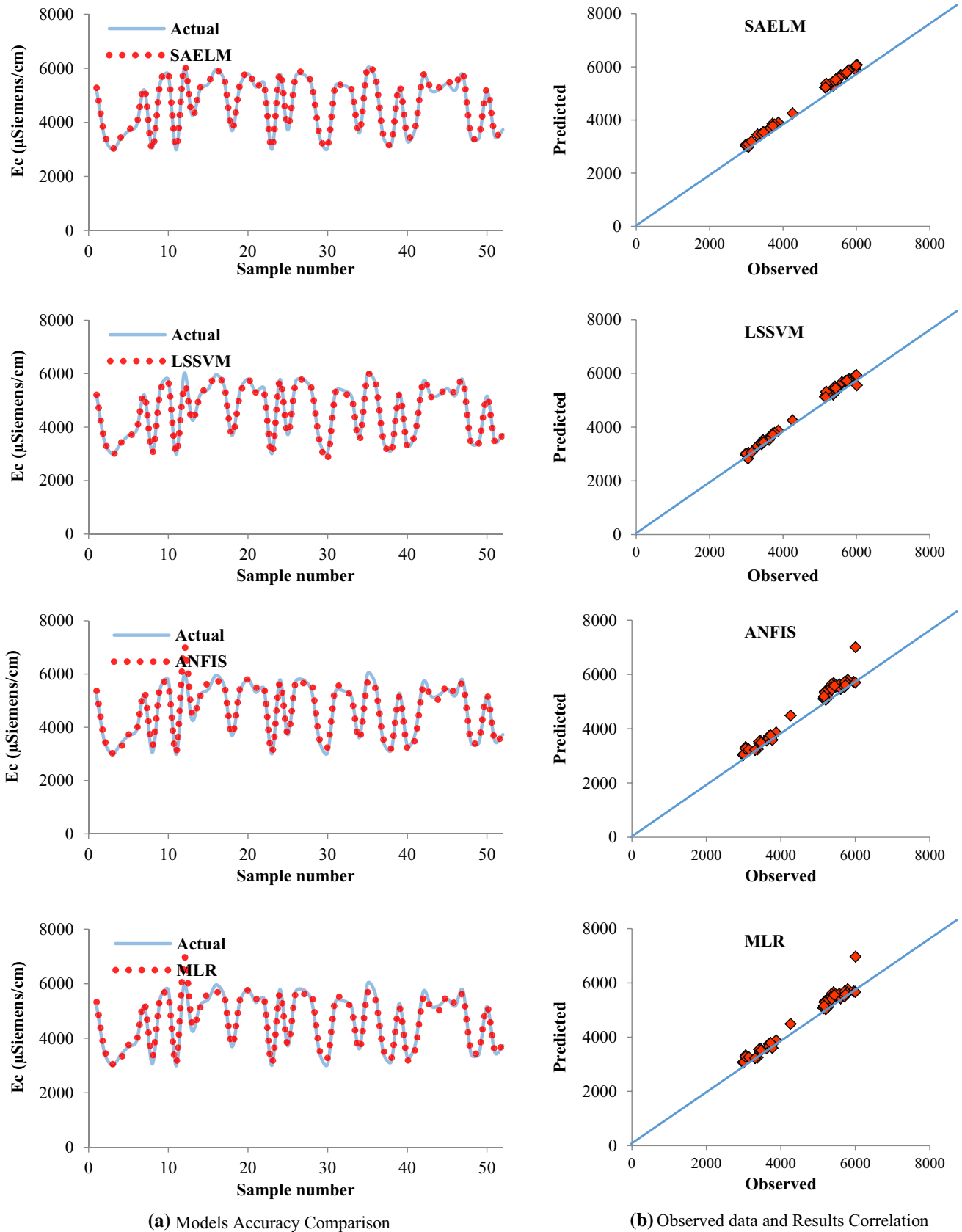


Fig. 7 Model performance in EC simulation

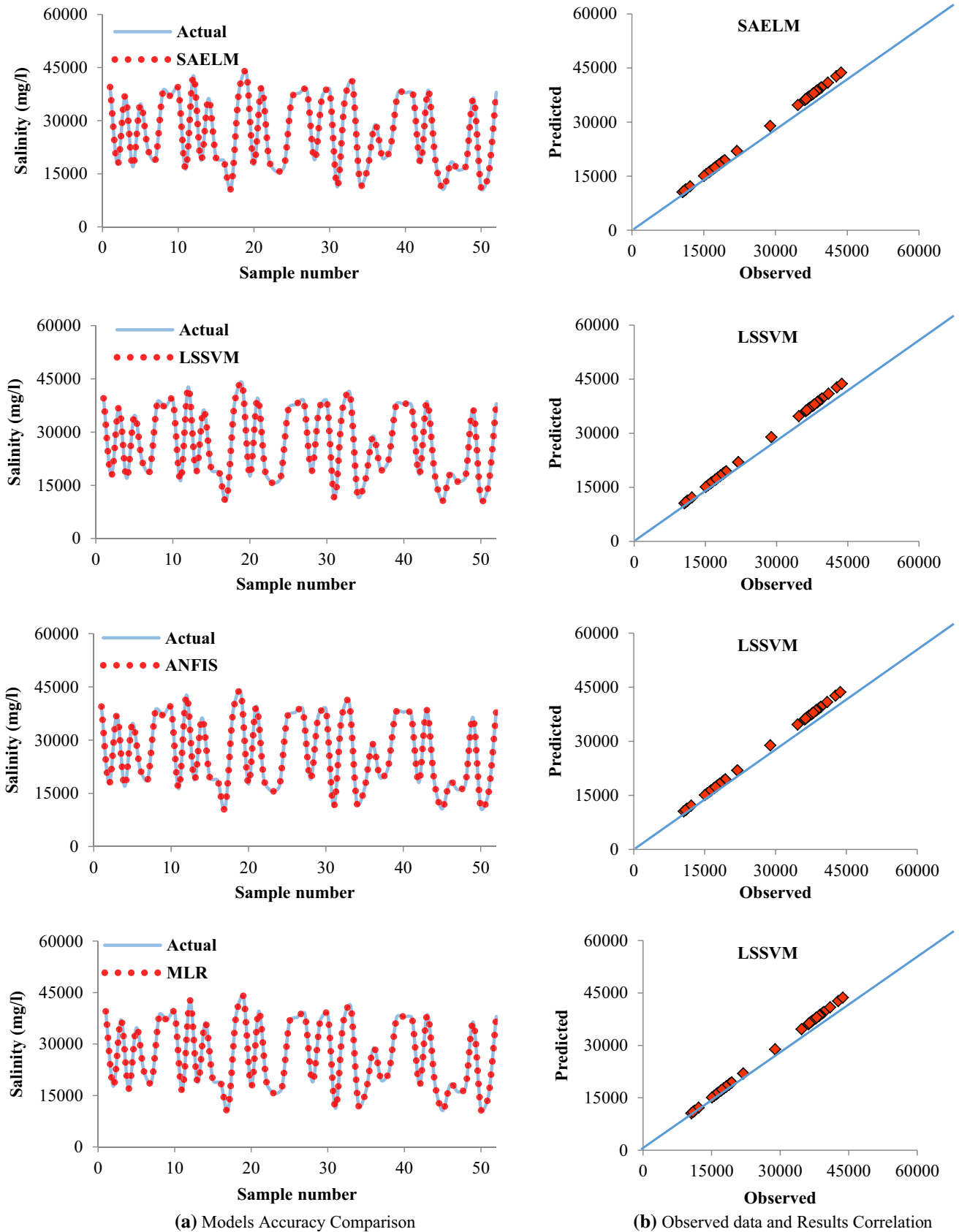


Fig. 8 Model performance in salinity simulation

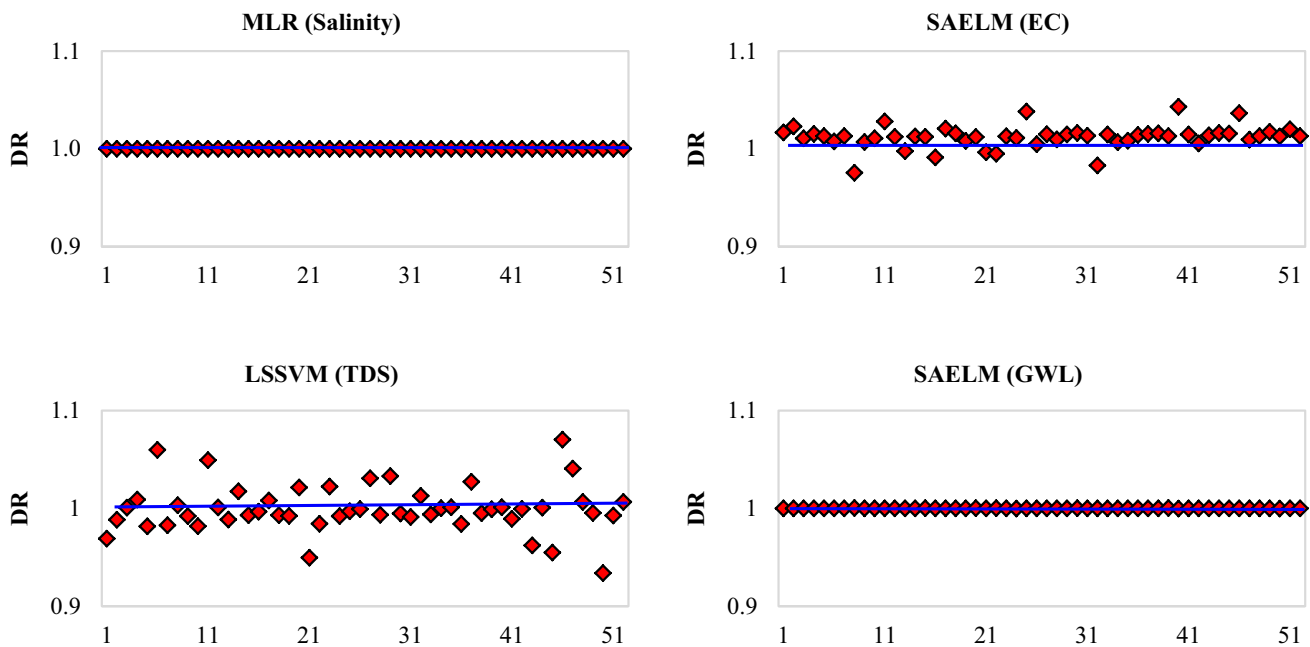


Fig. 9 Discrepancy ratio chart

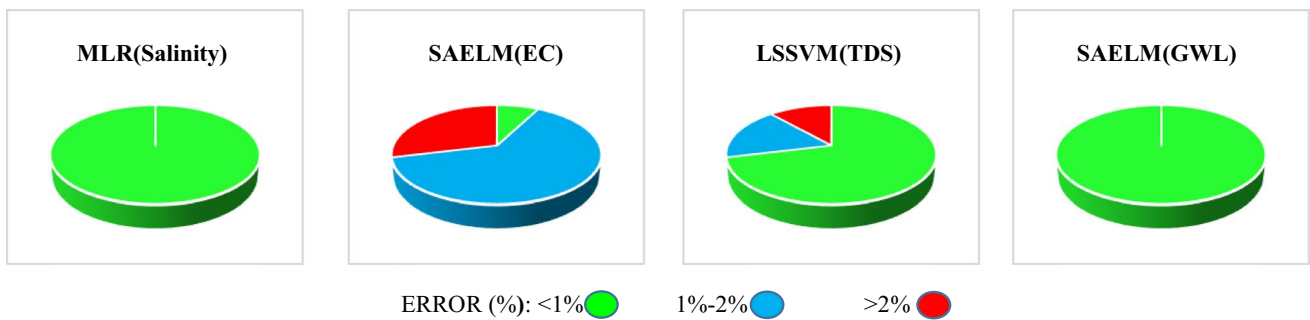


Fig. 10 Error percent distribution of models

Funding The author(s) received no specific funding for this work.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, 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 changes were made. 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/4.0/>.

References

Ahmadianfar I, Jamei M, Chu X (2020) A novel hybrid wavelet-locally weighted linear regression (W-LWLR) model for electrical conductivity (EC) prediction in surface water. *J Contam Hydrol* 232:103641. <https://doi.org/10.1016/j.jconhyd.2020.103641>

Arora S, Keshari AK (2021) ANFIS-ARIMA modelling for scheming re-aeration of hydrologically altered rivers. *J Hydrol* 601:126635. <https://doi.org/10.1016/j.jhydrol.2021.126635>

Asadi S, Amiri SS, Mottahedi M (2014) On the development of multi-linear regression analysis to assess energy consumption in the early stages of building design. *Energy Build* 85:246–255. <https://doi.org/10.1016/j.enbuild.2014.07.096>

Azimi S, Azhdary Moghaddam M (2020) Modeling short term rainfall forecast using neural networks, and gaussian process classification based on the SPI drought index. *Water Resour Manag* 34:1369–1405. <https://doi.org/10.1007/s11269-020-02507-6>

Bonakdari H, Gholami A, Mosavi A et al (2020) A novel comprehensive evaluation method for estimating the bank profile shape

- and dimensions of stable channels using the maximum entropy principle. *Entropy* 22:1–23. <https://doi.org/10.3390/e22111218>
- Čamdevýren H, Demýr N, Kanik A, Keskýn S (2005) Use of principal component scores in multiple linear regression models for prediction of Chlorophyll-a in reservoirs. *Ecol Modell* 181:581–589. <https://doi.org/10.1016/J.ECOLMODEL.2004.06.043>
- Chang CL, Chung SC, Fu WL, Huang CC (2021) Artificial intelligence approaches to predict growth, harvest day, and quality of lettuce (*Lactuca sativa* L.) in a IoT-enabled greenhouse system. *Biosyst Eng* 212:77–105. <https://doi.org/10.1016/J.BIOSYSTEMSENG.2021.09.015>
- Che Nordin NF, Mohd NS, Koting S et al (2021) Groundwater quality forecasting modelling using artificial intelligence: A review. *Groundw Sustain Dev* 14:100643. <https://doi.org/10.1016/J.GSD.2021.100643>
- Cristianini N, Shawe-Taylor J (2000) An Introduction to support vector machines and other kernel-based learning methods. *Introduct Support Vector Mach Other Kernel-Based Learn Methods*. <https://doi.org/10.1017/CBO9780511801389>
- Dibike YB, Velickov S, Solomatine D, Abbott MB (2001) Model induction with support vector machines: introduction and applications. *J Comput Civ Eng* 15:208–216. [https://doi.org/10.1061/\(ASCE\)0887-3801\(2001\)15:3\(208\)](https://doi.org/10.1061/(ASCE)0887-3801(2001)15:3(208))
- Elshafie A, Taha MR, Noureldin A (2006) A neuro-fuzzy model for inflow forecasting of the Nile river at Aswan high dam. *Water Resour Manag*. 213(21):533–556. <https://doi.org/10.1007/S11269-006-9027-1>
- Elkiran G, Nourani V, Abba SI (2019) Multi-step ahead modelling of river water quality parameters using ensemble artificial intelligence-based approach. *J Hydrol* 577:123962. <https://doi.org/10.1016/J.JHYDROL.2019.123962>
- Gholami R, Kamkar-Rouhani A, Doulati Ardejani F, Maleki S (2011) Prediction of toxic metals concentration using artificial intelligence techniques. *Appl Water Sci* 1:125–134. <https://doi.org/10.1007/S13201-011-0016-Z/TABLES/4>
- Guneshwor L, Eldho TI, Vinod Kumar A (2018) Identification of groundwater contamination sources using meshfree RPCM simulation and particle swarm optimization. *Water Resour Manag*. 324(32):1517–1538. <https://doi.org/10.1007/S11269-017-1885-1>
- Harris G (2009) Salinity. *Encycl Inl Waters*. <https://doi.org/10.1016/B978-012370626-3.00103-4>
- Heddam S, Sanikhani H, Kisi O (2019) Application of artificial intelligence to estimate phyococyanin pigment concentration using water quality data: a comparative study. *Appl Water Sci* 9:1–16. <https://doi.org/10.1007/S13201-019-1044-3/FIGURES/10>
- Huang GB, Zhou H, Ding X, Zhang R (2012) Extreme learning machine for regression and multiclass classification. *IEEE Trans Syst Man Cybern Part B Cybern* 42:513–529. <https://doi.org/10.1109/TSMCB.2011.2168604>
- Huang GB, Zhu QY, Siew CK (2004) Extreme learning machine: A new learning scheme of feedforward neural networks. *IEEE Int Conf Neural Networks - Conf Proc* 2:985–990. <https://doi.org/10.1109/IJCNN.2004.1380068>
- Huang GB, Zhu QY, Siew CK (2006) Extreme learning machine: Theory and applications. *Neurocomputing* 70:489–501. <https://doi.org/10.1016/J.NEUCOM.2005.12.126>
- 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*. <https://doi.org/10.1016/j.jhydrol.2020.125335>
- Kalteh AM (2014) Wavelet genetic algorithm-support vector regression (Wavelet GA-SVR) for monthly flow forecasting. *Water Resour Manag* 294(29):1283–1293. <https://doi.org/10.1007/S11269-014-0873-Y>
- Kalteh AM (2015) Improving forecasting accuracy of streamflow time series using least squares support vector machine coupled with data-preprocessing techniques. *Water Resour Manag* 302(30):747–766. <https://doi.org/10.1007/S11269-015-1188-3>
- Kheradpisheh Z, Talebi A, Rafati L et al (2015) Groundwater quality assessment using artificial neural network: a case study of Bahabad plain, Yazd, Iran. *Desert* 20:65–71. <https://doi.org/10.22059/JDESERT.2015.54084>
- Liang NY, Bin HG, Saratchandran P, Sundararajan N (2006) A fast and accurate online sequential learning algorithm for feedforward networks. *IEEE Trans Neural Netw* 17:1411–1423. <https://doi.org/10.1109/TNN.2006.880583>
- Majumder P, Eldho TI (2020) Artificial neural network and grey wolf optimizer based surrogate simulation-optimization model for groundwater remediation. *Water Resour Manag*. 342(34):763–783. <https://doi.org/10.1007/S11269-019-02472-9>
- Manu DS, Thalla AK (2017) Artificial intelligence models for predicting the performance of biological wastewater treatment plant in the removal of Kjeldahl Nitrogen from wastewater. *Appl Water Sci* 7:3783–3791. <https://doi.org/10.1007/S13201-017-0526-4/FIGURES/5>
- Nema MK, Khare D, Chandniha SK (2017) Application of artificial intelligence to estimate the reference evapotranspiration in sub-humid Doon valley. *Appl Water Sci* 7:3903–3910. <https://doi.org/10.1007/S13201-017-0543-3/TABLES/4>
- Parmar KS, Bhardwaj R (2014) Water quality management using statistical analysis and time-series prediction model. *Appl Water Sci* 4:425–434. <https://doi.org/10.1007/S13201-014-0159-9/FIGURES/3>
- Poursaeid M, Mastouri R, Shabanlou S, Najarchi M (2020) Estimation of total dissolved solids, electrical conductivity, salinity and groundwater levels using novel learning machines. *Environ Earth Sci*. 7919(79):1–25. <https://doi.org/10.1007/S12665-020-09190-1>
- Poursaeid M, Mastouri R, Shabanlou S, Najarchi M (2021) Modelling qualitative and quantitative parameters of groundwater using a new wavelet conjunction heuristic method: wavelet extreme learning machine versus wavelet neural networks. *Water Environ J* 35:67–83. <https://doi.org/10.1111/wej.12595>
- Qu X, Chen Y, Liu H et al (2020) A holistic assessment of water quality condition and spatiotemporal patterns in impounded lakes along the eastern route of China's South-to-North water diversion project. *Water Res* 185:116275. <https://doi.org/10.1016/J.WATRES.2020.116275>
- Sada SO, Ikpeseni SC (2021) Evaluation of ANN and ANFIS modeling ability in the prediction of AISI 1050 steel machining performance. *Heliyon* 7:e06136. <https://doi.org/10.1016/J.HELIYON.2021.E06136>
- Sapankevych N, Sankar R (2009) Time series prediction using support vector machines: a survey. *IEEE Comput Intell Mag* 4:24–38. <https://doi.org/10.1109/MCI.2009.932254>
- Sarkar J, Prottoy ZH, Bari MT, Al Faruque MA (2021) Comparison of anfis and ann modeling for predicting the water absorption behavior of polyurethane treated polyester fabric. *Heliyon* 7:e08000. <https://doi.org/10.1016/j.heliyon.2021.e08000>
- Serrano-Finetti E, Aliau-Bonet C, López-Lapeña O, Pallàs-Areny R (2019) Cost-effective autonomous sensor for the long-term monitoring of water electrical conductivity of crop fields. *Comput Electron Agric* 165:104940. <https://doi.org/10.1016/j.compag.2019.104940>
- Shahid ES, Ehteshami M (2015) Application of artificial neural networks to estimating DO and salinity in San Joaquin River basin. *New Pub Balaban* 57:4888–4897. <https://doi.org/10.1080/19443994.2014.995713>
- Solanki A, Agrawal H, Khare K (2015) Predictive analysis of water quality parameters using deep learning. *Int J Comput Appl* 125:29–34. <https://doi.org/10.5120/ijca2015905874>

- Sparks DL (2003) The chemistry of saline and sodic soils. *Environ Soil Chem*. <https://doi.org/10.1016/B978-012656446-4/50010-4>
- Tokachichu J, Das Gaddam TR (2021) Performance analysis of a transmission line connected with UPFC designed with three level cascaded H bridge inverter with generalized SVM technique using PI, FUZZY LOGIC, ANN and ANFIS controllers. *Mater Today Proc*. <https://doi.org/10.1016/j.matpr.2021.07.338>
- Vaheddoost B, Aksoy H (2018) Interaction of groundwater with Lake Urmia in Iran. *Hydrol Process* 32:3283–3295. <https://doi.org/10.1002/hyp.13263>
- Valyon J, Horvath G (2007) (PDF) Extended Least Squares LS-SVM. *World Acad Sci Eng Technol* 36:
- Vijay S, Kamaraj K (2021) Prediction of water quality index in drinking water distribution system using activation functions based ANN. *Water Resour Manag* 35:535–553. <https://doi.org/10.1007/s11269-020-02729-8>
- Walker D, Jakovljević D, Savić D, Radovanović M (2015) Multi-criterion water quality analysis of the Danube River in Serbia: A visualisation approach. *Water Res* 79:158–172. <https://doi.org/10.1016/J.WATRES.2015.03.020>
- Wang P, Yao J, Wang G et al (2019) Exploring the application of artificial intelligence technology for identification of water pollution characteristics and tracing the source of water quality pollutants. *Sci Total Environ* 693:133440. <https://doi.org/10.1016/J.SCITOTENV.2019.07.246>
- Wu W, Dandy GC, Maier HR (2014) Protocol for developing ANN models and its application to the assessment of the quality of the ANN model development process in drinking water quality modelling. *Environ Model Softw* 54:108–127. <https://doi.org/10.1016/J.ENVSOFT.2013.12.016>
- Yang X, Zhang H, Zhou H (2014) a hybrid methodology for salinity time series forecasting based on wavelet transform and NARX neural networks. *Arab J Sci Eng*. 3910(39):6895–6905. <https://doi.org/10.1007/S13369-014-1243-Z>
- Zhang Y, Gao X, Smith K et al (2019) Integrating water quality and operation into prediction of water production in drinking water treatment plants by genetic algorithm enhanced artificial neural network. *Water Res* 164:114888. <https://doi.org/10.1016/J.WATRES.2019.114888>
- Zhang Y, Wu L, Deng L, Ouyang B (2021) Retrieval of water quality parameters from hyperspectral images using a hybrid feedback deep factorization machine model. *Water Res*. <https://doi.org/10.1016/J.WATRES.2021.117618>
- Zhou Y (2020) Real-time probabilistic forecasting of river water quality under data missing situation: deep learning plus post-processing techniques. *J Hydrol* 589:125164. <https://doi.org/10.1016/J.JHYDROL.2020.125164>

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